

James J. Park
Laurence T. Yang
Changhoon Lee (Eds.)

Communications in Computer and Information Science

184

Future Information Technology

6th International Conference, FutureTech 2011
Loutraki, Greece, June 2011
Proceedings, Part I

Part 1

James J. Park Laurence T. Yang
Changhoon Lee (Eds.)

Future Information Technology

6th International Conference, FutureTech 2011
Loutraki, Greece, June 28-30, 2011
Proceedings, Part I

Volume Editors

James J. Park

Seoul National University of Science and Technology
Department of Computer Science and Engineering
172 Gongreung 2-dong, Nowon-gu, Seoul, 139-743, Korea
E-mail: parkjonghyuk1@hotmail.com

Laurence T. Yang

St. Francis Xavier University
Department of Computer Science
Antigonish, NS, B2G 2W5, Canada
E-mail: ltyang@stfx.ca

Changhoon Lee

Hanshin University
411 Yangсандong, Osan-si, 447-791, Korea
E-mail: chlee@hs.ac.kr

ISSN 1865-0929

e-ISSN 1865-0937

ISBN 978-3-642-22332-7

e-ISBN 978-3-642-22333-4

DOI 10.1007/978-3-642-22333-4

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011930821

CR Subject Classification (1998): C.2, H.4, H.3, I.2, D.2, H.5

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Message from the General Chairs

On behalf of the Organizing Committees, it is our pleasure to welcome you to the 6th International Conference on Future Information Technology (FutureTech 2011), which was held in Loutraki, Greece, June 28–30, 2011.

In past 25 years or so, information technology (IT) as influenced and changed every aspect of our lives and our cultures. Without various IT-based applications, we would find it difficult to keep information stored securely, to process information efficiently, and to communicate information conveniently. In the future world, IT will play a very important role in the convergence of computing, communication, and all other computational sciences and applications. IT will also influence the future world's various areas, including science, engineering, industry, business, law, politics, culture, medicine, and so on.

FutureTech 2011 added to the series of the highly successful International Conference on Future Information Technology (FutureTech 2010, Busan, Korea, 2010), previously held as International Symposium on Ubiquitous Applications & Security Services (UASS 2009, USA, 2009), UASS 2008 (Okinawa, Japan, 2008), UASS 2007 (Kuala Lumpur, Malaysia, 2007), and UASS 2006 (Glasgow, Scotland, UK, 2006).

This year we were proud not only to have a distinguished set of the following invited speakers, but also the distinguished support of James J. (Jong Hyuk) Park as the FTRA president.

- Habib F. Rashvand, PhD, School of Engineering, University of Warwick, UK
- J.C. (Hans) van Vliet, PhD, Department of Information Management and Software Engineering (IMSE), Vrije Universiteit, The Netherlands
- Peter Sloot, PhD, Full Professor of Computational Science, Scientific Director of the Informatics Institute, The Netherlands

We are sure you will find many exciting and stimulating papers in the proceedings, and we would like to thank the General-Vice Chair, Changhoon Lee, and the Program Chairs, Serge Chaumette, Jin Kwak, and Konstantinos Markantonakis, for collecting an extremely strong set of papers across a wide range of topics.

Finally, we want to extend our thanks to all our Chairs and Committees who are truly responsible for making this conference a success through their generous contributions of time and energy and are listed in the following pages. We also sincerely thank the EAPIC-IST 2011, ISCC 2011, ITSA 2011, F2GC 2011, MUST 2011, and SocialComNet 2011 Workshop Chairs for their kind cooperation.

Damien Sauveron
Sang-Soo Yeo

Message from the Program Chairs

Welcome to the proceedings of the 6th International Conference on Future Information Technology (FutureTech 2011), held in Loutraki, Greece, June 28–30, 2011.

Future information technology would stand for all of continuously evolving and converging information technologies, including digital convergence, multimedia convergence, intelligent applications, embedded systems, mobile and wireless communications, bio-inspired computing, grid and cloud computing, Semantic Web, user experience and HCI, security and trust computing and so on, for satisfying our ever-changing needs.

FutureTech 2011 intended to foster the dissemination of state-of-the-art research in all future information technology areas, including their models, services, and novel applications associated with their utilization.

We received 274 papers with the following topics from around the globe and after a rigorous review process, we accepted 123 papers, for a competitive acceptance rate of 45%.

- Track 1. Hybrid Information Technology
- Track 2. High Performance Computing
- Track 3. Cloud Computing
- Track 4. Ubiquitous Networks and Wireless Communications
- Track 5. Digital Convergence
- Track 6. Multimedia Convergence
- Track 7. Intelligent and Pervasive Applications
- Track 8. Security and Trust Computing
- Track 9. IT Management and Service
- Track 10. Bioinformatics and Bio-Inspired Computing

We believe that the selected papers provide novel ideas, new results, work in progress and state-of-the-art techniques in this field, stimulating future research activities.

We would like to express our appreciation to all the contributors and authors for the submissions to FutureTech 2011. Special thanks are due to the members of the Technical Program Committee and all the external reviewers for their invaluable help with the review process. Finally, we would like to thank everyone who participated in the conference, and acknowledge the members of the Organizing Committee for their contributions and in particular Leonie Kunz at springer for the great help and hard work.

Serge Chaumette
Jin Kwak
Konstantinos Markantonakis

FutureTech 2011 Organization

Steering Chairs

James J. (Jong Hyuk) Park	Seoul National University of Science and Technology, Korea
Laurence T. Yang	St. Francis Xavier University, Canada
Hamid R. Arabnia	The University of Georgia, USA

General Chairs

Damien Sauveron	University of Limoges, France
Sang-Soo Yeo	Mokwon University, Korea

General Vice-Chair

Changhoon Lee	Hanshin University, Korea

Program Chairs

Serge Chaumette	LaBRI, University of Bordeaux, France
Jin Kwak	Soonchunhyang University, Korea
Konstantinos Markantonakis	Royal Holloway University of London, UK

Steering Committee

Han-Chieh Chao	National Ilan University, Taiwan
Shu-Ching Chen	Florida International University, USA
Stefanos Gritzalis	University of the Aegean, Greece
Vincenzo Loia	University of Salerno, Italy
Yi Mu	University of Wollongong, Australia
Witold Pedrycz	University of Alberta, Canada
Wanlei Zhou	Deakin University, Australia
Young-Sik Jeong	Wonkwang University, Korea

Advisory Committee

Ioannis G. Askoxylakis	FORTH-ICS, Greece
Hsiao-Hwa Chen	National Sun Yat-Sen University, Taiwan
Jack J. Dongarra	University of Tennessee, USA
Javier Lopez	University of Malaga, Spain
Bart Preneel	Katholieke Universiteit Leuven, Belgium
Harry Rudin	IBM Zurich Research Laboratory, Switzerland

Waleed W. Smari	University of Dayton, USA
Doug Tygar	University of California Berkeley, USA
Shiuh-Jeng Wang	Central Police University, Taiwan
Doo-soon Park	SoonChunHyang University, Korea

Workshop Co-chairs

Sabrina De Capitani di Vimercati	Università degli Studi di Milano, Italy
Naveen Chilamkurti	La Trobe University, Australia

Publicity Chairs

Deok-Gyu Lee	ETRI, Korea
Theo Tryfonas	University of Bristol, UK
Karim El Defrawy	University of California, Irvine, USA
Sang Yep Nam	Kookje College, Korea
Young-June Choi	Ajou University, Korea

International Liaison Chairs

Jiankun Hu	RMIT University, Australia
Louis Marinos	ENISA, EU
Qishi Wu	University of Memphis, USA

Publication Chair

Jose Maria Sierra	Universidad Carlos III de Madrid, Spain

Local Arrangements Co-chairs

Pierre-Francois Bonnefoi	University of Limoges, France
Pierre Dusart	University of Limoges, France

Web and System Management Chair

Yang Sun Lee	Chosun University, Korea

Track Co-chairs

Track 1. Hybrid Information Technology

Umberto Straccia	ISTI - C.N.R., Italy
Malka N. Halgamuge	The University of Melbourne, Australia
Andrew Kusiak	The University of Iowa, USA

Track 2. High-Performance Computing

Waleed W. Smari	University of Dayton, Ohio, USA
Omer Rana	Cardiff University, UK
Satoshi Matsuoka	Tokyo Institute of Technology, Japan

Track 3. Cloud Computing

Geoffrey Fox	Indiana University, USA
Pascal Bouvry	University of Luxembourg, Luxembourg
Andrzej M. Goscinski	Deakin University, Australia

Track 4. Ubiquitous Networks and Wireless Communications

Ivan Stojmenovic	University of Ottawa, Canada
Tuomo Tuikka	Research Centre, Finland
Giannong Cao	Hong Kong Polytechnic University, Hong Kong

Track 5. Digital Convergence

Gabriel Jakobson	Altusys Corp., USA
Vladimir Zaborovsky	Polytechnic University of Saint-Petersburg, Russia
Johnnes Arreymbi	University of East London, UK

Track 6. Multimedia Convergence

Jonathan C.L. Liu	University of Florida, USA
Sofia Tsekeridou	Athens Information Technology, Greece
Chia-Chen Lin	Providence University, Taiwan

Track 7. Intelligent and Pervasive Applications

Hongli Luo	Indiana University-Purdue University Fort Wayne, USA
Isabelle Demeure	TELECOM ParisTech, France
Kuan-Ta Chen	Institute of Information Science, Academia Sinica, Taiwan

Track 8. Security and Trust Computing

Luis F.G. Sarmenta	MIT Media Lab, USA
Jose Maria Sierra	Universidad Carlos III de Madrid, Spain
Chan Yeob Yeun	Khalifa University of Science, Technology & Research (KUSTAR), UAE

Track 9. IT Management and Service

Youakim Badr	INSA-Lyon, France
Ying Jiang	Wuhan University, China
Hang-bae Chang	Daejin University, Korea

Track 10. Bioinformatics and Bio-Inspired Computing

Frederic Guinand	University of Le Havre, France
Sun-Yuan Hsieh	National Cheng Kung University, Taiwan
Sheng-Lung Peng	National Dong Hwa University, Taiwan

Table of Contents – Part I

Future Information Technology (24 Papers)

Whom to Convince? It Really Matters in BGP Prefix Hijacking Attack and Defense	1
<i>Yujing Liu, Bin Dai, Peidong Zhu, and Jinshu Su</i>	
Future Trends of Intelligent Decision Support Systems and Models	11
<i>Andrzej M.J. Skulimowski</i>	
Relation Extraction from Documents for the Automatic Construction of Ontologies	21
<i>Ikkyu Choi, Seungmin Rho, Young-Sik Jeong, and Minkoo Kim</i>	
Proactive Detection of Botnets with Intended Forceful Infections from Multiple Malware Collecting Channels	29
<i>Young Hoon Moon and Huy Kang Kim</i>	
Solving English Questions through Applying Collective Intelligence	37
<i>Dongjin Choi, Myunggwon Hwang, Byeongkyu Ko, and Pankoo Kim</i>	
Automatic Documents Annotation by Keyphrase Extraction in Digital Libraries Using Taxonomy	47
<i>Iram Fatima, Asad Masood Khattak, Young-Koo Lee, and Sungyoung Lee</i>	
IO-Aware Custom Instruction Exploration for Customizing Embedded Processors	57
<i>Amir Yazdanbakhsh, Mostafa E. Salehi, and Sied Mehdi Fakhraie</i>	
TSorter: A Conflict-Aware Transaction Processing System for Clouds . . .	67
<i>Po-Cheng Chen, Yen-Liang Su, Jyh-Biau Chang, and Ce-Kuen Shieh</i>	
New Secure Storage Architecture for Cloud Computing	75
<i>Sameera Abdulrahman Almula and Chan Yeob Yeun</i>	
An Adaptive WiFi Rate Selection Algorithm for Moving Vehicles with Motion Prediction	85
<i>Jianwei Niu, Yuhang Gao, Shaohui Guo, Chao Tong, and Guoping Du</i>	
A Radio Channel Sharing Method Using a Bargaining Game Model	93
<i>Jaesung Park, Yong-Hoon Choi, and Kilhung Lee</i>	

Load Balancing with Fair Scheduling for Multiclass Priority Traffic in Wireless Mesh Networks	101
<i>Neeraj Kumar, Naveen Chilamkurti, Jong Hyuk Park, and Doo-Soon Park</i>	
A Dependable and Efficient Scheduling Model and Fault Tolerance Service for Critical Applications on Grid Systems	110
<i>Bahman Arasteh and Mohammad Javad Hosseini</i>	
The Performance Evaluation of Heuristic Information-Based Wireless Sensor Network Routing Method	123
<i>Hyekyoung Jeon, Seungjin Han, Keewook Rim, and Junghyun Lee</i>	
A Study on the Spectral and Energy Efficient-Path Selection Scheme in Two-Hop Cellular Systems	131
<i>Se-Han Kim, Se-Jin Kim, Byung-Chul Kim, and Jae-Yong Lee</i>	
The Construction of Remote Microcontroller Laboratory Using Open Software	140
<i>Kwansun Choi, Saeron Han, Dongsik Kim, Changwan Jeon, Jongsik Lim, Sunheum Lee, Doo-soon Park, and Heunggu Jeon</i>	
LAPSE+ Static Analysis Security Software: Vulnerabilities Detection in Java EE Applications	148
<i>Pablo Martín Pérez, Joanna Filipiak, and José María Sierra</i>	
A Low-Power Wakeup-On-Demand Scheme for Wireless Sensor Networks	157
<i>Byung-Bog Lee, Se-Jin Kim, and Choong-Ho Cho</i>	
The Improved Space-Time Trellis Codes with Proportional Mapping on Fast Fading Channels	167
<i>Ik Soo Jin</i>	
Strategies for IT Convergence Services in Rural Areas	175
<i>Hyongsoon Kim, Kwang-Taek Ryu, Sang-Yong Ha, and Eunyoung Lee</i>	
Enlarging Instruction Window through Separated Reorder Buffers for High Performance Computing	183
<i>Min Choi, Jonghyuk Park, and Young-Sik Jeong</i>	
Smart Mobile Banking and Its Security Issues: From the Perspectives of the Legal Liability and Security Investment	190
<i>Se-Hak Chun</i>	
Future Green Technologies and Performances: A Case Study of the Korean Stock Market	196
<i>Se-Hak Chun and Seonmin Kim</i>	

Corporate Social Responsibility and Its Performances: Application to SRI (Socially Responsible Investing) Mutual Funds	200
<i>Joon Ho Hwang, Dong Han Kim, and Se-Hak Chun</i>	

IT Service and Cloud Computing (43 Papers)

Mobile Cloud e-Gov Design and Implementation Using WebSockets API	204
<i>Young-Hyuk Kim, Il-Kwon Lim, Seung-Gu Kang, and Jae-Kwang Lee</i>	
Noise Reduction in Image Using Directional Modified Sigma Filter	212
<i>Hye-Youn Lim, Mi-Ran Gu, and Dae-Seong Kang</i>	
An Approach to Real-time Region Detection Algorithm using Background Modeling and Covariance Descriptor	218
<i>Jong-Dae Park, Hye-Youn Lim, and Dae-Seong Kang</i>	
Optimization and Generation of Knowledge Model for Supporting Technology Innovation	224
<i>Seungsu Jun, Seungryong Lee, and Munbong Shin</i>	
An Application of SCORM Based CAT for E-Learning	230
<i>HwaYoung Jeong and Bong Hwa-Hong</i>	
Research of the Solar Orbit Compute Algorithm for Improving the Efficiency of Solar Cell	237
<i>Sang-Bong Byun, Eun-Kyung Kim, and Yong-Hwan Lee</i>	
A Learning System using User Preference in Ubiquitous Computing Environment	240
<i>HwaYoung Jeong and Bong Hwa-Hong</i>	
Introduction of the Art of Korean Traditional Culture: Multimedia based Pansori	248
<i>Dong-Keon Kim</i>	
A Study on Model Transformation Mechanism Using Graph Comparison Algorithms and Software Model Property Information	256
<i>Jong-won Ko, Hwa-young Jeong, and Young-jae Song</i>	
Video Service Algorithm using Web-Cached Technique in WLAN	265
<i>Iksoo Kim, Seoungyeol Lee, and Yoseop Woo</i>	
A Design and Implementation of Mobile Puzzle Game	273
<i>Seongsoo Cho, Bhanu Shrestha, Kwang Chul Son, and Bonghwa Hong</i>	
Implementation of AUV Test-Bed	280
<i>Soon T. Kwon, Woon Kyung Baek, and Moon G. Joo</i>	

Communication and Computation Overlap through Task Synchronization in Multi-locale Chapel Environment	285
<i>Bongen Gu, Weikuan Yu, and Yoonsik Kwak</i>	
Compensation System for RF System-on-Chip	293
<i>Sung-Woo Kim, Jee-Youl Ryu, Jae-Hwan Lim, Seung-Hun Park, Jung-Hoon Lee, and Min Yoon</i>	
A Study on the SNS (Social Network Service) Based on Location Model Combining Mobile Context-Awareness and Real-Time AR (Augmented Reality) via Smartphone	299
<i>Jang-Mook Kang and Bong-Hwa Hong</i>	
Design and Implementation MoIP Wall-Pad Platform for Home-Network	308
<i>Yong-kuk Jung, Min-Su Kweon, Kwang-Seon Heo, Dae-Sung Kim, and Young-Gyu Choi</i>	
Outlier Rejection Methods for Robust Kalman Filtering	316
<i>Du Yong Kim, Sang-Goog Lee, and Moongu Jeon</i>	
A Study on Receiving Performance Improvement of LTE Communication Network Using Multi-hop Relay Techniques	323
<i>Chan-Hong Park, Yong-Seok Choi, Young-Hwan Han, and Hyeon-Kyeong Seong</i>	
Implementation of Linux Server System Monitoring and Control Solution for Administrator	329
<i>Sung-Won Jang, Byeong-Ho Park, Sang-Ju Park, and Hyeon-Kyeong Seong</i>	
A Smart Personal Activity Monitoring System Based on Wireless Device Management Methods	335
<i>Jugeon Pak and Keehyun Park</i>	
A Study on Demodulation System Design of the VOR Receiver	343
<i>Jeong IL Park and Hwa Se Park</i>	
China's Electronic Information Policy	350
<i>Won-bong Lee</i>	
3D Content Industry in Korea : Present Conditions and Future Development Strategies	358
<i>Yunkyung Lee and Youngsik Kwak</i>	
A Study of Mobile Application Usage in After Service Management for Consumer Electronics Industry	364
<i>Wonkyu Ji and Xu Wang</i>	

A Study on the Mechanical Properties Analysis of Recycled Coarse Aggregate using Polymer Impregnated	371
<i>Hyun-gi Ryu and Jung-sik Kim</i>	
A Study on the Strength Improvement of Recycled Aggregates using Industrial Waste	378
<i>Jung-sik Kim and Hyun-gi Ryu</i>	
Design and Analysis of Optimizing Single-Element Microstrip Patch Antenna for Dual-Band Operations.....	387
<i>Dong-Hee Park and Yoonsik Kwak</i>	
An Analysis of U-Healthcare Business Models and Business Strategies: Focused on Life Insurance Industry	393
<i>Dong-Hwan Cho and Jin-Hyuk Hong</i>	
Outage Probability Evaluation of DSF-Relaying Equipped CDD Schemes for MC-CDMA Systems	401
<i>Kyunbyoung Ko and Choongchae Woo</i>	
On BER Performance of CDD-DF-Relay Scheme for MC-CDMA Systems	409
<i>Jeanyoung Jang, Choongchae Woo, and Kyunbyoung Ko</i>	
The Effect of Corporate Strategy and IT Role on the Intent for IT Outsourcing Decision	417
<i>Dong-Hwan Cho and Jin-Sung Kim</i>	
A Study on IT Organization Redesign with IT Governance: Focusing on K Public Corporation in Korea	426
<i>Dong Hwan Cho and Hyoung Seop Shim</i>	
MF-SNOOP for Handover Performance Enhancement	437
<i>Chi-Hyun Ahn, Hyoungchul Kim, and Jongjung Woo</i>	
The Influences of On-line Fashion Community Network Features on the Acceptance of Fashion Information	445
<i>Kieun Song, Sunjin Hwang, Yunsik Kim, and Youngsik Kwak</i>	
Exact Closed-form Outage Probability Analysis Using PDF Approach for Fixed-DF and Adaptive-DF Relay Systems over Rayleigh Fading Channels	456
<i>Jeanyoung Jang and Kyunbyoung Ko</i>	
From Trading Volume to Trading Number-Based Pricing at Home Trading System on Korean Stock Market	463
<i>Youngsik Kwak, Yunkyoung Lee, Jaeweon Hong, Wanwoo Cho, Ho Jang, and Daehyun Park</i>	

Design and Implementation of the Application Service System Based on Sensor Networks	469
<i>Yoonsik Kwak and Donghee Park</i>	
Software RAID 5 for OLTP Applications with Frequent Small Writes . . .	475
<i>Kijeong Khil, Dongho Kwak, Seokil Song, Yunsik Kwak, and Seungkook Cheong</i>	
The Effects of Project, Customer and Vendor Properties on the Conflict of IS Outsourcing Development	483
<i>Dong-Hwan Cho and Haeng-Nam Sung</i>	
A Study on the Split Algorithm of URL LIST Collected by Web Crawler	492
<i>IL-Kwon Lim, Young-Hyuk Kim, Seung-Gu Kang, and Jae-Kwang Lee</i>	
A Study on Authentication System Using QR Code for Mobile Cloud Computing Environment	500
<i>Dong-Sik Oh, Bong-Han Kim, and Jae-Kwang Lee</i>	
Research on the I/O Performance Advancement of a Low Speed HDD Using DDR-SSD	508
<i>Seung-Kook Cheong, Jae-Jin Jeong, Yong-Wan Jeong, Dae-Sik Ko, and Young-Hoon Lee</i>	
Neighborhood Evolution in MANET Experiments	514
<i>Julien Franzolini, Frédéric Guinand, and Damien Olivier</i>	
Author Index	523

Table of Contents – Part II

Social Computing, Network, and Services

A User Supporting Personal Video Recorder Based on a Generic Bayesian Classifier and Social Network Recommendations	1
<i>Benedikt Engelbert, Malte Blanken, Ralf Kruthoff-Brüwer, and Karsten Morisse</i>	
Local Community Finding Using Synthetic Coordinates	9
<i>Harris Papadakis, Costas Panagiotakis, and Paraskevi Fragopoulou</i>	
Virtual B-Tree Topology: A Self-Routing Topology for a Wireless Sensor Network	16
<i>Shan-Yang Wang, Tzu-Chiang Chiang, and Chu-Sing Yang</i>	
A Peer-to-Peer Social Network Overlay for Efficient Information Retrieval and Diffusion	24
<i>Salma Ktari and Artur Hecker</i>	
Good Friends, Bad News - Affect and Virality in Twitter	34
<i>Lars Kai Hansen, Adam Arvidsson, Finn Aarup Nielsen, Elanor Colleoni, and Michael Etter</i>	
Semantic Verification in an Open Collaboration Scenario	44
<i>Jorge Coelho and Mário Florido</i>	
Designing Persuasive Third Party Applications for Social Networking Services Based on the 3D-RAB Model	54
<i>Isaac Wiafe, Keiichi Nakata, and Stephen R. Gulliver</i>	
A Self-adaptive Clustering Scheme with a Time-Decay Function for Microblogging Text Mining	62
<i>Chung-Hong Lee and Chih-Hung Wu</i>	
Weighted Co-authorship Network Based on Forgetting	72
<i>Milos Kudelka, Zdenek Horak, Vaclav Snasel, and Ajith Abraham</i>	
Towards Usage-Centered Design Patterns for Social Networking Systems	80
<i>Dina Hussein, Ghada Alaa, and Ahmed Hamad</i>	
Ranked Clusterability Model of Dyadic Data in Social Network	90
<i>R.B. Fajriya Hakim, Subanar, and Edi Winarko</i>	

Forensics for Future Generation Communication Environments

Forensic Analysis of Residual Information in Adobe PDF Files	100
<i>Hyunji Chung, Jungheum Park, and Sangjin Lee</i>	
Digital Quantum Forensics: Challenges and Responses	110
<i>Richard E. Overill</i>	
Recovery Techniques for Deleted Email Items in Digital Forensic Context	115
<i>Chorong Jeong, Keun-gi Lee, and Sangjin Lee</i>	
Digital Evidence Discovery and Knowledge Management Issues Concerning Multimedia Computing Devices Utilizing GPS Navigation Services and Social Network Activities	123
<i>Hai-Cheng Chu, K.H. Chang, Yi-Da Wang, and Jong Hyuk Park</i>	
Website Risk Assessment System for Anti-Phishing	131
<i>Young-Gab Kim and Sungdeok Cha</i>	

Intelligent Transportation Systems and Applications

Delivering Real-Time Bus Tracking Information on Mobile Devices	139
<i>Bashir Shalaik and Adam Winstanley</i>	
Camel-Vehicle Accidents Mitigation System: Design and Survey	148
<i>Khaled Ragab, Mohammed Zahrani, and Asrar Ul Haque</i>	
A Hierarchical Architecture for Semantic Sensing Information Representation	159
<i>Rajani Reddy Gorrepati, Hyeon Park, Dong-Hwan Park, and Do-Hyeun Kim</i>	
An Integrated Process Method of Sensing Data for Multi-sensor Networks	167
<i>Jun-Pil Boo, Do-Hyeun Kim, Hee-Dong Park, and Paul Roe</i>	

Multimedia and Semantic Technologies

Relevance Feedback for Surveillance Video Retrieval at Object Level . . .	175
<i>Thi-Lan Le</i>	
Identifying the Common Elements of a Commercial Website by Semantic Ontology	183
<i>Tirayu Songvetkasem</i>	

Computational Techniques in Political Language Processing: AnaDiP-2011	188
<i>Daniela Gîfu and Dan Cristea</i>	
An Efficient Approach for Video Action Classification Based on 3D Zernike Moments	196
<i>I. Lassoued, E. Zagrouba, and Y. Chahir</i>	
Visual Information Based Argument Categorization for Semantics of Chinese Verb	206
<i>Haipeng Liu, Xiaojie Wang, and Yixin Zhong</i>	

Information Science and Technology

A Novel Remote Detection Method of Illegal Electricity Usage Based on Smart Resistance.....	214
<i>Byambasuren Bat-Erdene, Sang-Yep Nam, and Dong-Han Kim</i>	
Applied Method for One Source Multi Use (OSMU) in the Broadcasting Communication Convergence Environment	224
<i>Hyunmi Jang, Sengphil Hong, Kyongjin Kim, and Jae-Jung Kim</i>	
Finger Gesture Estimation for Mobile Device User Interface Using a Rear-Facing Camera.....	230
<i>Jun-Ho An, Jin-Hong Min, and Kwang-Seok Hong</i>	
An Adaptive Parameter Estimation Method for Wireless Localization Using RSSI Measurements	238
<i>Jinhyung Park, Hyunhun Cho, Seunghae Kim, Dowoo Park, Ahsol Kim, and Joongoo Park</i>	
Sidelobe Suppression Methods of Cubic-Phase Linear Chirp in Two-Ray Multipath Model for UWB Ranging	245
<i>Jinhyung Park, Hyunhun Cho, Seunghae Kim, Changbeom Lim, and Joongoo Park</i>	
Device Collaboration System Based on Context-Aware Architecture in Smart Environment	256
<i>Sook-Youn Kwon, Chan-Yung Choi, and Jae-Hyun Lim</i>	
Developing English Learning Contents for Mobile Smart Devices	264
<i>Seongwon Park, Kwangeak Kim, and Bong Gyou Lee</i>	
Pre-Test Analysis for First Experiences of Korean E-Voting Services....	272
<i>Ki Youn Kim, Dae Jung Kim, and Bong Gyou Lee</i>	
Analyzing User Satisfaction Factors for Instant Messenger-Based Mobile SNS	280
<i>Seongwon Park, Duckshin Oh, and Bong Gyou Lee</i>	

Analysis of the RPS System in Korea Based on SCP Framework	288
<i>Giseob Byun, Taisiya Kim, Ki Youn Kim, and Bong Gyou Lee</i>	
A Design of Short-Term Load Forecasting Structure Based on ARIMA Using Load Pattern Classification	296
<i>Kyoung-Mi Im and Jae-Hyun Lim</i>	
Comparative Analysis of Learning Effect on Lexical Recognition in the e-Learning and s-Learning	304
<i>JaeChoon Jo and HeuiSeok Lim</i>	
Predicting Text Entry for Brain-Computer Interface	309
<i>Saebyeok Lee and Heui-Seok Lim</i>	
A Safe Driving Support System for Motor Vehicles with DSRC	313
<i>Myung Hwa Hyun, Jisu Park, Jingon Shon, and Kwang Sik Chung</i>	
Changes and Directions of Mobile Technology and Infrastructure to Enhancing Public Education: Policy and Strategy Perspectives in South Korea	321
<i>JiYoung Choi, Hye Jeong Kim, and WonGyu Lee</i>	
A Cognitive Ability Measuring System with Cognitive Loading Task	325
<i>YeongWook Yang and HeuiSeok Lim</i>	
Automatic Music Emotion Classification Using Chords and BPM	329
<i>Sang-yong Park, Han-moi Sim, Mi-kyung Kwon, and Won-hyung Lee</i>	
The Real-Time Stereoscopic 3D Image Conversion of Photo and Video Contents	337
<i>Jun-Eui Lee, Sang-won Lee and Won-Hyung Lee</i>	
The Changes in Performance during Stress-Inducing Cognitive Task: Focusing on Processing Difficulty	345
<i>Jinyoung Song, Taewon Chung, Jinwon Kang, and Kichun Nam</i>	
The Changes in Performance during Cognitive Load Inducing Tasks	348
<i>Heejin Park, Yoonkyung Choe, Yoonhye Na, and Kichun Nam</i>	
Emotion Based Gesture Animation for Mobile Phone	352
<i>Haewon Byun</i>	
Interaction for Interactive Popup Book	357
<i>Haewon Byun</i>	
The Mobile ID Based Biometric Undeniable Signature Scheme	360
<i>SungHyun Yun</i>	
Automatic Metadata Conversion System with Ontology	367
<i>Min-Ho Lee, Hwa-Mook Yoon, Won-Kyung Sung, and Jae-Cheol Ryou</i>	

A Research Model for Evaluating the Success of the DCMS: A Digital Content Management System	369
<i>Sung-Ho Shin, Wha-Mook Yoon, and Won-Kyung Sung</i>	
User's Location Prediction System Using the Filtering with Correlation Coefficients Weight	371
<i>Kyoung-Mi Im, Chan-Yung Choi, and Jae-Hyun Lim</i>	
An Efficient String Searching Algorithm Based on Vowel Occurrence Pattern	379
<i>Kwang Sik Chung, Heon-Chang Yu, and Sung Ho Jin</i>	
A Flexible Cost-Based Privacy-Aware Data Integration System in Cloud	387
<i>Yuan Tian, Biao Song, and Eui-Nam Huh</i>	
An Applicable Model Enabling IPTV Service Delivery over Virtual Network	397
<i>Biao Song, Mohammad Mehedi Hassan, and Eui-Nam Huh</i>	
Virtualization of Access Network for Multihomed Wireless Thin Client	407
<i>Md. Motaharul Islam, Jun-Hyung Lee, and Eui-Nam Huh</i>	
New Optimization Techniques for Shear-Warp Volume Rendering	415
<i>Ki-Young Choi, Sung-Up Jo, and Chang-Sung Jeong</i>	
Author Index	423

Whom to Convince? It Really Matters in BGP Prefix Hijacking Attack and Defense

Yujing Liu, Bin Dai, Peidong Zhu, and Jinshu Su

School of Computer, National University of Defense Technology, Changsha 410073, China
liuyujing@nudt.edu.cn

Abstract. BGP prefix hijacking is one serious security threat to the Internet. In a hijacking attack, the attacker tries to convince as many ASes as possible to become infectors for redirecting data traffic to him instead of the victim. It is important to understand why the impact degree of prefix hijacking differs a lot in different attacks. In this paper, we present a trust propagation model to understand how ASes choose and propagate routes in the Internet; define AS Criticality to describe the ability of an AS for transmitting routing information; and evaluate impact of prefix hijacking attacks based on this metric. From the results of a large amount of simulations and analysis of real prefix hijacking incidents that occurred in the Internet, we find that only a few ASes have very high AS Criticality, and numerous ASes have very low Criticality. There is a tight relationship between the impact of attacks and the Criticality of infectors. For prefix hijacking attack, it is impactful to convince the most critical ASes to trust the false route forged by the attacker. And for prefix hijacking defense, it is effective to convince the most critical ASes to stick to the origin route announced by the victim.

Keywords: BGP prefix hijacking; Impact evaluation; Tier-1 AS; AS Criticality.

1 Introduction

The Internet consists of more than thirty thousand ASes (Autonomous Systems) nowadays. They communicate with each other using inter-domain routing protocol. The Border Gateway Protocol (BGP) is the only de facto inter-domain routing protocol in today's Internet. But because of its lack of security mechanism, the inter-domain routing system becomes vulnerable to a variety of malicious attacks. BGP prefix hijacking is one sort of them. There were many prefix hijacking incidents occurred in the Internet, which caused a large scale of outages in data reachability [1-3]. In a hijacking attack, the attacker tries to convince ASes to become infectors for redirecting data traffic to him instead of the victim. The more infectors there are, the larger impact an attack has.

On January 22, 2006, AS27506 wrongly announced the IP prefix 204.13.72.0/24 which belongs to AS33584 into the global routing system. By analyzing routing tables collected by Route Views Project [4], we found that there were 79.5% of the total recorded ASes, which believe the attacker, changing their routes for AS33584 into

hijacking routes toward AS27506. At the same time, AS27506 also announced the IP prefix 65.164.53.0/24, which belongs to AS20282. But there were only 17.5% of the ASes, which became infectors and changed their origin routes into hijacking routes. It is interesting and important to understand why the number of poisoned ASes differs a lot in different prefix hijacking attacks. This knowledge can lead us to a better understanding of how impactful attacks occur and how to improve network's resilience by applying impact-aware defense.

In this paper, we construct a trust propagation model to understand how ASes determine to choose the origin route or the forged one, and how the trust propagates in the Internet. From this process, our observation is that not every AS plays an equivalent role in inter-domain routing system. So we define the *Criticality* of an AS to measure how much routing information the AS is responsible for transmitting. From the estimation results, we find that only a few ASes have very high AS Criticality, and numerous ASes have very low Criticality. And all the critical ones with the highest Criticality values belong to Tier-1 AS set of the Internet. Based on this result, we evaluate the impact of prefix hijacking attacks by simulating ten thousand attack scenarios. The result shows that the attacker who convinces more Tier-1 ASes to trust the hijacking route will launch a more impactful attack. On the other side, deploying defense filters on the critical ASes first to prevent bogus routes from spreading is more efficient to gain high resilience of the network against BGP prefix hijacking attacks.

2 Background

BGP is a policy-based routing protocol. It controls the propagation of routing information among ASes by applying routing policies that are set locally according to business relationships. There are three major types of business relationships between distinct ASes: provider to customer, customer to provider and peer to peer. The route selection process depends on the export policies of upstream AS and the import policies of downstream AS. According to the export policies, an AS usually does not transmit traffic between any of its providers or peers, which is called Valley-free property [5]. According to the import policies, an AS applies a priority to every route learnt from its neighbors. If a BGP router receives routes to the same destination from different neighbors, it prefers route from customer over those from peer and provider; and it prefers route from peer over that from provider [5]. Metrics such as path length and other BGP attributes are used in route selection if the preference is the same for different routes.

Because of the lack of security mechanism, every BGP router has to believe the announcements received from other ones, no matter whether the message is creditable. As a normal situation before an attack, a legal AS announces its IP prefix to the Internet. The other ASes who have learned this origin route will send data traffic to the origin AS in the future. During the prefix hijacking attack, the attacker announces IP prefix which belongs to the victim network. Such bogus hijacking route propagates on the Internet, too. The ASes who choose to believe the forged route become infectors. Data traffic from those polluted ASes will be redirected to the attacker instead of the victim.

3 Related Work

Previous efforts on defending against BGP prefix hijacking can be sorted into three categories: preventions before the attack, detections during the attack and reactions after the attack. The impact evaluation of prefix hijacking is orthogonal to all the existing researches in the area [6, 7]. The result of impact evaluation can provide evidences for network operators that how and where to employ incremental deployment in the Internet. According to [6], ASes higher up in the routing hierarchy can hijack a significant amount of traffic to any prefix. In [7], researches show that direct customers of multiple tier-1 ASes are the most resilient networks against prefix hijacking, meanwhile, the most effective launching pads for attacks. The outcomes of the related work show the tight relation between AS topology and the spread scope of an illegal route. But there is still a significant lack of reasonable explanations why they are related and what guidelines it offers for defense against prefix hijacking.

Our paper exceeds the analysis of superficialities, figures out the root cause why the impact of prefix hijacking differs a lot when the attack is launched in a different topological location. Then, we provide a impact-aware defense policy against the attack, which can improve resilience of the Internet in an efficient way.

4 AS Criticality Based Evaluation of Prefix Hijacking Impact

BGP prefix hijacking is due to the unconditional trust between ASes. The impact is caused by the propagation of bogus routing information among ASes. In this section, we build a trust propagation model to study the flow of routing information in the Internet, measure the grade of AS that how crucial it is for transmitting routing information, and evaluate impact of prefix hijacking based on this AS Criticality.

4.1 Trust Propagation Model

In BGP, an AS innocently believes what neighbors have announced. This trust will propagate among ASes governed by routing policies. As mentioned in Section 2, this process can be formally described as (1).

$$r_{s,d} = Import_{s,d} \left(\bigcup_{n \in Neighbor(s)} Export_{n,d} \right) \quad (1)$$

$$Export_{n,d} = \begin{cases} Append(r_{n,d}, n) & \text{if the appended path is valid} \\ \emptyset & \text{if the appended path is not valid} \end{cases} \quad (2)$$

$$Import_{s,d}(R) = Prefer(r_i, r_{i+1}), \text{ for } 1 \leq i \leq |R|-1, r_i \in R \quad (3)$$

$$Prefer(r_i, r_j) = \begin{cases} r_i, & r_i.lp > r_j.lp \text{ or } (r_i.lp = r_j.lp \text{ and } |r_i.ap| < |r_j.ap|) \\ r_j, & r_i.lp < r_j.lp \text{ or } (r_i.lp = r_j.lp \text{ and } |r_i.ap| \geq |r_j.ap|) \end{cases} \quad (4)$$

In order to send data traffic to destination AS d , an AS s must choose a route $r_{s,d}$ to import from those that its neighbors export based on import policies $Import_{s,d}(R)$ and export policies $Export_{n,d}$, where n is one of the neighbors of s . According to import policies, an AS who has received multiple routes picks the best route from the candidate set R based on local preference $r.lp$ and AS path length $lr.apl$. The selected route was appended with the AS's number, represented as $Append(r_{n,d}, n)$ and exported to the AS's neighbors only if it satisfies Valley-free property. The routes to the announcing AS propagate in the inter-domain routing system in this way.

When the attacker a wrongly announces victim v 's IP prefix, whether an AS i becomes infected depends on which route it selects to send traffic, i.e., the origin route $r_{i,v}$ or the hijacking route $r_{i,a}$. The decision is made following the same routing process without checking the origin AS from which the route is announced. However, we can figure out the infected status by detecting which route AS i has chosen, like (5). In BGP routing table, if an AS becomes an infector, the origin AS of its AS path changes from victim to attacker, which is called a Multiple-Origin-AS conflict. Furthermore, we can define the impact of a prefix hijacking attack $I(a, v)$ as the percentage of ASes who has become infectors in the AS set $AsSpace$ which consists of all ASes in the Internet, like (6). $I(a, v)$ is considered to be an appropriate metric to describe the damage impact of an attack.

$$Infect(a, i, v) = \begin{cases} 1, & \text{if } r_{i,v} = r_{i,a} \\ 0, & \text{if } r_{i,v} \neq r_{i,a} \end{cases} \quad (5)$$

$$I(a, v) = \frac{1}{|AsSpace|} \sum_{i \in AsSpace} Infect(a, i, v) \quad (6)$$

4.2 AS Criticality

If an AS appears along the AS path from the source AS to the destination AS, then the traffic from the source to the destination will go through that AS. The more times a transit AS appears in different AS paths, the more traffic it is responsible for transmitting, the more critical it is in the inter-domain routing system.

According to this evidence, we define the *Criticality* of an AS as the fraction of AS paths from any source to any destination which contain it as a transit AS to describe how critical the AS is. Furthermore, we define Criticality of an AS set as the fraction of AS paths from any source to any destination which contain at least one AS in this set as a transit AS likewise. Formally, for every AS m in $AsSpace$, the Criticality of m - $C(m)$ is defined as (7). And for AS set M , the Criticality of it - $C(M)$ is defined as (8). In these functions, $\chi_{r_{s,d}.ap-\{d\}}(m)$ equals to 1 if m is one of the transit ASes in the AS path from source s to destination d , which is represented as $r_{s,d}.ap - \{d\}$. Otherwise, it equals to 0.

$$C(m) = \frac{1}{|AsSpace|^2} \sum_{s,d \in AsSpace} \chi_{r_{s,d}.ap-\{d\}}(m) \quad (7)$$

$$C(M) = \frac{1}{|AsSpace|^2} \sum_{s,d \in AsSpace} \min \left\{ 1, \sum_{m \in M} \chi_{r_{s,d}.ap-\{d\}}(m) \right\} \quad (8)$$

$$\chi_{r_{s,d}.ap-\{d\}}(m) = \begin{cases} 1, & \text{if } m \in r_{s,d}.ap - \{d\} \\ 0, & \text{if } m \notin r_{s,d}.ap - \{d\} \end{cases} \quad (9)$$

In order to evaluate the Criticality of every AS accurately, we use the AS Relationships data in CAIDA [8] as a data source and simulate the BGP route selection process to infer all the AS paths from any source to any destination in the Internet. The evaluated result of AS Criticality of every AS in the Internet is shown as Fig. 1.

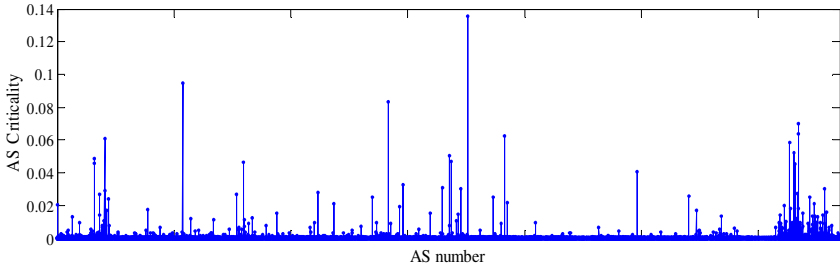


Fig. 1. Evaluation of AS Criticality

From Figure 1, we observe that only a few ASes have very high AS Criticality, while most ASes have very low AS Criticality. Furthermore, we find that those ASes which have high AS Criticality belong to the Tier-1 AS set^[9]. These 13 Tier-1 ASes are higher up in the routing hierarchy and locate on the dense core of the Internet. They peer with each other and have no providers. We consider those ASes as one set *Tier1* and measure the Criticality of this set. The result of $C(Tier1)$ is as high as 0.827, which means 82.7% of the total AS paths from any source to any destination traverse at least one Tier-1 AS. This indicates that the Tier-1 ASes play the most critical role in inter-domain routing system, because they transit a large fraction of BGP routing information and data traffic in the Internet.

We believe this phenomenon is caused by the routing policy configurations in BGP, as well as the topology of the Internet. We call this phenomenon *hinge-transmit* property of BGP routing. Next, we validate this property by BGP routing tables collected by Route Views project. We choose the snapshots of routing tables in 10 different times randomly and calculate the fraction of AS paths containing Tier-1 ASes. The validation results are shown in Fig. 2. The fraction of AS paths with hinge-transmit property are higher than 75% in all 10 cases.

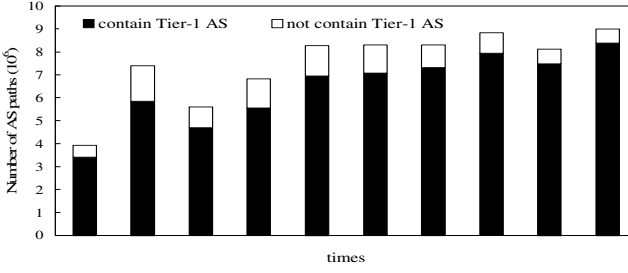


Fig. 2. Fraction of AS paths with hinge-transmit property

4.3 AS Criticality Based Evaluation

Because ASes in the Internet play such a different role in transmitting routing information, there must be significantly different impact of convincing different ASes to be infectors. As a transit AS, being infected means that it not only wrongly sends data traffic to the attacker itself, but also propagates this mistaken route to others. This ability of trust propagation is presented as Criticality of AS in this paper. So it is of great significance to evaluate the impact of prefix hijacking based on AS Criticality. We perform 10,000 simulations of BGP prefix hijacking attacks based on the AS topology inferred by CAIDA on January 2010. Victims and attackers are chosen from different topology locations considering different impacts of the attacks [7]. The distribution of impact is shown in Fig. 3.

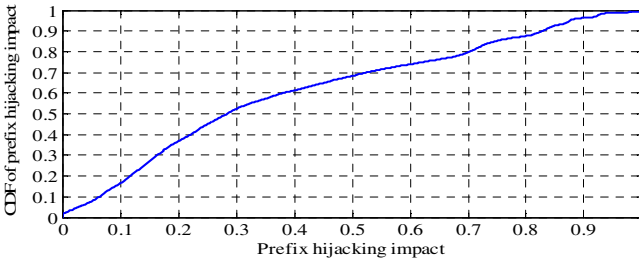


Fig. 3. Distribution of prefix hijacking impact

The uniform-like distribution curve indicates that there are many levels of impactful prefix hijacking attacks in these simulations and seems to be numerous factors affecting the results, such as the topological location or the routing hierarchy of victim and attacker. But our analysis shows a novel yet profound insight about this problem, which is the tight relationship between the impact of attacks and the Criticality of infectors. Since the Criticality of Tier-1 AS set is as high as 0.827, it is rational to consider Tier-1 ASes as the major dominant factor of amplifying the infected scope. Fig. 4 shows the approximate linear relationship between infected Tier-1 ASes and mean value of prefix hijacking impact in our simulations, which offers evidences for supporting our AS Criticality based evaluation method.

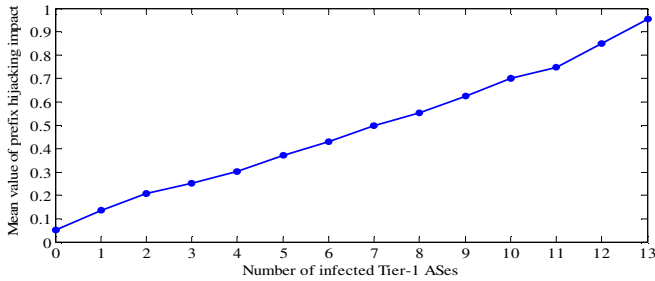


Fig. 4. Relationship between infected Tier-1 ASes and prefix hijacking impact

To sum up, if ASes with high Criticality are infected, due to wildly propagation of hijacking route in the network, the impact of this attack will be amplified a lot. In the Internet, Tier-1 ASes are the most critical ones. The attacker who convinces more Tier-1 ASes to trust the hijacking route will launch a more influential attack. We believe that this is the root cause of the matter why the impact of prefix hijacking differs a lot in different prefix hijacking attacks.

5 Better Understanding of Prefix Hijacking Attacks

There were some real prefix hijack incidents occurred in the Internet. Their impacts were recorded by the infectors' routing tables as MOAS conflicts. In this section, we verify our findings by analyzing BGP routing tables collected by Route Views, leading us to a better understanding of prefix hijacking attacks as we mentioned at the beginning of this paper.

On January 22, 2006, AS27506 hijacked 204.13.72.0/24, which belongs to AS33584. As shown in Fig. 5(a), Tier-1 AS701 and AS2914 were infected by the hijacking route first due to the shorter AS path length. Then this trust propagated along peer sessions among Tier-1 AS set, causing 9 of the 13 Tier-1 ASes to be infected according to the records. Most compromised Tier-1 ASes propagated this hijacking route wildly in the whole network, launching a prefix hijacking attack with impact as high as 0.795. At the same time, the IP prefix 65.164.53.0/24, which belongs to AS20282, was hijacked by AS27506, too. As shown in Fig. 5(b), Tier-1 AS2914 was infected because the hijacking route was announced by a customer, beating the original peer route learnt from other Tier-1 ASes. But the direct providers of victim preferred the origin route over the hijacking one according to the local preference, and other Tier-1 ASes tended to believe the origin one when they were of the same preference value and AS path length. There were only 2 of the 13 Tier-1 ASes became infectors. Consequently, the impact of this attack is as low as 0.175.

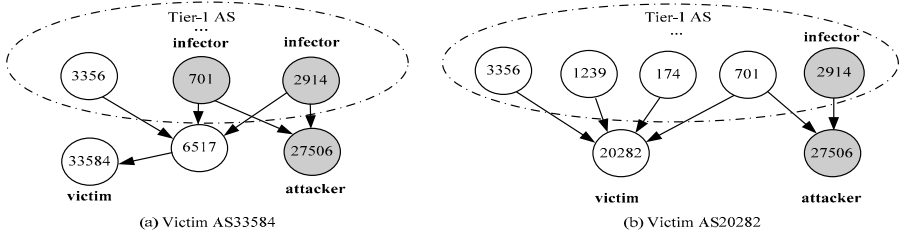


Fig. 5. Local topology of incidents

We randomly pick up 10 prefix hijacking incidents in the past and calculate the results, shown in Table 1. Surprisingly, we find that no matter whom the victim and attacker are or where they locate, we can simply estimate the impact of an incident approximately by calculating the number of infected Tier-1 ASes in the attack.

Table 1. 10 prefix hijacking incidents in the Internet’s history

id	attacker	victim	Number of infected Tier-1 ASes	impact
1	AS27506	AS20282	2	0.175
2	AS27506	AS33584	9	0.795
3	AS27506	AS7169	2	0.104
4	AS27506	AS33477	4	0.158
5	AS9121	AS19198	7	0.628
6	AS9121	AS30576	5	0.256
7	AS9121	AS31846	4	0.152
8	AS9121	AS30594	2	0.023
9	AS9121	AS668	2	0.096
10	AS9121	AS19281	3	0.119

6 Guidelines for Impact-Aware Defense against Prefix Hijacking

Currently, the efficient and practical approach to defend prefix hijacking may be the deployment of filters for hijacking route on the edge of the Internet. If a stub AS announces a path to any IP prefix that it does not own, the provider drops the announcement and stops the propagation of this hijacking route. Defensive filtering completely eliminates attacks by stubs, but does not prevent attacks launched by non-stubs, which is proved to be more impactful^[10]. From the result of prefix hijacking impact after we deploy filters on providers of all stubs in our simulations. We can see more than 80% of impact of attacks, which are launched by stub ASes, vanishes, but those launched by non-stubs are not affected.

In section 4, we show the fact that AS Criticality is the metric measuring the ability of an AS propagating routing information. It is interesting to know whether it is a more effective way to deploy filters based on AS Criticality. To this end, we deploy routing filters on Tier-1 ASes and Top 20 to Top 300 ASes ordered by their Criticality respectively. From the aspect of topology locations, these critical ones lay on the core

of the Internet. Fig. 6 shows the distribution of impact under these defenses, presenting their different effects. For instance, although defense on Tier-1 ASes can not totally eliminate attacks as much as defense above stub ASes, it mitigates attacks with impact more than 0.5 significantly. Similarly, other cases show the same effect on reducing more impactful attacks. The more deployed ASes there are, the less impact there remains in simulations. Especially, when the Top 300 ASes are protected from the hijacking routes, the impact of all the attacks are almost eliminated. It is an impact-aware defense policy by applying filtering mechanism on the core ASes.

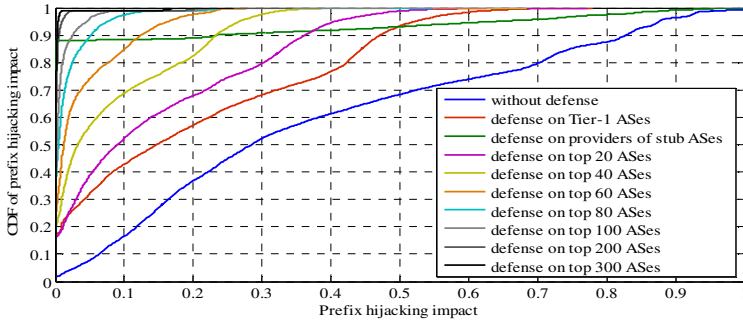


Fig. 6. Distribution of prefix hijacking impact under impact-aware defense

Fig. 7 presents the comparison of deploying cost and effect between defenses on core ASes and edge ASes. From the figure we can see that when the average impact is the same, the deploying scope of edge-defense is thousands of times of that of core-defense. Besides, the core-defense policy can mitigate the average impact of attacks from 1 to nearly 0 sharply if it is only deployed on 1% of the ASes all over the Internet. It is an efficient way to incrementally deploy impact-aware defense on core ASes based on their Criticalities to achieve high resilience of the whole network.

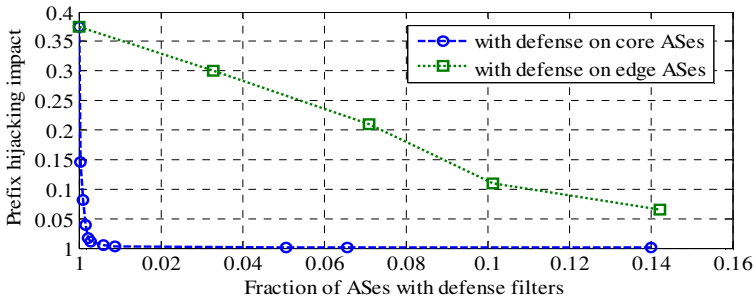


Fig. 7. Comparison between defenses on core ASes and edge ASes

7 Conclusion

In this paper, we commit ourselves to find the reason why different prefix hijacking attacks cause different impact. By analyzing the propagating behavior of routing information in the Internet, we find that it really matters that who is convinced as an infector. For prefix hijacking attack, it is impactful to convince the most critical ASes to trust hijacking route which is launched by attacker. On the other hand, for prefix hijacking defense, it is effective to convince the most critical ASes to stick to the origin route which is announced by victim. The most critical ASes are higher up in the routing hierarchy and located on the dense core of the Internet. The findings of our paper explain why there is a tight relation between the impact of prefix hijacking and AS topology. Such a deep understanding makes it easier to evaluate the overall damage once an attack occurs, and to provide guidance on how to improve the Internet's resilience by applying impact-aware defense against attacks.

Acknowledgement

This research is supported by the National High-Tech Research and Development Plan of China, under Grant No. 2009AA01Z432 and Grant No. 2009AA01A346; the National Natural Science Foundation of China, under Grant No. 61070199 and Grant No. 61003303; the aid program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province.

References

1. Popescu, A.C., Premore, B.J., Underwood, T.: Anatomy of a leak: AS9121. Renesys Corp. (2004)
2. Pakistan hijacks YouTube, http://www.renesys.com/blog/2008/02/pakistan_hijacks_youtube_1.shtml
3. Routing Instability. NANOG mail archives, <http://www.merit.edu/mail.archives/nanog/2001-04/msg00209.html>
4. Route Views Project Page, <http://www.route-views.org>
5. Gao, L.X.: On inferring autonomous system relationships in the Internet. *IEEE-ACM Transactions on Networking* 9, 733–745 (2001)
6. Ballani, H., Francis, P., Zhang, X.: A study of prefix hijacking and interception in the Internet. *ACM SIGCOMM Computer Communication Review* 37, 265–276 (2007)
7. Lad, M., Oliveira, R., Zhang, B., Zhang, L.: Understanding resiliency of Internet topology against prefix hijack attacks. In: *Proc. of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, Edinburgh, Scotland (2007)
8. CAIDA AS Relationships, <http://www.caida.org/data/active/as-relationships/>
9. Tier 1 network - Wikipedia entry, 1 network, <http://en.wikipedia.org/wiki/Tier>
10. Goldberg, S., Schapira, M., Hummon, P., Rexford, J.: How Secure are Secure Interdomain Routing Protocols? In: *Proc. of ACM SIGCOMM 2010*, New Delhi, India (2010)

Future Trends of Intelligent Decision Support Systems and Models

Andrzej M.J. Skulimowski^{1,2}

¹ Decision Science Laboratory, Chair of Automatic Control,
AGH University of Science and Technology, Al. Mickiewicza 30, 30-050 Kraków, Poland

² International Centre for Decision Sciences and Forecasting,
Progress & Business Foundation, Miechowska 5B, 30-041 Kraków, Poland
ams@agh.edu.pl

Abstract. The aim of this paper is to investigate, formulate, and analyse the general rules and principles that govern the evolution of key factors that influence the development of decision support systems (DSS) and models. In order to elaborate a model suitable for medium-term forecasts and recommendations, we have defined eight major elements of Information Society that characterise the evolution of the corresponding digital economy. The evolution of the overall system is described by a discrete-continuous-event system, where the mutual impacts of each of the elements are represented within state-space models. Technological trends and external economic decisions form inputs, while feedback loops allow us to model the influence of technological demand on IT, R&D, production, and supply of DSS. The technological characteristics of the product line evolution modelled in this way can provide clues to software providers about future demand. They can also give R&D and educational institutions some idea on the most likely directions of development and demand for IT professionals. As an example, we will model the evolution of decision-support systems and recommenders for 3D-internet-based e-commerce, and their impact on technological progress, consumption patterns and social behaviour. The results presented here have been obtained during an IS/IT foresight project carried out in Poland since 2010 and financed by the ERDF.

Keywords: Information Society Technologies, Foresight, Quantitative Trend Analysis, Scenarios, Discrete-Event-Systems, Decision-Support-Systems.

1 Introduction and Objectives

The rapid technology-driven evolution of the information society (IS) has raised a number of research, technological, economical, political, and social challenges. The IS development laboratory that has enabled us discover new aspects of IS modelling was the EU accession of the twelve New Member States in 2004 and 2007. The accelerated implementation of EU policies concerning the IS in these countries, part of the Community's cohesion policy priority goals, has allowed us to make a number of methodological observations on the modelling, foresight, and stimulating the development of the IS. This, in turn, has made it possible to draw more general conclusions

regarding the evolution of the IS in Europe and its future scenarios as well as to study the future development of selected information technologies (IT) and application areas. Some of these findings were first described in a series of reports [4] prepared for the FISTERA (Foresight of the Information Society in the European Research Era) project – a thematic network of twenty organisations led by the Institute for Prospective Technological Studies (IPTS) – DG JRC in Seville.

Other recent EU projects devoted to the investigation of the development of the Information Society include SEAMATE (Socio-Economic Analysis and Macromodelling of Adapting to information Technologies in Europe [13]) and ISCOM [5]. New trends, processes, and phenomena concerning the current and future state of the IS have been observed, and several case studies can be referred to in [4,12]. The progress achieved in [12], compared to earlier work on IS models, cf. [1,6,7,13], lies in the appropriate use of statistical data concerning IS indicators. The research projects mentioned above have shown i.a. that the sole use of classical econometric methods and narrative foresight descriptions have proved to be insufficient to get industry-oriented IT foresight results. Another observation, relevant to the scope and results of this paper, is that any individual IT is embedded in a complex technological, economic and social system in such a way that its evolution cannot be explained without investigating this system in a holistic way. This touches upon as well an important class of applications, the decision support systems (DSS), which has been the subject of prospective studies presented in this paper. This is why, despite the scarcity of published research results on the future trends of DSS, (cf. e.g. specific-application-oriented papers [8,7] or the bibliometric trend analysis in [3]), we can benefit from the general methods and results derived within the above mentioned projects for the IT – from the providers' point of view, and IS – from the users' one.

To sum up, the methods presented here as a background to elicit trends and elaborate scenarios of decision-support and decision-making systems can constitute an input to any IS/IT foresight exercise. We will also present the conclusions regarding the development of a cluster of new classes of e-commerce systems, coupled with recommendation engines, that have been elaborated within an IT foresight project [8]. The study of differentiated factor interactions, modelled in a different way, led to the application of modern modelling methods such as discrete-event-systems, multicriteria analysis, and discrete-time control. The results should be constructively applied to developing technological policies and strategies at different levels, from corporate to multinational. Trends and scenarios thus generated can be used to better understand the role of global Information Society Technologies (IST) development trends and to develop IS and IT policies in an optimal control framework.

To achieve the ultimate objective described above a few intermediate research goals have been defined. These include:

- An information system, consisting of an ontological knowledge base which stores raw data together with the appropriate technological models, trends and scenarios in the form of so-called proceedings containing data together with records of their step-by-step analyses and results. This information can be used as input to multicriteria rankings methods suitable for IT management and capable of generating recommendations for decision-makers as regards the prioritization of IT investments.

- A detailed analysis of several real-life industrial applications of decision support systems in selected technological areas submitted by industrial partners cooperating on the implementation of the project's results, including an analysis of technological trends and scenarios in areas such as virtual-reality-based e-commerce, expert systems, recommenders, m-health, neurocognitive technologies.

The achievement of the above goals should provide useful solutions to the technology management problems presented by the companies involved in the project, allowing them to apply the knowledge gained to set strategic technological priorities and formulate IT and R&D investment strategies. This is discussed further in Sec. 3.

2 Modelling Methodology

One of the research questions that could be posed by a user of a technological knowledge base investigated here could read as follows: *how the development of selected information technology depends on the global IT development processes and on integration of the IS around the world, driven by the global trends*. We investigate this question in more detail in the next Sec. for the case of decision-support methods and technologies. As regards the global environment, various factors must be considered such as falling telecommunication prices, exchange of information through the internet, rapid diffusion of information on innovations and technologies, the development of e-commerce, and free access to web information sources. The civil society evolution, driven by the growing availability of e-government services and related web content, has been taken into consideration as well. Finally, the psychological and social evolution of IT users, including all positive and negative i-inclusion phenomena has been taken into account as a set of feedback factors influencing the legal and political environment of the IS.

Due to the complex nature of the decision-support technologies, that rely strongly on the cognitive and social phenomena, it is difficult to create a technology evolution model that is clear, unambiguous and concise. One of the aforementioned FISTERA project's findings [12] was that the composite indicators based on user data rarely provide an adequate description of the technology parameter dynamics. Therefore when performing the research described in this paper, it was decided that the use of aggregates as the basis of forecasts and recommendations should be avoided. Instead, we have introduced a new class of input-output models that fit well into the specificity of this kind of technologies. In particular, we analyze different groups of potential users separately that might eventually explain the development of separate product lines. To cope with the high level of system's interconnectedness, we have defined eight major elements of an IS, such as population and its demographics, legal system and IS policies, ITs in personal and industrial use, etc. (cf. Fig.1) that can influence the technological evolution. The relations between different groups of users of DSS are described at this level only.

These elements correspond to the subsystems of the IS, and are related to the IST development trends evidenced in the past that are supposed to be able to effectively characterise their evolution. During analysis, each appears as a bundle of discrete events, continuous trends and continuous or discretised state variables. The evolution of the IST is then modelled as a discrete-continuous-event system, where the mutual

impacts of each of the elements are represented either in symbolic form, as generalised influence diagrams, or within state-space models. Some external controls, such as legal regulations and policies, are modelled as discrete-event controls, while the others, such as tax parameters or the central bank's interest rates are included in the discretised part of the model. Other exogenous (non-controlled) variables include exchange rates, energy prices, demographic structure, attitude towards IT-related learning and so on. Technological trends and external economic decisions form inputs, while feedback loops allow us to model the influence of technological demand on IT, R&D, production and supply of selected information technology or its products, as well as on overall GDP growth rates. Another lesson from the past trends is the model of adaptation of new versions of software to the progress in the development of processors, storage and peripheral devices. The technological characteristics of the product line evolution modelled in this way can provide clues to software providers about future demand. They can also give R&D and educational institutions hints on the most likely directions of development and demand for IT professionals.

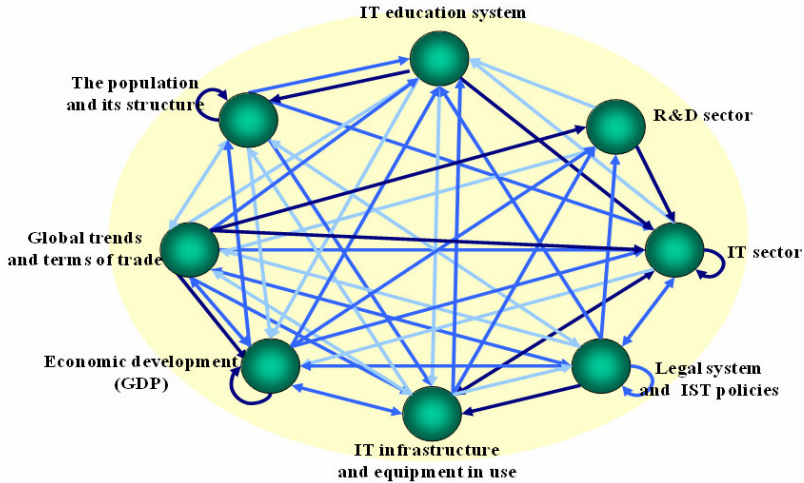


Fig. 1. A causal graph linking the major subsystems that can influence the development of an IT area: dark arrows denote strong direct dependence, medium once indicate average relevance of causal dependence, and light grey arrows denote weak direct dependence between subsystems

A causal graph of the underlying dynamical model derived from an expert Delphi [9] is presented in Fig.1 above. Only direct impacts, i.e. those which show immediately or within one modelling step are marked. The indirect impact may be obtained by multiplying the coincidence matrix associated with the impact graph by itself.

Let us recall that a *discrete-event system* can be described as a 5-tuple [11]:

$$P=(Q,V,\delta,Q_0,Q_m) \quad (1)$$

where Q – is the set of system states, V – the set of admissible operations, $\delta: V \times Q \rightarrow Q$ – the transition function governing the results of operations over states, Q_0 – the set of (potential) initial states of the process, Q_m – the set of final reference states. A pair of states $e := (q_1, q_2)$, such that $q_2 = \delta(v, q_1)$ will be termed *an event*. The set of all admissible events in the system (1) will be denoted by E . Following the above assumptions concerning the controlled discrete-event variables, the operations from V may be either controls, i.e. the decision-maker's actions over Q , or may occur spontaneously as the results of random processes. Further, we assume that there exists a set $X(Q)$ of quantitative or ordinal characteristics of states from Q , which can be deterministic, interval, stochastic, fuzzy etc. One of the coordinates of G can be (but does not need to be) identified with time. An *elementary scenario* s is a sequence of events $(e_1, \dots, e_p)$, such that if $e_i = (q_i, q_{i+1})$ then $e_{i+1} = (q_{i+1}, q_{i+1+2})$. In order to conform with the usual definition of scenarios, we define a *foresight scenario* as a cluster of elementary scenarios, where clustering is based on certain similarity rules applied to events.

After scaling the dynamics based on past observations, key technological, economic or social characteristic trends can be described quantitatively as solutions to discrete-time dynamical systems of the form

$$x_{t+1} = f(x_t, \dots, x_{t-k}, u_t, \dots, u_m, \eta_t, \dots, \eta_n) \quad (2)$$

where $x_t, \dots, x_{t-k}$ are state variables, $x_j = (x_{j1}, \dots, x_{jN}) \in \mathbb{R}^N$, $u_t, \dots, u_m$ are controls, and $\eta_t, \dots, \eta_n$ are external non-controllable or random variables. In the models analysed in [8], f has been linear non-stationary with respect to x , and stationary with respect to u and η .

Discrete-event and discrete-time control systems may jointly govern the evolution of causal systems, thus providing a tool to elicit trends and construct elementary scenarios, which appear as trajectories to (1)-(2). Consequently, when using (1)-(2) to generate optimal technological strategies or investment policies, the goals should be quantified and associated with events and elementary scenarios. A generalisation of the multicriteria shortest-path algorithm for variable-structure networks can then be applied to the optimal control of discrete-events [11] and discrete dynamical systems.

To complement the quantitative analysis, a new method of expert assessment has been developed, namely a dynamic generalisation of the SWOTC analysis (SWOT with “Challenges” as an additional element [12]) that also includes the detection of real options emerging in the system for its principal stakeholders and allows us to characterize the IS technologies development prospects more effectively.

The overall foresight process has been organized within the framework of expert systems, built into an ontological knowledge base coupled with autonomous web-crawlers and analytical engines. The knowledge base with its analytical capabilities and automatic knowledge acquisition, update and verification, can itself be regarded as an instance of a decision support system. It includes ontology management functionality, specifically ontology merging and splitting, time evolution, operations on metadata, data updating protocols as well as other data warehousing functionalities.

The technological focus areas of the foresight project, which are also reflected in the scope of knowledge gathered and processed by the above system, are listed below:

- Basic hardware and software technologies,
- Key IS application areas (e-government, e-health, e-learning, e-commerce),
- Expert systems, including decision support systems and recommenders

- Machine vision and neurocognitive systems,
- Molecular and quantum computing.

Out of the list above, the research on first four listed focus areas has been used to elicit development trends of decision support and autonomous decision-making systems for the next 15 years. The thematic databases store the above area-specific information, while a common data block contains interdisciplinary information to be used during thematic analyses, such as macroeconomic data, social characteristics (employment, education, demographics), geographic information, and other data potentially useful in providing decision-making support.

Any query from a client can be processed by making use of a common database, containing global trends and general development models, as well as context-dependent information related to the specific area of the query. Thus the development of decision support technologies, that is described here, could be extracted from the overall knowledge base by a small set of queries.

3 Future Intelligent DSS and Their Applications

The new approaches outlined above have made it possible to elicit trends and build technological scenarios concerning the decision support systems and to visualise their evolution. In particular, we have identified the key development topics of expert systems/DSS [9], listed according to the relevance scores (first best) as follows:

- recommenders for e-commerce (excluding banking and finance)
 - graphical (content-based) recommenders for multimedia
 - graphical (content-based) recommenders for 3D-e-commerce
 - recommenders able to judge on product portfolios (constrained by its value)
- recommenders for security and commodity trading
- intelligent intermediary agents for negotiations, partner matching, collaboration.

Among the key technologies, methods and models to be used in DSS, the following appeared in most of the experts' recommendations:

- GIS technologies, able to evaluate or elicit preferences as regards individual sites in a large region, making an intensive use of visualisation, and coupled with GPS
- Cognitive features of expert systems, allowing one to eliminate the negative consequences of decisions made in a hurry, or by a nervous decision-maker etc.

Some of the salient trends concerning the future DSS (until 2025) that have been identified in [9] and relate to the 3D-e-commerce applications, are listed below:

- the role and the sophistication degree of OR-based methods applied in DSS will increase; especially multicriteria optimization, uncertainty models and management,
- the class of decision problems regarded as numerically non-tractable will shrink,
- the DSS (including and starting from recommenders) will converge with search engines and intelligent data mining agents; the latter will complete missing data that might help in solving decision problems supplied in client's queries.

Although most information provided in pilot Delphi interviews was qualitative, some of the trends elicited could be characterised quantitatively, see Tab.1 below. The trends thus obtained allow us to characterize the selected technologies, rank and position the companies, countries or regions under review in terms of development of the particular technological area as well as to give hints on how to design tailored Delphi research to address more specific issues. The future characteristics concerning the DSS market are helpful in assessing the competitiveness of DSS suppliers as well as individual products that can be accomplished during an interactive benchmarking process, using DEA (cf. e.g. [14]) or other performance measures.

Table 1. Some quantitative DSS characteristics according to an expert Delphi (average values)

Trend description	Present state	Value in 2020	Value in 2025
Penetration of the mobile DSS in OECD countries	2%	50%	80%
Seeking advice from an online medical DSS (EU)	15%	45%	70%
DSS as a component of social media	5%	60%	95%
Share of DSS using multicriteria analysis (except simple weighting methods)	35%	50%	80%

The specification of key technologies, focus areas, methods and models allowed, in turn, to focus research on trends and scenarios concerning the selected objects. The scenarios can then be used to re-examine technological evolution principles in the knowledge base, forming thus a consistent interactive and adaptive evolution model.

The main user group of the foresight exercise outputs, recommendations, and future information system described in Sec. 2 are innovative IT companies seeking technological recommendations, advice concerning R&D priorities, as well as corporation from different sectors that invest in IT. Moreover, the global trends identified and technological characteristics of the IS evolution can provide clues to policy-makers as well as R&D and educational institutions on the key directions of development, and on the demand for IT professionals.

Foresight results can allow corporations to determine an adequate level of funds to allocate for IT investment over a relatively long-term period as part of the company's overall strategic decision making. For IT, this can range between 10 and 15 years, while for related R&D it can reach a planning horizon of 30 years [9]. Foresight outcomes can situate the IT project portfolio management and fund allocation strategies within the macroeconomic, political, technological and research environment by providing recommendations, relative importance rankings, trends and scenarios. More objective and quantifiable future technological and economical characteristics will enable us to define more appropriate policy goals and measures to implement. The quantitative characteristics of the technological evolution can provide direct clues to IT providers, specifically the DSS, as regards future demand for their products.

The implementation of IT foresight results in a company may be modelled by a hierarchical multicriteria decision problem that explores the results of external (foresight-based) advice with a set of internal criteria describing the preferences of shareholders as well as the degree of achievement of the company's long-term market and investment targets. The above-mentioned problem may admit various forms, depending on the company's needs and foresight outcomes available.

A real-life example of such industrial foresight user is a Cracow-based investment fund focused on 3D and virtual reality technologies for modern e-commerce applications that takes into account IT development trends and rankings of prospective products, technologies and markets elicited during a foresight exercise. At a higher decision-making level, dynamic ranking methods are used to rank corporate development policies, which concern the sector, size, or regional preferences regarding targeted markets or portfolio structure. At the lower decision-making level, rankings are implemented as investment rules, by assigning funds to specific undertakings. Each assignment is a function of time and of external logical variables, the latter representing the changes in higher-level ranking and the states of external socio-economical (including the financial markets) situation and research environments. The foresight scenarios are used to establish future investment rankings in an adaptive way. In particular, based on the feasible scenarios found at moment t_0 , the management of the fund can calculate corresponding future rankings for $t = t_0 + 1, t_0 + 2, \dots, t_0 + k$. This makes it possible to input into fund allocation planning more knowledge coming from systematically updated foresight results in the form of future recommendations and real options than when applying a usual static approach.

According to a needs analysis performed on the companies listed on the Warsaw Stock Exchange recently, the most needed IT foresight applications relate to new product development, while foresight information on technological prospects and consumer preferences is used to minimise investment risk and increase the company's competitiveness. The structure of respondents' interests is shown in Fig.2 below.

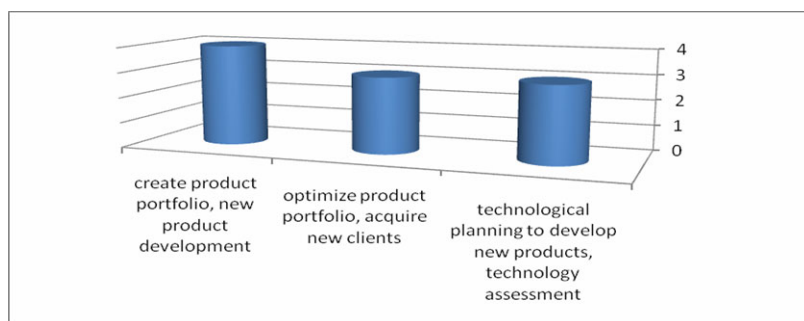


Fig. 2. The declared goals of 10 Polish IT companies seeking foresight results [9]

Apart from making the time order rational, financing IT and market expansion projects, higher-level investment policy ranking may also help to determine organisational structure and human resource characteristics, future budgetary needs, and actions to be taken when priorities change as a result of an external event.

4 Conclusions

Comparing quantitative and descriptive approaches to elicit technological trends and build scenarios, it is noticeable that the approach of extracting evolution rules prior to scenario analysis proves especially useful in case of converging information societies,

and converging technologies, such as DSS. Although the data used to derive the IS/IT trends presented in this paper origins mostly from Poland and other EU States which acceded in 2004 and 2007, the global IS/IT trends have been taken into account as well. A good coherence of forecasts and their ex-post verification five years later validates the modelling methods used first in [12] and confirms the adequacy of the general IT/IS evolution model outlined in Sec.2 for the analysis of global trends that influence the development of the digital economy in a country or region.

In particular, the recent developments of distributed, grid-, and cloud-computing-based decision support systems [2,10] indicate that after the first revolution that occurred in mid-80's, namely the migration from the mainframe-based to PC-based DSS, and after the second one, at the end of 90's, when web based DSS started to dominate and first common web recommenders were created, nowadays we face another challenging period in the development of this class of applications. It is likely to be characterized by an increased role of collective decision making tools, including social decision computing, decision grids and clouds, a growing relevance of cognitive features implemented in the DSS, advanced options allowing to express more creativity by the decision-makers, the use of sophisticated MCDM methods, virtual reality with all its attributes and a growing degree of realism, to list only a few.

The second conclusion refers to the methodology rather than to foresight results. Namely, the use of methods described in this paper shows that for the technologies changing as rapidly as DSS one can expect rational foresight results in form of trends, scenarios and rankings for the time horizon of about 15 years. This planning perspective should be sufficient for most corporate strategic IT development decisions and for all IT investment decisions by non-IT companies as the legal and real depreciation periods for this type of investment are much shorter. Finally, let us mention that in the area of decision support systems, the results of the ERDF-financed IT foresight project [9] could provide constructive and successful recommendations to companies interested in the development of novel e-commerce applications.

References

1. Antoniou, I., Reeve, M., Stenning, V.: The Information Society as a Complex System. *Journal of Universal Computer Science* 6(3), 272–288 (2000)
2. Bhargava, H.K., Power, D.J., Sun, D.: Progress in Web-based decision support technologies. *Decision Support Systems* 43(4), 1083–1095 (2007)
3. Eom, S.B.: Decision support systems research: current state and trends. *Industrial Management & Data Systems* 99(5-6), 213–220 (1999)
4. FISTERA project web page, <http://fistera.jrc.ec.europa.eu/> (accessed 12.2010)
5. ISCOM project's web page, <http://www.iscom.unimo.it> (accessed 12.2010)
6. Karacapilidis, N.: An Overview of Future Challenges of Decision Support Technologies. In: Gupta, J.N.D., Forgionne, G.A., Mora, M.T. (eds.) *Intelligent Decision-making Support Systems. Foundations, Applications and Challenges*, pp. 385–400. Springer, Heidelberg (2006)
7. Miller, R.A.: Computer-assisted diagnostic decision support: history, challenges, and possible paths forward. *Advances In Health Sciences Education* 14, 89–106 (2009)

8. Pereira, A.G., Quintana, S.C., Funtowicz, S.: GOUVERNe: new trends in decision support for groundwater governance issues. *Env. Modell. Software* 20(2), 111–118 (2005)
9. Scenarios and Development Trends of Selected Information Society Technologies until 2025, Interim Report 2010, Progress and Business Foundation (2011), <http://www.ict.foresight.pl>
10. Schwiegelsohn, U., et al.: Perspectives on grid computing. *Future Generation Computer Systems* 26, 1104–1115 (2010)
11. Skulimowski, A.M.J.: Optimal Control of a Class of Asynchronous Discrete-Event Systems. In: *Proceedings of the 11th IFAC World Congress, Automatic Control in the Service of Mankind*, Tallinn, Estonia, vol. 3, pp. 489–495. Pergamon Press, London (1991)
12. Skulimowski, A.M.J.: Framing New Member States and Candidate Countries Information Society Insights. In: Compano, R., Pascu, C. (eds.) *Prospects For a Knowledge-Based Society in the New Members States and Candidate Countries*, pp. 9–51. Publishing House of the Romanian Academy (2006)
13. SEAMATE report web page, http://www.pascal.case.unibz.it/retrieve/1165/seamate_d3_1.pdf (accessed 04.2011)
14. Zhu, J.: *Quantitative Models for Performance Evaluation and Benchmarking: DEA with Spreadsheets and DEA Excel Solver*. Springer, Kluwer Academic Publishers, Boston (2003)

Relation Extraction from Documents for the Automatic Construction of Ontologies

Ikkyu Choi¹, Seungmin Rho^{2,*}, Young-Sik Jeong³, and Minkoo Kim⁴

¹ Graduate School of Information and Communication,
Ajou University, Suwon, Republic of Korea
ikchoi@ajou.ac.kr

² School of Electrical Engineering, Korea University, Seoul, Republic of Korea
smrho@korea.ac.kr

³ Department of Computer Engineering,
Wonkwang University, Iksan, Jeonbuk, Republic of Korea
ysjeong@wonkwang.ac.kr

⁴ College of Information Technology,
Ajou University, Suwon, Republic of Korea
minkoo@ajou.ac.kr

Abstract. The Semantic Web relies on domain ontologies that structure underlying data enabling comprehensive and transportable machine understanding. It takes so much time and efforts to construct domain ontologies because these ontologies have to be manually made by domain experts and knowledge engineers. To solve this problem, there have been some researches to semi-automatically construct ontologies. In this paper, we propose a hybrid method to extract relations from domain documents which combines a named relation approach and an unnamed relation approach. Our named relation approach is based on the Snowball system. We add the generalized pattern method to their methods. In our unnamed relation approach, we extract unnamed relations using association rules and clustering method. We also recommend candidate names of unnamed relations. We experiment and evaluate the proposed method by using Ziff documents set offered by TREC.

1 Introduction

Today, the Web has a large amount of information and becomes an important expedient for communication between people and enterprises. The Semantic Web should bring structure to the content of Web pages, being an extension of the current Web, in which information is given a well-defined meaning. Thus, the Semantic Web will be able to support automated services based on these descriptions of semantics.

Ontology that is logical basis for the Semantic Web defines a common vocabulary for researchers who need to share information in a domain. Ontology includes machine-interpretable definitions of basic concepts in the domain and relations among them. A ‘concept’ is a set of properties which describe the features of the concept in

* Corresponding author.

the world. A ‘property’ has facet which restricts the range of the property. Ontology is written in the formal languages like RDF, DAML+OIL, and OWL for the machine to understand the meaning of the domain knowledge. But, there are needs for so much time and effort to construct domain ontologies because these ontologies have to be manually made by domain experts and knowledge engineers. To solve this problem, there have been some researches to semi-automatically construct ontologies. The two main parts of automatically constructing ontologies are term selection part and relation extraction part. Most of researches focused on relation extraction part. The relation extraction in these researches can be mainly divided by named relation approaches which use linguistic pattern information and unnamed relation approaches which use co-occurrence information among concepts.

Named relation approaches usually use the Hearst’s patterns [1] and manually defined patterns that could extract relations among concepts. Because it is difficult to define linguistic patterns for relations, some approaches proposed how to find automatically linguistic patterns using statistic information of contexts. The problem of these approaches does not consider variations of patterns that exist in real domain documents. Unnamed relation approaches mainly used association rules based on co-occurrence information among concepts. These approaches only offered simple information, for example, relation strength value between concepts according to association rules [8]. Therefore, domain experts and knowledge engineers have to decide the name of the relation.

To solve these problems, we propose a hybrid method to extract relations from domain documents which combines a named relation approach and an unnamed relation approach. Our named relation approach is based on the Snowball system. We merge a generalized pattern scheme into their method. In our unnamed relation approach, we extract unnamed relations using association rules and clustering method. Moreover we recommend the candidate name of the group to offer intuitive information to the user.

The paper is structured as follows. Section 2 briefly introduces the related works on semi-automatic construction of ontologies. Section 3 suggests our hybrid method that extracts relations from domain documents which combines a named relation approach and an unnamed relation approach. Section 4 describes the experimental setting and result. In the last section, we conclude our paper and describe the future direction.

2 Related Works

In this section, we briefly discuss some works related to the semi-automatic construction of ontology in domain documents. The two main parts of constructing ontologies are term selection part and relation extraction part. Most of researches focus on relation extraction and can be classified into two categories. One is unnamed relation approaches which use the linguistic characteristics of the selected language. Another is unnamed relation approaches which use the statistical characteristics like co-occurrence and term frequency. Following two sub sections cover more detail about these two approaches.

2.1 Named Relation Approaches

These approaches use linguistic syntax and meaningful patterns to extract relations from documents. Snowball [6] is a novel system which automatically finds patterns between location and organization from document collections. Snowball is initially given a handful of valid tuples of organization and location. In order to generate new patterns, Snowball gathers occurrences of the initial tuples from documents. Left, middle, right contexts associated tuples are expressed as a vector. Snowball generates 5-tuple $t = \langle l_c, tag_1, m_c, tag_2, r_c \rangle$ using left, middle, right context. In this case, tag1 is organization and tag 2 is location. Snowball clusters these 5-tuples using a simple single-pass bucket clustering algorithm [7] and the Match function which calculate the similarity between the 5-tuples. The centroid of each cluster becomes patterns. Using these patterns, Snowball finds sentences that contain an organization and location as determined by the named-entity tagger. The benefit of Snowball is that the method only requires valid tuples instead of fixed linguistic patterns and is independent on document languages. However, the method can only extract organization-location relation and relies on strict named-entity information.

2.2 Unnamed Relation Approaches

Unnamed relation approaches generally use data clustering methods and data mining methods based on statistic distribution information which generally contains term frequency, document frequency, and term dependency information [2] to extract relations from documents.

There are many researches to cluster documents from large documents set in Information Retrieval area. Clustering can be defined as the process of organizing documents into groups whose members are similar. In general there are two categories of clustering: non-hierarchical clustering and hierarchical clustering. Hierarchical clustering algorithms are preferable for detailed data analysis because the output groups of non-hierarchical clustering have a flat structure. In clustering methods, an important point is how to define concept. These researches can be distinguished to groups as the defining schemes for concepts. One creates the system of classification using hierarchical clustering that treats concept as a set of terms. This approach cannot offer intuitive information to the user, and have the problem of naming on generated clusters [4]. Another treats phrase in the document as concept and creates the system of classification using the dependency and statistic distribution among phrases [3]. However, this approach also has weak point that the relation of generated clusters is different from a taxonomy relation because its purpose is user efficiency for accessing information.

KAON system that was developed by Karlsruhe university and FZI research center is an assistant tool for creating ontology [5]. The relation extraction module of KAON system is also hybrid approach that uses linguistic pattern and generalized association rule. We will focus on the statistic method of KAON system. KAON system follows next steps for semi-automatically creating ontology. First, the system preprocesses documents using shallow parser and finds noun phrase for candidate concepts. Next, KAON system finds generalized association rules on selected phrases and shows the possibility of the relation between these phrases to domain experts. KAON system

represents the support and confidence value of the possible relation rather than directly representing the relation and leaves the naming process to domain experts.

In [8], Byrd divides relations among concepts into named-relations and unnamed-relations. This method extracts named-relations from documents based on linguistic patterns and unnamed-relations based on association rules. These relations are triples, containing two concepts, and either a relation strength (unnamed relations) or a relation name (named relations).

3 Our Hybrid Approaches for Extracting Relations

As we discussed in the previous section, these approaches have problems: In named relation approaches, it is hard to find out the patterns for relations and to cover realization forms of relation in a real document. Unnamed relation approaches do not offer intuitive information about relations but relation strength information. To solve these problems, we propose a hybrid method to automatically extract relations from domain documents which combines a named relation approach and an unnamed relation approach.

3.1 Named Relation Approach in Extracting Relations

In our system, we focused on taxonomy relation which is the one of the most fundamental relations of the concepts. There are major difficulties in finding named relation automatically. One of the most obvious difficulties in making validate patterns is diversity of a relation. To extract each relation, different patterns are needed. Making relevant patterns for each relation costs a lot of time and human labor. The other is realization forms of relation in a real document. Though semantics of sentence represent same relation, the way of representation is different in a document. For example, the sentence which contains meaning of hyponymy of X and Y can be wrote like “X is a Y” or “X is a [adjective] Y” or “X is a kind of Y”.

To solve these problems, we applied Snowball which automatically generates relation dependent patterns from a document collection. To consider the variety of patterns of the relation, we make generalized patterns like SP+PRF system [9]. For simplified patterns, we only consider core of a sentence. Some of terms in a sentence are removed and some of them are translated into other words. Rules that we used are listed in Table 1.

Table 1. Categorization of Terms

Terms/POS tag	Category	Example
is, are, am, was, were	BE	Is ->BE
Noun, Noun phrases	Not translated	Kodak-> Kodak
Adjective modifiers	ADJ	Good -> ADJ
Adverb modifiers	To be deleted	
Determiner	DT	the ->DT

3.2 Unnamed Relation Approach in Extracting Relations

In our system, we assume that the relationship between two concepts can be described by contexts that exist in same sentences. Therefore, contexts between two concepts are useful to explain the relation derived by two concepts. We find the unnamed relations between concepts using association rule method. The transaction unit of association rule method is contexts. We cluster contexts that exist between two concepts for assigning intuitive names to unnamed relations.

Our system for unnamed relation approach consists of 5 parts; Preprocessor, Context Extractor, Association Rule Miner, Pattern Clustering Module, and Relation Naming Module as described in Figure 1.

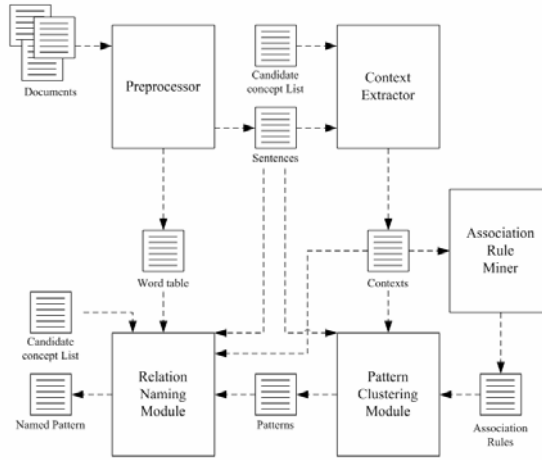


Fig. 1. The main processes of our unnamed relation approach

Five steps to extract relations

1. Preprocessor: extracts sentences from documents and
2. Context Extractor: reads the given candidate concept list. Then extract contexts from preprocessed documents. Finally it prints out concept pairs and contexts.
3. Association Rule Miner: creates association rules from extracted contexts.
4. Pattern Clustering Module: clusters contexts between two concepts and find patterns. To improve the speed of clustering, our system uses the simple single-pass clustering algorithm [7].
5. Relation Naming Module: assigns candidate names to generalized patterns by calculating the weight of context in generalized patterns.

3.3 Combining Method for Hybrid Approaches

We combine the result of named relation approach with the result of unnamed relation approach to get better results. The result of named relation approach has relation names that domain experts assigned. The names are very important information to our

system. Therefore, we combined the results based on named relation approach result. The four steps of combining the results are as follows.

1. Get a group in unnamed relation approach result
2. Find all similarities of selected group and all relation group in named relation approach
3. If highest similarity value is over a threshold then assign the selected group to the relation name in named relation approach result
4. Repeat 1-3 until no more group in unnamed relation approach result

4 Experiments and Evaluation

To verify the proposed method we use Ziff document set(Information from Computer Select disks-1989, 1990, copyrighted by Ziff Davis) offered by TREC(Text Retrieval Conference). The Ziff document set contains 785 files, and about 800MB size. The system selected 4,170,525 sentences and 387,192 words through its preprocessing. In this evaluation, we used concept list from C-value/NC-value method. We selected top 5,000 concept list which is arranged by the value of importance. To measure the precision and usefulness, we asked 5 persons who are familiar with computer domain, and calculated average precision value of their answer.

4.1 Named Relation Approach's Result and Evaluation

As shown in Table 2, generalized pattern method finds more tuples than Snowball. This is because that our method simplified patterns and these patterns contain concise information. This effects the calculation of match degree. Thus more tuples are selected than Snowball. Generalized Pattern (ADJ) means not deleting adjective because adjective is related to relation and may involve meaning in some sentences.

Table 2. Number of New Tuples in Taxonomy

	Number of New Tuples	Precision
Snowball	13006	28.9%
Generalized Pattern	14587	29.9%
Generalized Pattern (ADJ)	16809	30.8%

Generalized Pattern (ADJ) returned most number of seed tuples. Precision of each method are similar and did not reach to our expectation. This is due to the precision of the concept list from C/NC-value.

4.2 Unnamed Relation Approach's Result and Evaluation

Our system found 100,878 contexts from Ziff documents set using 5,000 concepts. Finally our system found 23,051 relations between concepts and found 192 generalized relations. Table 3 summarizes the evaluation result about extracted relations. It shows the accuracy of clustering on concept pairs: company-company, company-product, product-product and product-company. Summarizing the experimental result,

the usefulness of relation name is 58.33%, the expectation of useful relation is 55.75%. These values show that the relation name extracted by our system can be useful information for describing the relation between two concepts.

Table 3. Summary of the evaluation reslut of experiment

Items	Value
Number of extracted relations	192
Number of useful names	112
Average of usefulness	58.33%
Expectation of a useful relation	55.75%
Average precision of clustering	78.44%
Expectation of a correct cluster	74.69%

4.3 Combining the Named Relation’s Result with the Unnamed Relation’s Result

We estimated 3 combining tasks: Combining the snowball result with the unnamed relation approach result (CSS), Combining the generalized pattern result with the unnamed relation approach result (CGPS), Combining the generalized ADJ pattern result with the unnamed relation approach result (CGPAS). We evaluated the result based on precision how many concept pair is correct which in group is assigned by new relation name. Table 4 shows the results of our experiments. In this result, CGPAS method has most combined groups and precision is also high.

Table 4. Result of Combining the results (threshold = 0.2)

Experiment’s Methods	Number of Combined Group	Precision
CSS	38	51.7%
CGPS	62	58.8%
CGPAS	72	58.5%

5 Conclusions

In this paper we propose a hybrid method to extract relations from domain documents which combines a named relation approach and an unnamed relation approach. Our named relation approach is based on the Snowball system. We add a generalized pattern method into Snowball system. In our unnamed relation approach, we extract relations according to the three steps. First step is to select concept pairs using association rules which are between two concepts from document set. Second is to find patterns in selected each concept pair. Finally, we make pattern groups by using clustering method and recommend the candidate name of the group to offer intuitive information to the user.

Our contribution is next three points. First is to generalize patterns using soft matching method to recognize various context forms of the relation in a sentence. We produced recall evaluation value up using this point. Second is to group unnamed-relation into unnamed-group relation and assign useful relation names to

unnamed-group relations on 58.33% precision using our unnamed relation approach. Third is to propose a hybrid method to extract relations from domain documents which combines a named relation approach and an unnamed relation approach.

Acknowledgments

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract (UD060048AD).

References

1. Hearst, M.A.: Automatic Acquisition of Hyponyms from Large Text Corpora. In: Proceedings of the 14th International Conference on Computational Linguistics (1992)
2. Kim, H.-s., Choi, I., Kim, M.: Refining Term Weights of Documents Using Term Dependencies. In: Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 552–553 (2004)
3. Lawrie, D., Croft, W.B., Rosenberg, A.: Finding Topic Words for Hierarchical Summarization. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 349–357 (2001)
4. Lawrie, D.J., Bruce Croft, W.: Generating Hierarchical Summaries for Web Searches. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 457–458 (2003)
5. Maedche, A., Staab, S.: Semi-Automatic Engineering of Ontologies from Text. In: Proceedings of the 12th International Conference on Sw Engineering and Knowledge Engineering, SEKE 2000 (2000)
6. Agichtein, E., Gravano, L.: Snowball: Extracting Relations from Large Plain-Text Collections. In: Proceedings of the ACM International Conference on Digital Libraries, DL 2000 (2000)
7. Frakes, W.B., Baeza-Yates, R.: Information Retrieval: Data Structures and Algorithms. Prentice-Hall, Englewood Cliffs (1992)
8. Byrd, R.J., Ravin, Y.: Identifying and extracting relations from text. In: Proceedings of the 4th International Conference on Applications of Natural Language to Information Systems (1999)
9. Cui, H., Kan, M.-Y., Chua, T.-S.: Unsupervised Learning of soft patterns for generating definition. In: Proceedings of 13th International World Wide Web Conference (2004)

Proactive Detection of Botnets with Intended Forceful Infections from Multiple Malware Collecting Channels

Young Hoon Moon and Huy Kang Kim

Graduate School of Information Security, Korea University,
Anam-dong, Seongbuk-gu, Seoul, 136-713, Republic of Korea
{ymoon, cenda}@korea.ac.kr

Abstract. As the major role of Internet Service Providers becomes shifted from caring for their legitimate x-DSL subscribers and enterprise leased line users to protecting them from outside attacks, botnet detection is currently a hot issue in the telecommunications industry. Through this paper, we introduce efficient botnet pre-detection methods utilizing Honeynets with intended forceful infections based on different multiple channel sources. We applied our methods to a major Internet Service Provider in Korea, making use of multiple channel sources: Payloads from Spam Cut services, Intrusion Detection Systems, and Abuse emails. With our proposed method, we can detect 40% of real C&C server IPs and URLs before they are proven to be malicious sites in public. Also, we could find the C&C servers before they caused many victims during their propagation periods and, eventually, we will be able to shut them down proactively.

Keywords: Botnet Detection, Distributed Denial of Service, Honeynets, C&C Servers, Intended forceful infection.

1 Introduction

The impacts of emerging botnets have seriously affected the Internet community by sabotaging enterprise systems. Attackers can threaten Internet Service Providers (ISPs) with bandwidth depletion methods, mostly like Distributed Denial of Service (DDoS) attacks. Therefore, rapid botnet detection is very important in the communications industry because ISP's major role is not only providing network connectivity to their customers but also providing security measures to protect the customers from malicious attackers. Several studies distinguish botnets based on their communication channels and their propagation methods. These previous studies show good efficiency providing that the botnets utilize already-known methods of communication and propagation. If unknown malware change their communication protocols or propagation techniques, these protocol-based detections will be useless. In this paper, we introduce efficient botnet detection methods utilizing Honeynets with different multiple channel sources based on behavioral analysis that does not rely on protocols.

2 Related Work

IRC botnet detection methods are suggested and studied in many different ways. Recently, Peer to Peer (P2P) botnets have appeared and an advanced hybrid P2P botnet has been studied [8]. This approach would be good if the bot masters or botnet programmers followed the previous type of communications or propagation techniques. However, this ideal presumption is very dangerous and risky, because the result of a protocol-specific detection approach will bring unexpected false positives or false negatives. Evan Cooke [2] summarized botnets and their history suggesting a comprehensive approach for detecting the variety of C&C activities. Even so, it is not enough to detect rapidly changing botnet structures such as P2P architecture. James R. Binkley [3] pointed out the weaknesses of signature-based botnet detection and then proposed an anomaly-based algorithm using heuristic TCP scans with work weight attributes. As presented in [3], if communication is encrypted between botnets, it is no good for examining or classifying the encrypted tuples as needed for detection. W. Timothy [4] made an approach to detect botnets with traffic analysis and Honeynets. He eliminated uncorrelated tuples from a bunch of traffic, then grouped them as common communication patterns. Nevertheless, his approach still had the limitation of finding non-IRC botnets such as P2P communication botnets. Similar to [3], Guofei Gu [5] suggested network-based anomaly detection methods calculating temporal-spatial correlation within coordinated botnet communication traffic. This is a systematic approach for dealing with the main characteristic of C&C communication, propagation and network activities. Inversely, detecting them using an anomaly-based algorithm is not good enough for pre-detection because this method must wait until it has a certain amount of a network traffic queue to be analyzed. Julian B. Grizzard [6] made his point for revealing a flaw of traditional botnet detection methods fitted in C&C botnets by introducing P2P botnets. Unfortunately, he was a good introducer but not a good defender against the suggested P2P botnets and their behavioral characteristics.

3 Proposed Method

The proposed method has two basic simple modules: one for collecting botnet samples and the other for analyzing botnets based on behavioral traits. First, the concept for the module of collecting samples of botnets originated from the fact that recent botnets have propagated over millions of victim machines. We approached this method focusing on how we can get infected with recent botnets as much as possible. Then, we gathered all the suspicious URL links and attached files from international gateways providing that lots of botnet propagations occur between nations. Secondly, the suggested module of botnet behavioral analysis is designed to get certain meaningful information from the infected virtual machines (VMs). Most of all, finding C&C server information is the goal of this module from botnet behavioral observation.

3.1 Hybrid Honeynets with Intended Forceful Infection

The traditional Honeynets have been developed in two general ways based on their inducing methodologies. One is Low-Interaction Honeynet and the other is

High-Interaction Honeynet. The former, as we can infer from its name, suggests low attractions to botnets or malicious users by emulating virtual services or vulnerabilities. The latter suggests more attraction to those malicious attackers by responding to real hosts or machines. The major weakness of Low-Interaction Honeynet (LIH) is that they are easily detectable by botnets with simple checks using complex vulnerability checks or services. However, LIH have an advantage over High-Interaction Honeynet (HIH) considering that the infected machines cannot be involved with attacking other hosts or services. Similarly, HIH can easily get infected without worrying about emulating any virtual services, which is a major drawback of LIH management. In this paper, we will construct a hybrid Honeynet with intended forceful infections, which is a combination of the advantages of both LIH and HIH. The weakest point of LIH is the low infection rate compared to HIH. In other words, the mere attraction point of LIH is simply open services or unpatched vulnerabilities for attracting botnets and attackers. This passive attraction point is critical for Honeynets to gather substantive information from botnets. For example, LIH should have vulnerable services with specific system ports and maintain its current major vulnerabilities to get infected with all new botnets. In addition, LIH has to keep up with the state of the art of evasion techniques of botnets that try to avoid getting caught in Honeynets. Inversely, the biggest weakness of HIH is the malicious activities being executed by Honeynets. Because HIH resides in a real machine, it is more difficult to control them than to control LIH. To prevent these actual malicious activities from Honeynets, the system operators take appropriate actions such as installing virtual Intrusion Detection System (IDS) or Intrusion Prevention System (IPS) within Honeynets.

3.1.1 Advanced Honeynets with Forceful Infections

While prior botnet propagation methods are simply based on the newly found application vulnerabilities, recent ones are gradually based on social networking such as downloading files from P2P sites, executing the attached file from the known mailing lists. To utilize social networking, we need those URLs and binary files for intended infections on our VMs. We made our system emulating human interactions by executing programs from the lists of botnet candidates, gathering botnet samples from various sample channels. 6 botnet sample channels were provided by a major ISP in Korea: a URL Crawler, E-mail, User Inputs, Honeypots, IDS, and International Spam Cut Services. We put suspicious URL links or download paths from six channels to the managed queues then executed them on an MS Windows VM. After intended forceful infection, our system analyzed all the behaviors from the network layer to the application layer through pre-installed agents on VMs. The agent was designed to detect all the modification of files, registries, and DNS queries and to send that information to the main analysis servers. After a 2~3 minute initial analysis, the agent on the VM waits for 100 seconds for other abnormal activity then reverts the VM into a previous clean state for other observations. Fig. 1 shows our proposed system's overall architecture. Our system has 5 sub-systems – a Malicious Activity Analysis System, Pre-detection Management System, Anti-Virus Checking System, Botnet Sample Collectors and Honeynets with abnormal detectors. In the Malicious Activity Analysis System, intended infection is executed with virtual machines. Then the detected result will be given to a Pre-detection Management System.

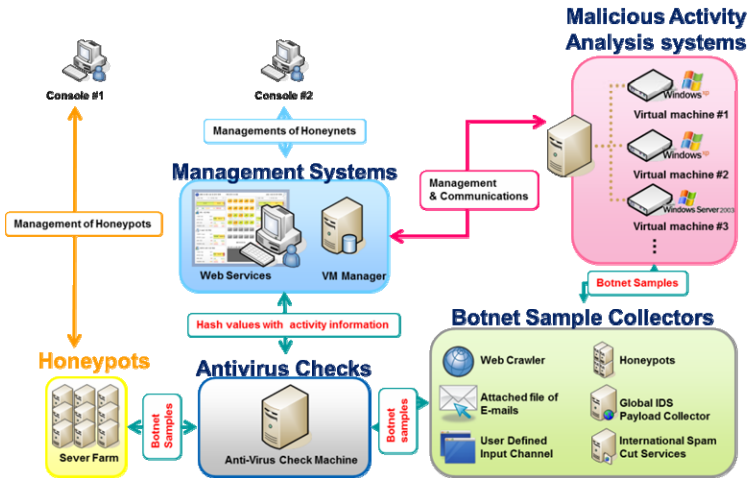


Fig. 1. System architecture of botnet pre-detection system

4 Simulations

We have tested our system in the real world since at the beginning of 2010. Our system extracted suspicious IPs and URLs from behavioral analysis agents residing on virtual machines. After 7-month run, we finally compare our lists with KORNET black lists made by CERT Team for responding bot attacks. After the IP and URL list match ups, we can get the hit rates of 36.2% and 40.1%. We gathered suspicious botnet samples from KORNET, the biggest Korean Backbone Network, for seven months with six assigned channels. E-mail from the POP3 protocol, International IDS and Spam Cut Service-named KAIS channels are the majority of input sources described in Table 1. However, user-related source channels such as URLs collected by crawlers and User-Defined channels show a low volume of sample collections. These two channels are used for special issues. For URL channels, it is waiting for user input for identifying phishing sites or other abuse cases. User-Defined channels are used for checking suspicious binary files or URLs. Therefore, both channels are not very often active compared with the others. For honeypot channels, interestingly, it gained a few sample collections at the beginning of the year. However, since June it gathered samples more actively than any other month. This is probably related with the increasing of CVE Candidates between May and June in 2010.

All binary samples collected from channels are delivered to 'Antivirus Check' machine for classification of malware by pattern matching analysis before they are analyzed in VMs. Most of samples are in the class of Trojan including Downloader, GameThief and Dropper. As shown in Table 2, 2,058 malware samples are identified and classified during antivirus checkups and 13,541 samples are not classified when found. These unknown samples are finally detected as malware by behavioral analysis agents on VMs. The volume of unidentified samples is approximately six times of identified samples. This figure can verify that our system is the very front line of fire

Table 1. The number of collected samples from six channels

Items Months	Collected Samples by Channels						Outputs		Total	
	URL	EMAIL	USER	HNYP	IDS	KAIS	Abnl Act.	Bot Act.	Outputs/Inputs	Rate(%)
JAN	1	44,276	87	-	8,620	98,111	4,122	457	151,095 / 63,6306	24%
FEB	-	13,918	2	1	3,526	68,114	2,937	117	85,561 / 29,6193	29%
MAR	3,693	3,614	1	1	484	49,540	3,499	691	57,333 / 20,3851	28%
APR	159	16,027	20	-	6,424	88,792	3,972	98	111,422 / 859,922	13%
MAY	-	12,321	6	7	3,530	86,292	1,296	16	102,156 / 22,5793	45%
JUN	-	13,361	8	59	2,886	75,808	1,317	19	92,122 / 178,1924	5%
JUL	-	16,687	11	125	2,806	98,980	1,222	15	118,609 / 215,5488	6%

in the malware detection battlefield. This can be explained by the time gap between the detection of botnets and updated signatures of detected botnets. This time gap is discussed in detail in 4.2.

4.1 Botnet Detection by Channel

Fig. 2 shows that most botnet detection comes from IDS and Spam Cut Service Channels during the test periods. This is somewhat different from the traditional botnet propagation methods utilizing system vulnerabilities and service scanning.

Table 2. Montly detection result from system with Antivirus checks

Month	Detection Result		Increase/Decrease	
	Known botnets	Unknown botnets	Known botnets	Unknown botnets
JAN	1296	7782	-	-
FEB	250	976	▼ 1046	▼ 6806
MAR	23	174	▼ 227	▼ 802
APR	316	2038	▲ 293	▲ 1864
MAY	78	910	▼ 238	▼ 1128
JUN	46	705	▼ 32	▼ 205
JUL	49	956	▲ 3	▲ 251

For example, after we analyzed an infected host, we detected a malicious binary executable that was hidden in a compressed obscene media file originating from a P2P site. With this, we assume that botnet masters are using social networks for propagating their bot programs by attracting users with interesting digital media such as pornography. They usually created media files with eye-popping titles regarding sex to attract Internet users and got them to download those files not even knowing that the compressed file included malicious programs. Therefore, if we can possibly include our sample channel with a P2P site, we can get more samples from them comparing with those of SPAM or URL links. This will be implemented in our future work.

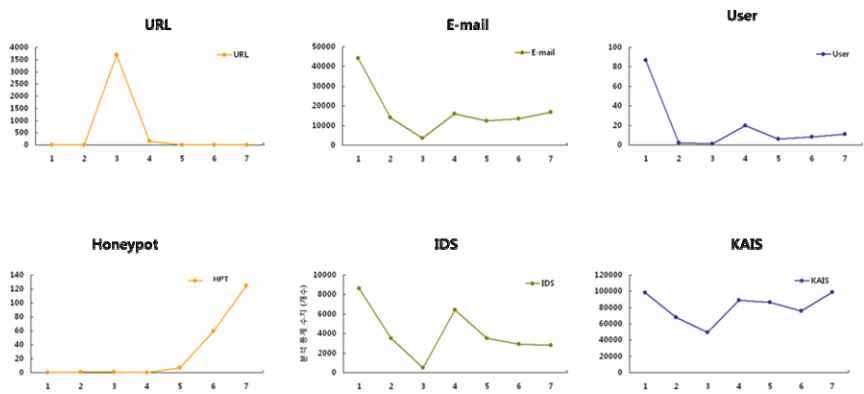


Fig. 2. Monthly analysis result of the collected samples from 6 channels

4.2 Pre-detection of Botnets

The most difficult part of measuring performance of our system was how to show that the data from our system is accurate or meaningful for detecting botnets. With intensive research, we could get malicious lists from a major ISP in Korea and then compare those with our data. The data we received from the ISP was the blocked C&C server URLs and IPs from January to July in 2010. More interestingly, to cope with large scale DDoS attacks, the Korean Government was operating a Botnet Response Team whose role was to analyze bot programs with dynamic and static analysis and to spread their results to major ISPs in Korea to not route those IPs and to not respond to botnet DNS queries. After cross-checking the data we got, from a C&C IP point of view, our system hit 370 IPs from 1021 IPs showing a 36.2% accuracy rate and from a C&C URL point of view, 810 URLs were exactly matched with the government list showing a 40.1% accuracy rate. Even though 40% seems to be a low accuracy rate, if we consider the large scale of DDoS attacks at around 100~200 Gbps of bandwidth, pre cutoffs of those communication channels between bots with control servers are overwhelmingly attractive solution for major ISPs handling everyday bandwidth depletion attacks or DDoS attacks.

Table 3. The number of Pre-detection of URLs before / after blocking

sc er & U s	ys r sc er & U s								e ec U s by H / cke U s by	
		6		12	1	18	21	24		
g	1	64	148	2	46	8	1	28	64 / 2 2	1
A	g		7		4					%
	18	1	16	66	2	8	1	28	81 / 2 2	4 1

In addition to the accuracy rate, we inspected these two lists, IPs and URLs, by comparing the dates of blocks from the ISP with those of detection from our system. We found that our system pre-detected malicious URLs and IPs before the ISP

blocked them for protecting its network. As mentioned in 4, there exists a time gap between detection by behavioral analysis and detection by static pattern analysis due to the time elapsed while the antivirus companies perform the static analysis. Fig. 3 shows the pre-detection result from our system and it shows 148 URLs and 80 IPs are detected 90 days before the ISP cut them off. This can be explained by the fact that there existed a time gap between botnet propagation and attacking victims. In this case, a three-month time gap is the most common from our simulation.

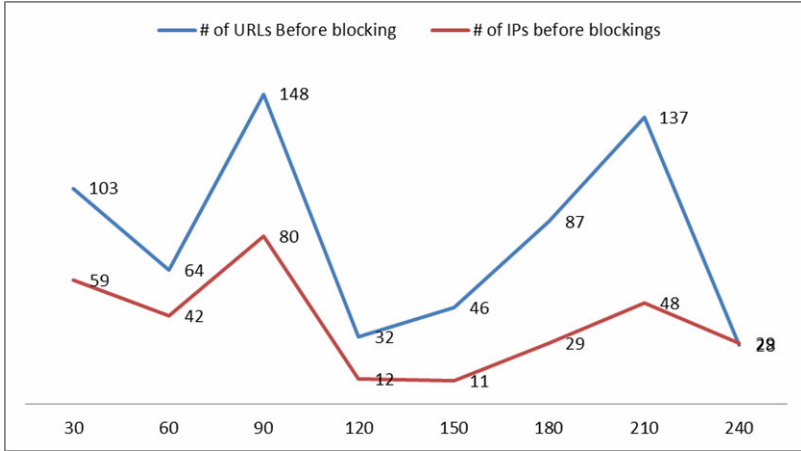


Fig. 3. Number of days of detection prior to the detected URLs or IPs being blocked

5 Conclusion

Botnet detection is an emerging big issue in various areas from major ISPs to government-owned infrastructure. Novel and traditional methods for detecting botnets have been suggested and studied. Signature-based or anomaly-based approaches have a limitation in detecting newly intellectualized and organized botnets to avoiding those detecting techniques by producing tons of variants or by hiding themselves in normal traffic. In this paper, we suggested an intended forceful infection method for analyzing and detecting botnets before their harmful action begins, especially DDoS attacks of randomized victims. After gathering suspicious botnet samples including download URLs and binaries from multiple channels located on nationwide network gateways, we then put those suspicious URLs and binaries on a VM for dynamic analysis. During the analysis process, preinstalled agent software on a VM monitored all the behaviors from network traffic to file system modification after intended infections. All the analyzed results from each VM are transferred to the main database servers and classified by pre-defined rules according to their behavioral traits. With a 7 month field test, we got over 40% accuracy comparing our results with C&C URLs and IP addresses from the list of a government agency. More interestingly, 84% of our detections occurred before the attack took place, compared with the date of blocking by ISPs. From this, we could proactively screen out botnet attack attempts by

isolating the C&C URLs and IPs from our detection system. This will be a great help to manage heavy bandwidth depletion attacks, which are critical for major ISPs.

For future works, we will seek other channels to increase accuracy of this system such as black list of P2P download sites, and we will evaluate which channel will influence the accuracy most significantly.

References

1. Watson, D., Riden, J.: The Honeynet Project: Data Collection Tools, Infrastructure, Archives and Analysis. In: WOMBAT Workshop on Information Security Threats Data Collection and Sharing, pp. 24–30 (2008)
2. Cooke, E., Jahanian, F.: The zombie roundup: Understanding, detecting, and disrupting botnets. In: Steps to Reducing Unwanted Traffic on the Internet Workshop, SRUTI 2005 (2005)
3. Binkley, J.R., Singh, S.: An Algorithm for Anomaly-based Botnet Detection. In: Steps to Reducing Unwanted Traffic on the Internet Workshop, SRUTI 2006 (2006)
4. Timothy Strayer, W., Walsh, R.: Detecting Botnets with Tight Command and Control. In: 31st IEEE Conference on Local Computer Networks (2006)
5. Gu, G., Lee, W.: BotSniffer: Detecting botnet command and control channels in network traffic. In: Proceedings of the 17th Conference on Security Symposium (2008)
6. Grizzard, J.B., Sharma, V., Nunnery, C., Kang, B.B.: Peer to Peer Botnets: Overview and Case Study. In: Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets (2007)
7. Dagon, D., Gu, G., Lee, C.P., Lee, W.: A Taxonomy of Botnet Structures. In: Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), pp. 325–339 (2007)
8. Wang, P., Sparks, S., Zou, C.C.: An Advanced Hybrid Peer-to-Peer Botnet. *IEEE Transactions on Dependable and Secure Computing* 7(2), 113–127 (2010)
9. Barford, P., Yegneswaran, V.: An Inside Look at Botnets. In: *Malware Detection. Advances in Information Security*, vol. 27, part III, pp. 171–191 (2007)

Solving English Questions through Applying Collective Intelligence

Dongjin Choi¹, Myunggwon Hwang², Byeongkyu Ko¹, and Pankoo Kim¹

¹ Dept. Of Computer Engineering Chosun University, Gwangju, South Korea

² Korea Institute of Science and Technology Information, Daejeon, South Korea
{Dongjin.Choi84, rhqudrb135}@gmail.com,
mgh@kisti.re.kr, pkkim@chosun.ac.kr

Abstract. Many researchers have been using n-gram statistics which is providing statistical information about cohesion among words to extract semantic information in web documents. Also, the n-gram has been applied in spell checking system, prediction of user interest and so on. This paper is a fundamental research to estimate lexical cohesion in documents using trigram, 4gram and 5gram offered by Google. The main purpose of this paper is estimating possibilities of Google n-gram using TOEIC question data sets.

Keywords: N-gram, Natural Language Processing, Semantics.

1 Introduction

N-gram has been applied to various fields for information processing to estimate semantic information in web documents and to analyze semantic relation between words. The n-gram information is a statistical data set collected and extracted from huge web document sets through analyzing frequency between adjacent words. The n-gram consists of bigram, trigram, 4gram and 5gram and the trigram is the most common data of the n-gram. The n-gram is based on probabilities of an adjacent words occurrence. For this reason, the n-gram can be a fundamental data set in natural language processing and word recommendation field. For instance, there is a difference and similarity between English and Chinese language. [1] compared how English statistically far apart from Chinese and how close they are using n-gram information. Also, the n-gram was applied to text processing [2] and user demands forecasts system [3] and so on. This paper focused on text processing using n-gram data sets provided by Google. Google n-gram contains approximately 1.1 billion words that occurred more than 40 times in documents set what Google had in 2006. Google has been using this n-gram data to query recommendation system. Moreover it is applied to speech recognition [4] and word recommendation text editor [5]. This paper estimates usability of each Google n-gram using one of the representative English language tests in Korea TOEIC as an experimental data set. The differences between each n-gram recall and precision rate will be presented.

The reminder of the paper is organized as follows. Section II describes what Google n-gram and TOEIC are and related works of this paper. A method to apply

Google n-gram data to TOEIC is described in Section III with examples. In Section IV, experimental results are given based on our forecasts system. Finally, Section V concludes with a summary of the results and gives challenging problems.

2 Related Works

2.1 Fundamental Data

Google N-gram. There is no doubt that Google is the most famous and representative search engine in the world. Google contains huge web pages concerning diverse fields such as news, conference and government information and so on. Lots of researcher and public users are using Google search engine to resolve their curiosities. Google has own strong retrieval strategy so that they can save lots of useful web documents. Google grasped n-gram data sets from various web documents and these n-gram data sets are provided by LDC (Linguistic Data Consortium)¹. The Google n-gram data sets consist of approximately 1.1 billion words occurred more than 40 times in web documents sets (about one trillion word terms). For this reason, we expect the reliability of results using Google n-gram is high. Table 1 shows statistics of each Google n-gram.

Table 1. The number of tokens of each n-gram

N-grams	Number of Tokens
Unigram	13,588,391
Bigram	314,843,401
Trigram	977,069,902
4gram	1,313,818,354
5gram	1,176,470,663
Total token	1,024,908,267,229
Total sentences	95,119,665,584

TOEIC. For more than 30 years, the TOEIC test has set the standard for assessing English-language listening and reading skills needed in the workplace. More than 10,000 organizations in 120 countries throughout the world trust the TOEIC test to determine who has the English language skill to succeed in the global workplace. Because of this fact, most of companies and universities in Korea are using TOEIC as a criteria to determine his (her) English ability. It is divided in Section I (Listening) with 4 kinds of parts and Section II (Reading) with 3 kinds of parts. The fifth part is a multiple-choice assessment that has incomplete sentences with four different possible answers. We use the fifth part as an experimental data set provided by one of famous English education web site named Hackers² to estimate usability of Google n-gram. Following table 2 gives examples of fifth part in TOEIC test.

¹ <http://www.ldc.upenn.edu>
² <http://www.hackers.co.kr>

Table 2. Examples of fifth part in TOEIC test

No.	Questions
1	According to the recent survey by the Transport Committee, the subway system is considered the most ----- means of transportation. (A) preference (B) preferred (C) preferring (D) prefer
2	The state commissioner said that signs ----- the exit route throughout the company premises should be posted. (A) indicating (B) indication (C) indicate (D) indicates
3	Meyers Shoes.com is well known for its widest ----- of women's and men's shoes and accessories. (A) display (B) selection (C) placement (D) position
...	...

This paper contains basic research for analysing lexical structure in web documents using n-gram data frequency. To test usability of Google n-gram, we made simple forecast system that chooses which word can be the best answer to the blank using trigram, 4gram and 5gram in Google n-gram. Therefore, the usability and precision rate will be evaluated.

2.2 Similar Works

N-gram is a collective intelligence frequency data among adjacent words so it has been applied to NLP (Natural Language Processing), Parsing and Terminology Extraction and so on. The n-gram can be described in two different ways. First is the gram by each character in sentences and second is gram by each word. For example, when we have a sentence "The best thing about the future is that it comes one day at a time" by Abraham Lincoln, each n-gram can be extracted as follows using first and second method.

→ The first method

trigram: {The, heb, ebe, bes, est, ...} #50

4gram: {Theb, hebe, ebes, best, estt, ...} #49

5gram: {Thebe, hebes, ebest, bestt, estth, ...} #48

→ The second method

trigram: {The best thing, best thing about, thing about the, ...} #13

4gram: {The best thing about, best thing about the, thing about the future, ...} #12

5gram: {The best thing about the, best thing about the future, thing about the future is, ...} #11

Because the total size of n-gram based on first method is bigger than second one, it takes more time to retrieve and calculate the n-gram data sets [6]. Also, it is not easy to match human spoken language with n-gram data because most of the n-gram data sets are written word from [6]. For example, onomatopoeia such as "um", "uh" is frequently used in spoken English even though it has no meaning but not in written English. Therefore, the n-gram data extracted from spoken and written English are different even they have the same meanings. To reduce the differences between spoken and written English, [6] suggests a method to merge 4gram data sets in spoken

English to trigram data set in written English. Besides, n-gram is applied to estimate lexical structure of documents to find semantics. [7] proposed a method to decide which words such as pronoun and preposition are suitable in sentences using Google n-gram data sets. Additionally, n-gram has been used to estimate noun compounds in specified domain documents to determine what keyword it is. The author of [8] presented a method to analyse which glossary term will be precisely represent documents in Grolier³ data sets collected by Mark Lauer based on bigram data frequencies. The biggest obstacle of Google n-gram is the size of data. The fourth and fifth grams are nearly 23GB each but trigram is 15GB. Lots of researches are based on trigram because bigram is too short to find semantics and fourth and fifth are too big to satisfy costing time. To overcome this limitation, [9] gave an idea to modify the threshold value which is a criteria when extract n-gram frequencies. As we can see from above researches, n-gram model has been dynamically used in various fields because it is reliable data set.

3 A Method to Apply Google n-Gram

This section describes a method to apply Google n-gram data sets to TOEIC test questions. There are 4 kinds of possible answers in fifth part of TOEIC. Questions are incomplete sentences with a blank. Therefore, there are at least four possible candidates for correct answer when using 5gram if the blank occurred at the front or the end. If the blank placed at the middle of sentences, the possible candidates for correct answer are up to twenty using 5gram. The number of candidates in a sentence using each n-gram was followed by given formula (1). Following formula (2) indicates total number of candidates in a sentence with four kinds of possible answer.

$$T = 2 \times n - 1, NG = T - n + 1. \quad (1)$$

where, T is total number of terms inputted in the system including blank, n is depth of n-gram and NG is total number of constructed n-gram.

$$TNG = NG \times 4. \quad (2)$$

where, TNG is the total number of constructed n-gram including 4 kinds of possible answers.

The position of the blank in sentences and depth of n-gram determine the number of candidates. Following table 3 shows examples of possible n-gram determined by position of the blank.

Table 4 gives examples of fifth part in TOEIC test with frequencies from Google 5gram data. The system compares the 5gram from given sentence with 5gram in Google. If these 5grams are matched, the word which has the highest frequency rate will be determined as an answer. This is based on the fact that the most frequently used sentences have the highest probability for correct answer. The system sum the frequencies of each possible answers and the highest one decided as a correct answer. This paper contains the basic research to assess usability of Google n-gram. The evaluation of recall and precision rate are based on n-gram frequencies provided by Google.

³ <http://en.wikipedia.org/wiki/Grolier>

Table 3. Example of candidates n-gram determined by position of the blank

Blank at the front	
Example	----- the store is understaffed right now, ~ (A) Although (B) Yet (C) Meanwhile (D) But
5gram	Although the store is understaffed, Yet the store is understaffed, Meanwhile the store is understaffed, But the store is understaffed $NG = T - n + 1 = 5 - 5 + 1 = 1$, $TNG = NG \times 4 = 4$
Blank at the end	
Example	~ after a colleague said the slides would be too -----. (A) distract (B) distracted (C) distractedly (D) distracting
4gram	would be too distract, would be too distracted, would be too distractedly, would be too distracting $NG = T - n + 1 = 4 - 4 + 1 = 1$, $TNG = NG \times 4 = 4$
Blank at the middle	
Example	~ within 90 days and ----- by the barcode on the box. (A) altered (B) adjusted (C) accepted (D) accompanied
Trigram	days and altered, days and adjusted, days and accepted, day and accompanied, and altered by, and adjusted by, and accepted by, and accompanied by, altered by the, adjusted by the, accepted by the, accompanied by the $NG = T - n + 1 = 5 - 3 + 1 = 3$, $TNG = NG \times 4 = 12$

Table 4. Example of fifth part in TOEIC and its 5grams

Question & Answer		~ in order to inform () about the purpose of ~ (A) themselves (B) them (C) that (D) it
5gram candidates	themselves	- in order to inform themselves 170 - order to inform themselves about 68 - to inform themselves about the 1858
	them	- in order to inform them 2825 - order to inform them about 720 - to inform them about the 8980 - inform them about the purpose 47 - them about the purpose of 302
	that	- in order to inform that 118
	it	- in order to inform it 158 - to inform it about the 467
		$NG = 5$, $TNG = 20$
	Sum	themselves: 2096 them: 12874 that: 118 it: 652

4 Evaluation and Results

The usability, efficiency and precision of Google n-gram are compared and evaluated with each n-gram through experiments in this section. For the evaluation, the n-gram

which had special character has been removed so total sizes of trigram, 4gram and 5gram data sets are approximately 24GB, 24GB and 13GB respectively. The TOEIC test sentences for the evaluation were provided by Hackers TOEIC which is one of popular English educational web sites. 200 questions were randomly chosen from the Hacker TOEIC and were compared with Google n-gram data sets. We made a simple system to automatically determine which word will be correct answer for the question shown in figure 1. This system needs two kinds of inputs. First one is an incomplete question sentence including blank and second one is four kinds of possible answers. The system compares every possible 5gram candidates extracted from question sentence with Google 5gram data and saves total frequency of each word if they are matched. Eventually, the system found eleven matched 5grams between question data and Google 5gram data shown in figure 1. The frequencies of each word are 2096, 12874, 229, 625. The system chose ‘them’ as a correct answer due to its highest frequency.

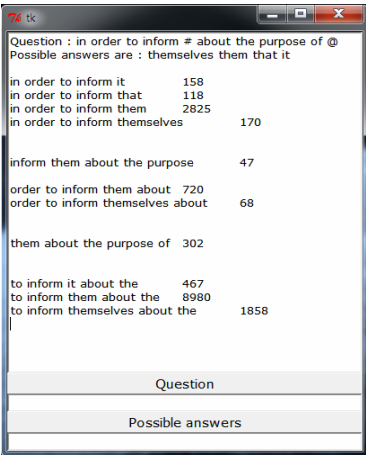


Fig. 1. Example of the system

Following table 5 indicates frequencies and summations of each word when system finds the answer using Google trigram and 4gram. According to our system, ‘procedure’ and ‘coordinating’ were chosen as answer and these are same with precise answers.

There are two types of question in TOEIC part five. First is a vocabulary question which stems of given example words are not same. Second is grammar question which stems of given words are the same. Moreover, in natural English language, it is hard to count the number of way to express English. In order to test whether our system satisfies in dynamic English expression or not, we need to compare recall, precision and F1 rate of vocabulary and grammar question based on following formula (3) and (4). Table 6 gives examples of these two types of question data.

Table 5. Example of fifth part in TOEIC and its 5grams

Using 4gram		
Question & Answer	... follow the standard # if they are @ ... (A) procedures (B)developments (C) categories (D) qualifications	
4gram candidates	- qualifications if they are	216
	- categories if they are	253
	- developments if they are	190
	- follow the standard procedures	575
	- procedures if they are	880
Sum	procedures: 1655 categories: 253	developments: 190 qualifications: 216
Using trigram		
Question & Answer	... charge of # the projects @ ... (A) collaborating (B)intending (C) pending (D) coordinating	
4gram candidates	- of coordinating the	25477
	- of intending the	112
	- charge of collaborating	117
	- charge of coordinating	10708
	- charge of intending	59
	- of pending the	91
	- coordinating the projects	412
Sum	collaborating: 117 pending: 91	intending: 171 coordinating: 36597

Table 6. Two types of Question in part five

A type of the vocabulary question
David Park was ----- to revise the draft of the contract, making it in the best interest of the company. (A) decided (B) intended (C) offered (D) instructed
A type of the grammar question
To help our staff better understand the nature of the meeting, the agenda was ----- a moment ago. (A) distribute (B) distributing (C) distributed (D) distribution

$$Recall = |A \cap B| / |A|, Precision = |A \cap B| / |B|. \quad (3)$$

where, A is the relevant set of sentences(n-gram) for the query, B is the set of retrieved sentences.

$$F1 \text{ measure} = 2 \times R \times P / (R + P). \quad (4)$$

where, R is the Recall rate and P is the Precision rate.

Table 7. Evaluation results

	Measurement	Vocabulary	Grammar	Total
5gram	Recall(%)	56.8	46.667	53
	Precision(%)	78.873	100	85.849
	F1-measure	66.041	63.636	65.538
4gram	Recall(%)	90.16	79.92	85
	Precision(%)	85.455	86.667	85.882
	F1-measure	87.746	881.504	85.438
Trigram	Recall(%)	100	98.611	99.5
	Precision(%)	75.781	85.915	79.397
	F1-measure	86.222	91.826	88.318
Trigram & 4gram	Recall(%)	100	97.436	99
	Precision(%)	83.607	86.842	84.848
	F1-measure	91.071	91.831	91.379

As we can see in table 7, the recall rate based on Google 5gram is only around 53%. In other word, the probability that five continuous words from TOEIC are matched with Google 5gram data is approximately 53%. When we have word set A and B consisted of five continuous words, two word sets are matched by 53%. Although the recall rate is low, the precision rate is nearly 86% which means that if the system found matched 5gram data, this data would close to answer with 86%. The recall rate based on Google 4gram was suddenly increased to around 85% and the precision rate was stayed in steady. The requirement of matching condition using 5gram was five words but 4gram is four words. It is simple to understand that the recall rate based on 4gram was increased to 85%. For the same reason, the recall rate of using trigram was increased to 99.5% which means that most of three continuous words from TOEIC test are placed in Google trigram data sets. However, the precision rate was decreased to 79.397%. It means that there are too many matched trigram data between TOEIC test and Google trigram. Because system based on the logic that it choose the word which has the highest frequency. But the way of natural English language expression is so much dynamic so the word with highest frequency is not always the answer. This is the reason why the precision rate of using trigram is lower than using 4gram and 5gram. It is best if answer word had the highest frequency but is not. To overcome this limitation, we combined 4gram and trigram together due to the fact that 4gram has the highest precision rate with smaller size than 5gram and trigram has the highest recall rate than others. The procedure to find answer has two steps that the system chooses an answer based on 4gram at first. If the system can't find answer, try again using trigram. We believe that this combined method can improve the recall and precision rate both of all. The table 7 supports this point that the performance rates are improved. The first graphs of figure 2 show the recall, precision and F1 rate of vocabulary questions and second graphs give the result of grammar questions. The last one is the final result graphs using trigram and 4gram together to find the answer for TOEIC test.

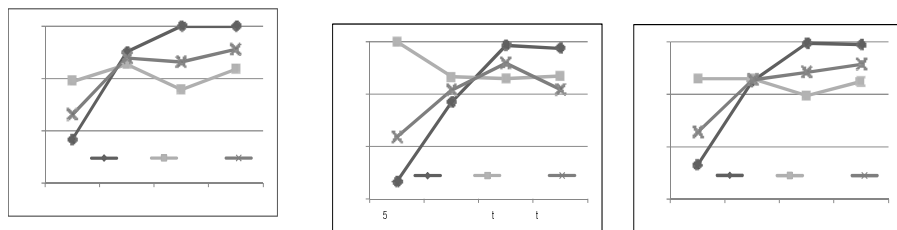


Fig. 2. First Graphs is result using vocabulary, second is grammar questions and third is total result

5 Conclusion and Future works

We evaluated a usability of Google n-gram data through applying n-gram data to part five in TOEIC test sentences provided by Hackers TOEIC. The testing results are based on frequencies that Google has been providing. We have not applied any statistic model to the system yet, but the results of recall and precision rate are reliable when the system using both trigram and 4gram. It has a limitation that the system is not able to find answer if there is no matched n-gram data in Google n-gram. This is means that the candidate n-gram from TOEIC test has to be placed in Google n-gram data to find the answer. To overcome this limitation, we are in progress to apply probability model such as HMM (Hidden Markov Model) and so on. Also, we need to apply our method to another English test data such as TOEFL and IELTS and so on to improve its objectivity and usability. We expect that Google n-gram data has a huge potential that applicable to the query recommendation system, automatic text completion, spell checking and natural language processing and so on. The obstacle for using Google n-gram is the size of data that is too huge. Even we filtered the special character of data, the size of 4gram is nearly 13GB and it is not only unacceptable for real time system but also needs lots of costing time and maintenance efforts. For these reason, the method to reduce size of n-gram is required. We believe that it is possible to reduce the size when we build specified n-gram data to fit user personal interests or characteristics.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0011656).

References

1. Yang, S., Zhu, H., Apostoli, A., Cao, P.: N-gram Statistics in English and Chines: Similarities and Differences. In: International Conference on Semantic Computing, pp. 454–460 (2007)
2. Brown, P.F., de Souza, P.V., Mercer, R.L., Pietra, V.J.D., Lai, J.C.: Class-Based n-gram Models of Natural Language. *Computational Linguistics* 18(4), 467–479 (1992)
3. Su, Z., Yang, Q., Lu, Y., Zhang, H.: WhatNext: A Prediction System for Web Requests using N-gram Sequence Models. *Web Information Systems Engineering* 1, 214–221 (2000)

4. Khudanpur, S., Wu, J.: A Maximum Entropy Language Model Integrating n-grams and Topic Dependencies for Conversational Speech Recognition. In: Proceedings of ICASSP 1999, pp. 553–556 (1999)
5. Hwang, M., Choi, D., Choi, J., Lee, H., Kim, P.: Text Editor based on Google Trigram and its Usability. In: UKSim 4th European Modelling Symposium on Computer Modelling and Simulation, pp. 12–15 (2010)
6. Siu, M., Ostendorf, M.: Variable N-Grams and Extensions for Conversational Speech Language Modeling. *IEEE Transactions on In Speech and Audio Processing* 8(1), 63–75 (2000)
7. Bergsma, S., Lin, D., Goebel, R.: Web-Scale N-gram Models for Lexical Disambiguation. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence, pp. 1507–1512 (2009)
8. Nakov, P., Hearst, M.: Search Engine Statistics Beyond the n-gram: Application to Noun Compound Bracketing. In: Proceedings of the 9th Conference on Computational Natural Language Learning, pp. 17–24 (2005)
9. Siivola, V., Pellom, B.L.: Growing an n-gram language model. In: Proceedings of 9th European Conference on Speech Communication and Technology (2005)

Automatic Documents Annotation by Keyphrase Extraction in Digital Libraries Using Taxonomy

Iram Fatima, Asad Masood Khattak, Young-Koo Lee, and Sungyoung Lee

Department of Computer Engineering, Kyung Hee University, Korea
{iram.fatima, asad.masood, sylee}@oslab.khu.ac.kr,
ykleee@khu.ac.kr

Abstract. Keyphrases are useful for variety of purposes including: text clustering, classification, content-based retrieval, and automatic text summarization. A small amount of documents have author-assigned keyphrases. Manual assignment of the keyphrases to existing documents is a tedious task, therefore, automatic keyphrase extraction has been extensively used to organize documents. Existing automatic keyphrase extraction algorithms are limited in assigning semantically relevant keyphrases to documents. In this paper we have proposed a methodology to assign keyphrases to digital documents. Our approach exploits semantic relationships and hierarchical structure of the classification scheme to filter out irrelevant keyphrases suggested by Keyphrase Extraction Algorithm (KEA++). Experiments demonstrate that the refinement improves the precision of extracted keyphrases from 0.19% to 0.38% while maintains the same recall.

1 Introduction

Keyphrases precisely express the primary topics and theme of documents and are valuable for cataloging and classification [1,2]. A keyphrase is defined as a meaningful and significant expression consisting of a single word, e.g., information, or compound words, e.g., information retrieval. Manual assignment and extraction of keyphrases is resource expensive and time consuming. It requires a human indexer to read the document and select appropriate descriptors, according to defined cataloguing rules. Therefore, it stimulates the need for automatic extraction of keyphrases from digital documents in order to deliver their main contents.

Existing approaches for keyphrase generation include keyphrase assignment and keyphrase extraction [3, 4]. In keyphrase assignment keyphrases are selected from a predefined list of keyphrases, thesaurus or subject taxonomy (i.e., Wordnet, Agrovoc) [4]. While in later approach all words and phrases included in the document are potential keyphrases [5, 6]. Phrases are analyzed on the basis of intrinsic properties such as frequency, length, and other syntactic information. The quality of the generated keyphrases by the existing approaches has not been able to meet the required accuracy level of applications [7,8].

The extraction algorithm used in this paper, KEA++, applies a hybrid approach of keyphrase extraction and keyphrase assignment [7-9]. KEA++ combines advantages of both, while avoiding their shortcomings. It makes use of a domain specific taxonomy to assign relevant keyphrases to documents. Limitation of this approach is that output keyphrases contain some irrelevant information along with the relevant

ones. For example, out of five keyphrases, two might fit well, while the remaining three have no semantic connection to the document (discussed later in the case study). The focus of this paper is to improve the semantic alignment procedure by exploiting different hierarchical levels of taxonomy. The proposed methodology is a novel approach of refinement, and comprises two major processes: (a) extraction and (b) refinement. KEA++ (Key Phrase Extraction Algorithm) [7-9] has been adopted for extracting keyphrases. The refinement process refines the result set of keyphrases returned by KEA++ using different levels of taxonomy. It detects the semantic keyphrases that are closer to human intuition as compared to KEA++. Experiments have been performed on dataset of 100 documents collected from the Journal of Universal Computer Science (JUCS¹). Experimental results show better precision (0.45) of proposed methodology in comparison to the precision (0.22) of KEA++ at the third level of the ACM Computing Classification² while maintaining the same recall.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 explains the proposed methodology of automatic keyphrase refinement. Results from JUCS dataset are given in Section 4. Conclusion together with possible future work discusses in section 5.

2 Related Work

Keyphrase extraction is a process to gather useful information from documents that help in describing the true content of the documents. KEA [10,11] identifies candidate phrases from textual sequences defined by orthogonal boundaries and extract relevant ones based two feature values for each candidate: the (1) TF x IDF measure, and (2) the distance from the beginning of the document to the first occurrence of a phrase. GenEx uses the genetic algorithm which is based on 12 numeric parameters and flags [12, 13]. This keyphrase extraction algorithm has two main components: (1) Genitor and (2) Extractor. Genitor is applied to determine the best parameter settings from the training data. Extractor combines a set of symbolic heuristics to create a ranked list of keyphrases.

The next approach is to use Natural Language Processing (NLP) tools in addition to machine learning, therefore the A.Hulth algorithm [14] compares different methods to extract candidate words and phrases like NP chunking, Parts of Speech (PoS) pattern matching, and trivial n-gram extraction. Candidates are filtered on the basis of four features: (1) term frequency, (2) inverse document frequency, (3) position of the first occurrence, and (4) PoS tag. In keyphrase assignment, a predefined set of keyphrases called the controlled vocabulary is used to describe the characteristics of documents in order to find the appropriate keyphrases, rather than individual phrases within them [15,16,17].

KEA++ is a hybrid of keyphrase assignment and keyphrase extraction [7-9]. It can involve taxonomy in extracting keyphrases from documents. Keyphrase selection is based on the computation of the naïve based statistical model and relations within the taxonomy. KEA++ takes a document, along with the taxonomy, as input for key-

¹ http://www.jucs.org/jucs_16_14

² <http://www.acm.org/about/class/1998/>

phrase extraction. KEA++ extracts terms from the documents (i.e., not explicitly mentioned in the document but existing in the taxonomy) by relating them to the terms of the taxonomy. The results of controlled indexing are highly affected by the parameter settings [18, 19]. The major parameters affecting the results are: vocabulary name, vocabulary format, vocabulary encoding, max. length of phrases, min. length of phrase, min. occurrence and no of extracted keyphrases.

The quality of the generated keyphrases by existing algorithms is inadequate, and they need to be improved for their applicability in real world applications. Some of the existing approaches use the taxonomy's hierarchy, yet it can be utilized in a more effective way. The results of KEA++ returned relevant keyphrases along with noise. In order to filter out the irrelevant information from the returned keyphrases of KEA++, there is a need for some refinement methodology that reduces the noise from the returned results of KEA++.

3 Proposed Methodology

Proposed methodology processes the returned results of KEA++ [7-9] by exploiting different hierarchical level of taxonomy. It involves two main steps: (a) extraction and (b) refinement. Extraction is a prerequisite of refinement. The refinement process is based on refinement rules. Refinement rules are applied to the set of keyphrases returned by KEA++ after the customized parameter settings. We set the vocabulary name parameter to the ACM computing classification in the SKOS format using UTF-8 encoding. It is used for the implementation and testing purpose of our algorithm, while our contribution is adoptable for other classification systems. The remaining refinement parameter settings of KEA++ are: *Max. Length of Phrases*: After analyzing the ACM Computing Classification, we set the value of this parameter to five words. *Min. Length of Phrase*: The minimum phrase length is one word in ACM taxonomy (i.e., hardware), which is the top level. We set the value of this parameter to two words because setting the value to one word provides many irrelevant keyphrases. *Min. Occurrence*: KEA++ recommends two words for this parameter in long documents. *No. of Extracted Keyphrases*: If the value of this parameter is less than ten words, for example four words, then KEA++ returns the top four keyphrases from the results it computes. These keyphrases might not be relevant. Other parameter settings as mentioned above can affect the results of this parameter.

3.1 Refinement Rules

These rules emphasize the importance of different levels/facts and their associated semantic relation in the training and semantic keyphrase extraction process. The basic idea filtered out semantic keyphrases according to the most related levels and available relations within different levels of taxonomy applying following rules:

Rule I: Adopting the Training Level: The training level is the hierarchical level of the taxonomy, adjusted for manually extracted keyphrases in documents. We adopt the KEA++ training level during the refinement process to extract the refined set of semantic keyphrases. The effective usage of the remaining rules depends on the accurate value of the training level of the taxonomy.

Rule II: Preserving the Training Level Keyphrases: We only preserve keyphrases aligned on the training level. KEA++ results have keyphrases that belong to different levels in the taxonomy. In addition to training level keyphrases, it might have upper level keyphrases and lower level keyphrases which do not contain information as relevant as the training level keyphrases. This rule selects the most relevant keyphrases from the resulting set of KEA++.

Rule III: Stemming the Lower Level General Keyphrases: In the ACM Computing Classification, there is the *general* category of keyphrases on each level of the hierarchy. If a keyphrase is aligned on a lower level than the training level (e.g., C.2.3.0), and associated with the general category in the lower level, then we stem the lower level keyphrase to its training level (e.g., C.2.3) keyphrases. This rule helps in extracting the maximum possible information from the lower level keyphrases in the presence of training level keyphrases.

Rule IV: Preserving the Lower Level Keyphrases: If the result set of KEA++ contains no training level keyphrases, then we preserve the lower level keyphrases from the result set of KEA++. This rule identifies the relevant keyphrases in the absence of training level keyphrases. In this case, lower level keyphrases represent the documents alignment on more accurate nodes, which belong to more specific keyphrases in the taxonomy.

Rule V: Identifying and Preserving the Training Level Equivalent Keyphrase: Different keyphrases aligned to separate categories of the ACM Computing Classification can be semantically equivalent, e.g., *Control Structures and Microprogramming* (B.1) is equivalent to *Language Classifications* (D.3.2). The procedure is carried out by first identifying the training level keyphrases from the set of upper level keyphrases. If the upper level has equivalent keyphrases of the training level, then preserve the training level keyphrase in the refined result set and discard the upper level keyphrase. But before discarding them, it helps in preserving any training level keyphrases equivalent to an upper level keyphrase.

Rule VI: Removing Redundant Keyphrases: After stemming the lower level general keyphrases and identifying and preserving the training level equivalent keyphrases, the result might contain redundant keyphrases (i.e., C.2.3, C.2.3, D.4.5). Remove the redundant keyphrases from the set of refined keyphrases (i.e., C.2.3, D.4.5).

3.2 Refinement Algorithm

The algorithm describes the flow of refinement rules that is illustrated in Algorithm 1. Extraction of the semantic keyphrases is the essential requirement of the refinement process. First of all parameters of the extraction algorithm KEA++ are set with respect to keyphrases' length in the taxonomy and length of the documents. Secondly train KEA++ on the set of documents using taxonomy. Then apply KEA++ on actual documents (data). Adopting the training level for the refinement rules has primary

Input: Training:

- (a) Set the parameters of KEA++ by keeping in view the keyphrase length in the taxonomy and documents type
- (b) Documents along with their keyphrase and taxonomy

Dataset for Extraction:

- (a) Documents with unknown keyphrases

Output: Set of refined keyphrases

```

1: TrainLevel ← KEA++ TrainLevel //(Rule I)
2: resultSet [] ← returned keyphrases by KEA++[]
3: resultSet [] ← level labels (Resultset [])
4: for resultSet[] > empty do
5:   if (resultSet(training level)) then
6:     if (keyphrase level = lower level keyphrases) then
7:       processSet[] = preserving lower level keyphrases //Rule II
8:     else
9:       processSet[] ← identifying and preserving training level equivalent //Rule V
10:      processSet[] ← remove redundant keyphrases //Rule VI
11:      refineSet[] ← processSet[]
12:    end if
13:   else
14:     if (keyphrase level = training level) then
15:       refineSet[] ← processSet[]
16:     else
17:       if (keyphrase level = upper level) then
18:         processSet[] ← identifying and preserving training level
19:           equivalent keyphrases //Rule V
20:       else
21:         processSet[] ← stemming lower level general keyphrases //Rule III
22:       end if
23:       processSet[] ← remove redundant keyphrases //Rule VI
24:       refineSet[] ← processSet[]
25:     end if
26:   end if
27: end for
27: return refineSet[] //refine result set of semantic keyphrases.
  
```

Algorithm 1. Refinement Algorithm

importance because it guides the remaining rules in their process. The keyphrases returned by KEA++ is processed to get its level label in the taxonomy. Identify level labels is required before applying the refinement rules because they represent the hierarchical order of the keyphrases as described in steps 1 to 3 of Algorithm 1. If the KEA++ result has training level keyphrase then these training level keyphrases are retained in the result set as shown in steps 5 to 12 of Algorithm 1. Lower level keyphrases are stemmed to their training level keyphrases and kept in the result set if they are associated with the general category at the lower level in taxonomy. Otherwise lower level keyphrases are discarded. Upper level keyphrases are handled according to Rule-V and discarded after indentifying and preserving their equivalent keyphrases from taxonomy which belong to the same level of training level keyphrases. If the initial result does not contain any training level keyphrases then lower level keyphrases of the result are preserved and added in the final refined result. Upper level keyphrases in the initial result are handled according to Rule-V and discarded after indentifying and preserving their equivalent keyphrases from taxonomy which belong to the same level of training level keyphrases. This process is executed from steps 13 to 21 of the algorithm. Finally redundant keyphrases are removed from the final refined set of keyphrases.

3.3 Case Study

To focus more on the refinement process proposed in this paper, a case study is presented in which training models is trained on third level of ACM Computing Classification.

Table 1 illustrates the information about the documents used in the evaluation (available on the web). The first column of Table 2 represents the semantic keyphrases returned by KEA++ after applying the parameters proposed in the refinement process.

Table. 1. Sample document

Title	Passive Estimation of Quality of Experience
Identification Key:	JUCS, Vol. 14, Issue 5, year 2008
Manual Alignment:	C.2.3 (Network Operations), C.4 (Performance of System)

Extracted semantic keyphrases align the document on five nodes of the ACM Computing Classification, while document is manually aligned on two nodes. Extracted keyphrases include both relevant and noise/irrelevant keyphrases.

Table 2. Results of the refinement algorithm

KEA++ Results	Level Label	Refined Results
Network Management	C.2.3.0	
Distributed Functions	G.3.2	G.3.2
Network Operations	C.2.3	C.2.3
Approximate Methods	I.4.2.1	
Data Structures	E.1	

We select the level labels of the keyphrases from the ACM Computing Classification as shown in the second column of Table 2. Keyphrases with their associated level labels show alignment of the document with different depths in the ACM Computing Classification. The refined results are calculated after applying the rules of the refinement algorithm. The results are quite improved in that they include an exact match with one relevant keyphrase. The whole process of refinement involves following steps. After identifying the level labels of keyphrases, the refinement algorithm checks whether the level labels of the keyphrases contain the training level. If it has the training level (C.3.2 and G.3.2), then it preserves these training level keyphrases. Now it identifies whether the result set has upper level keyphrases. As it has upper level keyphrases (E.1), the rule V is applicable here. The existence of low level keyphrases belongs to a general category (C.2.3.0), so it is stemmed to the training level (C.2.3) keyphrase. In the end, the algorithm removes the redundant keyphrases (C.2.3, C.2.3) and declares the result set as the final refined result set (C.2.3, G.3.2), as shown in the third column of Table 2. This case study explains that lower level keyphrase extraction is not significant when compared to training level keyphrases.

4 Results and Evaluation

In this section, the results of manual annotation, KEA++, and the proposed refinement algorithm are compared. The precision of the refinement algorithm is evaluated on various hierarchical levels as provided in the manual annotations of the dataset. Evaluation contains both (a) *keyphrase based evaluation* and (b) *document based*

evaluation. We modeled the ACM Computing Classification in SKOS format³. Dataset⁴ from (JUCS) has a key file along with the document file which contains manually assigned keyphrases. Two experiments have been performed on dataset of 100 documents. Experiment I is based on 65 documents, 50 for training and 15 for testing purpose. Experiment II is performed on 100 documents, 70 documents were used for training and 30 were used for testing. To evaluate the precision of the system, documents which were used as testing documents in the first experiment were replaced with the training documents.

4.1 Keyphrase Based Evaluation

It is used to estimate the performance in terms of returned keyphrases. It is further categorized as (a) keyphrases returned per average number of documents, and (b) total returned keyphrases. In the former category, we compare results among (1) manual annotation, (2) KEA++, and (3) the refinement rules. Fig. 1 (a) and (b) show that the number of KEA++ returned keyphrases lies between 0 to 7, 0 to 13, respectively. Manual annotation of the same documents varies from 1 to 4, 1 to 5, respectively. Keyphrases returned by the refinement algorithm range from 0 to 3. It shows that the refinement algorithm reduces the number of keyphrases returned against an average number of documents as compared to KEA++ and closer to the manual annotation. However, it does not affect the precision of correctly aligned documents, as shown in the next subsection.

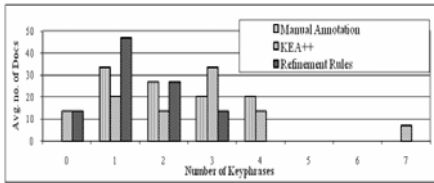


Fig. 1(a). Keyphrases returned per average no of documents of experiment I

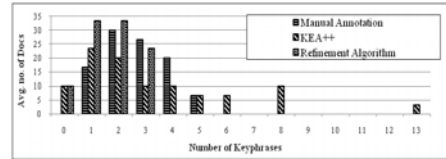


Fig. 1(b). Keyphrases returned per average no of documents of experiment II

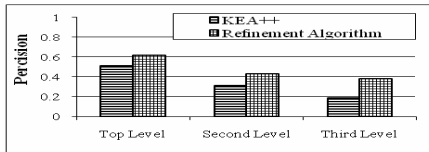


Fig. 2(a). Precision against total keyphrases returned of experiment I

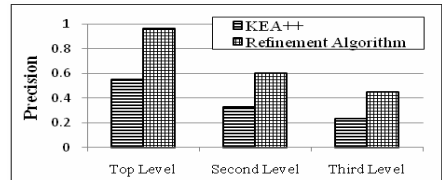


Fig. 2(b). Precision against total keyphrases returned of experiment II

³ http://uclab.khu.ac.kr/ext/ACM_Computing_Classification.rar

⁴ [http://uclab.khu.ac.kr/ext/Dataset_Journal_of_Universal_Computer_Science\(JUCS\).rar](http://uclab.khu.ac.kr/ext/Dataset_Journal_of_Universal_Computer_Science(JUCS).rar)

Total returned keyphrases compares the precision, and recall of both KEA++ and the refinement rules. Fig. 2 (a) and (b) show the precision of both algorithms. In the case of KEA++, the precision 0.19 and 0.23, while the refinement algorithm shows more precise results, with values of 0.38 and 0.45, respectively. The recall comparison is illustrated in Fig. 3 (a) and (b) and the recall of KEA++ and the refinement algorithm are the same in the case of both experiments.

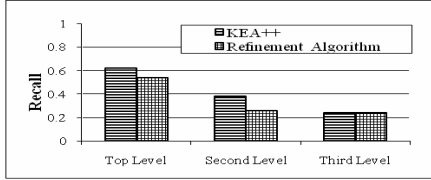


Fig. 3(a). Recall against total keyphrases returned of experiment I

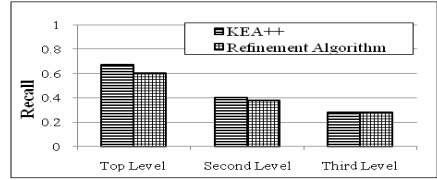


Fig. 3(b). Recall against total keyphrases returned of experiment II

4.2 Document Based Evaluation

The document based evaluation verifies the performance of both algorithms against correctly aligned documents. We do not consider the recall calculation as the number of documents is the same in both cases. This evaluation criterion is further categorized as (a) the totally matched result and (b) the approximate matched result. The totally matched result contains all of the manually annotated keyphrases of the particular document, while the approximate matched result comprises a subset of manually annotated keyphrases of the particular document.

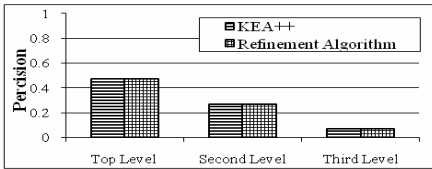


Fig. 4(a). Precision of totally matched results of experiment I

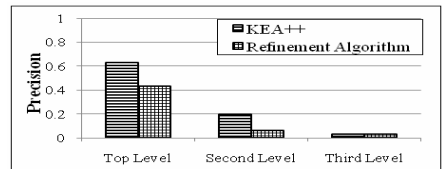


Fig. 4(b). Precision of totally matched results of experiment II

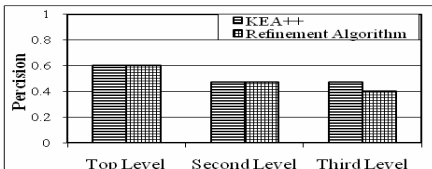


Fig. 5(a). Precision of approximate matched results of experiment I

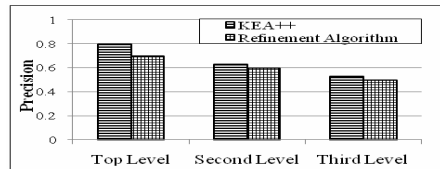


Fig. 5(b). Precision of approximate matched results of experiment II

The totally matched result is a more conservative approach because it ignores the approximately aligned documents. Figure 4 (a) and (b) illustrate the precision for the totally matched results, the precisions is same on the third level of taxonomy. Furthermore, the refinement rules returned a reduced number of keyphrases. Figure 5 (a) and (b) show the precision of both approaches for the approximate matched results. Due to the reduced number of keyphrases per average number of documents, the precision is comparatively lower on the third level of the taxonomy.

Table 3. Precision, recall, and F-measure statistics

Experiments	Results of [9].	Experiment. I		Experiment. II	
Algorithms	KEA++	KEA++	Refined Results	KEA++	Refined Results
Documents	200	65		100	
Avg. # of Manual Annotation	5.4	2.27		2.35	
Precision	0.28	0.19	0.38	0.23	0.45
Recall	0.26	0.24	0.24	0.28	0.28
F-measure	0.25	0.21	0.29	0.25	0.34

Table 3 shows precision, recall, and F-measure statistics, in results of [9], the precision, recall, and F-measure of KEA++ are 0.28, 0.26, and 0.25, respectively, while the average number of manual annotation is 5.4 per document in the dataset of 200 documents. The precision, recall, and F-measure of KEA++ on our experiments are different. Obviously, the precision and recall is affected by a change in the number of documents in the dataset, and the average number of manual annotations per document in each dataset. In the case of the refinement algorithm, the precision has been improved in all performed experiments while the recall is same as shown in Table 3.

5 Conclusion and Future Work

An exponential growth of electronic documents requires extraction of keyphrases for semantic alignment. The proposed algorithm processes the results returned by KEA++ and removes irrelevant keyphrases by exploiting the hierarchical structure of taxonomy to achieve better accuracy. The refinement algorithm provides the functional flow to the refinement rules. To evaluate the methodology, JUCS dataset is used in different experiments and results show obvious improvement in the precision as compared to KEA++, while maintaining the same recall. Currently the focus was on a single training level in applying the refinement algorithm so in future this refinement algorithm can be extended to involve more than one training level while executing the refinement algorithm in order to achieve more accurate results and to make the methodology scalable.

Acknowledgement. This research was supported by the MKE (Ministry of knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency)" (NIPA-2009-(C1090-0902-0002)).

References

1. Liu, Z., Li, P., Zheng, Y., Sun, M.: Clustering to Find Exemplar Terms for Keyphrase Extraction. In: *ACL Proceedings of Empirical Methods on Natural Language Processing*, Singapore, pp. 257–266 (2009)
2. Frank, E., Paynter, G.W., Witten, W.I.H., Gutwin, C., Nevill-Manning, C.G.: Domain specific keyphrase extraction. In: *Sixteenth International Joint Conference on Artificial Intelligence*, Sweden, pp. 668–673 (1999)
3. Roberto, O., Pinto, D., Tovar, M.: BUAP: An Unsupervised Approach to Automatic Keyphrase Extraction. In: *23rd 5th International Workshop on Semantic Evaluation ACL*, Sweden, pp. 174–177 (2010)
4. Barker, K., Cornacchia, N.: Using noun phrase heads to extract document keyphrases. In: *Conference on Artificial Intelligence, Canadian*, pp. 40–52 (2000)
5. Jacquemin, C., Bourigault, D.: Term extraction and automatic indexing. *The Oxford Handbook of Computational Linguistics*, pp. 559–616 (2003)
6. Jones, S., Paynter, G.: Automatic extraction of document keyphrases for use in digital libraries: evaluation and applications. *J. of the American Society for Information Science and Technology* 53(8), 653–677 (2002)
7. Medelyan, O., Witten, H.I.: Thesaurus Based Automatic Keyphrase Indexing. In: *Joint Conference on Digital Libraries, USA*, pp. 296–297 (2006)
8. Medelyan, O., Witten, H.I.: Semantically Enhanced Automatic Keyphrase Indexing, *WiML* (2006)
9. Medelyan, O., Witten, H.I.: Thesaurus-based index term extraction for agricultural documents. In: *6th Agricultural Ontology Service workshop at EFITA, Portugal* (2005)
10. Witten, H.I., Paynter, G.W., Frank, E., Gutwin, C., Nevill-Manning, C.G.: KEA: Practical automatic keyphrase extraction, pp. 254–256 (1999)
11. Witten, H.I., Paynter, G.W., Frank, E.: Gutwin: Kea: Practical automatic keyphrase extraction. *Design and Usability of Digital Libraries*, 129–152 (2005)
12. Turney, P.D.: Learning algorithms for keyphrase extraction. *Information Retrieval*, 303–336 (2003)
13. Turney, P.D.: Coherent keyphrase extraction via web mining. *J. of the American Society for Information Science and Technology*, 434–439 (2003)
14. Hulth, A.: Improved automatic keyword extraction given more linguistic knowledge. In: *Empirical Methods in Natural Language Processing*, pp. 216–223 (2003)
15. Kim, S.N., Kan, M.Y.: Re-examining automatic keyphrase extraction approaches in scientific articles. *MWE*, 9–16 (2009)
16. Barker, K., Cornacchia, N.: Using noun phrase heads to extract document keyphrases, *Canadian*, pp. 40–52 (2000)
17. Paice, C., Black, W.: A three-pronged approach to the extraction of key terms and semantic roles. In: *Recent Advances in Natural Language Processing*, pp. 357–363 (2003)
18. Fatima, I., Khan, S., Latif, K.: Refinement Methodology for Automatic Document Alignment Using Taxonomy in Digital Libraries. In: *ICSC USA*, pp. 281–286 (2009)
19. El-Beltagy, S.R., Rafea, A.: KP-Miner: A keyphrase extraction system for English and Arabic documents 34(1), 132–144 (2009)

IO-Aware Custom Instruction Exploration for Customizing Embedded Processors

Amir Yazdanbakhsh¹ and Mostafa E. Salehi²

¹ Islamic Azad University, Bandarabbas Branch, Iran

² Islamic Azad University, Qazvin Branch, Qazvin, Iran
amir.yazdanbakhsh@gmail.com, mersali@ut.ac.ir

Abstract. This paper describes a methodology for identifying custom instructions for critical code segments of embedded applications, considering the available data bandwidth constraint between custom logic and the base processor. Our approach enables designers to optionally constrain the number of input and output operands for custom instructions to reach the acceptable performance. We describe a design flow to establish the desired performance. We study the effects of input/output constraints and registerfile read/write ports on overall speedup of the system. Our experiments show that, in most cases, the solutions with the highest merit are not identified with relaxed input/output constraints. Results for packet-processing benchmarks covering cryptography, lookup, and classification show speed-up up to 40%.

Keywords: Custom instruction exploration; Extensible processors; IO constraint; Packet-processing applications; Network processors; Register file exploration.

1 Introduction

Embedded systems are special purpose systems which perform specific tasks with predefined performance requirements. Using a general purpose processor for such systems usually results in a design that does not meet the performance demands of the application. On the other hand, ASIC design cycle is too costly and too slow for the embedded application market. Recent developments in customized processors significantly improve the performance metrics of a general purpose processor by coupling it with an application specific hardware.

Designers carefully analyze the characteristics of the target application and fine tune the implementation to achieve the best performance. The most popular strategy is to build a system consisting of a number of specialized application-specific functional units coupled with low-cost and optimized general-purpose processors as a based processor with basic instruction set (e.g. ARM [1] or MIPS [2]). The base processor is augmented with custom-hardware units that implement application-specific instructions.

There are a number of benefits in augmenting the core processor with new instructions. First, the system is programmable and support modest changes to the application, such as bug fixes or incremental modifications to a standard. Second, the computationally intensive portions of applications in the same domain are often similar in structure. Thus, the customized processors can often be generalized to have applicability across a set of applications.

In recent years, customized extensible processors offer the possibility of extending the instruction set for a specific application. A customized processor consists of a microprocessor core that is tightly coupled with functional units (FUs) that facilitates critical parts of the application to be implemented in hardware using a specialized instruction set. In the context of customized processors, hardware/software partitioning is done at the instruction level. Given the application code, basic blocks of the application are transformed into data-flow graphs (DFGs), where the graph nodes represent operations similar to those in assembly languages, and the edges represent data dependencies between the nodes. Instruction set extension exploits a set of custom instructions (CIs) to achieve considerable performance improvements by executing the hot-spots of the application on hardware.

Extension of the instruction set with new custom instructions can be divided into instruction generation and instruction selection phases. Given the data-flow graph (DFG) code, instruction generation consists of clustering some basic operations into larger and more complex operations. These complex operations are entirely or partially identified by subgraphs which can cover the application graph. Once the subgraphs are identified, these are considered as single complex operations and they pass through a selection process. Generation and selection are performed with the use of a guide function and a cost function respectively, which take into account constraints that the new instructions have to satisfy for hardware implementation.

Partitioning an application into base-processor instructions and custom instructions is done under certain constraints. First, there is a limited area available in the custom logic. Second, the data bandwidth between the base processor and the custom logic is limited, and the data transfer costs have to be explicitly evaluated. Next, only a limited number of input and output operands can be encoded in a fixed-length instruction word. The speed-up obtainable by custom instructions is limited by the available data bandwidth between the base processor and the custom logic. Extending the core registerfile to support additional read and write ports improves the data bandwidth. However, additional ports result in increased registerfile size and cycle time. This paper presents a systematic approach for generating and selecting the most profitable custom instruction candidates. Our investigations show that considering the architectural constraints in the custom instruction selection leads to improvements in the total performance.

The remainder of this paper is organized as follows: in the following section, we review some existing work and state the main contributions of this paper. Section III describes the overall approach of the work. In Sections IV, we discuss about the experimental setup, and in Section V we provide experimental results for packet-processing applications. Finally, the paper concludes in Section VI.

2 Related Work

The bandwidth of registerfile is a limitation in improving the performance of customized processors. Many techniques have been proposed to cope with this limitation. The Tensilica Xtensa [3] uses custom state registers to explicitly move additional input and output operands between the base processor and custom units. Use of shadow registers [4] and exploitation of forwarding paths of the base processor [5] can improve the data bandwidth.

Kim [6] developed two techniques for reducing the number of registerfile ports. The techniques are based on a pre-fetch technique with pre-fetch buffer and request queue, and delayed write-back technique. Park [7] proposed two different techniques for reducing the read/write ports of the registerfiles including bypass hint technique and the register banking to decrease the required read and write ports, respectively. Karuri [8] developed a cluster-based registerfile to overcome the number of access port limitations of the registerfile. The proposed technique in [8] is adapted from the idea of VLIW architectures, and uses more area in comparison with conventional registerfiles. All of the investigated techniques tried to deal with registerfile IO constraints by changing the registerfile architecture. On the other hand, there are some techniques that address the CI selection.

Sun et al. [9] imposed no explicit constraints on the number of input and output operands for custom instructions and formulated custom instruction selection as a mathematical optimization problem. Atasu et al. [10] introduced constraints on the number of input and output operands for subgraphs and showed that exploiting constraint-based techniques could significantly reduce the exponential search space. The work of Atasu et al. showed that clustering based approaches (e.g. [11]) or single output operand restriction, (e.g. [12]) could severely reduce the achievable speed-up using custom instructions. In [13], Biswas et al. proposed a heuristic which is based on input and output constraints. This approach does not evaluate all feasible subgraphs. Therefore, an optimal solution is not guaranteed. In [14], Bonzini and Pozzi derived a polynomial bound on the number of feasible subgraphs while the number of inputs and outputs for the subgraphs are fixed. However, the complexity grows exponentially as the input/output constraints are relaxed. In [15], Yu and Mitra enumerate only connected subgraphs having up to four input and two output operands and do not allow overlapping between selected subgraphs. Atasu [16] proposed a technique for custom instruction identification. The proposed technique generates a set of candidate templates by using ILP and the best CIs are selected based on the Knapsack model. Beside this technique, the authors did a study on the effect of the registerfile ports effect on the performance of the selected CIs. Additionally, the researches that work on CI identification techniques, have shown that increasing the number of input and output ports of the selected CI, improves the performance.

In this paper, we target customizable architectures for exploring the data bandwidth between the base processor and the custom logic considering the available registerfile ports. Our main contributions are as follows: 1) a novel methodology for extracting custom-instructions, which optionally constrains the number of input and output operands of custom instructions and explicitly evaluates the data-transfer costs in terms of latency 2) an evaluation of the impact of registerfile port constraints on the execution cycle count of the packet-processing benchmarks.

3 The Proposed Framework

Fig. 1 depicts the complete flow of the proposed IO-aware framework for obtaining the optimum set of custom instructions in terms of speedup. The CI generation and selection evaluates system performance considering registerfile with different number of read and write ports. Our proposed framework explores different configurations in terms of number of read and writes ports and finally selects and generates the set of custom instructions for that registerfile configuration.

The custom instruction generation and selection framework contains two main parts: match enumeration and template generation. Match enumeration algorithm is based on binary search tree [17] and explores all possible custom instructions considering the following constraints for each custom instruction:

- 1) The number of input/output ports should be less or equal to the allowed input/output ports.
- 2) Each custom instruction should not include any memory operations such as Load or Store instructions.
- 3) The identified custom instruction should be convex.

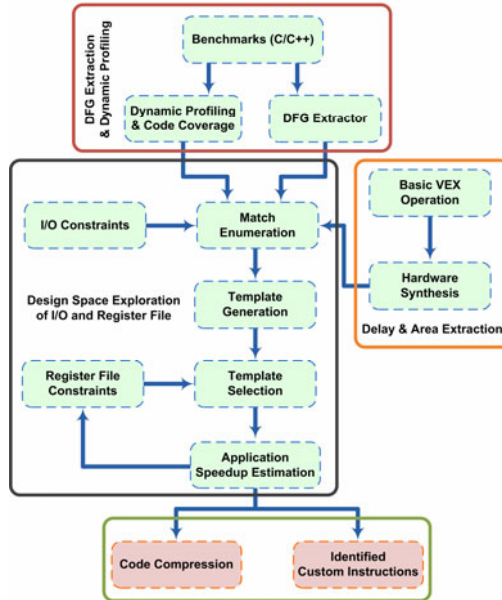


Fig. 1. The Proposed IO-Aware Framework

After this part all structurally equivalent valid custom instructions are categorized into isomorphic classes, called templates. Each template is assigned a number that shows the performance gains of that template regarding the number of read and write ports of registerfile. According to our assumptions, the following formula [24] demonstrates the speedup estimated for each template:

$Speedup_{Template} =$

$$I_{tr} * \left(SW_{Latency} - [HW_{Latency}] - \left\lceil \frac{(CI_{in} - RF_{in})}{RF_{in}} \right\rceil - \left\lceil \frac{(CI_{out} - RF_{out})}{RF_{out}} \right\rceil \right)$$

The evaluated $Speedup_{Template}$ is assigned to each template and then, the objective is to find the maximum weighted independent set out of these templates.

4 Experimental Setup

We define a single-issue baseline processor based on MIPS architecture including 32 general-purpose 32-bit registers in a 5-stage pipeline. We do not constrain the number of input and output operands in custom instruction generation. However, we explicitly account for the data-transfer cycles between the base processor and the custom logic if the number of inputs or outputs exceeds the available registerfile ports. CHIP [16] assumes two-cycle software latencies for integer multiplication instructions and single-cycle software latencies for the rest of the integer operations. Since the latency of arithmetic and logical operations are considerably different, assuming single-cycle latency for all of the logical and arithmetic instructions may lead to non-optimal results. We assume a more accurate and fair value for latency of different operations. In software we assume single-cycle latency for each operation excluding memory operations. However, in hardware, we evaluate the latency of arithmetic and logic operators by synthesizing them with a 90-nm CMOS process and normalize them to the delay of the MIPS processor.

VEX is composed of many components that their main objective is to compile, simulate, and analyze C programs for VLIW processor architectures. VEX has also the capability to extract DFGs and CFGs from C/C++ programs. We use this possibility to extract CFGs from packet-processing applications extracted from PacketBench [20]. The extracted CFGs are converted to an intermediate format that is known for our custom instruction selection framework. On the other hand, with the help of *geov* and *gprof* [22] in conjunction with GCC, the code coverage and the number of iterations of basic blocks of codes are calculated in the dynamic execution of domain-specific benchmarks. These numbers and the intermediate format are processed by our custom-instruction-selection framework to find a set of CIs that increase the performance of packet-processing benchmarks. The selected applications are IPv4-radius and IPv4-trie as RFC1812-compliant look-up and forwarding algorithms [19], a packet-classification algorithm called Flow-Class [20], internet protocol security (IPSec) [21] and message-digest algorithm 5 (MD5) as payload-processing applications.

The accumulated software latencies of a custom instruction candidate subgraph estimate its execution cycles in a single-issue processor. The hardware latency of a custom instruction as a single instruction is approximated by the number of cycles equal to the ceiling of the sum of hardware latencies over the custom instruction critical path. The difference between the software and the hardware latency is used to estimate the speedup. Input and output or IO violations are taken into account by the penalties in the fitness function. We do not include division operations in custom instructions due to their high latency and area overhead. We have also excluded

memory access instructions from custom instructions, to avoid nondeterministic latencies due to the memory system. We assume that given RF_{in} read ports and RF_{out} write ports supported by the registerfile, RF_{in} input operands and RF_{out} output operands can be encoded within a single-instruction word.

5 Experimental Result

In this paper, we use input/output constraints to control the granularity of the custom instructions and to capture structural similarities within an application. Our motivation is that applications often contain repeated code segments that can be characterized by the number of input and output operands. When the input/output constraints are tight, we are more likely to identify fine-grain custom instructions. Relaxation of the constraints results in coarse-grain custom instructions (i.e., larger data-flow sub-graphs). Coarse-grain instructions are likely to provide higher speed-ups, although at the expense of increased area.

We have modeled and synthesized registerfiles with different input and output ports and have compared the area overhead of these registerfiles across the standard MIPS registerfile. We introduce the $(RFin, RFout)$ notation as the (number of read ports, number of write ports) to recognize different configuration of registerfiles. Based on this definition MIPS registerfile which has two read ports and one write port is defined as a $(2, 1)$ registerfile. The area overhead of increasing the read and write ports of the registerfile are evaluated by synthesizing different configurations of registerfile in terms of read/write ports with 90nm CMOS process. These values which are normalized to the area of MIPS registerfile (i.e. $(2, 1)$) are shown in Fig. 2.

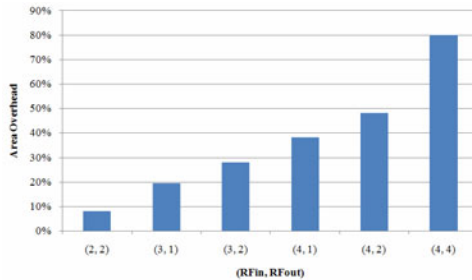


Fig. 2. Area Overhead of Increasing Input and Output Ports of Different Registerfile Vs. $(2, 1)$ Registerfile

In embedded processors memory modules are the major source of power consumption and also impose hardware cost. Therefore, reduction in the code can improve both the memory cost and power consumption. The custom instructions can be used as a strategy to reduce the application code size. The code compression that can be achieved by a set of custom instructions is calculated by following formula:

$$\text{Code Compression} = \frac{\sum_{i=0}^n (\text{Number of Nodes in } CI_i - 1)}{\text{Number of Operation in the Application}}$$

The numerator represents the number of operations that will be saved by a set of custom instructions and the denominator shows the number of all operations of an application without any custom instructions. The code compression of the representative applications is shown in Fig. 3. IPSec have the highest code compression among the other packet processing applications. It means that the identified CIs in IPSec have compact more nodes into one custom instruction. This high code compression is because IPSec includes many logical operations (as indicated in [23]) that can be easily merged into one custom instruction.

In Fig. 4, we analyze the effect of different input and output constraints (i.e. CI_{in} , CI_{out}) on the achieved speedup of custom instructions versus registerfiles with different (RF_{in} , RF_{out}) for a payload-processing (i.e. MD5) and a header-processing application (i.e. IPv4-trie). As shown in this figures when the custom instructions are generated based on $(CI_{in}, CI_{out}) = (3, 2)$ and the registerfile constraint is $(3, 2)$, the speedup is higher than when $(CI_{in}, CI_{out}) = (\infty, \infty)$ and $(RF_{in}, RF_{out}) = (3, 2)$. Therefore, both of (CI_{in}, CI_{out}) and (RF_{in}, RF_{out}) should be considered for achieving the best speedup (indicated by an arrow in for MD5 application). Another important observation is that application speedup is almost saturated in $(CI_{in}, CI_{out}) = (6, 3)$ and $(RF_{in}, RF_{out}) = (6, 3)$ and the application speedup of this point is about 5 to 10% higher than $(CI_{in}, CI_{out}) = (3, 2)$ and $(RF_{in}, RF_{out}) = (3, 2)$. Furthermore, the former achieves this improvement with a greater area overhead compared to the latter configuration.

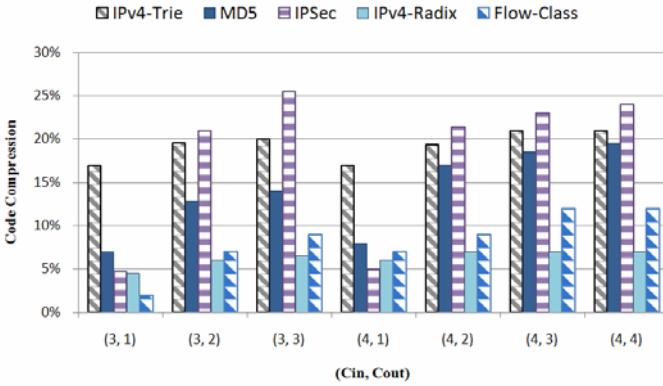


Fig. 3. The percentage of code compression for the representative applications

The overall application speedup of the representative applications versus $(3, 2)$, $(4, 2)$, and $(4, 4)$ registerfiles is also summarized in Fig. 5. As shown, IPv4-trie, IPSec, and MD5 have up to 30% speedup. However, as evaluated in 23, Flow-class/IPv4-radix has the most memory/branch instructions that lead to lower performance improvements when enhanced with custom instructions.

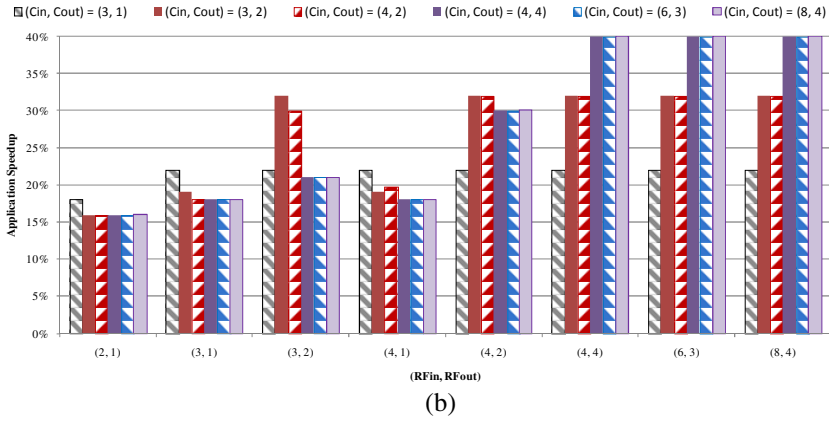
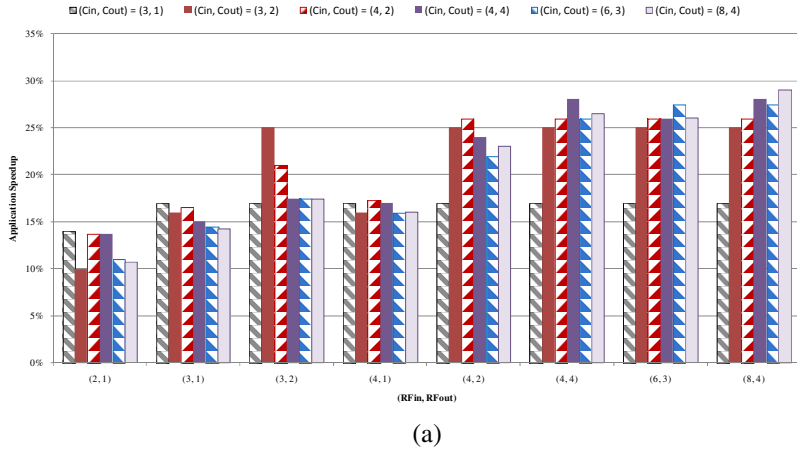


Fig. 4. Effect of IO constraints and Registerfile Ports on the Achieved Performance Improvement for a header- and payload-processing application a) MD5, b) IPv4-trie

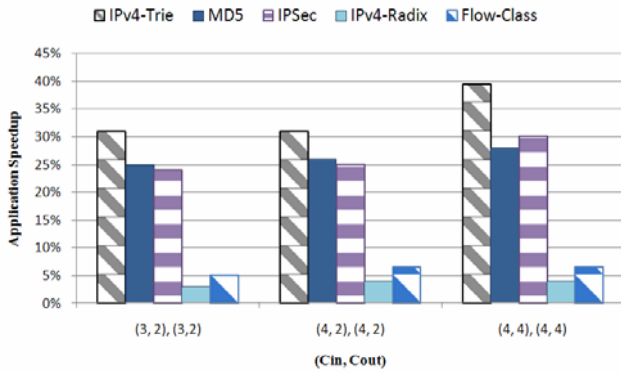


Fig. 5. Application Speedup of different packet-processing applications versus three registerfile configurations

6 Conclusion

We have presented a methodology for exploring custom instructions for critical code segments of packet processing applications, considering the registerfile IO constraint between custom logic and the base processor. Our experiments show that, in most cases, the solutions with the highest merit are not identified with relaxed input/output constraints. Results for packet-processing benchmarks covering cryptography, lookup, and classification are shown, with speed-ups up 40% and code compression up to 25%. It is also shown that applications that include high logical operations such as IPSec and MD5 are the most appropriate candidates for custom instruction identification. The structure of the program also affects the application speedup that can be obtained by custom instructions. As stated in 23 IPv4-Radix has more logical and arithmetic operations than IPv4-Trie that inclines it to get high improvement in performance when augmented by valuable custom instructions, but many load and branches happens between the logical and arithmetic operations. This reason cause the custom instruction identification algorithm not find worthy CIs. In compare to IPv4-Radix, the IPv4-Trie has more memory operation but because of the program structure, the custom instruction identification algorithm can find many appropriate and worthy custom instructions that improve the performance of the IPv4-Trie more than the improvement that is obtainable by CIs for IPv4-radix. On the other hand for the applications those have many branch and memory operations another strategy, such as branch predication and increasing memory ports, may help to increase their performance.

References

1. ARM, The architecture for the digital world, <http://www.arm.com>
2. MIPS technologies Inc., <http://www.mips.com>
3. Gonzalez, R.E.: XTENSA: A configurable and extensible processor. *IEEE Micro* 20, 60–70 (2000)
4. Cong, J., et al.: Instruction set extension with shadow registers for configurable processors. In: *Proc. FPGA*, pp. 99–106 (February 2005)
5. Jayaseelan, R., et al.: Exploiting forwarding to improve data bandwidth of instruction-set extensions. In: *Proc. DAC*, pp. 43–48 (July 2006)
6. Kim, N.S., Mudge, T.: Reducing register ports using delayed write-back queues and operand pre-fetch. In: *Proceedings of the 17th Annual International Conference on Supercomputing*, pp. 172–182 (2003)
7. Park, Powell, M.D., Vijaykumar, T.N.: Reducing register ports for higher speed and lower energy. In: *Proceedings of the 35th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 171–182 (2002)
8. Karuri, K., Chattopadhyay, A., Hohenauer, M., Leupers, R., Ascheid, G., Meyr, H.: Increasing data-bandwidth to instruction-set extensions through register clustering. In: *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 166–177 (2007)
9. Sun, F., et al.: A scalable application-specific processor synthesis methodology. In: *Proc. ICCAD*, San Jose, CA, pp. 283–290 (November 2003)

10. Atasu, K., Pozzi, L., Ienne, P.: Automatic Application-Specific Instruction-Set Extensions under Microarchitectural Constraints. In: Proc. of the 40th Annual Design Automation Conference, pp. 256–261. ACM, Anaheim (June 2003)
11. Baleani, M., et al.: HW/SW partitioning and code generation of embedded control applications on a reconfigurable architecture platform. In: Proc. 10th Int. Workshop HW/SW Codesign, pp. 151–156 (May 2002)
12. Alippi, C., et al.: A DAG based design approach for reconfigurable VLIW processors. In: Proc. DATE, Munich, Germany, pp. 778–779 (March 1999)
13. Biswas, P., et al.: ISEGEN: Generation of high-quality instruction set extensions by iterative improvement. In: Proc. DATE, pp. 1246–1251 (2005)
14. Bonzini, P., Pozzi, L.: Polynomial-time subgraph enumeration for automated instruction set extension. In: Proc. DATE, pp. 1331–1336 (April 2007)
15. Yu, P., Mitra, T.: Satisfying real-time constraints with custom instructions. In: Proc. CODES+ISSS, Jersey City, NJ, pp. 166–171 (September 2005)
16. Atasu, K., Ozturan, C., Dundar, G., Mencer, O., Luk, W.: CHIPS: Custom hardware instruction processor synthesis. *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems* 27, 528–541 (2008)
17. Pozzi, L., Atasu, K., Ienne, P.: Exact and Approximate Algorithms for the Extension of Embedded Processor Instruction Sets. *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems* 25, 1209–1229 (2006)
18. Fisher, J.A., Faraboschi, P., Young, C.: *Embedded Computing: A VLIW Approach to Architecture, Compilers and Tools*. Elsevier Morgan Kauffman, New York (2005)
19. Baker, F.: Requirements for IP version 4 routers. RFC 1812, Network Working Group (June 1995)
20. Ramaswamy, R., Wolf, T.: PacketBench: A tool for workload characterization of network processing. In: Proc. of IEEE International Workshop on Workload Characterization, pp. 42–50 (October 2003)
21. Kent, S., Atkinson, R.: Security architecture for the internet protocol. RFC 2401, Network Working Group (November 1998)
22. The GNU operating system, <http://www.gnu.org>
23. Salehi, M.E., Fakhraie, S.M.: Quantitative analysis of packet-processing applications regarding architectural guidelines for network-processing-engine development. *Journal of System Architecture* 55, 373–386 (2009)
24. Yazdanbakhsh, A., Salehi, M.E., Safari, S., Fakhraie, S.M.: Locality Consideration. In: *Exploring Custom Instruction Selection Algorithms*. In: ASQED 2010, Malaysia, pp. 157–162 (2010)

TSorter: A Conflict-Aware Transaction Processing System for Clouds^{*}

Po-Cheng Chen¹, Yen-Liang Su¹, Jyh-Biau Chang², and Ce-Kuen Shieh¹

¹ Institute of Computer and Communication Engineering,
Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan
{pcchen, xpsteven, shieh}@ee.ncku.edu.tw

² Department of Digital Applications, Leader University, Tainan, Taiwan
andrew@mail.leader.edu.tw

Abstract. The high scalability feature of cloud storage systems benefits many companies and organizations. However, most available cloud storage systems lack for providing a full transaction processing support that is really needed by many daily-use applications such as on-line ticket booking. Although a few cloud-based transaction processing systems have been proposed, they achieve barely satisfactory throughput when the conflict-intensive workload is performed. In this context, this paper presents a cloud-based transaction processing system called "TSorter" that uses a conflict-aware scheduling scheme for achieving the high throughput when the conflict-intensive workload is performed. Moreover, Tsorter uses a data caching and an affinity-based scheduling schemes to improve the per-node performance. The experiment results indicate that Tsorter achieves the high throughput, irrespective of the workload types (i.e. the conflict-intensive workload or the conflict-free workload).

Keywords: cloud computing; transaction processing; database.

1 Introduction

Recently various cloud storage systems (CSSs) such as HBase [1], Dynamo [2], Amazon S3[3] and SimpleDB [4] have been adopted for building diverse large-scale web services. These services can benefit from the excellent scalability and the high throughput of CSSs. However, most available CSSs lack for providing a full transaction processing support that is necessary for many daily-use applications such as auction services, payment services, course enrollment systems and e-business services. Consequently, a few cloud-based transaction processing systems (TPSs), for example CloudTPS [5, 6], transactional HBase (THBase) [7] and G-Store [8], have been proposed. Although these systems can provide both the high scalability feature and the transaction processing support, they still have two imperfections.

First, the transaction schedulers of existing cloud-based TPSs are conflict-unaware. In other words, these systems have not yet comprehensively considered the

^{*} The authors are grateful to the National Science Council of Taiwan for the financial support. (This research was funded by contract NSC99-2221-E426-007-MY3).

conflict-intensive workload. In such a system, the conflicting transactions may be dispatched to the different processing nodes, and thus the distributed concurrency control is needed to avoid the conflict serializability violation. Generally, the distributed concurrency control incurs the overhead of the transaction processing. However, the conflict-intensive workload is commonly appeared in the daily-use applications such as course enrollment systems and ticket booking systems. For example, different transactions that involve booking the same seat (or enrolling the same course) absolutely conflict with each other.

Moreover, according to the research results appeared in literature [9], the read latency of many CSSs is higher than conventional database management systems; this situation is due to the coarse-grained index mechanism adopted by CSSs. Specifically, the index mechanism of HBase [1] and BigTable [10] only contains information about nodes in which the replicas of a given data item are stored. The remainder step of finding the data item relies on searching each file that possibly stores the data item. Searching a data item causes the high read latency. Thus, the performance of existing cloud-based TPSs suffers from such a high read latency.

In fact, the research on the cloud-based TPS is still in its infancy. In this context, we proposed a novel cloud-based TPS called "TSorter". The most significant contribution of TSorter is that TSorter can achieve high throughput, irrespective of the workload types (i.e. the conflict-intensive workload or the conflict-free workload). Additionally, TSorter improves the per-node performance when processing volumes of transactions; it also schedules the conflicting transactions under the time ordering.

The remainder of this paper is organized as follows: Section 2 briefly review the major related work, while Section describes the design of TSorter. Section 4 details the implementation of TSorter and also evaluates its performance under various workloads. Finally, section 5 presents some remarkable conclusions and indicates the intended direction of future research.

2 Related Work

THBase [7] is an extended feature of HBase for providing an experimental transaction support. It uses an optimistic concurrency control scheme to ensure the serializability property of a transaction schedule. It optimistically assumes that the conflict between multiple transactions is seldom. Therefore, it processes transactions without locking the data items that are accessed by the transactions. If there is no conflict, it commits all processed transactions. Otherwise, it resolves a conflict by allowing one transaction to commit successfully and aborting the remainder transactions. However, if conflicts happen often, the high abort rate hurts system throughput significantly. In other words, THBase is not suitable for the conflict-intensive workload.

CloudTPS [5, 6] is based on the Peer-to-Peer (P2P) architecture. CloudTPS splits an incoming transaction into multiple sub-transactions; it uses a time timestamp-based concurrency control scheme in which each transaction has a unique and incremental timestamp; all sub-transaction of the same transaction has the identical timestamp. All incoming sub-transactions of a virtual node are sorted by timestamp, and committed one by one. However, only if all sub-transactions of a transaction have to be committed successfully is the transaction allowed to be committed. Otherwise, the entire

transaction is aborted. Obviously, when a conflict-intensive workload is performed, the high abort rate hurts the entire system throughput.

3 Design

TSorter aims to be a high throughput TPS no matter what type of workload is perform. It tries its best to dispatch the conflicting transactions into the same node for efficiently serializing them; and to dispatch the non-conflicting transactions into the different nodes for concurrently processing them (a conflict-aware scheduling scheme). Moreover, TSorter uses both a data caching and an affinity-based scheduling schemes to minimize the cost of the remote data access, and thus improving the per-node performance.

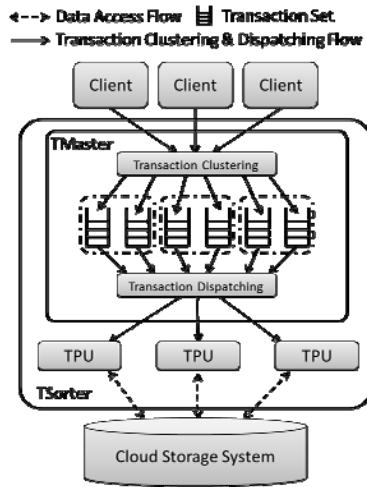


Fig. 1. System overview of Tsorter

3.1 System Overview

As shown in Fig. 1, TSorter consists of a centralized TMaster and multiple transaction processing units (TPUs). All transaction submitted from client nodes are scheduled by TMaster. TMaster classifies an incoming transaction under one of a transaction set (T-Set) according to the conflict and affinity relation between the incoming transaction and the T-Sets. Each T-Set could be regarded as a FIFO queue, and therefore transactions in the same T-Set could be processed sequentially. Transactions in multiple T-Sets are then dispatched to a proper TPU. TPUs are responsible for processing transactions. Moreover, TSorter adopts a cooperative data caching mechanism to lower the data access latency. Each TPU maintains its local cache; these local caches are collected to form a distributed cooperative cache that is a write-through cache. To decrease the overhead of maintaining the cache coherence, a certain data item is allowed to be cached by only one TPU.

3.2 Conflict Relation and the Affinity Relation

To complete the transaction clustering, TMaster has to examine the conflict relation and the affinity relation between incoming transactions and TPUs. Therefore, TMaster inspects the information about the data items accessed by each incoming transaction; TMaster can obtain the information by examining the key enumeration of each incoming transaction because each data item stored in the underlying CSS is a key-value pair.

Considering the conflict relation between incoming transactions and T-Sets, TMaster tries its best to make the transactions in the same T-Set more conflicting and to make the transactions in the different T-Sets less conflicting. To determine the conflict relation, TMaster respectively compares the key enumeration of an incoming transaction with the key set of each of T-Sets. The key set of a T-Set is the union of the key enumeration of the scheduled transactions within the T-Set. If a key appears in both the incoming transaction and a certain T-Set, the transaction conflicts with the T-Set.

Moreover, considering the affinity relation between incoming transactions and the cached data items in each TPU, TMaster tries its best to places an incoming transaction in the T-Set that handled by the TPU that the transaction has a close affinity with. To determine the affinity relation, TMaster respectively compares the key enumeration of an incoming transaction with the associated key of cached data in a given TPU.

3.3 Transaction Clustering

According to the conflict relation and the affinity relation between an incoming transaction and T-Sets, the transaction clustering of the incoming transaction could be one of the following cases:

1) The incoming transaction conflicts with one or more T-Sets.

In the first case, if the incoming transaction conflicts with only one T-Set, TMaster classifies the incoming transaction under the T-Set. Otherwise, if the incoming transaction conflicts with more than one T-Set and these T-Sets are processed by the same TPU, we called this situation the intra-TPU conflict relation. On the other hand, if the conflicting T-Sets are processed by the different TPU, we called this situation the inter-TPU conflict relation.

In the case of the intra-TPU conflict relation, TMaster selects a T-Set from all conflicting T-Sets, and classifies the incoming transaction under the selected T-Set. After that, TMaster logically merges the other T-Sets into the selected T-Set. Having done so, the processing of the conflicting transactions dispatched into the TPU is serialized.

Moreover, In the case of the inter-TPU conflict relation, TMaster uses a conditional wait mechanism to serialize the conflicting transactions across different TPU. Specifically, TMaster selects a T-Set from all conflicting T-Sets, and classifies the incoming transaction under the selected T-Set. TMaster inserts a conditional wait into the selected T-Set, and then dispatches the transaction into the T-Set. After that, TMaster inserts signals into the remainder conflicting T-Sets processed by other TPUs. These signals could be regarded as pseudo-transactions; they also processed by

the TPUs. Once all signals are processed, the corresponding conditional wait can be removed. Besides, when a T-Set is blocked by a conditional wait, TMaster can still dispatch other scheduled transactions in the non-blocked T-Set. This design can prevent that a TPU becomes idle.

- 2) *The incoming transaction does not conflict with any scheduled transaction but has the affinity relation with one or more T-Sets.*

In the second case, TMaster selects a T-Set processed by the TPU that has highest affinity relation with the incoming transaction, and classifies the transaction into the T-Set. As mentioned before, a certain data item is allowed to be cached by only one TPU for simplifying the system design. Thus, if a transaction has the affinity relation with more than one TPU, TMaster invalidates the related data cache in the remainder affinity TPUs.

- 3) *The incoming transaction has neither the conflict relation nor the affinity relation with any T-Set.*

In the last case, TMaster takes dynamic load balancing into account. Specifically, TMaster selects a T-Set processed by the TPU that the fewest number of transactions are dispatched to.

4 Implementation and Performance Evaluation

TSorter was programmed by using Python programming language (v. 2.6.2); Python's Tornado, a high performance web framework was also introduced in TSorter. Tornado uses the non-blocking socket to avoid the I/O waiting time, and improving the performance and scalability for serving thousands HTTP connections simultaneously. Besides, HBase (v. 0.20.5) was adopted as the underling cloud storage system in TSorter because it can guarantee the serializability at the row-level. Moreover, HDFS (v. 0.20.2) [15] was adopted to be the data back-end of the HBase instance. Additionally, Memcached (v. 1.2.2) [16] was adopted as the write-through cache to relieve the high read latency of HBase.

To explore the practicality and benefit of TSorter, the transaction processing throughput of TSorter was compared with that of an existing cloud-based TPS, THBase [7]. The transaction processing throughput was evaluated by processing different synthesized workloads. The synthesized workloads reproduce various application behaviors, including the conflict-free workload or the conflict-intensive workload.

All experiments were executed on eight nodes, each of which ran Linux (v. 2.6.28 x86_64) on two 2.4GHz AMD Opteron Dual Core Processor 280 with 4GB RAM, 250 GB HDD, and Gigabit Ethernet NIC. One of them was adopted as a synthesized workload generator. Besides, another node was adopted as master (i.e. TMaster of TSorter), and the remainder six nodes was adopted as slave (i.e. TPUs of TSorter).

Moreover, the synthesized input data for all experiments were two tables: user table and share table. User table represents the per-user data such as the student information in a course enrollment system or the consumer information in a ticket booking

system. Share table represents the dataset possibly accessed by all transactions such as the remaining amount of the commodity or the train ticket.

User table contains one million rows. Each row contains two column families: value column and payload column. Value column stores a random generated integer, and payload column stores about 510 bytes dummy data. Thus, user table approximately occupies 1.6 TB disk space. Share table contains ten thousands rows. Each row only contains value column, which stores a random generated integer.

4.1 Conflict-Intensive Workload

Typical examples that represent conflict-intensive workload are the course enrollment system and the ticket booking system. The synthesized workload used in the experiment reproduces the following scenario: one million consumers intensively buy M commodities. In this scenario, the transactions can be clustered by the data item recorded in the share table. The transactions in the same cluster certainly conflict with each other. In such a workload, each transaction respectively reads a data item from

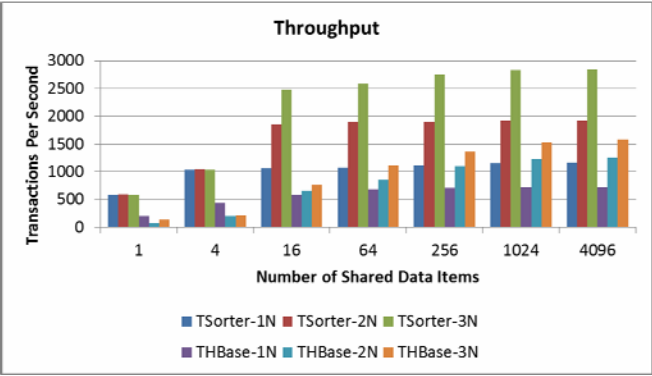


Fig. 2. System throughput (conflict-intensive workload)

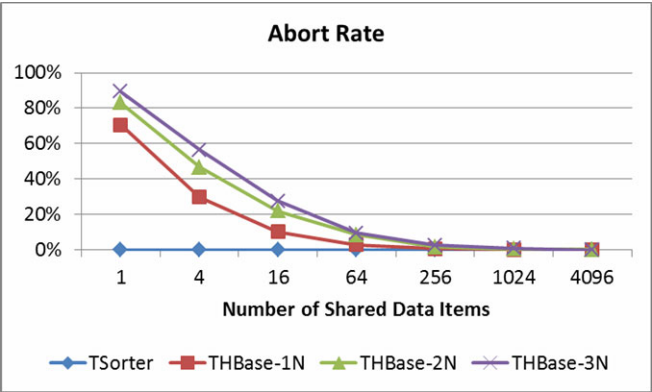


Fig. 3. Abort rate of transaction commitment (conflict-intensive workload)

the user table and the other data item from the share table, and writes the processing results. The data item from the user table is randomly selected from one million rows, and the data item from the share table is randomly selected from the first M rows (the value M indicates the number of the shared data items).

In the first experiment, we executed this workload by using TSorter and THBase with different number of nodes. Fig. 2 illustrates the experiment results derived from running the conflict-intensive workload. In Fig. 2, each experiment configuration is denoted as "System-xN", where x means the total number of nodes used by the TPS, i.e. TSorter or THBase. Fig. 2 suggests that TSorter had superior throughput compared to THBase. In fact, the frequency of transaction conflict tended to increase in such a workload when the workload that involves fewer shared data item is executed. Fig. 3 shows that the abort rate of the transaction commitment was close to 90% when THBase was adopted and the workload involves only one shared data item. Besides, the abort rate tended to increase when more nodes were used. On the contrary, the abort rate of TSorter was 0%. TSorter avoids the occurrence of the unsuccessful commit by adopting the conflict-aware scheduling algorithm, and therefore outperforms THBase. Although the abort rate of THBase was close to 0% when the number of shared data items was more than 256, the throughput of THBase was still less than that of TSorter. These results should be explained by considering the procedure of the data read operation. In TSorter, the newest data records can be accessed quickly from the data cache. On the other hand, THBase accesses data directly from HBase. Thus, THBase suffers the overhead of finding the newest data item because of the append-based update model of HBase.

4.2 Conflict-Free Workload

In this experiment, we executed the synthesized conflict-free workload in which each transaction only accesses a randomly selected data item from the user table. Each transaction does not conflict with the other transaction.

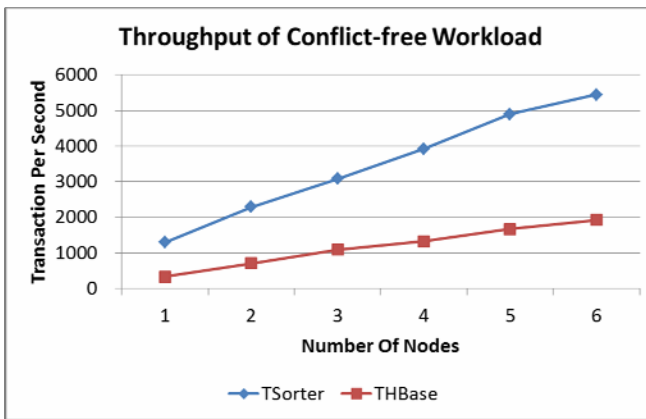


Fig. 4. System throughput (conflict-free workload)

Fig. 4 illustrates the experiment results derived from running the conflict-free workload. It suggests that Tsorter still had superior throughput compared to THBase even if THBase is suitable for processing the conflict-free workload and the impact of the cache mechanism of TSorter is insignificant in the case of running the conflict-free workload. These results should be explained by considering the poor performance of the data read operation of THBase.

5 Conclusion

We proposed a novel cloud-based TPS, which we designated as "TSorter". TSorter is based on the popular open-source cloud database, HBase [1]. The experimental results suggest that TSorter had superior throughput compared to THBase and zero abort rate of transaction commitment. However, the current implementation of TSorter is still a proof-of-concept version. TMaster is a single-point-of-failure of this prototype. In a future study, we might parallelize the single TMaster into multiple T Masters.

References

1. <http://hbase.apache.org/>
2. Bala, V., Duesterwald, E., Banerjia, S.: Dynamo: a transparent dynamic optimization system. In: Proceedings of the ACM SIGPLAN 2000 Conference on Programming Language Design and Implementation, pp. 1–12. ACM, Vancouver (2000)
3. <http://aws.amazon.com/s3/>
4. <http://aws.amazon.com/simplifiedb/>
5. Wei, Z., Pierre, G., Chi, C.: CloudTPS: Scalable transactions for Web applications in the cloud. Technical Report IR-CS-053, Vrije Universiteit, Amsterdam, The Netherlands (February 2010), http://www.globule.org/publi/CSTWAC_ircs53.html
6. Wei, Z., Pierre, G., Chi, C.: Scalable transactions for web applications in the cloud. In: Sips, H., Epema, D., Lin, H.-X. (eds.) Euro-Par 2009. LNCS, vol. 5704, pp. 442–453. Springer, Heidelberg (2009)
7. <http://hbase.apache.org/docs/r0.20.5/api/org/apache/hadoop/hbase/client/transactional/package-summary.html>
8. Das, S., Agrawal, D., Abbadi, A.E.: G-Store: a scalable data store for transactional multi key access in the cloud. In: Proceedings of the 1st ACM Symposium on Cloud Computing, pp. 163–174. ACM, Indianapolis (2010)
9. Cooper, B.F., Silberstein, A., Tam, E., Ramakrishnan, R., Sears, R.: Benchmarking cloud serving systems with YCSB. In: Proceedings of the 1st ACM Symposium on Cloud Computing, pp. 143–154. ACM, Indianapolis (2010)
10. Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., Gruber, R.E.: Bigtable: A Distributed Storage System for Structured Data. ACM Transactions on Computer Systems 26, 1–26 (2008)

New Secure Storage Architecture for Cloud Computing

Sameera Abdulrahman Almula and Chan Yeob Yeun

Khalifa University of Science, Technology and Research, Sharjah Campus,
United Arab of Emirates

{Sameera.almulla,cyeun}@kustar.ac.ae

Abstract. The rapid development of technology has resulted in an increase in the need for computing assets and many organizations are moving in the direction of outsourcing their storage and computing needs. This new trend of outsourcing resources and processes has been called “Cloud Computing”. The challenge is to provide security for data stored in public cloud, where users have reduced control of their own information. In this paper, we are going to propose a new security architecture for customers to ensure that their data is secured by deploying both symmetric and asymmetric encryption in an environment that allows for lawful interception.

Keywords: Cloud Computing, Secure Storage, Lawful Interception, Encryption.

1 Introduction

There are several cloud models available in the market, an agreed upon framework of cloud services described as Software-as-a-Service (SaaS), Platform-as-a-service (PaaS) and Infrastructure-as-a-Service (IaaS) (collectively referred to as “SPI”) [1],[2] and [3]. Since security measures will differ in each framework [4] and [5], in this paper, we will focus on cloud based storage supplied as IaaS.

Our goal is to enhance cloud security in one aspect, namely storage, by satisfying the security requirements including confidentiality, integrity, data segregation and authentication while facilitating LI.

The LI process for IP based communication is performed on the traffic between two communicating entities. Once the Law Enforcement Agency (LEA) has granted warrant to intercept the communication, a packet sniffing tool will be placed at the internet Service Provider (ISP) of the suspected entity. Later, the sniffed data will be used for digital forensics analysis [6]. In addition, sniffing tools are very useful to analyze the sniffed network traffic and determine its behaviors and trends, however, it is a challenging task to extract individual user’s activities. The main barrier that an LEA will face is the encrypted traffic.

In this paper, we propose a new architecture to perform LI on the encrypted storage rather than the traffic without compromising user’s credentials such as the encryption sub-key and yet decrypt the suspicious evidence. Unlike the network traffic, intercepting user’s information at rest in the cloud environment consumes less time to gather and analyze the suspected user’s information. The proposed solution ensures user’s data isolation which facilitates the LI process.

The main contribution of this proposal is to provide secure outsourced storage in the cloud where the clients are assured that their data is protected (The service provider has no information regarding the content of any client's data) and yet support lawful interception.

The rest of the paper is organized as follows: section 2 discusses the related work. The new Secure Storage Architecture and its four phases described in section 3. Section 4, we will analyze the security of the proposed architecture and an overview on its implementation in section 5. Finally, the conclusion and future work will be covered in section 6.

2 Related Work

S. Kamara and K. Lauter in [7], discussed the architecture of Virtual Private Storage Services (VPSS). Their aim was to satisfy cyber security objectives such as Confidentiality, Integrity, Availability (CIA), whereas our goal is to support CIA and additional cloud based security requirements such as data segregation, legal investigation and lawful interception. C. Wang and *et al.* [8], discussed security using third party auditing protocol independent of data encryption, where users are not involved in data integrity checking every time they upload or download data from the cloud. Auditing and masking for digitally signed data implies that even a trusted Third Party Auditor (TPA) cannot intercept the data content. Section 2.1 and 2.2 discuss related main two components of the proposed architecture namely secret sharing scheme and LI.

2.1 Shamir Secret Sharing Scheme (SSSS)

SSSS is able to overcome limitations of sharing the secret keys among the group of participants. As described in [9] it consists of two main protocols: shares initialization, distribution and reconstruction protocols.

- Shares Initialization and Distribution Protocol

Let t, w be positive integers, such that $t \leq w$ and the threshold (t, w) are defined as a technique of sharing a secret key (K) among set of w participants (P) in a way that any t participants can reconstruct the value of K , but no set of $(t-1)$ participants can do so. The value of K is chosen by Trusted Third Party (TTP) which does not belong to (P). TTP will distribute the part of K called *shares*.

- Shares Reconstruction Protocol

In order to reconstruct the K , a subset of the participants B belongs to P want to combine their shares. TTP can be involved in this process. If $|B| \geq t$ then the secret K can be computed. Whereas, if the $|B| < t$, then secret K cannot be computed.

As stated by A. Shamir [9], threshold schemes best suit applications where “*mutually suspicious participants with conflicting interest must cooperate*”. In cloud computing environment which supports LI, users, Cloud Service Providers (CSP) and LEA, collectively present the mutually suspicious with conflicting interests who need to cooperate. In the next section, we will discuss reasons for emphasizing providing cloud based services such as storage which support LI.

2.2 Lawful Interception (LI)

Lawful Interception (LI) [10] is the lawfully authorized interception and monitoring of communication over the internet or telecommunication systems pursuant of an order of jurisdiction, to obtain necessary forensics evidence. For interception to be lawful, it must be done according to national laws and after receiving proper authorization from competent authorities. In this section, we will discuss the LI requirements as stated by International Telecommunication Union (ITU) in [10].

- Lawful Interception Requirements

The ITU specifies the general requirements for LI which are:

- The system should provide transparent interception, where the intercepted subject (A user, group of users, a computer or other equipment acting on behalf of the users) is not aware of the interception process.
- The provided services to other users who are not involved in the interception should not be affected by the LI process.
- A distinct separation between LI network and public usage network to avoid unnecessary interception of users who should not be intercepted.
- The LI should be done within the specified dates and times as provided in the LI order.

In the proposed architecture, we emphasize on supporting LI while providing secure storage; because as stated in [11], governments have delayed or reject the rollout of telecommunication and IP based technologies because of the dissatisfaction with the country LI obligations.

The Internet Engineering Task Force (IETF) in [10] has refrained from recommending lawful interception as they argued LI complexity may reduce security. In our proposed solution, LI functionality will be provided without increasing the complexity, yet preserving client's privacy.

3 New Secure Storage Architecture for Cloud Computing

In order to encrypt their information prior to uploading to the cloud, users should set a password (secret key) which is used to generate secret shares. In the proposal, the user's *secret key* will be changed periodically on a monthly basis that is Periodic Encryption Key (PrEK) or upon the users' request, and new *shares* are re-produced, encrypted and uploaded along with the encrypted file.

There is an issue of managing the PrEK and how to link the encrypted file with its corresponding key. Managing the PrEK can be achieved by Trusted Third Party (TTP). The main role of the TTP is to manage the PrEK for each user and maintain the consistency of the encrypted file with its encryption key. The proposal consists of four phases:

3.1 Phase I: Key Generation and Distribution

Using SSSS, the PrEK is divided into three Secret Shares with a threshold of (2, 3) meaning two shareholders should combine their shares to decrypt the information.

These shares will be distributed as UserA secret share (SS_A), Cloud Service Provider share (SS_{CSP}) and Law Enforcement Agency share (SS_{LEA}). User A will log on to the service using username and authentication password. Using the API, he will browse the local file to be encrypted, and then type the encryption password (used to generate PrEK). Finally, the encrypted file will be uploaded along with the corresponding secret shares. Since the files are encrypted on the client side, it will enhance the system performance and maintain client privacy (Fig. 1 describes the secret shares upload protocol).

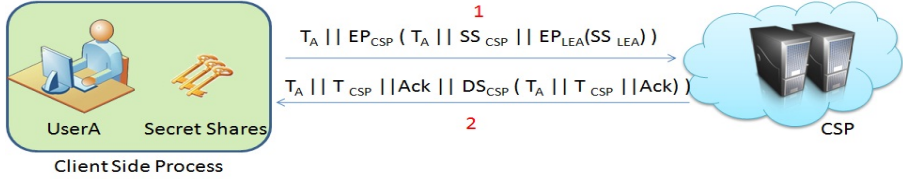


Fig. 1. Secret shares upload protocol

T_A : Time stamp generated by user A. EP_{CSP} : Asymmetric encryption using CSP's public key
 T_{CSP} : Time stamp of CSP. DS_{CSP} : Digital signature using CSP's private key.

- Protocol Steps

1. $T_A || EP_{CSP} (T_A || SS_{CSP} || EP_{LEA}(SS_{LEA}))$

The combination of user A time stamp (T_A), the CSP secret share (SS_{CSP}) and the LEA encrypted secret share ($EP_{LEA}(SS_{LEA})$) will be encrypted using CSP's public key and uploaded to its server.

2. $T_A || T_{CSP} || Ack || DS_{CSP} (T_A || T_{CSP} || Ack)$

To authenticate the CSP, the response message from the CSP will be signed using its private key. User A can verify the signature using CSP's public key. Hence, user A will be confident that the message is uploaded to the correct CSP.

Note: the Key generation and distribution is discussed separately to ease protocol analysis. However, it is performed along with the file upload.

3.2 Phase II: Encryption and Upload

Using PrEK, the files will be encrypted using Advance Encryption Standard (AES) with key size of 256 bit and upload it to the cloud. Encrypted files are sent along with its IDentifier (F_{ID}) and its digest (Fig. 2. describes the file encryption and the upload architecture).

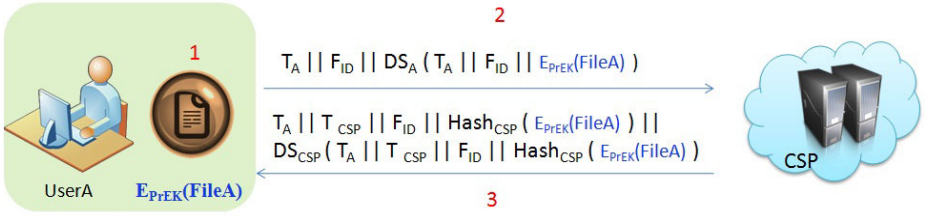


Fig. 2. File encryption and upload

T_A : Time stamp generated by user A.

DS_A : Digitally signed using A's private key.

T_{CSP} : Time stamp of CSP.

DS_{CSP} : Digitally signed using CSP's private key.

E_{PrEK} : Symmetric Encryption using PrEK.

F_{ID} : File Identifier.

$Hash_{CSP} (E_{PrEK}(FileA))$: The digest of the encrypted file A generated by CSP.

• Protocol Steps

1. Using API, the file will be encrypted with PrEK.

2. $T_A || F_{ID} || DS_A (T_A || F_{ID} || E_{PrEK}(FileA))$

File ID, user generated time stamp (T_A) and the encrypted file will be signed using user A's private key $DS_A (T_A || F_{ID} || E_{PrEK}(FileA))$. CSP can verify the signature using user A's public key, if the verification process was successful CSP will accept user A's file.

3. $T_A || T_{CSP} || F_{ID} || Hash_{CSP} (E_{PrEK}(FileA)) || DS_{CSP} (T_A || T_{CSP} || F_{ID} || Hash_{CSP} (E_{PrEK}(FileA))$

The response to the uploaded message (in step 2) will include the digest of the encrypted file which is generated by the CSP $Hash_{CSP} (E_{PrEK}(FileA))$. User A can verify the uploaded digest with the replied one to check its integrity and that the file was successfully uploaded without alteration. Signing the timestamps (both T_A and T_{CSP}), F_{ID} and its digest by the CSP's private key, will ensure that the file was uploaded to the correct CSP.

3.3 Phase 3: Download and Decryption

This occurs when the user requests to retrieve or download the encrypted file which is stored in the CSP server. The request message includes the desired file ID. Based on the FileID, the CSP will look for the file in the database and send the encrypted file to the user. Then, userA first will download and save the file on their PC's and then using Secure Storage API, decrypt the file using the same encryption/decryption key PrEK used to encrypt it (Fig. 3. describes the file download and decrypt).

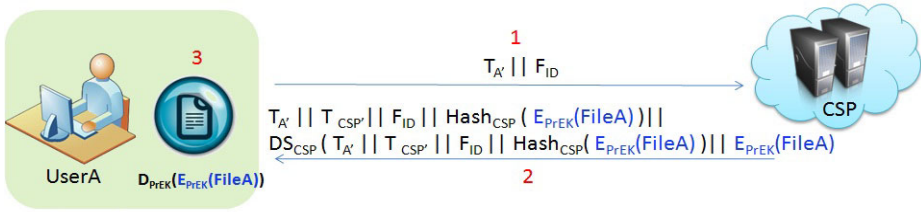


Fig. 3. Download and decrypt

$T_{A'}$: New time stamp of the userA.

DS_{CSP} : Digitally signed using CSP's private key.

$T_{CSP'}$: New time stamp of the CSP.

$Hash_{CSP}(E_{PrEK}(FileA))$: The digest of the encrypted fileA generated by UserA.

E_{PrEK} : Symmetric Encryption using PrEK.

F_{ID} : File Identifier.

D_{PrEK} : Symmetric Decryption using PrEK

• Protocol Steps

1. User A will send a request message to the CSP by specifying which file he wants to download by sending the corresponding F_{ID} .
2. $T_{A'} || T_{CSP'} || F_{ID} || Hash_{CSP}(E_{PrEK}(FileA)) || DS_{CSP}(T_{A'} || T_{CSP'} || F_{ID} || Hash_{CSP}(E_{PrEK}(FileA))) || E_{PrEK}(FileA)$

The encrypted file will be sent to the user along with the CSP signature. CSP signs the request message time stamp, the response message time stamp ($T_{A'} || T_{CSP'}$), the requested fileID and the digest of the encrypted file. After successful file download, user A will decrypt the file ($D_{PrEK}(E_{PrEK}(FileA))$) and get the original file.

3.4 Phase IV: Lawful Interception

In case of the lawful interception, the proposed security architecture will facilitate the Law Enforcement agency to intercept illegal user's activity in the cloud. This can be achieved because of the SSSS deployment (Fig. 4. describes the LI protocol) where

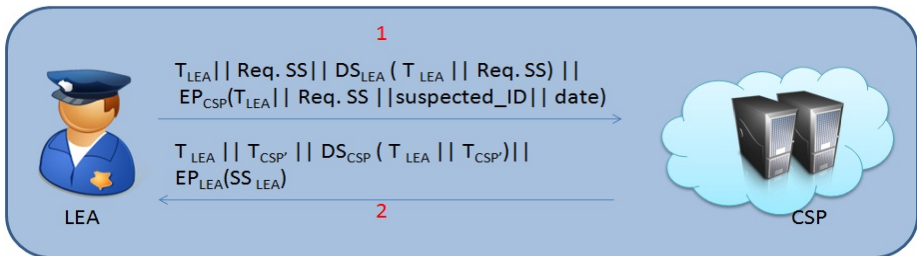


Fig. 4. The Lawful Interception Architecture

T_{LEA} : Time stamp of the LEA.	DS_{CSP} : Digital signature using LEA's private key.
T_{CSP} : Time stamp of the CSP.	DS_{CSP} : Digital signature using CSP's private key.

- Protocol Steps

- 1 $T_{LEA} || \text{Req. SS} || DS_{LEA} (T_{LEA} || \text{Req. SS}) || EP_{CSP}(T_{LEA} || \text{Req. SS} || \text{Suspected\_ID} || \text{date})$

LEA will request their part of the secret shares in the case of the LI. Since the request and the LEA timestamp is signed by the LEA private key, CSP can verify the signature to ensure that the request is established by the LEA. It is essential to encrypt suspected user's information, because if the message was intercepted then the user will know that his/her files have been intercepted (which violates one of the LI system requirements). Also, this information is sent because the PreK is not fixed for the same user and to ensure that the correct user's share sent to the LEA.

- 2 $T_{LEA} || T_{CSP} || DS_{CSP} (T_{LEA} || T_{CSP}) || EP_{LEA}(SS_{LEA})$

CSP will send the LEA encrypted secret share corresponding to the suspected identifier and the date specified in step 3. LEA will verify the signature to ensure that the received encrypted secret share is sent from the desired CSP. Once the LEA gets their part of the secret share, they can perform the LI in co-operation of the CSP.

4 Security Analysis

Unlike the traditional network, security analysis of cloud computing require additional evaluation criteria besides the confidentiality, integrity, and authentication such as data segregation and investigative issues. In this section, we will discuss these security requirements in the new architecture, how to achieve them and provide countermeasures of possible attacks.

- Confidentiality, Integrity, Authentication and Authorization

Both symmetric encryption (AES 256bits) for storage and asymmetric encryption (RSA) for sub-keys used to achieve confidentiality. Time stamps used to provide mutual authentication. To preserve storage integrity hash function (SHA1) was deployed. Based on Role Base Access Control (RBAC), the CSP administrator, LEA and users have different authorization levels, functionality and interface according to the provided credentials such as username and password.

- Data segregation

In the cloud computing, all client information is stored in a shared environment, CSP must provide data segregation for data at rest. All uploaded files are encrypted which increases the isolation level, where only the client (encryption key owner) can decrypt their storage. In the case of lawful interception, LEA can access the encrypted files for a particular user and for a predetermined strict date and time. The encryption will ensure the segregation in both cases.

- Investigative support

Illegal activities can be subject for investigation. It is important that the CSP supports and facilitates the investigation process. However, because of the cloud's dynamic nature, it is impossible to investigate such behavior. Deploying SSSS in the new security architecture will facilitate the investigation process without exposing unsuspected users information to the LEA.

- Lawful Interception

Using SSSS, the new security architecture was able to support LI without increasing the system complexity and yet securing clients credentials. The latter was achieved by adopting the periodic encryption keys rather than static keys. Even, if the LI was performed and the clients secret key was exposed it will be for a specific period of time such as one month. It will preserve clients privacy even if it is compromised (which is not the case) during the process of LI.

- Attacks and Countermeasures

Mutual authentication is an important security requirement which needs to be satisfied in lawfully intercepted applications. For example, in the case where a suspected client impersonates an innocent user and uploads file on his behalf. After performing LI, this evidence will be attached to an innocent client.

Digital signatures, provides sender's authentication and prevent him from repudiating the action. Thus, it prevents Man-In-The-middle attacks and impersonation of legitimate user. Time stamps, ensure message freshness and prevent from reply attacks. Preventing denial of service attack is challenging task, however, having time-stamps will ease detection of such attacks.

5 Implementation of Proposed Architecture

In this section, we discuss the implemented web interface snapshots. The main components used to implement the proposed architecture are the web server, MySQL and the internally developed application called Kloud- Security package (developed using C++).

- Encryption and File Upload

Once the client is authenticated, he/she required to download the Kloud-security package. After that, client can browse the file on their PC, enter the password, then encrypt the file and upload it. Concurrently, the corresponding sub-keys will be encrypted and uploaded. Fig. 5 and Fig. 6 describe the process of file encryption and upload.

- Decryption and File Download

On file download, the list of all encrypted files uploaded by particular client will be displayed along with its digest values as shown in Fig. 7.

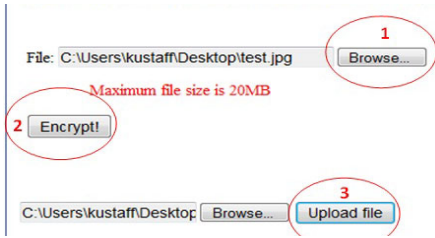


Fig. 5. File encryption and upload

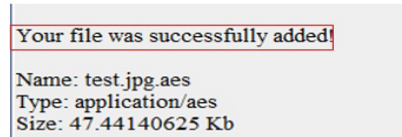


Fig. 6. Confirmation message of successfully uploaded file

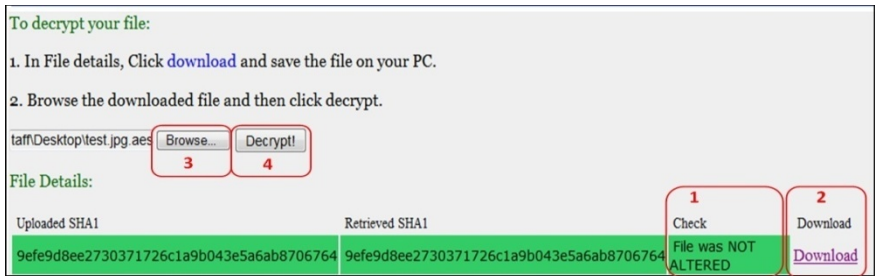


Fig. 7. File download steps

- Lawful Interception

To perform LI, LEA must enter the suspected client name and date of file upload. Once the request is submitted, a list of all uploaded files of the suspected client will be displayed. Then, both CSP and LEA presenters should be available to perform the LI. All their actions will be traced and recorded (one of LI requirements). Fig. 8 and Fig. 9 shows the reconstructed secret after the sub keys entered into the system.

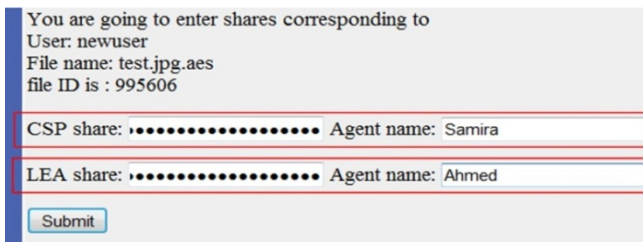


Fig. 8. Agent's sub-keys entry interface

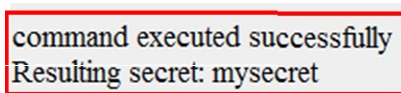


Fig. 9. The resultant encryption key

6 Conclusion and Future Work

In this paper, we have described new secure cloud storage architecture which satisfies security requirements. These security requirements are achieved by deploying symmetric, asymmetric encryption, digital signatures and hash function. The environment allows for Lawful Interception by deploying SSSS. It is a great challenge to deploy any security architecture for Cloud Computing services because of the dynamic nature of the cloud environment. In the future, we will improve the architecture to perform online secret sharing and extract more forensic related information of the storage.

References

1. Marther, T., Kumaraswamy, S., Latif, S.: Cloud Security and Privacy. O'Reilly, Sebastopol, ISBN: 978-0-596-802769
2. An Oracle White Paper in Enterprise Architecture, Architectural Strategies for Cloud Computing (August 2009)
3. Furlani, C.: Cloud Computing: Benefits and Risks of Moving Federal IT into the Cloud, National Institute of Standards and Technology (January 2010)
4. Cloud Security Alliance CSA, Security Guidance for Critical Areas of Focus in Cloud Computing (April 2009)
5. European Network and Information Security Agency, Cloud Computing: Benefits, risks and recommendations for information security (November 2009)
6. Broadway, J., Turnbull, B., Slay, J.: Improving the Analysis of Lawfully Intercepted Network Packet Data Captured For Forensic Analysis. In: The Third International Conference on Availability, Reliability and Security (2008)
7. Kamara, S., Lauter, K.: Cryptographic Cloud Storage. In: The Proceedings of Financial Cryptography: Workshop on Real-Life Cryptographic Protocols and Standardization. Published by Microsoft Researcher (January 2010)
8. Wang, C., Chow, S., Wang, Q., Ren, K., Lou, W.: Privacy-Preserving Public Auditing for Secure Cloud Storage. In: The Proceeding of IEEE INFOCOM 2010, San Diego, USA, March 14-19 (2010)
9. Shamir, A.: How to Share a Secret. Communications of the ACM 22 (November 11, 1979)
10. International Telecommunication Union, Technical Aspects of Lawful Interception, Technology Watch Report (May 2008)
11. Branch, P.: Lawful Interception of the Internet. Australian Journal of Emerging Technologies and Society 1(1) (2003)

An Adaptive WiFi Rate Selection Algorithm for Moving Vehicles with Motion Prediction

Jianwei Niu¹, Yuhang Gao¹, Shaohui Guo², Chao Tong¹, and Guoping Du¹

¹ School of Computer Science and Engineering, Beihang University, Beijing 100191, China

² School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

{niu Jianwei, tongchao, duguoping}@buaa.edu.cn, sguo@cs.cmu.edu

Abstract. With the proliferation of Wi-Fi devices, many cities have been covered by Wi-Fi network; it is increasingly important for people in moving vehicles to be able to access the Internet using Wi-Fi devices. Since the signal strength varies dramatically for Wi-Fi devices in moving Vehicles, it is necessary for the data-rate of Wi-Fi devices to be adjusted dynamically. In this paper, we propose an Adaptive Wi-Fi Data-rate Selection (AWDS) algorithm based on motion predication. By detecting the signal strength of APs, Wi-Fi devices on the move are able to predicate the motion model of vehicles, and to select data rate of Wi-Fi devices accordingly. In this way, Wi-Fi rate selection will be more consistent with its wireless surroundings in the future period of time. Experiment results demonstrate that the AWDS algorithm outperforms the rate selection algorithm built in Android G1.

Keywords: Vehicle-mounted WiFi devices; Signal Strength; Rate Selection; Mobility Prediction.

1 Introduction

The original design goal of WiFi technology is to provide wireless-access services for users in static environments, and the typical devices are laptops. With the popularity of WiFi technology and the wide application of WiFi devices, WiFi modules have been easily integrated with cell phones, PDA, wearable computers and other mobile devices. The miniaturization of mobile devices boosts up the mobility of WiFi devices, which changes the application scenarios. It is reasonable to suppose that users hope to surf the Internet in a moving vehicle by using their smart phones with WiFi interfaces. This kind of highly dynamic surroundings will bring up some serious problems for the WiFi communications. Even if users move within the range of the same access point, the signal strength and interference will change dramatically.

This paper targets to tackle the issue of WiFi rate selection with Wi-Fi devices in mobile environment. WiFi rate selection can directly affect the bandwidth and transmission efficiency of Wi-Fi communications. Usually, high transmission rate corresponds to low communication distance, and vice versa. It is apparent that high transmission rate is able to improve communication bandwidth and reduce transmission time. The optimal case is that with changes of distance between Wi-Fi devices

and access points, one algorithm can always select the optimal transmission rate to ensure the maximum communication bandwidth.

The main contributions of our paper are: 1) model the distances between moving vehicles and access points, and predict the location of access point based on changes of signal strength; 2) propose our algorithm to dynamically adjust the WiFi data rate based on the predicted distance between the mobile devices and access points according to our model; 3) evaluate the performance of our approach.

2 Related work

The PHY layer of 802.11 provides three different transmission rate: 802.11b supports 1~11Mbps, 802.11a supports 6~54 MBps, and 801.11g supports 1~54Mbps. Each one has different characteristics of anti-interference and transmission distance. The algorithm which depends on actual channel environment to select the appropriate transmission rate is called adaptive algorithm. 802.11 does not define adaptive algorithms explicitly for specified transmission rate, and usually different manufacturers design their own algorithms.

According to the selection mode, the existing rate selection algorithms can be divided into two categories: one is based on statistical data, and the other is based on channel SNR.

Typical adaptive algorithms based on statistical data are ARF [1], AARF [2], CARA [3], ONOE [4], SampleRate [5]. This kind of algorithms assesses the quality of channel by counting data packets, and then determines to increase or decrease of data rate based on the evaluation results.

Algorithms based on SNR estimate SNR via signal strength, compare the corresponding threshold for each transmission rate with estimated SNR, and thus get an optimized rate. In general there are two ways to get SNR: one is that the receiver sends back the SNR of sender, then the sender gets channel SNR; the other is that the sender gets SNR by data packets from the receiver, according to the symmetry of radio channel [6], the sender can view it as current channel's SNR. To improve the accuracy of channel estimating, CHARM [7] takes into account the percentage of RSSI (Radio Signal Strength Indicator) from different times; in other words, the closer the interval between RSSI and estimated time is, the higher weight RSSI should have. This kind of algorithms which estimates signal loss has a very high accuracy; normally the error is less than 2dB.

The algorithms mentioned above are all general adaptive algorithms; however, they are not the optimal solutions targeting at the movement pattern of vehicles. Due to fast movement of vehicles, probably rate selections based on statistical results cannot present the real-time status of channel environment, because statistic-based algorithms have to take lots of time to evaluate channel's quality. Similarly, algorithms based on SNR, in order to ensure the result of analysis, spend a lot of time on analyzing SNR which also hamper their adaption to channel environment. In this paper, we propose a WiFi rate selection algorithm which predicts the motion trend of moving vehicles as the criteria to select WiFi rate. It is able to estimate the distance between vehicles and APs by analyzing changes of signal strength in mobile surroundings, and correspondingly adopt the optimized data rate according to our proposed rate

selection table. As a result, it mitigates the jitters owing to transient signal variations, improves the stability of rate adaption and the bandwidth of data transmission.

3 Data Test

At the beginning of this research, we did lots of empirical tests to analyze the performance of channel and obtained the optimal rate selection table, which provides empirical support for the AWDS algorithm.

As shown in Fig. 1, one laptop was laid on the roadside which connected to the AP via an Ethernet connection. Also there was one smartphone in the car which connected to the laptop with its Wi-Fi interface via the AP. The car traversed the road at fixed speed within the coverage area of the AP, and the smartphone recorded experimental results. To avoid interference factors, like channel interference, all experiments were conducted in an open field.

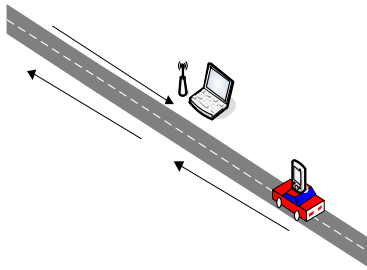


Fig. 1. Experiment scenario

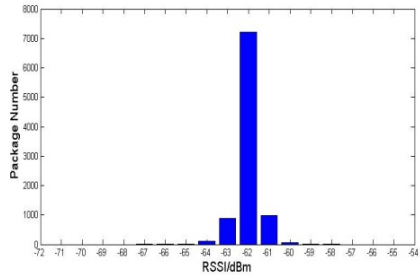


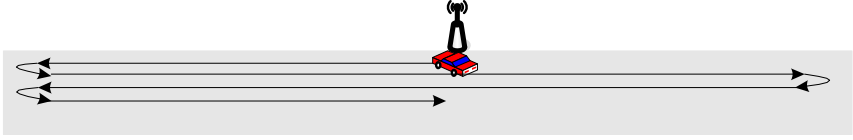
Fig. 2. RSSI distribution

From the attenuation model of radio signal, it can be concluded that the radio signal gradually decays as the communication distance increases. In our test, we used RSSI provided by the Wi-Fi interface of smartphone to obtain signal strength. Via the AP, the laptop continually sent 50 KB UDP data packets to the smartphone which recorded RSSI value and timestamp of data packets. To analyze RSSI's actual characteristics, we adopted two types of measurement method respectively.

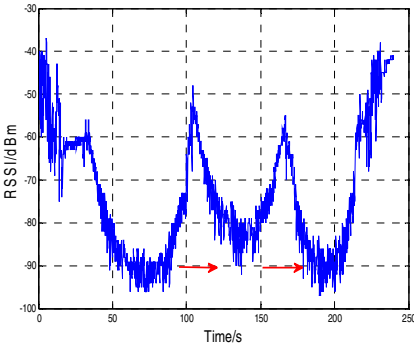
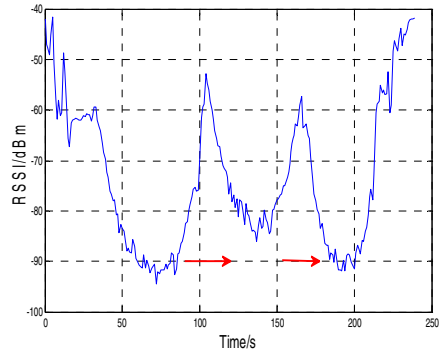
Fixed Location Measurement. Firstly we kept the smartphone and the laptop staying still, and measured the signal strength 30 meters away from AP. The result is shown in Fig. 2.

It can be seen from Fig.2 that although both devices remained stationary, the value of RSSI still has a minor swing due to interference and measurement error; 99.6% data drop within [-63,-61] range, and the average value is -62. There a few RSSI values with high offset do exist, however RSSI value keep stable generally.

Mobile Measurement. As shown in Fig. 3, the car with the smartphone traveled backwards and forwards for three times, and corresponding test results is shown in Fig. 4.

**Fig. 3.** Vehicle route

As can be seen in Fig. 4, the signal strength varies with the change of distance between the car and AP. Due to measurement errors and noise interferences, the signal strength in a short time, however, fluctuates greatly.

**Fig. 4.** RSSI variations**Fig. 5.** RSSI variations (average values for one second)

Further, we calculate the average values of RSSI each second, and the results are shown in Fig. 5. As a result, the fast fluctuation is greatly smoothed. Therefore, it can be approximately regarded as one process that contains patterns. The signal strength increase and decrease periodically with the variations of distance.

4 Our Rate Selection Algorithm

4.1 Algorithm concept

We assume that when one Wi-Fi device in a moving vehicle visits a fixed AP, once the connection is established between the device and one AP, the vehicle will approach or depart from the AP at a relative fixed speed. During this process, the distance is the major factor causing changes of radio signal strength. Just as what we discussed from previous tests, to a great extent the distance can determine the rate selection. Therefore, taking the distance as the basis for rate selection, one adaptive algorithm can own advantages as follows:

- Overcoming the problem that data rate changes due to channel contention;
- With the change of distance, reasonably increasing or decreasing the data rate;
- Avoiding fluctuations due to continually testing higher data rate;
- Maximizing data rate and transmission bandwidth.

4.2 Calculation of Distance between WiFi Devices and APs

The key part of this algorithm is to calculate the distance from WiFi devices to AP. For distance calculation, the only available reference for measurement is RSSI. According to RSSI, we can estimate the distance between one WiFi device and one AP. Paper [8] studied the function relationship between RSSI and distance. To get distance, RSSI measurement utilizes two types of statistical model: signal strength model [9] and space localization model [10]. The general calculation process is as follows.

$$P_d = P_0 - 10n \lg \frac{d}{d_0} \quad (1)$$

P_d is the signal strength in d position, n is the distance loss factor, ranging from 2 to 4, P_0 is the signal strength corresponding to d_0 position. Without loss of generality, we can set $d_0 = 1$ m, so

$$P_d = A - 10n \lg d \quad (2)$$

A can be viewed as the power of received signal when it transmits 1 meter. In practical, A and n should be set by experience values. From Eq. (2), we can obtain the relationship between P_d and d .

$$d = 10^{\frac{A-P_d}{10n}} = f(P_d) \quad (3)$$

We assume that normally vehicles move straightforward when passing APs on the roadside. Further, within the coverage of an AP, vehicles take uniform linear motion, as shown in Fig. 6.

Supposing that vehicles move along the positive X-axis, and we measure RSSI $s_i (i=1,2,\dots)$ at $t_i (i=1,2,\dots)$, and thus predict a series of samples (s_i, t_i) . Based on Eq. (3), RSSI can be converted into a series of distance $d_i (i=1,2,\dots)$, as shown by Eq. (4).

$$d_i = f(s_i) = 10^{\frac{A-s_i}{10n}} \quad (4)$$

Table 1. Optimal rate selection table

Distance Range	Optimal Rate
0~10 Meters	54Mbps
10~15 Meters	48Mbps
15~20 Meters	36Mbps
20~30 Meters	24Mbps
30~45 Meters	18Mbps
45~65 Meters	12Mbps
65~70 Meters	9Mbps
> 70 Meters	6Mbps

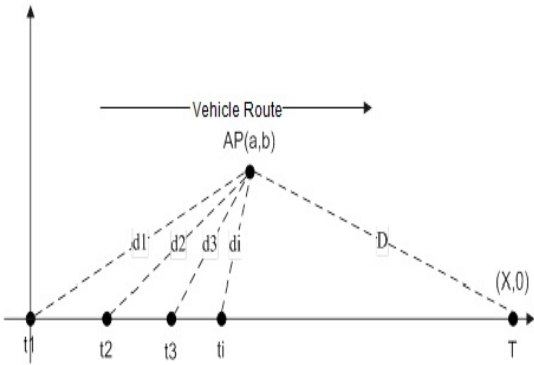


Fig. 6. Vehicles moving along AP

According to Eq. (4), our algorithm can dynamically select the optimal data rate by utilizing Table 1.

4.3 Algorithm Procedures

When receiving one data packet, the AWDS algorithm executes steps as follows:

- The algorithm records the signal strength and time of data packet, and save it to its signal strength list;
- The algorithm calculates the average of signal strength each one second;
- According to the algorithm described in Section 4.2, the algorithm calculates the distance between WiFi devices and APs;
- Based on the distance, the algorithm select the data rate based on Table 1.

4.4 Performance and Discussion

We compared the performance of G1 algorithm and our proposed algorithm, and the results are showed in Fig. 7 and Fig.8.

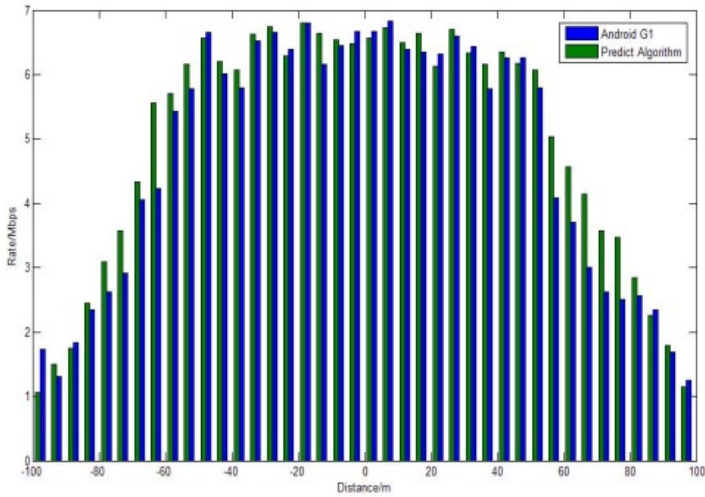


Fig. 7. Rate comparison between Android G1 and our algorithm

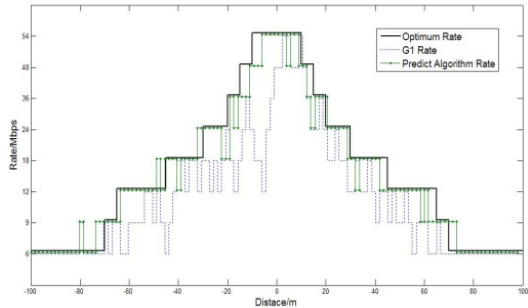


Fig. 8. Rate comparison among optimal rate, G1 and our algorithm

As shown in Fig. 7, in the interval of $[-80, -60]$ and $[60, 80]$, comparing with Android's algorithm, the data rate of our algorithm is higher. In the interval of $[-60, 60]$, these two algorithms produce the similar results. In terms of sending rate, as shown in Fig. 8, our algorithm enhances the stability and reduces rate fluctuation. For vehicle-mounted Wi-Fi devices, the adaptability to channel of our algorithm is more accurate than the existing algorithm built in Android G1.

Our algorithm with mobility prediction is designed for vehicle-mounted WiFi devices visiting fixed APs. It has the following advantages:

- Predicating relative position of vehicles according to variation trend of signal strength, so as to more accurately estimate variation pattern of signal strength;
- Avoiding the rate vibration problem owing to the transient signal variations.

5 Conclusions

By analyzing the process how moving vehicle-mounted Wi-Fi devices visit APs at the roadside, this paper reveals that state-of-the-art algorithms are not qualified due to their slow adaptability to the change of channel status. This paper proposes an adaptive Wi-Fi Data-rate selection algorithm for moving vehicles. It firstly calculates the distance between a vehicle and its connected AP according to the change of signal strength, and then selects the optimal data rate using our proposed *Rate Matching Table*. Experiment results demonstrate that the AWDS algorithm outperforms the adaptive algorithm of rate selection built in Android G1.

Acknowledgment

This work was supported by the Research Fund of the State Key Laboratory of Software Development Environment under Grant No. BUAA SKLSDE-2010ZX-13, the National Natural Science Foundation of China under Grant No. 60873241, the National High Technology Research and Development Program of China (863 Program) Granted No.2008AA01Z217, the Fund of Aeronautics Science granted No. 20091951020.

References

1. Kamerna, A., Monteban, L.: WaveLAN-II: a high performance wireless LAN for the unlicensed band. *Bell Labs Technical Journal* 2(3), 118–133 (1997)
2. Lacagem, M., Manshaei, M.H., Turetli, T.: IEEE802. 11 Rate Adaptation: a Practical Approach. In: *ACM MSWM 2004*, Venice, Italy, pp. 126–134. ACM, NewYork (2006)
3. Kim, J., Kim, S., Choi, S., Qiao, D.: CARA: collision-aware rate adaptation for IEEE 802. 11 WLANs. In: *IEEE INFOCOM 2006*, Barcelona, Spain, pp. 1–11. IEEE press, USA (2006)
4. MADWIFI. Multiband Atheros driver for WiFi, <http://sourceforge.net/projects/madwifi/>
5. Bicket, J.: Bit-rate selection in wireless networks. Technical report, MIT: Department of EECS (2005)
6. Tai, C.T.: Complementary Reciprocity Theorems in Electromagnetic Theory. *IEEE Trans. on Antennas and Propagation* 6(8), 675–681 (1992)
7. Judd, G., Wang, X.H., Steenk, I.: Extended Abstract: Low-overhead Channel-aware Rate Adaptation. In: *ACM MobiCom 2007*, Montreal Canada, pp. 354–357. ACM, USA (2007)
8. Fang, Z., Zhao, Z., Guo, P., Zhang, Y.: Analysis of Distance Measurement Based on RSSI. *Chinese Journal of Sensors and Actuators* 20(11), 22–31 (2007)
9. Shen, X., Wang, Z., Jiang, P.: Connectivity and RSSI based localization scheme for wireless sensor networks. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) *ICIC 2005*. LNCS, vol. 3645, pp. 578–587. Springer, Heidelberg (2005)
10. Zhou, Y., Li, H.: Space Localization Algorithm Based RSSI in Wireless Sensor Networks. *Journal on Communications* 30(06), 1–18 (2009)

A Radio Channel Sharing Method Using a Bargaining Game Model

Jaesung Park¹, Yong-Hoon Choi^{2,*}, and Kilhung Lee³

¹ Dept. of Internet Information Eng., University of Suwon,
Gyeonggi-do, 445-743 Korea
jaesungpark@suwon.ac.kr

² Dept. of Information and Control Eng., Kwangwoon University,
Seoul, 139-701 Korea
yhchoi@kw.ac.kr

³ Dept. of Computer Science and Engineering,
Seoul National University of Science and Technology
Seoul, 139-743 Korea
khlee@snut.ac.kr

Abstract. Network sharing between MNO (Mobile Network Operator) and MVNO (Mobile Virtual Network Operator) imposes a new design challenge on devising a fair and efficient radio channel allocation method. Partitioning the amount of radio channels statically between MNO and MVNO is a simple way to share radio resources. However, since input loads of the operators are usually asymmetric, the resource utilization and the system capacity in terms of the number of accepted users are limited if the resources are not dynamically shared between MNO and MVNO. In this paper, we employ a bargaining game theory to provide a framework for radio channel sharing strategy between these two types of operators. In the proposed scheme, the operators bargain the radio channels so as to increase the resource utilization by allocating unused channels of an operator to the other operator needing more channels. Furthermore, the proposed scheme allocates radio resources in a fair manner according to the number of services of operators so as to prevent an operator from monopolizing radio channels.

Keywords: Bargaining Game Theory, Radio Channel Sharing.

1 Introduction

There has been an increasing interest in network sharing for reducing communication cost by lower capital expenditure in infrastructure investment and reduced operational cost in the long run. In Korea, the relevant laws have been revised to lay a foundation for introducing the Mobile Virtual Network Operator (MVNO). The MVNO is a wireless network operator that does not have license to use frequency spectrums necessary to provide mobile telecommunications service. However, MVNO offers such service

* This paper corresponds to Yong-Hoon Choi.

by utilizing the network of the existing Mobile Network Operator (MNO) who owns frequency and equipment such as a base station. In this new environment where competing operators share network resources, an important issue has emerged on the development of a scheme for efficient management of radio resources that have the extremely high scarcity value.

The best practice to share resource among operators is that the MNO and the MVNO sign a contract so as to allow the MVNO to get static allocation of resources from the MNO in a certain amount. However, as offered load is different among operators at the same point of time, an operator with high offered load cannot utilize the resource that is assigned to the other operator with low offered load in the complete partitioning (CP) method that enables static allocation of radio resource between the MNO and the MVNO. This undermines the efficiency of resource utilization. Therefore, it is required to have a dynamic resource sharing method in consideration of offered load for each operator that shares the base station. In this method, excessive offered load by one operator may lead to resource monopoly, undermining the quality of service provided by other operator. This possible case should be prevented.

This paper suggests the radio channel sharing model that uses the Nash Bargaining Solution (NBS) in the environment that the MNO and the MVNO share radio resource of base station. The suggested scheme allocates radio resource according to offered load of each operator, which improves the utilization of radio resources. In addition, the proposed scheme guarantees a fair resource allocation among operators, which prevents an operator from monopolizing resources to degrade the minimum service level of the other operator.

In this paper, the Section 2 presents the cases of applying game theory to radio resource management, the Section 3 suggests the radio resource sharing model that uses the NBS between the MNO and the MVNO, the Section 4 examines the validity of the suggested scheme by making the numerical analysis, and the Section 5 presents the conclusion and the direction for future study.

2 Related Works

The game theory has been used in many fields such as medium access control, congestion control, and power control in wired and wireless networks [1]. The purpose of such studies is to find out the Nash equilibrium point by using the non-cooperative game theory. However, it was found that the Nash equilibrium point was not Pareto optimal [2]. Consequently, management of radio resources with the non-cooperative game may result in squandering of radio resource.

For this reason, many studies have been conducted recently on applying the cooperative game theory to management of radio network resource by pursuing coalition among game players. All of such studies suggested schemes that allocated limited resources to multiple users based on the fact that solution of cooperative game has the characteristics of Pareto optimality and axiomatic fairness [6]. In [3], authors proposed the coalition scheme among multiple wireless access networks to provide users with high bandwidth. In the meantime, reference [4] suggested a method to manage bandwidth of wireless relay node by utilizing cooperative game. In [5], authors proposed the cooperative game model for spectrum allocation between cognitive radios (CR).

In this paper, we employ a bargaining game to devise a radio channel sharing scheme between a MNO and a MVNO. The bargaining game provides an analytic framework to derive a fair and optimal operational point. There are a few well-known bargaining solutions such as Nash, Raifa-Kalai-Smorodinsky, and modified-Thomson. These bargaining solutions are said to be axiomatic bargaining solutions because they ignore bargaining process by assuming some desirable and fair properties, defined by axioms, about the outcome of resource bargaining process. Among the bargaining solutions, we adopt the Nash bargaining solution because it is easily applied to our radio channel sharing problem.

3 Radio Channel Sharing Model

3.1 Nash Bargaining Solution

In the bargaining game, each participant tries to maximize its utility by arbitration. A utility function is defined to quantitatively represent the utility of a participant when the participant receives a portion of resources as a result of the bargaining game [6]. To formally describe a bargaining solution, we define following notations.

- n : The number of game participants
- $u_i \in \mathbb{R}$: Utility function of a game participant i .
- U : Set of all possible bargaining results.
- d_i : Utility that a game participant i would receive when the bargaining fails and there is no cooperation among the participants.
- $\vec{d} = (d_1, \dots, d_n)$: Bargaining failure point

The purpose of a bargaining game is to find a fair and efficient solution when is $(U, \vec{d})$ is given. Nash analytically showed that a solution called Nash Bargaining Solution (NBS) satisfying the following four axioms exists.

- Invariance: the arbitrated value is independent of the unit and origin of the utility measurement.
- Independence of irrelevant alternatives: If $U_2 \subseteq U_1$, $u^* \in U_2$ when the solution of $(U_2, \vec{d})$ is u^* , then u^* is the solution of $(U_1, \vec{d})$.
- Pareto optimality: No participant can be better off without making the other worse off.
- Symmetry: If U is symmetry about the axis $\vec{u} = (u_1, \dots, u_n)$ and $\vec{d}$ is on the axis, then the solution point is also on the axis.

The NBS of a bargaining game $(U, \vec{d})$ ($\vec{u}^* = (u_1^*, \dots, u_n^*)$) is obtained when the following equation is maximized.

$$\vec{u}^* = \arg \left\{ \max_{\vec{u}} \left(\prod_{i=1}^n (u_i - d_i) \right) \right\} \quad (1)$$

3.2 Dynamic Channel Sharing Method

This paper considers the environment where channel resource of a base station is shared between the MNO and a number of the MVNOs who have the contract with the MNO. Mobile telecommunications service is provided in voice, video and text, but operators still earn profit mostly from voice service. In consideration of the situation, this paper examines the environment where the MNO and the MVNO share radio voice channel to provide voice service.

Voice sources may have some difference in bit rates due to encoding methods and have the different ways of bandwidth allocation to accept various types of networks [8]. However, voice traffic has the characteristic of constant bit rate (CBR) at the packet level and guarantees voice service quality by allocating the fixed amount of radio channel resource to each voice call in the network. The purpose of this study is to suggest a model to share resource of base station by using the NBS. Therefore, on the assumption that the bandwidth allocated to voice call is constant without loss of generality, the amount of base station resource is expressed on the basis of the channel bandwidth that supports voice call. In other words, if the voice channel resource of base station is C , this means that it is possible to support the C number of voice calls.

For analysis of base station load due to voice traffic, it is assumed as shown in [7] that voice call arrival rate of MNO subscriber follows the Poisson distribution of average λ_0 while call holding time follows the exponential distribution of average μ_0 . In addition, it is supposed that mean call arrival rate (λ_i) and mean call holding time (μ_i) of MVNO i comply with the Poisson distribution and the exponential distribution respectively. If the amount of resource allocated to MNO is a_0 , and the amount of resource allocated to MVNO i is a_i , utility function of each operator depending on the allocated resource amount can be expressed in terms of the call admission probability as below.

$$u_i = 1 - p_i(a_i). \quad (2)$$

$p_i(a_i)$ means call blocking probability (CBP) of operator i based on the allocated resource amount a_i . Since we are considering a voice call, the CBP of each operator is expressed as below based on the Erlang-B formula.

$$p_i(a_i) = \frac{E_i^{a_i}/a_i!}{\sum_{j=0}^{a_i} E_i^j/j!}. \quad (3)$$

Here, Erlang $E_i = \lambda_i \mu_i$ means the offered load of operator i . Furthermore, the call admission probability is 0 ($d_i=0, i=0,1,\dots,n$) for the operator who does not participate in the bargaining game for base station voice channel sharing. As a result, based on the equation (1), the amount of base station resource that is allocated to each operator can be obtained as below considering the offered load of each operator.

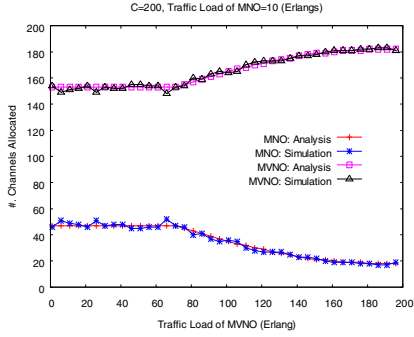
$$(a_0^*, \dots, a_n^*) = \arg \left\{ \max_{\vec{a}} \left(\prod_{i=1}^n u_i \right) \right\} \quad (4)$$

4 Experiment Results and Discussions

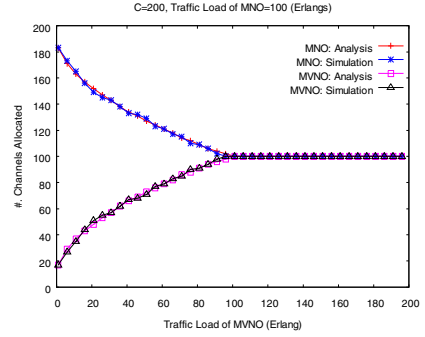
In this Section, simulation experiments are conducted to verify accuracy and validity of the base station radio channel sharing model suggested in this study. For convenience of explaining the experiment results, the resource allocation results depending on offered load of MNO and MVNO were illustrated in the case where the MNO shares the channel resource of base station with one MVNO. The voice channel resource of the base station was set at $C=200$ while the mean holding time of voice call of each operator was set at three minutes. The offered load of each operator was adjusted by changing the voice call arrival rate.

The Fig. 1 illustrates amounts of voice channels, which are allocated to each operator according to increase in offered load of MVNO when the MNO has the high or low offered load. The Fig. 2 illustrates service quality of each operator depending on allocated amount in terms of call admission probability. As shown in the figures, the results of mathematical analysis correspond with the results of simulation experiment. According to the Fig. 1-(a) and Fig. 2-(a), when the offered load of the MNO was low, the call admission rate of the MNO was 100% even with the low amount of resource allocated to the MNO. Therefore, the suggested scheme increases the channel resource that is allocated to the MVNO as most of the resource is allocated to the MVNO and the offered load of the MVNO increases. As a result, the suggested scheme utilizes resource more efficiently by allowing the MVNO to use the channel resource that is not used by the MNO, rather than allowing static allocation of channel resource to the MNO and the MVNO. For example, if a half of the channel resource of base station was assigned to the MNO and the MVNO in the static allocation method, the call admission rate of the MVNO decreased to 54.6% when the offered load of the MVNO took up 90% of the total resource amount of the base station. However, the suggested scheme guaranteed the call admission rate of the MNO at 99.5% and at the same time, maintained the call admission rate of the MVNO as high as 94.8%.

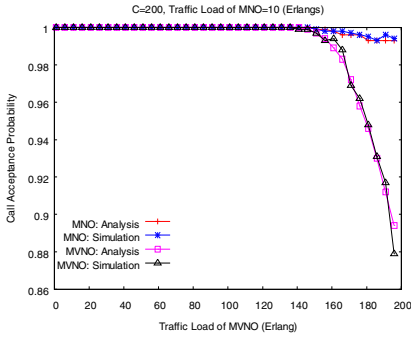
The suggested resource allocation scheme produces good results even when MNO has the high offered load rate. As shown in the Fig. 1-(b), when the offered load of MNO is 100Erlang and the offered load of MVNO is 100Erlang or lower, the suggested scheme allocates more resource to the MNO. However, when the offered load of MVNO is 100Erlang or higher, the total offered load of the base station exceeds the base station channel capacity C . In this case, the suggested scheme allocates the base station channel capacity to the two operators equally at the 1:1 ratio. In other words, when the offered load rate of the MNO is $C/2$ or lower, the suggested scheme allocates the base station resource equally to operators even if the offered load of the MVNO increases excessively. This prevents the case where excessive increase in offered load of a certain operator undermines quality of service that other operator provides. As shown in the Fig. 2-(b), when the offered load of the MVNO is low, both of the operators show the call admission rate that is close to 100%. When the offered load of the MVNO is 100Erlang or higher, the call admission rate of the MNO is kept constant, whereas that of the MVNO plummets due to excessive offered load.



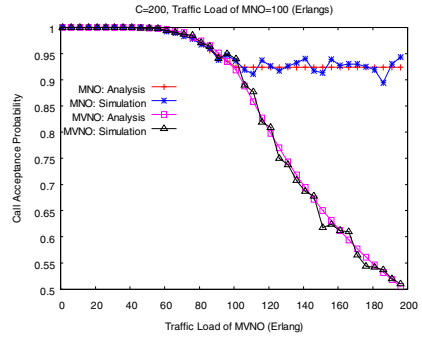
(a) When load of MNO is low



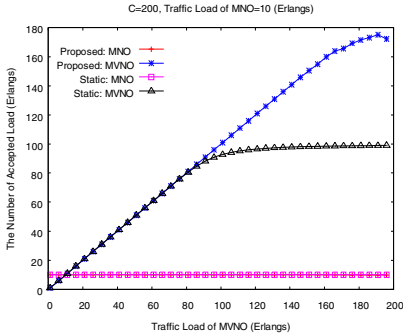
(b) When load of MNO is high

Fig. 1. Allocated resources

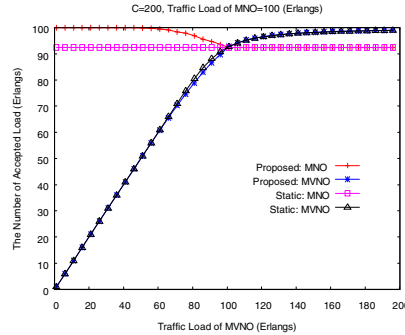
(a) When load of MNO is low



(b) When load of MNO is high

Fig. 2. Call admission probabilities

(a) When load of MNO is low



(b) When load of MNO is high

Fig. 3. The number of acceptable users

The Fig. 3 illustrates the results of comparison of the number of accepted users by each operator in case of between static allocation of the base station resource to MNO and MVNO at the 1:1 ratio and application of the suggested scheme. This experiment was conducted to compare the efficiency of the suggested resource sharing scheme with that of the static scheme. When the offered load of the MNO is low, the static allocation scheme and the suggested scheme accept all of the offered loads of the MNO and accept the offered loads of the MVNO, which are 90Erlang or lower (Fig. 3-(a)). However, when the offered load of the MVNO exceeds 90Erlang, the static scheme accepts the offered load of the MVNO that is 100Erlang at the maximum. On the contrary, the suggested scheme accepts all of the offered loads of the MVNO by allocating the channel resource, which is not used by the MNO, to the MVNO. For the same reason, when the load of the MNO has the high value of 100Erlang and the load of the MVNO is low, it is possible to increase the acceptable users of the MNO by using the suggested scheme, rather than the statics scheme. When the sum of the offered loads of the two operators exceeds the channel resource of base station, which means the case of overload, the suggested scheme shows the same performance as that of the static scheme that allocates base station resource at the 1:1 ratio.

5 Conclusions

In this paper, the cooperative game theory was used to suggest the management model that allocated voice channel resource of base station to MNO and MVNO in a fair and efficient way. The mathematical analysis and the simulation experiment were conducted to verify validity of the model. The suggested model increased utilization of channel resource by allowing operator with high load to use resource of operator with low load in consideration of offered loads of operators that shared base station resource. As a way to conduct future study, the suggested model has been expanded to a model that enables MNO and MVNO to share base station resource, which aims at transmitting data with different characteristics such as voice, data and video call on top of voice call. In addition, study has been underway to expand the suggested model to a model that enables efficient and asymmetric allocation of base station resource to MNO and MVNO.

References

1. Altman, E., Boulogne, T., El-Azouzi, R., Jimenez, T., Wynter, L.: A survey on networking games in telecommunications. *Computers & Operations Research* 33, 286–311 (2006)
2. Dubey, P.: Inefficiency of Nash equilibria. *Mathematics of Operational Research* 11(1), 1–8 (1986)
3. Antoniou, J., Kourkoutsidis, I., Jaho, E., Pitsillides, A., Stavrakakis, I.: Access network synthesis game in next generation networks. *Computer Networks* 53, 2716–2726 (2009)
4. Zhang, Z., Shi, J., Chen, H., Guizani, M., Qiu, P.: A cooperation strategy based on Nash bargaining solution in cooperative relay networks. *IEEE Trans. on Vehicular Technology* 57(4), 2570–2577 (2008)
5. Attar, A., Nakhai, M., Aghvami, A.: Cognitive Radio Games for Secondary Spectrum Access Problem. *IEEE Trans. on Wireless Communications* 8(4), 2121–2131 (2009)

6. Osborne, M., Rubinstein, A.: A course in game theory, pp. 117–132. The MIT Press, Cambridge (1994)
7. Zhang, Y., Xiao, Y., Chen, H.: Queuing analysis for OFDM subcarrier allocation in broadband wireless multiservice networks. *IEEE Trans. Wireless Communications* 7(10), 3951–3961 (2008)
8. Navarro, E., Mohsenian-Rad, A., Wong, V.: Connection admission control for multi-service integrated cellular/WLAN system. *IEEE Trans. on Vehicular Technology* 57(6), 3789–3800 (2008)

Load Balancing with Fair Scheduling for Multiclass Priority Traffic in Wireless Mesh Networks

Neeraj Kumar¹, Naveen Chilamkurti², Jong Hyuk Park³, and Doo-Soon Park⁴

¹ School of Computer Science & Engineering, SMVD University, Katra (J&K), India

² Department of Computer Engineering, LaTrobe University, Melbourne, Australia

³ Department of Computer Science & Engineering,

Seoul National University of Science and Technology, (SeoulTech), Korea

⁴ Division of Computer Science & Engineering, SoonChun Hyang University, Korea

nehra04@yahoo.co.in, n.chilamkurti@latrobe.edu.au,

parkjonghyuk1@hotmail.com

Abstract. Wireless Mesh Networks (WMNs) are emerging as a key technology for providing cost effective, resilient and adaptive services for the next generation networks. WMNs provide a number of applications to end users such as Video on demand (VoD), Voice over IP (VoIP), high speed internet access, video conferencing etc. To provide Quality of Service (QoS) to these applications, load balancing and fair scheduling are important metrics. In this paper, we propose a novel strict priority based fair load balancing (SPFLB) algorithm for multi class priority traffic in WMNs. The available capacity (AC) and Load Index (LI) of the link is calculated and based upon these values cost evaluation function (CEF) is defined. The performance of the proposed algorithm is evaluated using simulations on ns-2 with respect to various metrics. The results obtained show that the proposed algorithm is quite effective for balancing the load of multiclass traffic flows.

Keywords: Load balancing, fairness, traffic flow, priority traffic.

1 Introduction

Over the recent years, Wireless mesh networks (WMNs) are emerging as a new technology to provide cost effective services to the end users because it is self configured and self healing, low maintenance cost, and have easy deployment. A WMN combines the fixed network (backbone) and mobile network (backhaul). The nodes in a WMN often act as both relays, forwarding traffic to or from other mesh nodes, or providing localized connectivity to mobile or pervasive wireless devices, such as laptops, desktops and other mobile clients [1]. For each WMN, there are mesh gateways (MG), mesh routers (MRs) and mesh clients (MCs). Every node in WMNs acts as a router, forwards the packet to other nodes. Some of these routers may act as gateways which are directly connected to internet.

Recent research shows that Routing and MG selection are two key issues in determining the overall network performance with respect to throughput and capacity of WMNs [2, 3]. This is due to the fact that if many MCs select the same MGs, then the

selected gateway may be overloaded resulting the limiting access of capacity of WMNs. But on the other hand, a prioritized traffic mechanism can overcome this difficulty by blocking some traffic in a particular time interval and forward the traffic to MG which has higher priority. This will reduce the load on MG and increase the performance of the overall system. Currently, load balancing and QoS is addressed by the differentiated services (DiffServ) architecture [4] and fairness among TCP flows is addressed by various congestion control mechanisms. The load balancing mechanism is based upon avoiding network congestion and achieving fair rate allocation amongst the TCP flows [5, 6]. Packet marking bits can also be used by routers for congestion avoidance [7, 8].

To address the load balancing and routing issue, in this paper, we propose a novel mechanism for load balancing and fairness among traffic flows in WMNs. The proposed mechanism is used to estimate the link capacity and overloading of Internet fixed lines (if any). A strict priority fair load balancing (SPFLB) algorithm is proposed. The main objective of this algorithm is to classify the traffic into multiple flow classes and set the priority which distributes the traffic load among multiple gateways to ensure evenly utilization of network resources.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 describes network model and problem formulation. The proposed algorithm is described in section 4. Section 5 presents the simulation environment with results and discussion. Finally, Section 6 concludes the paper.

2 Related Work

Problem of load balancing has been studied widely in recent years. Cardellini et al. review the state of the art in load balancing techniques on distributed Web server systems [9]. Bryhni et al. present a comparison of load balancing methods for scalable Web servers [10]. Schroeder et al. [11] overview the clustering technology for Web-server clusters. Many proposals have applied queuing theory on single-hop ad hoc networks [12–14]. In [15], the authors analyze the average end-to-end delay and maximum achievable per-node throughput using G1/G/1 queuing networks. The proposal in [15] does not take the flow-level behaviors in consideration. Moreover, proposals in [16, 17] have developed queuing models to analyze the network capacity of heterogeneous WMNs. These papers extend those analytical studies to incorporate packet losses in the channel modeling. Recently Ancillotti et. al.[18] have formulated the load balancing problem as capacity estimation problem with capacity calculation for each node. The authors proposed a novel load aware route selection algorithm and showed 240% throughput improvement over the existing solutions.

The major challenge in all applications of WMNs is to provide QoS support and fair flow allocation. Traffic flow is studied with respect to fairness by defining fairness level parameter in [19]. By varying the fairness level from zero to infinity, a spectrum of fair rates is derived in which proportional fairness [20] and max–min fairness [21] are studied. The authors have also demonstrated a fair end-to-end window-based control which is globally asymptotically stable. The window-based algorithm uses packet round trip time (RTT) information and does not require feedback information from the routers which is different from the algorithms defined in [22, 23].

3 Network Model and Problem Formulation

The proposed network model for SPFLB for multiclass traffic is shown in figure 1. The wireless mesh backbone is formed by several MRs as shown in figure 1. These MRs takes the MCs requests and maintain different priority queues for multi class traffic. Let a WMN is represented by $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is the set of nodes/MCs and $E = \{e_1, e_2, \dots, e_n\}$ is the set of edges/communication links in the network. Let set of traffic flows are F , set of links are L . Every flow corresponding to a particular traffic has a priority from the priority set $P = \{p_1, p_2, \dots, p_n\}$ $0 \leq p_i \leq 1$. This priority decides its scheduling preference over the other flows from different network traffic class. Each flow is also associated with rate index $RI = \{r_1, r_2, \dots, r_n\}$ and delay estimation (DE). Before allocating a flow to a link l , its load index (LI) is calculated. LI is a measure of how much load that link has at a particular time interval, i.e., $LI = (p_i * \sigma_i), 1 \leq i \leq n$. Where σ_i is the request arrival rate from MCs. Define the service time and mean load of the link as: $E(\delta) = \mu, \beta = \sigma\mu$, where δ is the serving time, μ is the service rate.

Each link has bounds on AC which is a measure for upper and lower threshold on link capacity. As soon as the load on a particular link crosses these bounds, the load has to transfer to another link dynamically at run time.

Define the cost evaluation function for a link as

$$(CEF) = \frac{(\sum_L LI)}{(\sum_L F^f + \sum_L AC)} \dots \dots \dots (1), \text{ where } f \text{ is the fairness among the traffic flows}$$

F. Hence the combined objective of the proposed scheme reduces to $\min(CEF)$.

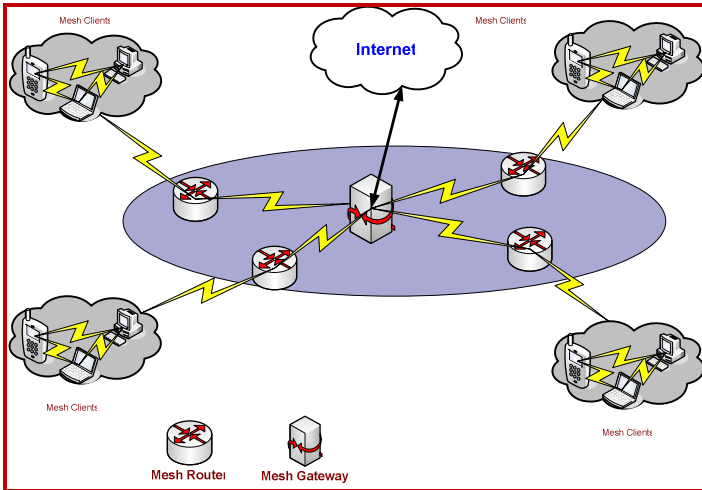


Fig. 1. Network Model and structure of router in proposed system

4 Proposed SPFLB Algorithm

As shown in figure1, the traffic from MCs comes through MRs to MGs. To balance the traffic load from multiple MCs, SPFLB algorithm is proposed. The algorithm considers multiple traffic flows and assigned a flow to link using Load Assignment Matrix (LAM). The LAM consists of entries of LI. There are two phases in the proposed algorithm:

- Calculation of capacity of the link and construction of ACM
- Assignment of load to a particular link based upon the values of LI and AC

4.1 Calculation of Capacity of the Link and Construction of ACM

To assign the load to a particular link, the capacity of the link gives a good estimation for load balancing. As the traffic flows come continuously in WMNs from different traffic classes, an estimation of AC is used for construction of ACM. As the capacity of the links changes with respect to time, a matrix ACM is constructed by different values of ACs. The value of AC can be calculated by total number of traffic flows,

i.e., $AC_{ij} = \frac{LI_{ij}}{F}, 1 \leq i \leq j \leq n$. The values in ACM are arranged as (AC_{ij}) ,

where $AC_{ij} \in E, 1 \leq i \leq j \leq n$

Table 1. Notations and their meaning dividing the LI

Notations	Meanings
V	Set of vertices
E	Set of edges
F	Total number of traffic flows
L	Set of links
P	Priority set
R	Rate of flows
RI	Rate index
LI	Load Index
AC	Available capacity
DE	Delay estimation
f	Fairness among traffic flows
CEF	Cost evaluation function
g	Total traffic intensity
K	Total number of traffic flows
B_1, B_2	Lower and upper bounds
TTL	Time to live
thr	Threshold for time to live value

4.2 Assignment of Load to a Particular Link Based Upon the Values of LI and AC

Once the AC of each link is calculated, then the available load of traffic flows is arranged according to the values of AC and LI. As the traffic demands are served, LAM is constructed containing the entries of LI for a particular link. The construction of LAM is as (LI_{ij}) .

As the new traffic flows come from MCs, these are placed in higher or lower layers of LAM depending upon the metric LI. In the initial step, all the traffic flows requests are accepted from MCs. The priority of each incoming flow is checked. If the priority is for real time flow, the link having minimum value of CEF is taken and flow is allocated to this link else these flows are sorted in decreasing order of RI values.

At each step of iteration, the existing flow allocation is increased by increasing the value of LI by one and decrement the value of AC by one. To provide the fairness among traffic flows, at each round of iteration the value of f is calculated. The value of fairness is between 0 and 1 ($0 \leq f \leq 1$), with 1 as complete fairness while 0 as unfairness. For simplicity, we have considered mainly two types of traffic flows as real time and elastic having probability as p and p' . A lower bound B_1 and upper bound B_2 can guide the MRs to block a particular traffic or not. The upper and lower bounds can be found using Erlang B ($M/M/k/k$) as follows:

$$B_2(k, g) = \frac{g^k / k!}{\sum_{i=0}^k g^i / i!} \dots\dots\dots(2), \text{ } k \text{ is number of traffic flows, } g \text{ is traffic intensity.}$$

Similarly, the lower bound on AC can be modeled as $M/M/k/D$ queue as:

$$B_1(k, g, D) = \frac{g^D P_0}{k^{D-k} k!}, \text{ where } D = D + k * CEF \dots\dots\dots(3)$$

$$P_0 = \left(\sum_{n=0}^{k-1} \frac{g^n}{n!} + \sum_{n=k}^D \frac{g^n}{k^{n-k} k!} \right)^{-1}$$

The expressions in equations (2-3) give a rough estimation of upper and lower bound.

5 Results and Discussion

To evaluate the proposed SPFLB algorithm as shown below, we have used network simulator ns-2 [24] with respect to metrics such as response time and throughput.

a) Impact on response time

Figure 2 shows the comparison of the response time in the proposed SPFLB and non priority based algorithms such as multipath [25] and cost sensitive [26]. With an increase in number of MCs requests, the response time also increases. But the results obtained show that the proposed SPFLB algorithm outperforms the non priority based

algorithms [25-26] with respect to response time. This is due to the fact that proposed algorithm calculates AC and LI of each link before allocation of flow to a particular link. This results in less congestion over a particular link and reduction in collision among the competing flows. Hence the response time of the proposed SPFLB algorithm reduces compared to its counterpart.

SPFLB algorithm

Input: $G = (V, E)$, traffic demands TD, CEF of a link

Output: Strict priority and fair load balancing

Step1: Initialization of AC and LI

For $i=1,2,\dots,L$ **do**

$TTL \leftarrow thr, LI \leftarrow 0, F_i \leftarrow 0, DE \leftarrow 0,$

End for

repeat

accept the requests from MCs

Calculate the AC of each link by dividing LI with number of incoming traffic flows

Calculate the bounds B_1 and B_2 as defined in equations (2) and (3) above

If $(B_1 \leq AC \leq B_2)$

If $(P = p)$

Choose the link with minimum value of CEF as $l \leftarrow \min(CEF)$

Assign the traffic flow to l as $l \leftarrow F, F_i \leftarrow 1$

Decrement the available capacity as $AC \leftarrow AC - 1$

Increment LI as $LI \leftarrow LI + 1, ACM \leftarrow ACM + 1, LAM \leftarrow LAM + 1$

$DE \leftarrow DE + TTL$

Else

Sort the traffic flows in decreasing order of RI

Assign the traffic flow to l as $l \leftarrow F, F_i \leftarrow 1$

Decrement the available capacity as $AC \leftarrow AC - 1$

Increment LI as $LI \leftarrow LI + 1$

$ACM \leftarrow ACM + 1, LAM \leftarrow LAM + 1$

Else

Discard the traffic flow

Until $(F \neq \phi)$

b) Impact on throughput

The impact of the proposed SPFLB algorithm on throughput is shown in figure 3. As shown in figure 3, the throughput of the proposed SPFLB algorithm is higher than the non priority based load balancing algorithms such as multipath [25] and cost sensitive [26]. This due to the fact that the proposed algorithm has strict priority for real time traffic compared to elastic or non real time traffic. Hence the proposed algorithm has higher throughput than its counterpart. Moreover, the flows are assigned to a particular link based upon the CEF which is a combination of LI and AC. At each step of the

iteration of SPFLB algorithm, fairness is also preserved by including the fairness factor f in CEF value. Hence due to these factors throughput in the proposed algorithm is higher than a non priority based algorithms [25-26].

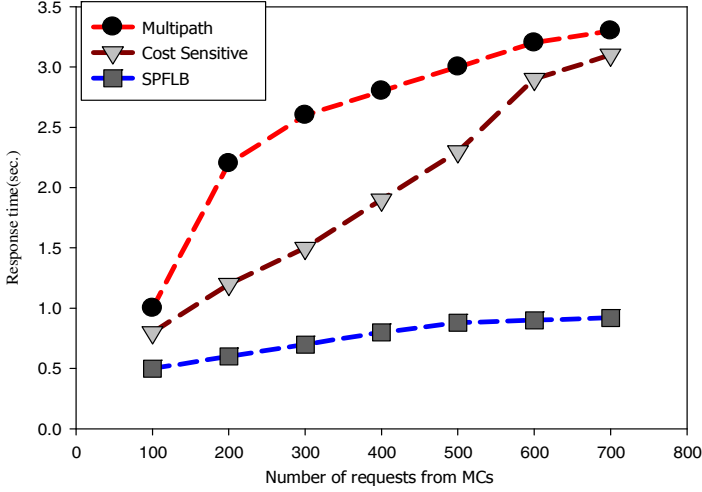


Fig. 2. Comparison of response time in proposed SPFLB and non priority based algorithms

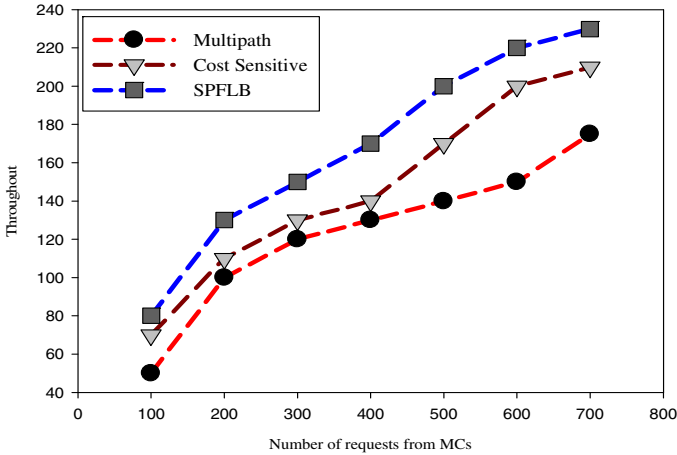


Fig. 3. Comparison of throughput in proposed SPFLB and non priority based algorithms

6 Conclusions

In this paper, we propose a strict priority fair load balancing (SPFLB) algorithm for assigning the different traffic flows to links. To assigning a traffic flow to a link, CEF

is evaluated which is a combination of AC and LI. Moreover, to achieve fairness among the different traffic flows fairness factor is also included in CEF. Two types of traffic flows are considered as real time and elastic. As the flows are allocated to links, two types of matrices are constructed as ACM and LAM based upon the values of AC and LI. The performance of the proposed SPFLB algorithm is evaluated using metrics such as response time, and throughput among different traffic flows. The results obtained show that the proposed algorithm is better than a non priority based scheme with respect to response time and throughput.

References

1. Akyildiz, F., Wang, X., Wang, W.: Wireless Mesh Networks: a survey. *Journal of Computer Networks* 47(4), 445–487 (2005)
2. Liu, Liu, Z., Towsley, D.: On the capacity of hybrid wireless networks. In: *Proc. of IEEE INFOCOM 2003*, vol. 2, pp. 1543–1552 (2003)
3. Zou, P., Wang, X., Rao, R.: Asymptotic capacity of infrastructure wireless mesh networks. *IEEE Trans. Mobile Computing* 7(8), 1011–1024 (2008)
4. Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W.: An architecture for differentiated services. IETF Network Working Group RFC 2475 (December 1998)
5. Kelly, F., Maulloo, A., Tan, D.: Rate control for communication networks: shadow price proportional fairness and stability. *J. Oper. Res. Soc.* 49, 237–252
6. Mo, J., Walrand, J.: Fair end-to-end window-based congestion control. *IEEE/ACM TON* 8(5), 556–567 (2000)
7. Floyd, S., Jacobson, V.: Random early detection gateways for congestion avoidance. *IEEE/ACM TON* 1(4), 397–413 (1993)
8. Floyd, S., Fall, K.: Promoting the use of end-to-end congestion control in the internet. *IEEE/ACM TON* 7(4), 458–472 (1999)
9. Cardellini, V., Colajanni, M., Yu, P.S.: Dynamic load balancing on Web-server systems. *IEEE Internet Computing* 3(3), 28–39 (1999)
10. Bryhni, Klovning, E., Kure, Q.: A comparison of load balancing techniques for scalable web servers. *IEEE Network*, 58–64 (2000)
11. Schroeder, T.: Scalable Web server clustering technologies. *IEEE Network*, 38–45 (2000)
12. Alizadeh-Shabdiz, F., Subramaniam, S.: A finite load analytical model for IEEE 802.11 distributed coordination function MAC. In: *Proc. ACM WiOpt 2003*, France (2003)
13. Özdemir, M., McDonald, A.: An M/MMGI/1/K queuing model for IEEE 802.11 ad hoc networks. In: *Proc. IEEE PE-WASUN 2004*, Venice, Italy, pp. 107–111 (2004)
14. Tickoo, O., Sikdar, B.: Modeling queuing and channel access delay in unsaturated IEEE 802.11 random access MAC based wireless networks. *IEEE/ACM Trans. Networking* 16(4), 878–891 (2008)
15. Bisnik, N., Abouzeid, A.: Queuing network models for delay analysis of multihop wireless ad hoc networks. *Ad Hoc Networks* 7(1), 79–97 (2009)
16. Bruno, R., Conti, M., Pinizzotto, A.: A queuing modeling approach for load-aware route selection in heterogenous mesh networks. In: *Proc. of IEEE WoWMoM 2009*, Greece (2009)
17. Bruno, R., Conti, M., Pinizzotto, A.: Capacity-aware routing in heterogeneous mesh networks: an analytical approach. In: *Proc. of IEEE MsWiM 2009*, Tenerife, Spain (2009)
18. Ancillotti, E., Bruno, R., Conti, M., Pinizzotto, A.: Load-aware routing in mesh networks: Models, algorithms and experimentation. *Computer Communications* (2010)

19. Mo, J., Walrand, J.: Fair end-to-end window-based congestion control. *IEEE/ACM TON* 8(5), 556–567 (2000)
20. Kelly, F., Maulloo, A., Tan, D.: Rate control for communication networks: shadow price proportional fairness and stability. *J. Oper. Res. Soc.* 49, 237–252 (1998)
21. Charny, A.: An algorithm for rate allocation in a packet-switching network with feedback, M.A. Thesis. MIT, Cambridge, MA (1994)
22. Low, S.H., Lapsley, D.E.: Optimization flow control I: basic algorithm and convergence. *IEEE/ACM TON* 7, 861–875 (1999)
23. Paganini, F., Wang, Z., Low, S.H., Doyle, J.C.: A new TCP/AQM for stability and performance in fast networks. In: *Proc. of IEEE INFOCOM* (April 2003)
24. Fall, K., Varadhan, K. (eds.) *NS notes and documentation*, The VINT project, LBL (February 2000), <http://www.isi.edu/nsnam/ns/>
25. Hu, X., Lee, M.J.: An efficient multipath structure for concurrent data transport in wireless mesh networks. *Computer Communications* 30, 3358–3367 (2007)
26. Zeng, F., Chen, Z.-G.: Cost-Sensitive and Load-Balancing Gateway Placement in Wireless Mesh Networks with QoS Constraints. *Journal of Computer Science and Technology* 24(4), 775–785 (2009)

A Dependable and Efficient Scheduling Model and Fault Tolerance Service for Critical Applications on Grid Systems

Bahman Arasteh and Mohammad Javad Hosseini

Islamic Azad University, Tabriz Branch, Tabriz, Iran

B_arasteh@iaut.ac.ir

Islamic Azad University, Sufian Branch, Sufian, Iran

Smj.hosseini@gmail.com

Abstract. The grid system is a framework with heterogeneous remote resources and a hazardous environment. Hence, the reliability and performance must be considered as a major criterion to execute the safety-critical applications in the grid. This paper proposes a model for job scheduling and fault tolerance service in the grid to improve dependability with respect economic efficiency. Dynamic architecture of the scheduling model leads to reduce resource consumption. The proposed fault tolerance service consists of failure detection and failure recovery. A three layered detection service is proposed to improve failure coverage and reduce the probability of false negative and false positive states. Checkpointing technique with an appropriate graining size is proposed as recovery service to attain a tradeoff between failure detection latency and performance overhead. Analytical approach (Markov approach) is used to analyze the reliability, safety and economic efficiency of proposed model in the presence of permanent and transient faults.

Keywords: Grid Scheduler, Dependability, Fault Tolerance Services, Economic Efficiency.

1 Introduction

Grid computing is a heterogeneous environment that enables computational resource sharing among many organizations in the world. Grid computing as an efficient distributed system with large number of heterogeneous resources and parallel processing capability is a framework for executing heavy and computationally intensive jobs in a parallel manner with reasonable costs [3]. On the other hand, computation intensive applications such as molecular sample examining, simulation of airplane and research concerning nuclear boiling need many hours, days or even weeks of execution. These application are the ones in which failure are not acceptable and may lead to catastrophes. Timely results and high level of reliability are the main constraints in the mission oriented applications. Responsiveness and dependability are the main measures in the high available applications; availability and safety are the main factors in the long mission applications. In Addition to timing

constraints, the real time applications need the high level of dependability like reliability and safety. With respect these features the grid system can exploit to execute the Safety-critical applications like long mission oriented applications and real-time applications which need high degree of performance and dependability.

On the other hand failure occurrence in each component of grid environment is a rule not an exception. Resource failure might fail the associated replicas. The resource in this paper refers to the computing resources. Hence, an efficient resource management and job scheduling mechanisms are needed to attain required quality of services [38]. In addition to service time, the dependability and its related criteria such as *reliability*, *safety* and *availability* must be considered in the grid resource selection and job scheduling. Resource selection and job scheduling in the grid are complex processes because the resources of grid are dynamic, heterogeneous, distributed and can enter and leave the grid at any time. It must be noted that the middleware and tools of grid computing like Globus and Condor-G do not provide a general failure handling technique. Therefore, the reliability, safety and availability should be considered as well as performance criteria such as service time and efficiency of resources in the resource selection and job scheduling. From the user's point of view in the grid application layer the payment is the other quality factor.

2 System Model

2.1. Abstract Model of System

The resources in the grid environment can be categorized into computational resources, storage resources, network resources and application resource. Each site has a resource manager which collects the status information of the machines like workload, availability and reliability and manages the resources. Meta computing Directory Service (MDS) [35] is used to gather resource information and resource searching in the Globus. The machines of each site are connected through a high bandwidth network with low communication overhead. The communication latency between sites can be considerable. Resource manager is one of the grid middle ware components. Resource discovery, resource selection and job scheduling are the main functions of resource management service. Needed resources of a submitted job can

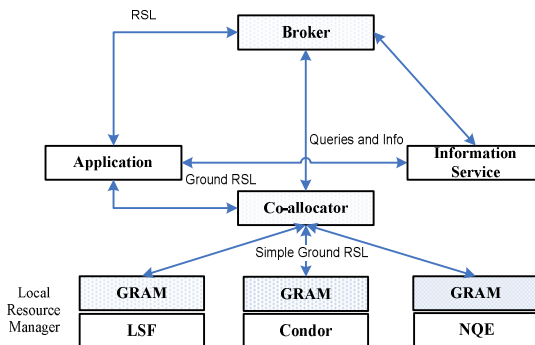


Fig. 1. An Overview of Resource manager in the Globus

be managed manually by users or automatically by resource management component in the grid middleware like Legion [33] and Condor-g [34]. Figure 1 shows an overview of resource management component with respect Globus [16, 35, 36]. The information service provides information about status of the resources. The co-allocator coordinates the resources of the multiple sites. Grid resource allocation management (GRAM) [16] is responsible for managing local resources. RSL refers to resource specification language [16, 36].

2.2 Fault Model

The resources may enter and leave the grid at any time. Hence, the grid is a hazardous environment and resource failure is a common event and not an exception. On the other hand, probability of fault, error and failure in each remote resource and the network framework is not negligible. Failures might happen during many stages of this process as a consequence of many software and hardware faults. The focus of this paper is on the *Resource failure and Local environment failure during the job execution*. A resource failure is occurred when a resource services is stopped because of *unavailability or resource crashes*. Many transient and permanent faults can lead to resource crashes which are listed in the following. *Fault in the host machine's resources like CPU or memory, Fault in the software layers of the host machine like operating system of the host and Fault in the transition channels* are fault model in this paper. The other type is the timing failure. Late results of a computation are not acceptable in some safety-critical applications. Hence, time threshold monitoring service is needed. One of the assumptions in this paper is the correctness and fault freeness of the submitted jobs which are replicated by scheduler.

3 Related Works

Fault-tolerance mechanisms can be implemented on different layers of grid computing. Fault tolerance mechanisms can be embedded in the *application layer* [29]. Using fault tolerance mechanisms into designing phase of scheduling system is the main focus of this approach, like N-version programming and Recovery- Blocks mechanisms. Fault-tolerant version of MPI is one of these mechanisms [30]. Exploiting fault-tolerance mechanisms into different layers of grid computing to handle relevant faults of each layer is the other technique. [31]. Developing an *external Fault-tolerance service* is the other type of gaining dependability in the grid environment [32]. This service can be implemented in the Globus and other toolkits.

Component Replication [22], Job Replication [23] and Data Replication [24] are Deferent replication methods in the grid computing. In the recently published papers these methods are used in the grid scheduling. Replication methods can be used in static or dynamic type. In the dynamic methods the resources such as component, job and data can vary during job execution. After error detection, dynamic fault tolerance methods use additional resources to tolerate detected error during job execution. Dynamic methods try to improve resource utilization and consequently reduce the cost of services in the economic grids. On the other hand, extending software fault tolerance methods to treatment hardware and software faults has been focused in the recent papers because of flexibility, scalability and portability of these methods.

Some classic and hybrid software fault tolerance techniques are discussed and verified with respect to dependability and performance aspects in previous papers [12][21][22]. Improvement of reliability with low overhead in performance and minimum redundancy are the main goals in the relevant papers [22, 23, 24, 25, 26 and 27]. The other significant point in the relevant papers which is considered in this paper is development of highly reliable output selection services like acceptance test (AT) or majority voter with low run time overhead. An output selection module (AT, voter or comparator) should be simple, effective and highly reliable to ensure that the anticipated fault is detected and non faulty state is not detected incorrectly as faulty (false negative). In some relevant papers the acceptance tests are considered perfectly reliable which it is not a reasonable and practical assumption in some critical job scheduling. An attempt has been made to use recovery blocks [2][8], consensus voting [2][12], Acceptance test [21] and transparent checkpoints in order to gain a required level of reliability, safety and availability with a reasonable cost in a economic grid system. The present scheduling model can handle hardware faults and resource failures during job execution on the grid. It integrates the advantages of fault masking, fault tolerance and dynamic replication techniques which lead to improve reliability, safety, performance and resources consumption. It uses a hybrid output selection service which decreases the possibility of false positive and false negative state. This service can be used for long mission, high available and soft real-time applications in which temporary incorrect outputs for a moment are acceptable.

3.1 A Classic and Reliable Model for Grid Scheduler

After a critical job is submitted through a host machine couldn't schedule it because of needed resources and job deadline, the grid resource management service are invoked to select needed resources and schedule it. The scheduler produces K replicas of it and dispatches them to appropriate selected machines that guarantee performance and reliability requirements. In order to Search needed resources the Meta computing Directory Service (MDS) in the Globus toolkit is invoked which provides the status information about resources and return the candidate resources. After resource selection the resource selection and scheduling services are invoked.

Passive redundancy of computing machines is one of classic model to develop fault tolerance services in grid systems. The majority voting is invoked by the scheduling services to compare the results of parallel replicas. The voter is developed as a services and its reliability affects on the total system reliability. This scheduler could at most tolerate $K/2$ resource failures. In addition, the Markov model has been used for reliability modeling and quantification of classic scheduling model. The occurrences of failures are discrete and the parameter λ is the failure rate and shows failure occurrences per time unit. The value of λ is assumed constant. $R_s(t)$ is the reliability of scheduling system. By means of markov model and Laplace transformation the following equations can be extracted.

$$R_s(t) = 3e^{-2\lambda t} - 2e^{-3\lambda t}, \quad MTTF = \int_0^{\infty} R(t)dt, \quad MTTF_{TMR} = \int_0^{\infty} R(t)dt = \int_0^{\infty} (3e^{-2\lambda t} - 2e^{-3\lambda t})dt$$

This model tolerates at most $K/2$ failures from k allocated machines. The TMR scheduling model can tolerate single resource failure during a job execution. According to the figure 2, the reliability of TMR model for grid scheduling service is

higher than basic scheduling model for a short mission jobs. In other word, this scheduling model is a good choice for scheduling short mission and real-time jobs with short deadline in the grid systems. But, the overall reliability for a long time interval is smaller than basic scheduling model. Hence, this model by means of three host machines can tolerate one failure. Mean time to failure is another dependability factor for grid scheduling system.

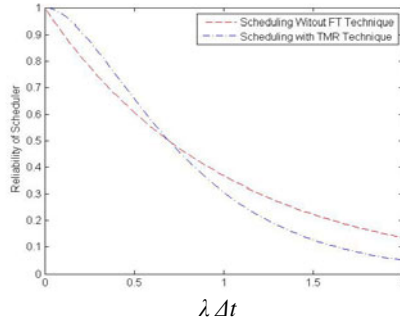


Fig. 2. Reliability of Scheduling Model without Fault Tolerance (FT) Technique and Scheduling Model with TMR Technique

$$MTTF: \text{Mean time to failure} \quad , \quad MTTF_{TMR} = \int_0^{\infty} R(t)dt = \int_0^{\infty} (3e^{-2\lambda\Delta t} - 2e^{-3\lambda\Delta t})dt$$

$$MTTF_{TMR} = 5/6\lambda < 1/\lambda \quad , \quad MTTF_{TMR} < MTTF_{Basic}$$

This model needs K host machine to start job execution simultaneously. Hence, it leads to increase the waiting time and on the other hand it is not economical in the economic grid. The other significant point is that in the NMR scheduling model the reliability of voter has important role in the overall reliability and Using perfectly reliable voter in the distributed systems commonly is practically impossible or needs high development cost and complexity.

4 Proposed Model

4.1 Architecture of Proposed Model

Our proposed model consists of two main components. The first is scheduling and the next is failure handling service. The resource search, resource selection and allocation as scheduling process are the main functions of resource manager in the grid systems. The following picture shows an overview of proposed model. Based on needed resources of the submitted job the MDS in the Globus tool is invoked and it finds the set of candidate resources. The needed degree of dependability and performance and remaining deadline of the job is important to discover the candidate resource.

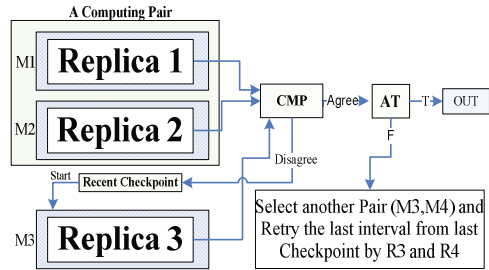
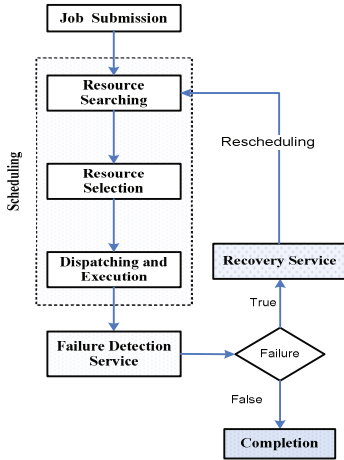


Fig. 3. An Overview of Proposed Model

Fig. 4. Scheduling model based on dynamic and active redundancy

In the resource discovery and resource selection algorithm the following parameters must be considered: Performance of the machines (workload and Length of the Local Queue), Dependability of the machine (Reliability, availability), Locality of the machine (The resources with high degree of locality impose low communication and performance overhead).

$$Resource\ Selection\ Criteria = \frac{Depenability * Locality}{Workload}$$

The optimal scheduling algorithm selects resources with hopes that the job will be completed with respect the remaining deadline. Some genetic and heuristic methods are proposed to estimate the dependability and performance of the resource before dispatching the job [38, 39]. The focus of this paper is on the failure detection and recovery after dispatching the job and during job execution. In the next step, the scheduler generates k replicas of the job and selects a computing pair consists of two ready machines from form the selected candidates. Two replicas from the K are assigned to the candidate pairs by the scheduler. This model requires two computing node at first as a computing pair and. Further resources are needed when the failure of an allocated resource is detected. Therefore, it has lower resource consumption and because of low needed resource at first it can start the submitted job quickly and reduces the total finish time. Hence, this scheduling model optimizes the dependability, performance and economic efficiency.

4.2 Failure Handling Services

The failure handling services consists of failure detection and recovery services. The selected pair starts to run the assigned replicas simultaneously and during the execution the detection service must monitor the resources and running job. The proposed model uses a three detection layer to improve detection coverage and

consequently total reliability. Middle results as Checkpoints are created concurrently with the running jobs and are used by recovery service. Figure 4 shows an overview of the model.

The other point that must be noted is a tradeoff between economic and reliability aspects. Using each redundancy technique such as job replication, component replication and data replication in the grid scheduling system increases the reliability and consequently cost of grid services. Therefore, a tradeoff is required between cost and dependability of services. On the other hand, using these techniques in the grid scheduling system involves resource and time overheads that must be noted in the overall performance. The previously mentioned classic model needs a high degree of resource redundancy to gain a desirable reliability. Therefore, the user must pay high costs for the requested services. In the long-life and highly available applications, the scheduling model based on classic redundancy technique demands more computing resources and consequently high service cost. Hence, this scheduling model is not economical for users which need high availability, safety and MTTF for their requested services. Scheduling, voting and checkpoint services must be implemented in the middleware and tools of grid like Globus.

4.2.1 Resource Failure Detection Mechanism

Resource crashes, resource unavailability and communication breakdown can lead to failure occurrence. The following are considered as the main question in the detection services: *Are the running machines available and online? Are the mid results correct? Are the time thresholds considered?* To answer these questions, the scheduler needs to detect errors and failure of the host node while running the jobs. This section of the paper deals with the detection failure during job execution before the deadline. The coverage refers to the probability of detecting an error induced by an arbitrary fault. It is difficult and more complex to have perfect coverage. In the proposed model, detection mechanisms are intended to detect resource failures through combination of function level detection and periodic diagnosis mechanisms. A diagnosis module, comparator and acceptance tests, as a function of detection services, are invoked periodically to verify mid results. The proposed schema failure detection covers both timing failure and content failure. In the timing failure the determined deadline for the computation of the result or delivery of the requested service is not met. Checking deadline constraints is an important factor in failure detection mechanisms especially in the real time applications. In timing failures detection, the detection service monitors time thresholds or checks time-out and deadline constraints of running. When the content failure occurs, the content of information delivered at the service interface is deviated from the specified functions [3]. The proposed model like heartbeat mechanism [25] detects timing and content failure during job execution. The detection service determines a time interval based on the remaining deadline of the job as a heartbeat signals period. This time must be smaller than the dead line of the job. When the scheduler assigns two replicas to the candidate pair, the machines start to run the replicas simultaneously for a time slice. After first interval execution the middle results of two replicas are sent to the scheduler. The detection service compares the received mid results which produced by the replicas in the first interval.

If the produced results in the first interval agree then the acceptance test (AT) is invoked. Using comparator and AT together enables detectors to detect two resource

failures. In this step, if the AT passes the results then the results will be stored as a last checkpoint. By saving at each checkpoint the state of the task in a reliable storage, the need to restart it after fault detection is avoided. If the results couldn't pass the test, the AT returns false. In this condition, the system is in the faulty state and must be recovered. We assume that the submitted jobs are perfectly reliable and there is not any software permanent fault in the jobs. This model can detect and tolerate two resource failures. After detection of this type of failure by AT, the recovery services are invoked. Therefore, the scheduler selects another ready computing pair by the mentioned selection algorithms and retries the last interval using the last stored checkpoint. The detection service stores this information in a database for statistical analysis and resource reliability estimation in the future. Figure 5 shows an overview of the detection and recovery techniques when two replicas were failed as a consequence of two hardware faults.

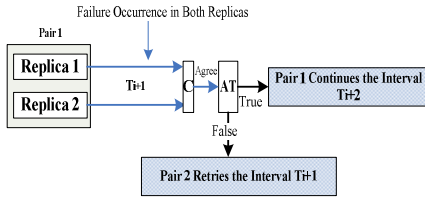


Fig. 5. The scheduler can detect two hardware faults in the pair1 and selects the other pair from ready candidate pairs by the selection algorithm.

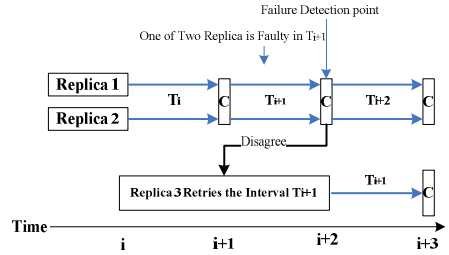


Fig.6. The scheduler retries the interval T_{i+1} on replica3 and at time $i+1$ by comparing the produced result of replica3 with the results of active replicas in the failure detection point, the faulty replica is detected.

If the produced results in the last interval disagree then a resource failure is detected and the states of the active replicas are stored at last interval. In this condition the scheduler select a machine and assigns a spare replica to it and retries the last interval from the last stored checkpoints. Both of the machines in the active pair do not stop during execution of the spare replica on the spare machine. Simultaneously the task continues forward on the active machines (pair) for the next interval. It leads to improve the resource utilization and performance. At the next checkpoint, the produced state of the spare machine is compared with the stored states in the failure point. Then the 2-out-of-3 decision determines the fault free replicas. In this step AT is applied to the outcome of the agreed independent replicas after comparison to confirm validity of the selected fault free replicas. The wrong agreed results can be diagnosed by AT. If the AT accepts the agreed states, then the scheduler stores it as a current checkpoint. At last the failed machine in the pair is released and the other active machine with spare replica makes a new active pair. Figure6 shows an overview of this recovery technique. Therefore, if the results of the pair disagree, a spare replica execution on a spare machine and AT together are used to diagnose the faulty resource. Using *three layer detection service* improves the scheduling reliability by using comparator, AT and spare replica execution. This technique reduces the probability of false negative and false positive states. Based on

submitted job features, different AT algorithms can be used. High coverage, short run-time and low development costs are the main criteria for selecting algorithms for AT. If the submitted job consists of a smaller code segment with logical or mathematical constraints, then the satisfaction of requirements [21] is regarded as the effective algorithm for AT. Conversely, if there are some pre-computed constraints such as pre-computed results, expected sequences of job states or other expected relationship in the running job, then the reasonableness test [21] is effective AT in this case [19].

4.2.2 Recovery Service and Graining Size

The checkpointing process is saving the status of running replicas and active machines in a reliable storage which is used as a recovery technique in the proposed recovery service. This model focuses on the system level checkpointing which is supported by Condor and Libckpt. The other type of checkpointing is application level checkpointing which is supported by some middleware and tools like BOINC and XtremWeb. The information saved at the checkpoints includes the mid results of replicas and state of host machine during job execution. In each interval the active machines in the pair sends the mid results to scheduler as heartbeat data. If a machine couldn't send in the determined time threshold a timing failure or halt will be detected. There needs to be a trade-off between the frequency of checkpoints and performance measurements. Hence, one of the main parameters that affect the dependability and performance is the granularity. Granularity is considered the time interval between checkpoints. To determine the graining size the following factors must be considered: *transparency, performance overhead consists of time and space and Failure detection latency.*

Transparent checkpointing process does not require the users or resource owner's efforts. Fine grains increase the number of checkpoints and consequently it leads to high performance overhead and low detection latency. On the other hand, coarse grains decrease the number of checkpoints and performance overhead and consequently it leads to high error detection latency. Therefore, the detection service should adjust the graining size with respect to the needed dependability, performance and deadline of the job. Optimal checkpointing intervals reduce the execution time and the needed space for saving the checkpoints. Check pointing time overhead consists of mid results transferring time, comparison time and saving time. In some systems with special purposes, two types of checkpoints can be defined. One checkpoint is used to invoke detection methods and the other one is defined as a store checkpoint. Increment of checkpoint frequency leads to increase performance overhead and decrease error detection latency. Sayori [20] have proposed an equation to find optimal checkpoint intervals. If the scheduler has the following parameters in the submitted job, then it can find the optimal checkpointing interval. T shows Checkpointing Interval and C shows Check pointing overhead which is comprised of Comparison Time and AT Time. λ_i is the constant rate of error occurrence in the machine i.

$\lambda_i = \lambda_j = \lambda$, For all $i, j > 0$

$$T^* = \frac{C}{2} \left[\sqrt{1 + \frac{2}{\lambda C}} - 1 \right]$$

5 Performance and Dependability Evaluation

5.1 Reliability and Safety Analysis

In order to evaluate the reliability and safety of proposed scheduling model, we use an analytical approach. This model improves the percentage of executed jobs successfully with respect the needed dependability before deadline. Markov model and fault tree have been using to dependability analysis of software and hardware systems [12, 17, 18, 28]. We have used the markov approach to analysis of reliability and safety of proposed scheduling model. Figure 8 shows the reliability of proposed scheduling model.

5.2 Performance and Resource Consumption analysis

Performance and Resource Consumption Analyses. The proposed fault tolerant (FT) scheduling model has a dynamic architecture. This model needs just two host machines with needed performance and reliability at first step but the NMR needs seven host machines to start. Hence, this model does not lead to starvation and reduces the waiting time. After a failure is detected during a job execution the scheduler selects the other candidate machine and does not waste the resources before failure occurrence. Therefore, dynamic architecture of proposed model reduces total service time and improves resources efficiency. This model improves the percentage of accepted requested resources and services in the grid systems. At the worst case the proposed scheduler can tolerate three host machine failures by means of five host machines during a job execution. On the other hand the NMR based scheduler can tolerate three host machine failures by means of seven host machines .Hence, the new model has lower needed resources in average case and this feature improves efficiency and reduces the service cost. The result of analysis shows that the proposed model has higher reliability and consequently safety than the NMR based scheduler. MTTF and availability is the other dependability factors which are optimized by using this model.

Economic Efficiency. Our proposed model considers the economic efficiency in the scheduling process. The resource cost or payment for requested resources are the economic efficiency from the grid user's point of view. As mentioned previously, this model needs two machines to start the job and allocates further resource after failure detection dynamically. Hence, the average needed resource is lower classic fault tolerance scheduling model which are shown in the figure9. The experiments show that the proposed model has the low average payment. On the other hand the penalties that are paid to the users by resource and service providers have been decreased because the percentage of failed jobs.

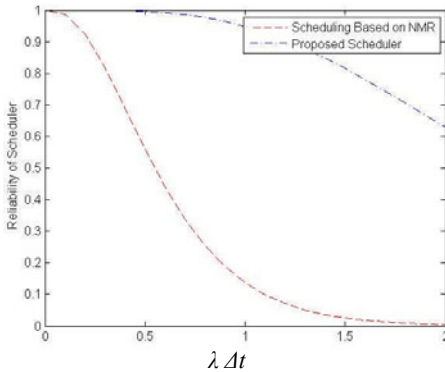


Fig. 7. Reliability of Proposed Scheduling model and NMR based Scheduling

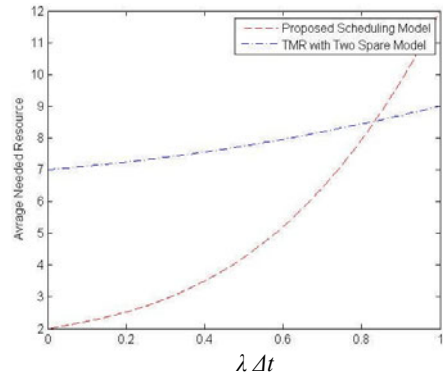


Fig. 8. The average needed resource to tolerate three failures during a job scheduling

6 Conclusions

This paper proposes a scheduling model with a fault tolerance service which improves the reliability, safety and availability. Dynamic architecture of the model leads to reduce resource consumption and improve economic efficiency. The proposed fault tolerance service consists of failure detection and failure recovery. The proposed three layered detection service leads to improve failure coverage and reduce the probability of false negative and false positive states. The recovery service uses checkpointing techniques in the system or application level with an appropriate time interval to attain a tradeoff between failure detection latency and performance overhead. Low waiting time to start the job is the other improvement of this work. Three layered detection technique masks the failure of AT and Comparator.

References

1. Armoush, Salewski, F., Kowalewski, S.: Efficient Pattern Representation for Safty Critical Embeded Systems. In: International Conference on Computer Science and Software Engineering (CSSE 2008) (2008)
2. Athavale, A.: Performance Evaluation Hybrid voting Schemas, Masters thesis, North Carolina State University, Department of Computer Sience (1989)
3. Avizienis, A., Laprie, J., Randle, B., Landwehr, C.: Basic Concepts and Taxonomy of Dependable and Secure Computing. IEEE Transaction on Dependable and Secure Computing, 11–33 (2004)
4. Bouteiller, A., Desprez, F.: Fault Tolerance Managment For Hairarchiecal GridRPC Middleware. In: Cluster Computing and The Grid (2008)
5. Huedo, E., Montero, S., Llorente, M.: An Experimental Framework for Executing Applications in Dynamic Grid Environments. ICASE Technical Report (2002)
6. Goto, H., Hasegawa, Y., Tanaka, M.: Efficient Scheduling Focusing on the Duality of MPL Representatives. In: Proc. IEEE Symp.

7. Shan, H., Olike, L.: Job Superscheduler Architecture and Performance in Computational Grid Environments. In: SC 2003 (2003)
8. Foster, I.: The Anatomy Of the Grid: Enabling Scalable Virtual Organization. International J. Super Computer Applications, 15–18 (2001)
9. Gehring, J., Preiss, T.: Scheduling a Metacomputer with Uncooperative Sub-schedulers. In: Proc. JSSPP 1999, pp. 179–201 (1999)
10. Shin, K.G., Lee, Y.: Error Detection Process-Model, Design and Its Impact on Computer Performance. IEEE Transaction on Computers c-33(6) (June 1984)
11. Chepten, M., Claeys, A., Dhoet, B., DE Turck, F., Demeester, P., Vanrolleghem, P.A.: Adaptive Task Checkpointing and Replication: Toward Efficient Fault Tolerant Grid. IEEE Transaction on Parallel and Distributed Systems 20(2), 180–190 (2009)
12. Lyu, M.: Handbook of Software Reliability Engineering. McGraw-Hill and IEEE Computer Society Press, New York (1996)
13. Zhang, L.: Scheduling Algorithms for Real Time Application on Grid Environment. In: Proceedings of IEEE Real Time System Symposium. IEEE Computer Society Press, Los Alamitos (2002)
14. Globus resource allocation Manager (GRAM) 1.6, <http://www.globus.org/gram>
15. Shooman, M.L.: Reliability of Computer Systems and Networks: Fault Tolerance, Analysis, and Design. John Wiley & Sons, Inc., Chichester (2002), 0-471-29342-3 (Hardback); ISBNs: 0-471-22460-X
16. Johnson, W.: Design and Analysis of Fault-Tolerant Digital Systems. Addison-Wesley Publishing Company, Inc., Reading (1989) 0-201-07570-9
17. Hecht, H.: Fault-Tolerant Software. IEEE Transaction On Reliability R-28(3), 227–232 (1979)
18. Nakagawa, S., Okuda, Y., Yamada, S.: Reliability Modeling, Analysis and Optimization, vol. 9, pp. 29–43. World Scientific Publishing Co. Pte. Ltd, Singapore (2006)
19. Pullum, L.: Software Fault Tolerance Techniques and Implementations. Artech House, Inc., Norwood (2001) ISBN: 1-58053-137-7
20. Arshad, N.: A planning-based approach to failure recovery in distributed systems. A thesis submitted to the University of Colorado in partial fulfillment of the requirements for the degree of Ph.D (2006)
21. Townend, P., Xu, J.: Replication - based fault tolerance in a grid environment, As part of the e-Demand project at University of Leeds, Leeds, LS2 9JT, UK (2004)
22. Antoniu, G., Deverge, J., Monnet, S.: Building fault-tolerant consistency protocols for an adaptive grid data-sharing service. IRISA/INRIA and University of Rennes 1, France (2004)
23. Jain, A., Shyamasundar, R.K.: Failure detection and membership management in grid environments. In: Proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing, GRID 2004, pp. 44–52 (2004)
24. Krishnan, S., Gannon, D.: Checkpoint and restart for distributed components in XCAT3. In: Proceedings of the Fifth IEEE/ACM International Workshop on Grid Comp. GRID (2004)
25. Choi, S., Baik, M., Hwang, C., Mingil, J., Yu, H.: Volunteer availability based fault tolerant scheduling mechanism in desktop grid computing environment. In: Proceedings of the Third IEEE International Symposium on Network Computing and Applications, NCA (2004)
26. Laprie, J.C., Arlat, J., Beounes, C., Kanoun, K.: Definition and Analysis of Hardware and Software Fault-Tolerant Architectures. Computer C-23, 39–51 (1990)

27. Medeiros, R., Cirne, W., Brasileiro, F., Sauve, J.: Faults in grids: Why are they so bad and what can be done about it? In: Fourth International Workshop on Grid Computing, p. 18 (2003)
28. Fagg, G.E., Dongarra, J.J.: FT-MPI: Fault tolerant MPI, supporting dynamic applications in a dynamic world. In: Dongarra, J., Kacsuk, P., Podhorszki, N. (eds.) PVM/MPI 2000. LNCS, vol. 1908, pp. 346–354. Springer, Heidelberg (2000)
29. Thain, D., Livny, M.: Error scope on a computational grid: Theory and practice. In: 11th IEEE International Symposium on High Performance Distributed Computing, p. 199 (2002)
30. Défago, X., Hayashibara, N., Katayama, T.: On the design of a failure detection service for large-scale distributed systems. In: Proceedings International Symposium Towards Peta-BitUltra-Networks, pp. 88–95 (2003)
31. Grimshaw, A., Wulf, W.: Legion—a view from 50,000 feet. In: Proceedings of 5th IEEE Symposium on High Performance Distributed Computing (1996)
32. Frey, J., Foster, I., Livny, M., Tannenbaum, T., Tuecke, S.: Condor-G: A Computation, Management Agent for Multi-Institutional Grids. University of Wisconsin, Madison (2001)
33. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers, Los Altos (1998)
34. Foster, I., Kesselman, C.: Globus: a metacomputing infrastructure toolkit. *Int. J. Supercomputer Appl.* 11 (2) (1997)
35. Foster, I., Roy, A., Sander, V.: A quality of service architecture that combines resource reservation and application adaptation, In: 8th International Workshop on Quality of Service (2000)
36. Abawajy, J.H.: Robust Parallel Job Scheduling on Service-Oriented Grid Computing. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3483, pp. 1272–1281. Springer, Heidelberg (2005)
37. Bin, Z., Zhaohui, L., Jun, W.: Grid Scheduling Optimization Under Conditions of Uncertainty. In: Li, K., Jesshope, C., Jin, H., Gaudiot, J.-L. (eds.) NPC 2007. LNCS, vol. 4672, pp. 51–60. Springer, Heidelberg (2007)

The Performance Evaluation of Heuristic Information-Based Wireless Sensor Network Routing Method

Hyekyoung Jeon¹, Seungjin Han², Keewook Rim³, and Junghyun Lee¹

¹ Department of Computer Science and Information Technology, Inha University,
253 Younghyun4-dong, Nam-gu, Incheon 402-751, Republic of Korea

² School of Information and Media, KyungIn Woomen's College,
101 Gyesan-gil, Gyeusan-dong, Gyeyang-gu, Incheon, 407-740 Republic of Korea

³ Department of Computer and Information Science,
SunMoon University, 336-708, Asan-si, Chungnam Republic of Korea
jkh7010@nate.com, softman@kic.ac.kr, rim@sunmoon.ac.kr,
jhlee@inha.ac.kr

Abstract. With recent technological advance, wireless sensor networks are often used in data collection and surveillance. One of the objectives of research on routing methods in wireless sensor networks is maximizing the energy life of sensor nodes that have limited energy. Among basic routing methods, the method using location information is efficient because it requires less information for calculation in route setting than flat routing and hierarchical routing. Because it utilizes distance, however, sensor nodes' energy utility may go down. In this study, we tried to even energy use in a wireless sensor network by giving a weight to the transition probability of ACS(Ant Colony System), which is commonly used to find the optimal path, based on the amount of energy in a sensor and the distance of the sensor from the sink. The proposed method showed improvement by 46.80% on the average in energy utility in comparison with representative routing method GPSR (Greedy Perimeter Stateless Routing), and its residual energy after operation for a specific length of time was 6.7% more on the average than that in ACS

Keywords: Sensor Network, Routing, Energy, Heuristic.

1 Introduction

Recently with the development of semiconductor, nano technology, micro-sensors, and wireless technology, wireless sensor networks are often used to watch over surrounding environment and collect data. Wireless sensor networks are applied commonly to areas that cannot be observed directly by humans for a long period such as battlefields, wild animal habitats, and natural disaster areas[1,2,3]. In such environments, tens of or even tens of thousands of sensor nodes are deployed. Because sensor nodes have limited power supply, many studies have been conducted in order to extend the life of sensor nodes as long as possible[1,4,5,6,7].

Randomly scattered sensor nodes communicate with one another and build a wireless sensor network, and they set routing paths in order to transmit collected data to

the sink node. Several methods have been invented for maximizing the life of power supply in sensor networks with limited power supply in order to maintain routing paths efficiently and as long as possible[8,9,10].

This study proposed a wireless sensor network routing algorithm using a variation of location-based routing that chooses the node with the shortest distance between the neighbor node and the sink[5], and ACS, one of heuristic techniques for finding the optimal solution[15]. ACS is an algorithm that increases the amount of pheromone for a path chosen by the leading ant so that the following ant chooses the optimal path. Because the ant algorithm can be used without additional information as long as local information, namely, the amount of pheromone is known, it is efficient for sensor networks with limited memory. The fact that ACS uses only local information is similar to the fact that routing is possible in location-based routing if the location of neighbor nodes is known.

This study proposed a routing method that finds the optimal path by choosing the next node to participate in routing among neighbor nodes using the location information of neighbor nodes necessary for location-based routing and transition probability used in ACS. By the proposed method, we applied the distance parameter used in existing transition probability as the amount of energy in sensor nodes, calculated weight from the distance from the sink and neighbor nodes, and managed the residual amount of energy in sensor nodes efficiently, and through these processes, extended the overall network life.

2 Related Work

2.1 Location-Based Routing

The routing and data transmission protocol of wireless sensor networks is largely divided into flat routing, hierarchical routing and location-based routing according to network structure.

Location-based routing uses the location information of sensor nodes in order to send data to a desired specific part of the network rather than to the entire network. Accordingly, each node can be designated using its relative position. The distance to neighbor nodes is measured using the intensity of received signal, and their relative coordinates are obtained by exchanging distance information with neighbor nodes. What is more, the location of a sensor node can be obtained directly from a satellite using GPS. Location-based routing saves energy by putting sensor nodes in a sleep state if they are not sending or receiving data[11,12].

This study compared the proposed method with GPSR among several existing location-based routing methods. For routing to the destination node, GPSR uses Greedy forwarding, which determines the forwarding of packets gradually using the location of the neighbor node receiving the packets and information on the destination of the packets[5].

2.2 ACS(Ant Colony System)

ACS solved the problem of local optimization in the ant algorithm using probability distribution called transition probability [13]. In ACS, ants choose a node at random according to initial rules, and each ant chooses the next node to visit according to the state transition rule. Through this process, ants change the amount of pheromone in each visited node according to the local updating rule. When all the ants have finished the search process through repeating the task, they again change the amount of pheromone according to the global updating rule. Then, in order to choose the optimal node, each ant completes a search path using heuristic information such as the amount of pheromone and transition probability.

ACS adjusts itself well to unpredictable environmental changes. Thus, it is applicable to the environment of wireless sensor networks like sensor networks where network topology is changed frequently.

3 The Proposed Method

This chapter proposed a method for setting an energy-efficient routing path in a sensor network based on the ant algorithm. When choosing the next node, the proposed algorithm calculates energy cost for sending a packet to a neighbor node and uses the energy cost in the transition probability of ACS[14].

$$ECost_{ij} = \frac{E_{TX}^{ij}}{E_i} + \frac{E_{RX}^{ij}}{E_j} \quad (1)$$

E_i is the amount of energy in node i , E_j the amount of energy in neighbor node j , E_{TX}^{ij} the amount of energy for sending a packet from node i to neighbor node j , and E_{RX}^{ij} the amount of energy for neighbor node j to receive the packet. Refer to Equation (2) and Equation (3) for detailed calculation of energy for transmission.

$$\begin{aligned} E_{TX}^{ij}(l, d) &= E_{TX-elec}^i(l) + E_{TX-amp}^i(l, d) \\ &= \begin{cases} lE_{elec}^i + l\epsilon_{fs}d^2, & d < d_0 \\ lE_{elec}^i + l\epsilon_{mp}d^4, & d \geq d_0 \end{cases} \end{aligned} \quad (2)$$

$$E_{RX}^{ij}(l) = E_{RX-elec}^j(l) = lE_{elec}^j \quad (3)$$

d is transmission distance, l is the number of bits to be transmitted, and the electric energy of nodes ($E_{TX-elec}^i$, $E_{RX-elec}^j$) depends on digital coding, modulation, filtering, and signal spreading. fs indicates the free space channel model (energy loss of the square of distance), and mp the multipath channel model (energy loss of the 4th power of distance). d_0 determines whether to use the free space channel model or the multipath channel model in calculating the amount of energy consumption. The value of d_0 is calculated as follows.

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \quad (4)$$

The threshold value of d_0 calculated by Equation (4) is around 87m.

This study added weight according to node location in calculating transition probability. The reason for adding weight is to find the optimal solution using the distance parameter by calculating the distance of the current node from its neighbor nodes and the sink. For example, the distances from neighbor nodes a, b and c and sink s are as follows.

Table 1. A portion of routing table's attribute

NodeID	Coordinate	Distance	Rank	Weight
source	(80, 120)	0		
a	(5, 30)	117	2	0.162
b	(20, 25)	112	1	0.311
c	(10, 10)	130	3	0.120

The distance between two nodes is calculated using $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. Because the node with the shortest distance is considered to be closest to the sink, it is given a higher weight in calculating transition probability. Weight is calculated as follows.

$$\omega_i = \frac{NSD_i}{\sum_{j=1}^{nn} NSD_j \times Rank_i} \quad (5)$$

$\sum_{j=1}^{nn} NSD_j$ is the sum of distances between neighbor nodes and the sink, and NSD_i is the distance between neighbor node i and the sink. nn is the number of neighbor nodes. $Rank_i$ is the rank of neighbor node i in consideration of its distance to the sink, and the rank is higher when the distance is shorter.

In the existing ant group system, transition probability is calculated as follows.

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [1/d_{ij}]^\beta}{\sum_{l \in J_i^k} [\tau_{il}(t)]^\alpha [1/d_{il}]^\beta} \quad (6)$$

d_{ij} is the distance between city i and city j , and $\tau_{ij}(t)$ means the amount of pheromone between the two cities. α is a parameter defining the influence of pheromone, and β is a parameter defining the influence of visibility ($1/d_{ij}$) between i and j . J_i^k is a set of neighbor cities that leading ant k has visited. This study set $\alpha=1$ and $\beta=1$ the optimal values suggested in [17] and [18], and transition probability used in the existing ant group system was revised as follows.

$$p_k(i, j) = \frac{[\tau_{ij}]^\alpha [ECost_{ij}]^\beta \omega_i}{\sum_{l \in nn_i} [\tau_{il}]^\alpha [ECost_{il}]^\beta \omega_j} \quad (7)$$

The proposed method provides an improved algorithm by changing transition probability through replacing $1/d_{ij}$ with energy cost $ECost_{ij}$ and adding weight ω_i .

According to the transition probability, each sensor routes packets to the sink node. When routing is completed and packets arrive at the sink node, all the links involved in the routing change their pheromone value through global update. If global update is not made, data are transmitted only through the nodes included in the initial optimal path, which results in rapid energy exhaustion in the nodes and, consequently, an energy hole or a disabled part of the network[15].

In ACS, both local update and global update are performed[16]. In this study as well, energy holes were prevented through local update and global update. Local update changes values as follows. In case of local update, if the next node is chosen and packets are sent to the node, the table in the current node is changed as follows. $\Delta\tau^k$ is the sum of energy costs until the current node on the path.

$$\tau'_{ij} = \tau_{ij} + \Delta\tau^k \quad (8)$$

Global update, which is triggered when the sum of energy costs on the path reaches the preset threshold value, change the pheromone value of the corresponding path so that other nodes are given an opportunity to be chosen. The threshold value is obtained by calculating the average amount of pheromone of the nodes on the path and then subtracting initial value 1 from the amount. This value is compared with the product of the amount of energy consumed in data transmission and the length of the routing path, and if it is larger global update is performed. The pheromone update equation for global update is as follows.

$$\tau''_{ij} = (1 - \rho)\tau_{ij} + \Delta\tau'^k \quad (9)$$

ρ ($0 \leq \rho \leq 1$) is the factor of evaporation for controlling the amount of pheromone, and it is set to 0.5 as suggested in [17] and [18]. $\Delta\tau'^k$ is the cost of the routing path for sending packets to the sink node. This calculates the cost of only the path through which packets are transmitted, and uses $\frac{1}{\sum_{(i,j) \in RP_k} ECost_{ij}}$. Here, RP_k is the path through which packets are transmitted.

4 Experiments

In this study, we conducted an experiment as follows in order to compare the performance of the proposed algorithm. First, we compared the proposed method first with representative location-based protocol GPSR, and then with ACS. In comparison between GPSR and the proposed method, we assumed that network life ends if the residual amount of energy becomes 0 or lower in any of the network sensors. In comparison between ACS and the proposed method, we measured the residual amount of energy after transmitting packets for the same length of time. The system for the experiments was built with Intel Core2 Duo 1.8GHz CPU, 2GB RAM, and Visual Basic 6.0. Table 2 shows parameters and their values set in the experiments.

The size of network was set to 100m $\times$ 100m, and the coordinates of the sink node were set to (1,1). The number of sensor nodes was set to 100 and they were deployed at random. The radius was set to 20m. It was assumed that each node knows the value of the sink and that the location of nodes can be obtained through GPS. In addition, we assumed that the sink node has an infinite amount of energy.

Table 2. Parameters and Values in using Test

Parameters	Values
Network Grid	(0m,0m) ~ (100m,100m)
Sink	(1, 1)
Number of Nodes	100
Transmission Range	20m
Packet Size	8 bytes
Initial Node Energy	10mJ/battery
Energy for Transmission	0.021mJ/bit
Energy for Reception	0.014mJ/bit

The figures below present the results of each method. Figure 1 is the residual amount of energy after the execution of the proposed method and ACS for the same length of time. As in the figure, the proposed method has a larger residual amount of energy than ACS, and uses energy more evenly than GPSR. The residual amount of energy after operation for a specific length of time was around 6.7% larger on the average in the proposed method than in ACS.

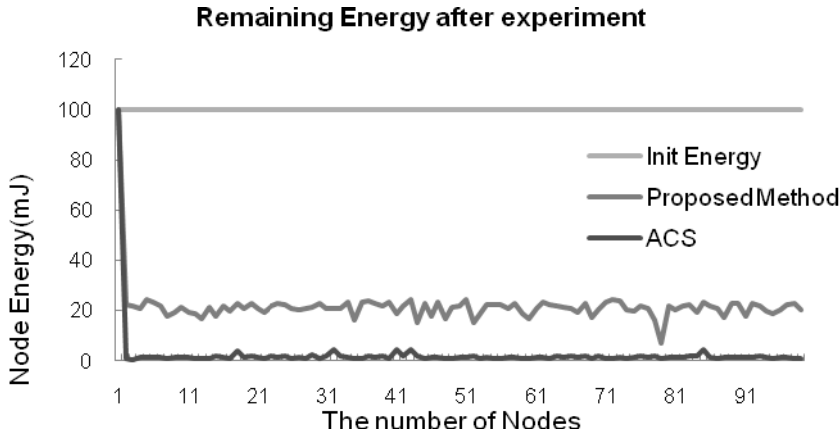


Fig. 1. Remaining energy after experiment using ACS and Proposed Method

Figure 2 compares the proposed method with GPSR through an experiment, which was ended when residual energy became 0 or lower in any of the sensor nodes. For the proposed method, the experiment appears to have terminated as node 79 had a negative value. In addition, energy was consumed evenly among the nodes in the proposed method, but in GPSR the amount of energy decreased rapidly in the nodes involved in routing. As in the figure, network life ended as nodes 23, 35, 40, 43, 54 and 81 showed a large difference from the other nodes in the amount of energy. Accordingly, we can see that location-based routing of GPSR can route with less information but it may lower the efficiency of energy use. In case of the proposed method,

the average residual amount of energy was 4.39mJ out of initial energy 100mJ, showing an energy use rate of 95.61%, but in GPSR, it was 55.25mJ, showing an energy use rate of 44.75%.

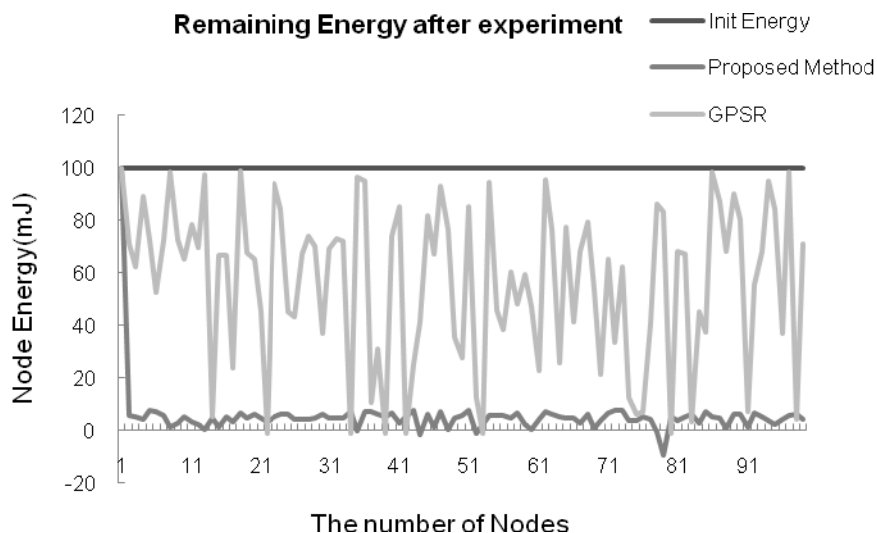


Fig. 2. Remaining energy after experiment using ACS and Proposed Method

5 Conclusion

This study proposed a method for finding the optimal path using location information and heuristic information such as the amount of pheromone and transition probability in the ant group system. The proposed method solved the problem of fast energy exhaustion in location-aware routing, and searched the entire network to find the optimal path in consideration of direction toward the sink. According to experiment results, the proposed method improved energy utility by 46.80% and reduced the residual amount of energy after path finding and data transmission for the same length of time by 6.7% compared to existing methods.

In future research, we plan to examine energy efficiency with a mobile sink node instead of a fixed one, and to build networks more robust against external attacks by applying encryption techniques to the proposed method.

Acknowledgements

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) Support program supervised by the NIPA(National IT industry Promotion Agency) (NIPA-2010-C1090-1031-0004).

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Communications Magazine* 40, 102–114 (2002)
2. Anastasi, G., Conti, M., Francesco, M., Passarella, A.: Eergy conservation in wireless sensor networks: A survey ad Hoc Networks. 7(3), 537–568 (2009)
3. Szewczyk, R., Osterwil, E., Polastre, J., Hamilton, M.: A Mainwaring Habitat Monitoring With Sensor Networks. *Communications of the ACM* 47(6), 34–40 (2004)
4. Chen, B., Jamieson, K., Balakrishnan, H., Morris, R.: Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. In: *Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking* (July 2001)
5. Karp, B., Kung, H.T.: Greedy perimeter stateless forwarding for wireless networks. In: *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 2000)*, pp. 243–254 (2000)
6. Kastern, O.: Energy consumption. ETH-Zurich, Swiss Federal Institute of Technology Technical Report, http://www.inf.ethz.ch/~kasten/research/bathtub/energy_consumption.html
7. Stemm, M., Katz, R.H.: Measuring and reducing energy consumption of network interfaces in hand-held devices. *IEICE Transactions on Communications* E80-B(8), 1125–1131 (1997)
8. Sohrabi, K.: Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications* 7(5), 16–27 (2000)
9. Younis, M., Youssef, M., Arisha, K.: Energy-aware routing in cluster-based sensor networks. In: *Proceedings of the 10th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, MASCOTS 2002* (2002)
10. Schurgers, C., Srivastava, M.B.: Energy efficient routing in wireless sensor networks. In: *The MILCOM Proceedings on Communications for Network-Centric Operations: Creating the Information Force* (2001)
11. Xu, Y., Heidemann, J., Estrin, D.: Geography-informed energy conservation for ad-hoc routing. In: *Proceedings of 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 2001)*, pp. 70–84 (2001)
12. Liu, X., Huang, Q., Zhang, Y.: Comb, needles, haystacks:balancing push and pull for discovery in large-scale sensor network. In: *Sensys 2004* (2004)
13. Gambardella, L.M., Dorigo, M.: Ant Colony System: A Cooperative Learning approach to the Traveling Salesman Problem. *IEEE Transactions on Evoutionery Computation* 1(1) (1997)
14. Rappaort, T.: *Wireless Communications: Principle & Practice*. Prentice-Hall, Englewood Cliffs (1996)
15. Funke, S.: Topological Hole Detection in Wireless Sensor Networks and its Applications. In: *Workshop on Discrete Algorithms and Methods for MOBILE Computing and Communications*, pp. 44–53 (2005)
16. Dorigo, M., Blum, C.: Ant colony optimization theory: A survey. *Theoretical Computer Science* 344, 243–278 (2005)
17. Dorigo, M., Maniezzo, V., Colomi, A.: Positive FeedBack as a search strategy. Report 91-106 (1991)
18. Bullnheimer, B., Hartel, R.F., Straub, C.: A New Rank Based Version of the Ant System - A Computational Study. Working Paper no. 1, Department of Management of Science (1997)

A Study on the Spectral and Energy Efficient-Path Selection Scheme in Two-Hop Cellular Systems

Se-Han Kim¹, Se-Jin Kim², Byung-Chul Kim³, and Jae-Yong Lee³

¹ RFID/USN Research Division, ETRI, Daejeon, Korea
shkim72@etri.re.kr

² Dept. of Computer and Information Science, Korea University, Korea
kimsejin@korea.ac.kr

³ Dept. of Infocom Eng., Chungnam National University, Daejeon, Korea
{byckim, jy1}@cnu.ac.kr

Abstract. In this paper, a spectral and energy efficient-path selection scheme is proposed to enhance downlink system throughput and reduce transmission energy consumption in two-hop cellular systems. In the proposed scheme a *base station* (BS) determines either a single-hop or two-hop path that uses less resource. The simulation results show that the proposed scheme outperforms a conventional path selection scheme which uses high *Signal to Interference and Noise Ratio* (SINR) strength based path selection in terms of system throughput and energy consumption and provides an insight into choosing the appropriate RS position and optimal frequency reuse factor.

Keywords: OFDMA, Cellular multihop network, Path selection, Relay station, USN.

1 Introduction

Cellular multihop network have been proposed as an attractive solution for next generation wireless communication since they enhance throughput and/or extend cell coverage using multihop *relay stations* (RSs) [1-7]. In cellular multihop networks, however, the resource management and path selection schemes are considerably more complex than those of conventional cellular systems because a *base station* (BS) shares its wireless resource with RSs and determines the optimal path for connecting with *mobile stations* (MSs). In [8], an optimal path selection scheme between single- and two-hop services was proposed in a cellular multihop network for highway deployment, where the BS and RSs were deployed along the road as roadside units. The proposed scheme in [8] outperformed other schemes in terms of system throughput but the proposed scheme can be used only for the highway environment. In [9], a path selection scheme for cellular multihop networks was introduced using a cost metric that indicated the effectiveness of the radio resource of a link for data transmission and a non-transparent frame structure that is described in [6]. In the simulation, a BS and several RSs were randomly deployed in a cell, and the BS chose paths to MSs through single- or multi-hop using the cost metric. However, the authors assumed the BS knows all link qualities of multi-hop paths to MSs and ignored interference and frequency reuse.

In this paper, we propose a novel *downlink* (DL) path selection scheme to enhance system throughput and reduce transmission energy consumption in two-hop Cellular networks with two-dimensional topology based on *orthogonal frequency division multiple access* (OFDMA) and *time division duplex* (TDD). The key idea of the proposed scheme is that the BS chooses the paths to MSs using the spectral efficiency between single-hop and two-hop paths using *channel quality information* (CQI) and the *modulation and coding scheme* (MCS) option in a transparent environment. The simulation results show that the proposed scheme outperforms a conventional path selection scheme which uses stronger *Signal to Interference and Noise Ratio* (SINR) for path selection in terms of system throughput and energy consumption.

2 The Proposed Path Selection Scheme

2.1 System Model

The system topology and frame structure [5,6,7,10] of cellular multihop network based on OFDMA-TDD are shown in Fig. 1. We assume the system consists of hexagonal cells of radius R and the cell coverage (C) is obtained via $C = 6\sqrt{3} \cdot R^2 / 4$. Also, a BS is located at the center of each cell and surrounded by six fixed RSs that are placed at a distance of D_{RS} from the BS. In the frame structure, the BS divides the timeline into contiguous frames each of which includes a DL and an *uplink* (UL) subframe. Then, the DL and UL subframes are further divided into zones to support BS-RS/MS communication, *access zone* (AZ), and to support RS-MS communication, *relay zone* (RZ). During the DL subframe, the BS transmits data to both MSs and RSs in AZ, and the RSs then subsequently relay the received data to the MSs in RZ. The BS is in silent mode during RZ. During the UL subframe, the BS and RSs receive data from the MSs in different AZ, and the RSs then relay the received data to the BS in RZ. Also, the technique of frequency reuse is considered to improve the overall network capacity and spectral efficiency. We assume that the *frequency reuse factor* (FRF), N_{FRF} , is always 1 for AZ because the BS only transmits to MSs within the BS region and RSs, whereas the RSs use different FRFs (i.e., 1, 2, 3, and 6) for RZ, and they are grouped from G_1 to G_6 , as is shown in Table 1. Each group uses bandwidth which is allocated by (total bandwidth / N_{FRF}).

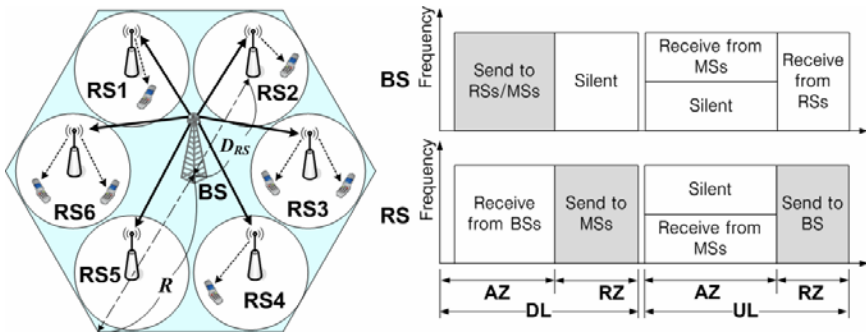


Fig. 1. System topology (left) and frame structure (right)

Table 1. Frequency reuse patterns in RZ

N_{FRF}	Frequency reuse patterns
1	G_1 (RS 1, 2, 3, 4, 5, 6)
2	G_1 (RS 1, 3, 5), G_2 (RS 2, 4, 6)
3	G_1 (RS 1, 4), G_2 (RS 2, 5), G_3 (RS 3, 6),
6	Each RS uses a different frequency ($G_1 \sim G_6$)

2.2 SINR Model in Cellular Multihop Networks

In our proposed SINR model, the interferences come mainly from two sources: intra-cell interference (I_{intra}^{zone}) and inter-cell interference (I_{inter}^{zone}). I_{intra}^{zone} is caused by BS and/or RSs using the same channel within a cell, whereas I_{inter}^{zone} is caused by other cells. When we assume L BSs are placed in a given area and each BS is surrounded by M RSs, the SINR of the MS or RS serviced by the BS can be expressed as (1).

$$\text{SINR}_{\text{BS-RS/MS}} = \frac{S_{BS_i}}{P_N + I_{intra}^{zone} + I_{inter}^{zone}}, \quad (1)$$

where, S_{BS_i} is the received signal power from a BS in the i -th cell ($1 \leq i \leq L$) and P_N is the white noise power.

A BS is in silent mode for RZ, whereas I_{intra}^{AZ} and I_{inter}^{AZ} for AZ can be written as (2).

$$I_{intra}^{AZ} = 0, \quad I_{inter}^{AZ} = \sum_{l=1, l \neq i}^L S_{BS_l}. \quad (2)$$

On the other hand, the SINR of MSs serviced by an RS from the j -th surrounding RSs ($1 \leq j \leq M$) in the i -th cell ($S_{RS_{i,j}}$) can be expressed as (3).

$$\text{SINR}_{\text{RS-MS}} = \frac{S_{RS_{i,j}}}{P_N + I_{intra}^{zone} + I_{inter}^{zone}}. \quad (3)$$

The RSs are in receive-mode for AZ, whereas I_{intra}^{RZ} and I_{inter}^{RZ} for RZ can be written as (4).

$$I_{intra}^{RZ} = \sum_{m=1, m \neq j}^M S_{RS_{i,m}}, \quad I_{inter}^{RZ} = \sum_{l=1, l \neq i}^L \sum_{m=1}^M S_{RS_{l,m}}. \quad (4)$$

2.3 Path Selection and Resource Allocation for the Proposed Scheme

To analyze the system performance, we draw a grid of squares in the cell coverage with sides of length 10m and measure the SINR of every junction using (1) and (3) to draw an SINR distribution of the cell. In the conventional scheme, the BS determines the path to MSs using high SINR strength from the BS or RSs. Thus, in this scheme,

the shape of each RS region is almost circular, such as the topology in Fig.1. In the proposed scheme, however, the BS determines a spectral efficiency based path between single-hop and two-hop communications according to the amount of resource allocation. The BS first calculates the required number of slots in both single-hop (ψ_{1-hop}) and two-hop (ψ_{2-hop}) paths using the received MSs' CQI. ψ_{1-hop} and ψ_{2-hop} can be written as (5).

$$\psi_{1-hop} = \frac{\zeta}{\gamma_{B-R}} \quad \psi_{2-hop} = \frac{\zeta}{\gamma_{B-R}} + \frac{\zeta}{\gamma_{R-R}} \quad (5)$$

where, γ_{BS-MS} , γ_{BS-RS} , and γ_{RS-MS} are respectively the number of bits per slot¹ for BS, RS and MS links and ζ is the amount of data from the BS to an MS.

Then, the BS determines the single-hop path when ψ_{2-hop} is higher than ψ_{1-hop} , whereas the BS decides the two-hop path in another case. The shape of each RS region is semi-circular because some MSs in the RS region communicate with the BS through a single-hop path instead of a two-hop path. We assume that RSs in both schemes periodically report the CQI of RS-MS links to the BS thus the BS knows the transmit rate. The logic flow and expected service region of the BS and RSs in the proposed scheme are as shown in Fig.2.

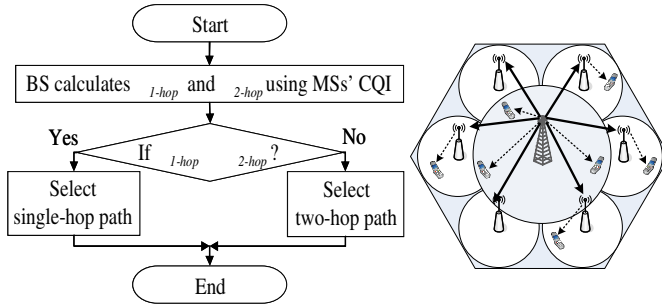


Fig. 2. The logic flow and expected service region of the BS and RSs in the proposed scheme

Then, we calculate the required number of slots per second for the i -th cell (ξ_i) under the given traffic density (ρ) that is Mbps/km². The number of slots per second for AZ ($\xi_{i,AZ}$) can be written as (6) and (7).

$$\begin{aligned} \xi_{i,A} &= \xi_{i,A} + \xi_{i,A-R} \\ &= \rho \cdot \sum_{A} \frac{\delta_{B-i} \delta_a}{A} + \rho \cdot \sum_{R} \frac{\delta_{R-i} \delta_a}{R}, \end{aligned} \quad (6)$$

¹ The slot is a unit for data transmission.

where, ξ_{BS_MS} and ξ_{i,AZ_RS} are respectively the numbers of slots per second for BS to MS and BS to RS links and the MCS option has N levels. $\delta_{BS_i}^n$ and $\delta_{RS_{i,m}}^n$ are respectively the numbers of points of the n -th MCS level for the BS and the m -th RS, and δ_{Total} is the number of total points of the i -th cell. R_{AZ}^n and R_{RZ} are the number of bits per slot of the n -th MCS level for BS-MS and BS-RS communications in AZ, respectively.

The number of slots per second for RZ ($\xi_{i,RZ}$) is obtained by (7).

$$\xi_{i,RZ} = \frac{\rho}{\alpha} \cdot \sum_{m=1}^M \sum_{n=1}^N \frac{C \cdot \delta_{RS_{i,m}}^n / \delta_{Total}}{R_{AZ}^n}, \quad (7)$$

where, $\alpha = M / N_{FRF}$.

Therefore, ξ_i can be represented as (8).

$$\xi_i = \xi_{i,AZ} + \xi_{i,RZ}, \quad \xi_i \leq \xi_{Total}, \quad (8)$$

where, ξ_{Total} is the number of total slots available for the DL in a cell.

Consequently, the maximum system throughput in bps of i -th cell (T_i) can be written as (9).

$$= \frac{\rho}{\delta_T} \cdot \left(\sum_{= } \delta + \sum_{= } \sum_{= } \delta \right) \quad (9)$$

where, ρ_{MAX} is maximum ρ according to (8)

2.4 Energy Consumption for Data Transmission

The energy consumptions per slot for data transmission in the BS ($\mathcal{E}_{BS}$) and an RS ($\mathcal{E}_{RS}$) are calculated by (10).

$$\mathcal{E}_{BS} = P_{BS} / \xi_{Total}, \quad \mathcal{E}_{RS} = P_{RS} / \xi_{Total}, \quad (10)$$

where, P_{BS} and P_{RS} are the transmission power per second for the DL using ξ_{Total} in the BS and an RS, respectively.

The average transmission power per bit (E) is calculated by

$$E = (\mathcal{E}_{BS} \cdot \xi_{i,AZ} + \mathcal{E}_{RS} \cdot \xi_{i,RZ} \cdot \alpha) / T_i. \quad (11)$$

3 Performance Evaluation

We evaluate the DL performance of the proposed scheme and compare it to that of the conventional scheme in terms of the ratio of single-hop service, maximum throughput, and energy consumption using a Monte Carlo simulation. In order to investigate

the performance parameters, we performed ten independent simulations. We assume the given total area is covered by a target cell and 36 cells in three-tiers; 6, 12 and 18 cells are placed in the 1st, 2nd, and 3rd tiers, respectively, and the center cell is observed. R is 1km, and the P_{BS} and P_{RS} are 20W and 5W, respectively. We use COST 231 Walfisch-Ikegami model for the access link (non-line-of-sight) and COST 231 model (excluding the rooftop-to-RS diffraction loss) for the relay link (line-of-sight); both are used for path loss models in [11]. Log-normal shadow fading is only considered for the access links with zero mean and a standard deviation of 8.0 dB. The performance evaluations are based on ω , which is defined as D_{RS}/R , and the key parameters are given in Table 2. We use the MCS option with a bit error rate less than 10^{-6} , and the MCS table is shown in Table 3 [12]. We modify the cost metric scheme in [10] into our simulation environments using the frame structure as shown in Fig. 1, and the simulation results are similar to those of the proposed scheme when $N_{FRF}=6$.

Table 2. System parameters

Parameter	Value
Carrier frequency	2.3 GHz
Bandwidth	10 MHz
Traffic density	Uniform distribution
TDD frame length	5ms
Number of sub-carriers	768
Number of total slots	768*24*200
Number of symbols for DL/frame	24
Antenna height	BS: 30m, RS: 10m, MS: 2m
P_N	-174dBm/Hz

Table 3. MCS level for DL

Level	Modulation	Coding rate	bits/slot	Value
1	QPSK	1/2 (6x)	1	-7.2 < SINR
2		1/2 (4x)	1	-5.1 < SINR
3		1/2 (2x)	1	-1.2 < SINR
4		1/2 (1x)	1	3.9 < SINR
5		3/4	1.5	5.7 < SINR
6	16QAM	1/2	2	11.0 < SINR
7		3/4	3	16.0 < SINR
8	64QAM	2/3	4	27.5 < SINR
9		5/6	5	34.5 < SINR

Fig.3 presents the ratio of single-hop service vs. ω with different FRFs. The ratios of a single-hop service for the conventional scheme are lower than those of the proposed scheme because some MSs in the RS region directly communicate with the BS through a single-hop path in the proposed scheme. The order of the ratio of single-hop service is FRF 1, 2, 3, and 6 because the spectral efficiency of a two-hop path increases as the value of FRF increases.

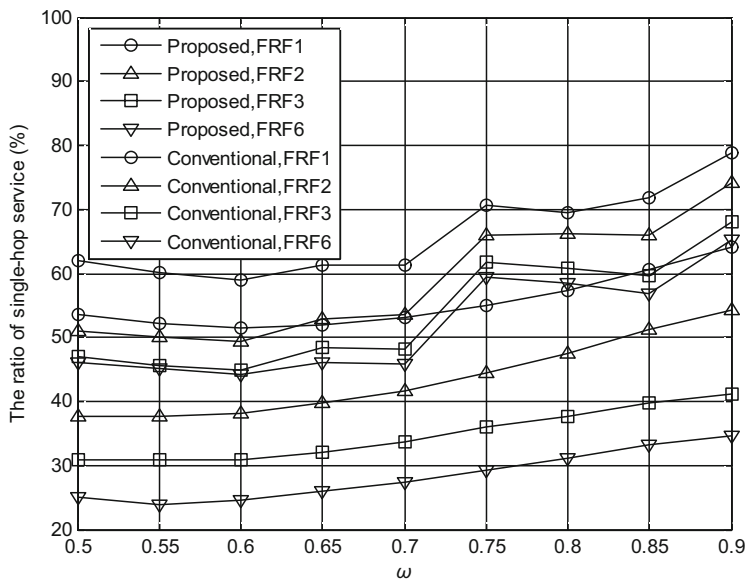


Fig. 3. The ratio of a single-hop service

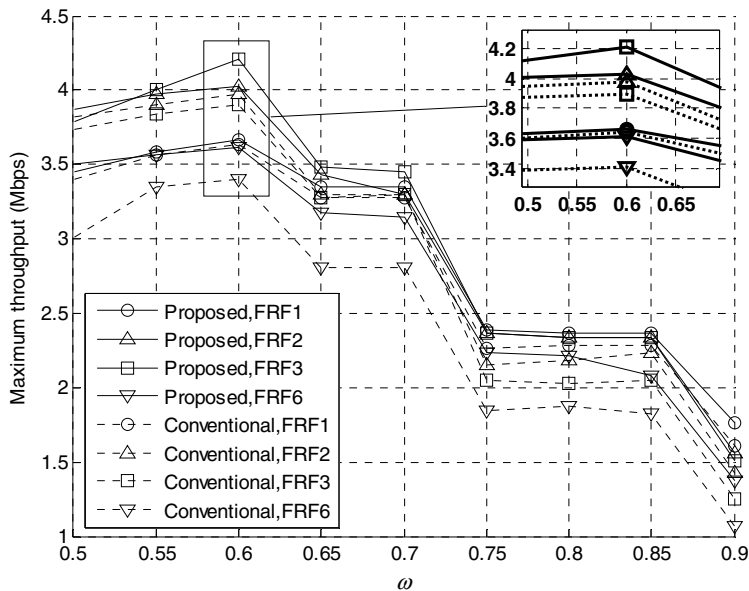


Fig. 4. Maximum throughput

Fig.4 describes the results of maximum throughput vs. ω with different FRFs. The results are highly affected by ω that are four types of MCS levels, QPSK rate 3/4 ($0.5 \leq \omega \leq 0.6$), QPSK rate 1/2 ($0.65 \leq \omega \leq 0.7$), QPSK rate 1/2 with repetition 2 ($0.75 \leq \omega \leq 0.85$), and QPSK rate 1/2 with repetition 4 ($0.9 \leq \omega$), for BS-RS communications, and thus the throughput highly decreases at 0.65, 0.75, and 0.9. The results of the proposed scheme are better than those of the conventional scheme, and FRF 3 of the proposed scheme achieves the highest system throughput of about 4.2 Mbps at $\omega=0.6$.

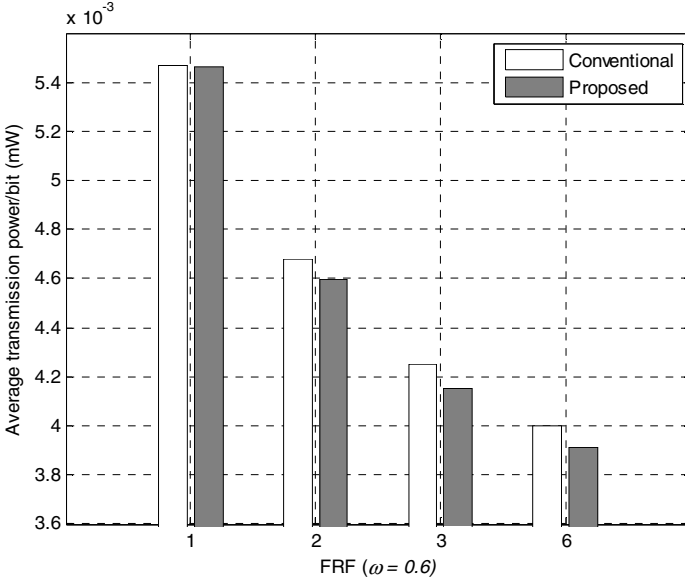


Fig. 5. Energy consumption ($\omega=0.6$)

Fig.5 shows the energy consumption of average transmission power per bit when ω is 0.6. The energy consumptions of the conventional and proposed schemes are almost the same in FRF 1, but those of the proposed scheme are lower than those of the conventional scheme in other cases. The reason is that the ratios of single-hop service in the proposed scheme are higher than those of the conventional scheme and the proposed scheme has a higher throughput. Also, the energy consumption decreases when the FRF increases because the number of RSs that transmit at once decreases when the FRF increases. Consequently, The FRF 3 of the proposed scheme has not only the highest throughput but also relatively low energy consumption comparing to that of the conventional scheme.

4 Conclusion

In this paper, we proposed a spectral and energy efficient-path selection scheme to enhance the system throughput and reduce transmission energy consumption for downlink in two-hop cellular systems. Via the simulation results, we showed that the

proposed scheme outperformed the conventional scheme and optimal RS position with FRF. The optimal ω was approximately 0.6, and FRF 3 was a good choice to achieve the highest throughput and low energy consumption for the system parameter values considered in our study. For future work, we plan to study an improved path selection scheme considering load aware and cooperative communication.

Acknowledgement

This work was supported by the Industrial Strategic technology development program, 10037299, Development of Next Generation Growth Environment System) funded by the Ministry of Knowledge Economy (MKE, Korea)"

References

1. WWRF/WG4 Relaying Subgroup: Relay-based Deployment Concepts for Wireless and Mobile Broadband Cellular Radio, White Paper (2003)
2. Walke, B., Pabst, R.: Relay-based Deployment Concepts for Wireless and Mobile Broadband Cellular Radio, WWRF/WG4 Relaying Subgroup, White Paper (2004)
3. IST WINNER II: Relaying concepts and supporting actions in the context of CGs, D3.5.1-3 (2007)
4. Chen, K.-C., Roberto, J., De Marca, B.: Mobile WiMAX. Wiley, Chichester (2008)
5. Genc, V., Murphy, S., Yu, Y., Murphy, J.: IEEE 802. 16j Relay-based Wireless Access Networks: An Overview. IEEE Wireless Communications 15(5), 56–63 (2008)
6. IEEE Standard 802.16j-2009: IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification (2009)
7. Peters, S.W., Heath, R.W.: The future of WiMAX- Multihop relaying with IEEE 802.16j. IEEE Communications Magazine 47(1), 104–111 (2009)
8. Ge, Y., Wen, S., Ang, Y.-H., Liang, Y.-C.: Optimal Relay Selection in IEEE 802.16j Multihop Relay Vehicular Networks. IEEE Transactions on Vehicular Technology 59(5), 2198–2206 (2010)
9. Wang, S.-S., Yin, H.-C., Tsai, Y.-H., Sheu, S.-T.: An Effective Path Selection Metric for IEEE 802.16-based Multi-hop Relay Networks. In: IEEE Symposium on Computers and Communications 2007, pp. 1051–1056 (2007)
10. IEEE Standard 802.16e-2005, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems (2006)
11. IEEE 802.16j-06/013r3: Multi-hop Relay System Evaluation Methodology (Channel Model and Performance Metric) (2007)
12. Yoon, D., Cho, K., Lee, J.: Bit Error Probability of M-ary Quadrature Amplitude Modulation. In: IEEE VTC-Fall 2000, vol. 5, pp. 2422–2427 (2000)

The Construction of Remote Microcontroller Laboratory Using Open Software

Kwansun Choi¹, Saeron Han¹, Dongsik Kim¹, Changwan Jeon¹, Jongsik Lim¹,
Sunheum Lee², Doo-soon Park³, and Heunggu Jeon⁴

¹ Department of Electrical Communication Engineering, SoonChunHyang University, Korea

² Department of Information Communication Engineering, Soonchunhyang University, Korea

³ Department of Computer Software Engineering, Soonchunhyang University, Korea

⁴ Department of Electrical Engineering, Anyang University, Korea

gkstofhs@paran.com, ultrabangbuje@hanmail.net,

{cks1329, dongsik, jeoncw, jslim, shlee}@sch.ac.kr

Abstract. A microcontroller can be considered a self-contained system with a processor, memory and peripherals. It is used in automatically controlled products and devices by programming. Therefore it is adopting with the essential subject in our department. We implemented a remote laboratory for 80196KC microcontroller. The remote laboratory provides functions with which learners can perform experiments by remote-controlling circuits and equipment related to experiments and practices on the Web and gets data and image information, so through it learners can make realistic experiments at any time and in any place. Although student can not physically touch any equipment, they can confirm the operation process of the 80196KC microcontroller by observing the result of experiment transferred through the web camera. The remote laboratory module based on Internet open sources such as JAVA Web start technology, FTP technology and Telnet technology in server-client environment can be implemented at low cost. It has the object of introducing remote experimentation, offering to students the opportunity to run a C code for 80196 KC microcontroller. Our remote Lab, therefore, is a useful tool that allows students to make real experiments with 80196 KC microcontroller. It will be used an auxiliary lab for teaching and learning microcontroller.

Keywords: Remote laboratory, Virtual Laboratory, Microcontroller, 80196KC, Java Web Start, Java FTP, Java Telnet.

1 Introduction

According to a development of internet technology, the virtual educational system and remote laboratory originates. There are many studies of a web-based virtual laboratory and remote laboratory specially. A virtual education system is performed at virtual space where is not physical space. It supplies more chances which students will attend to educational lectures without time and space limitations.

Discussions, lectures, evaluations are accomplished in the offline classroom. But in case of a virtual education, they accomplished in the virtual classroom [1-2].

The virtual laboratory is based on real education. The system is developing to complement a status of learning on campus. A study of engineering experiment is classified two groups of a virtual laboratory and a remote laboratory. A virtual laboratory is implemented by software same as equipments and circuit elements used at real experimental environment, but remote laboratory is performed and accessed equipments which is installed at the remote site. The remote laboratory has potential to complement hands-on practical labs and provides common utilization of high price equipments and flexibility for students to access the labs off campus and at their convenient time [3-5].

We implement a web-based remote laboratory which educates 80196 KC microcontrollers which are being used plentifully in industry sector. It is very important and essential for educator to perform experiments on it in order to learn the concept of microcontroller. In general many students perform experiment on microcontroller with experiment kits only in campus [6-8].

But our proposed system which applies the Internet supplies a possibility of experimenting outside campus. Therefore the students can perform experiment on microcontroller any time and any where using the system [9].

The virtual laboratory implements an equipment using software. There is a possibility of treating an experimental subject softly by modifying software. The equipments are not demanded anymore and the expense of system construction holds few [10].

On the other hand the remote laboratory will be able to acquire an actual data from remote place. It is supplied with a possibility of getting the measurement result which agrees with a theoretical value. It has merits to use the shared equipments of the various campus without time and space limitation.

A number of remote laboratories have been set up by some universities around the world. These offer remote access to laboratory equipment and experimental setups via the Internet. This paper[11] presents the implementation of a web-based laboratory experiment on a coupled tank apparatus, a Multiinput–Multioutput (MIMO) system. The web-based laboratory has been developed to serve students and staff in the Department of Electrical Engineering at the National University of Singapore(NUS). The laboratory is an educational tool for teaching students the basic principles and methodology in performing a series of experiments on a coupled tank apparatus at any time and from any location through the Internet.

The remote laboratory [12] is composed of Emulator System Vibration System Wind Tunnel. This remotely accessible laboratory allows the general public to perform experiments from anywhere with an internet connection by simply submitting an experiment request through this website. Requests are submitted to the experimental queue and are then performed in order on the automated equipment set up at the Remote Laboratory at Stevens Institute of Technology. The experimental results are recorded and can be viewed online as soon as the experiments are finished.

The remote laboratory [13] at Maribor University is based on two well-known commercially available software packages i.e. MATLAB/Simulink and LabVIEW and a custom-made hardware i.e. DSP based controller. The remote laboratory for circuit theory developed H&S Group at the Soochunhyang University using LabVIEW and DAQ controls passive elements at a distance [14].

In this paper, we propose remote laboratory not using commercial packages of LabVIEW and VEE. The system implements using by JAVA web start, JAVA FTP, JAVA Telnet which is composed of remote compile system and execution system.

The system does not have purchasing cost owing to sharing with hardware and software. Thus the system has the advantage what kind of purchase expense does not expend. The paper continues, in chapter II, with web-based remote hardware control. Chapter III contains the configuration of remote compile system. Chapter IV contains the description of our system. Chapter V contains Usability Measurement. Finally, the conclusion and future work are presented in chapter VI.

2 Web-Based Remote Laboratory

In this paper, we choose the 80196KC among many microcontrollers. A microcontroller is a device similar to the microprocessor but distinguishing features of the microcontroller chip is the inclusion, on a single chip, of all the resources that permit the IC to serve as a controller in a system or an instrument. As the microcontroller has peripherals built-in facilities within the chip itself has significant advantages, when used for building stand alone front ends and controllers.

The main features of the MSC 80196 include a large on-chip memory, Register-to-register architecture, three operand instructions, Bus controller to allow 8 or 16 bit bus widths, flat addressability of large register files. The microcontroller has an on-chip ALU, 4 channel 10 bit analog-to-digital converter (ADC), 3channel 8 bit pulse width modulator (PWM), 16bit watchdog timer, 4 16bit software timers, hardware multiply and divide, and 8 KB of on-chip ROM, full duplex serial port and so on.

The microcontroller 80196KC is commonly used in hard disk drives, modems, printers, pattern recognition and motor control. The proposed remote laboratory for 80196KC microcontroller is illustrated in Figure 1.

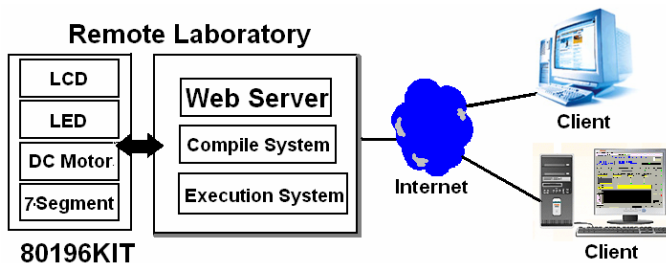


Fig. 1. Model of remote microcontroller laboratory

The educators edit source program for microcontroller experiment at the client PC. The source program is saved and is uploaded to the remote compile system. The uploaded source program is compiled through Telnet connection. After the compile ends and the execution file creates. The execution file is transmitted to the client and the transmitted file is stored. The file is transmitted to the microcontroller execution system. After the execution file is transmitted, executive command transmission system makes it execute in the experiment kit. The students are able to confirm an execution result by camera because the execution process at real-time is transmitted in client PC. After executing the user finishes learning and initializes the experimental kit.

3 Remote Compile System and Configuration

The remote lab enables remote experimental devices to be controlled by local learners, compensating for the lack of reality in the virtual experiment. The block diagram of the remote lab is shown in Figure 2. The Server system is composed of Compile module, 80196 Execution Module, 80196 KITs, Telnet Server and FTP Server, etc. The functions of major modules are as follows.

The remote compile system transmits source file to the server, compile it and creates execution file. The execution file is transmitted to the microcontroller training kit by serial communication. The transmitted file .runs in the microcontroller training kit. Therefore the remote compile system is a module which executes the microcontroller remotely.

In general student performs microcontroller experiments by connecting training kit with his local computer in the real experiment. The first step is a compiling process of source program and second step is linking process of object program. The created execution file is transmitted to the microcontroller training kit using RS232 serial communication. Students can confirm a result of execution directly in front of the training kit. In order to experiment microcontroller, student attend to the laboratory which equipped with microcontroller training kit certainly. Therefore in order to solve the problem, we implemented a remote compile system. The proposed system is composed of File Transmission/reception module, Compile module, text file input/output module and Source edit module.

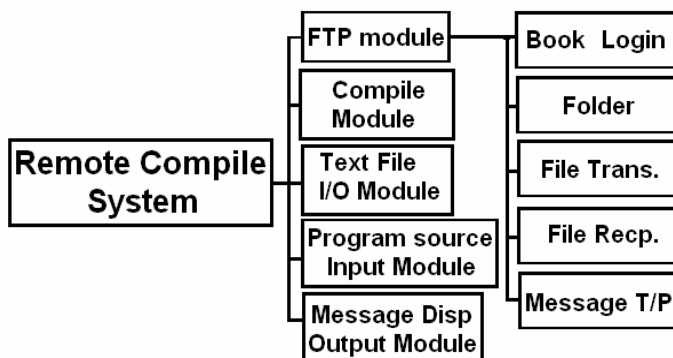


Fig. 2. Configuration of remote compile system

4 Case of Remote DC Experiment

When a learner has access to the web page for remote laboratory, client modules under the environment of Java Web START open the window as shown in Fig.10. The Client system is composed of program input module, file send module and control input command module. To write a program, a learner should click the first button named source writing. Then, the learner gets the source input window.



Fig. 3. First view of remote lab

The source input module is composed of the text input module and file save module. The text input module helps a learner to write a source program and the file save module saves the source file on the client computer. The file send module sends the local source file to the server to compile and execute the source file. Compile module in server generates the execution file by compiling and linking the code received from the client. The compile module returns the compiling message to the client. After completion of compilation and link, the 80196KC execution modules take control over the 80196KC system connected to the server computer by RS-232, execute the execution code, and return execution results to the clients through web camera.

In compliance with source program students study the contents which DC motors operate from DC motor experiments. The main function of source program is as like the figure 4. The program source is explained. It initializes a training kit board and is assigned address to port A, B, C. Finally it is assigned the value 0x80 to the control word register which determines input/output direction of the port A, B, C. The students confirm the result through web camera which a DC motor rotates with clockwise direction for constant period, after rotates counter clockwise direction for constant period infinitely.

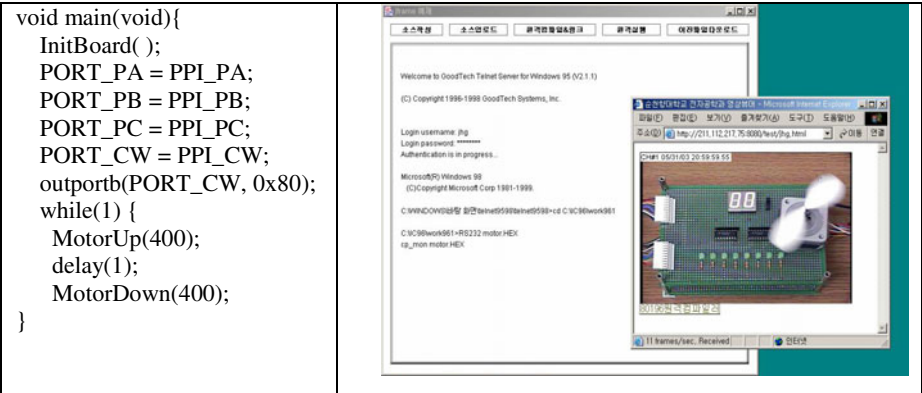


Fig. 4. Result of DC Motor Experiment

5 Usability Measurement

The microcontroller is a very important subject nowadays in area like Electric/Electronic Engineering, Computing Science. Various remote lab for

microcontroller for training is originated. Our proposed web-based remote lab offers to students the opportunity to run c code for microcontroller, and is used an auxiliary lab for teaching microcontrollers.

The usability of a web-based microcontroller laboratory is a function of system design and is determined by various factors, but we focused on ease of use, quality of the learning materials, effectiveness of remote laboratory, coverage of the contents and system responsiveness. A survey questionnaire that has been developed based on these issues is summarized in Table1. Students were asked to rate the usability of the web-based combined laboratory on a five-point scale, as follows: 1-very poor; 2-poor; 3-satisfactory; 4-good; and 5-very good. The web-based remote laboratory is provided to the students enrolled in a microcontroller course in addition to onsite lecture and experiment to compensate for the lack of the time allowed for the course. From 30 to 40 students enrolled in the course took part voluntarily in the survey from 2006 to 2009 year. Table 2 gives the average percentages per three years of students who answered the 5 different aspects of the web-based laboratory as very good, good, or satisfactory. The three years average of the students rated the Q1, Q2, Q3, and Q5 to be satisfactory, good or very good exceeded 85%. But The three years average of the students rated the Q4 to be satisfactory, good or very good exceeded 70%. The web-based remote laboratory needs to provide more diverse contents related to the topics. The student's experience in the web-based remote laboratory considerably reduced the time for the onsite experiment and the given experiment can be finished in the given time, otherwise extra time would be needed. Therefore, our proposed laboratory is very useful to enhance the quality of the onsite experiment courses or can be used as online education tool for the 80196KC microcontroller experiment stand alone. The proposed system allows flexibility for students to access a range of laboratory experiment at any time and any where there is an internet connection.

Table 1. Questionnaire used to measure the usability of the web-based remote laboratory

	On a scale of 1 to 5 rate: (1=very poor, 2=poor, 3=satisfactory, 4=good, 5=very good.
Q1	Was the web-based remote laboratory easy to use?
Q2	Were the laboratory components good enough to help you better understand the concepts and principles of learning materials?
Q3	Was the remote laboratory helpful to conduct the real-laboratory experiment?
Q4	Was the web-based combined laboratory self-contained enough to study alone?
Q5	How was the available material for the laboratory?

6 Conclusion and Further Works

In this paper, we proposed and implemented a remote laboratory to settle problems of a virtual laboratory which students can't experiment with real equipments. The proposed laboratory is remotely accessible to real training kits in distance microcontroller

education. Without commercial tools as like LabVIEW and VEE, the proposed system uses technologies such as Java Web Start, Java FTP, Java Telnet communication.

Therefore our system is implemented at low cost, and is effectively applicable to engineering experiment education in various areas related to real-time hardware control. Thus authorized users who are allowed to have access to both labs using a web browser no longer need to have their own 80196KC microcontroller-related experiment devices and software locally. Although clients can not physically touch any equipment, they can confirm the operation process of the 80196 KC microcontroller by observing the result of experiment transferred through the web camera. It shows a possibility of a remote laboratory which be remotely controlled for examples like DC/Servo motor, graphic/text LCD, 7-SEGMENT, LED and sensors. Our system, therefore, will be a useful tool that allows students to make real experiments with 80196KC microcontroller and be a very effective educational tool because the remote laboratory helps learners easily understand the programming method and the process of complex experimental operations about the 80196KC microcontroller. It will be an auxiliary laboratory for teaching microcontrollers, a very important subject nowadays in area like Electric/Electronic Engineering. In the future, we will develop a remote education system which offers existing text and sound contents and flash animation, develop improvement of transmission method, various devices and equipments controlled remotely remote controlling device value, and web based hybrid education system which is enriched by creative multimedia contents.

References

1. Jarc, D.J., Feldman, M.B., Heller, R.S.: Accessing the Benefits of Interactive Prediction Using Web-based Algorithm Animation Courseware. In: SIGCSE 2000, pp. 377–381 (2000)
2. Kim, D., Lee, S., Choi, K.: Practical Implementation of A Web-based Virtual Laboratory in the Area of Electrical Engineering. In: IASTED International Conf. on Computers & Advanced Technology in Education (2001)
3. Nakano, H., et al.: Distance Education System for Interactive Experiments on Electric Circuits over the Web. In: ISIMADE 1999, pp. 113–116 (1999)
4. Salzmann, C., Latchman, H.A., Gillet, D., Crisalle, O.D.: Requirements for Real-Time Laboratory Experimentation over the Internet. In: ICEE, Rio de Janeiro, Brazil (August 1998)
5. Ko, C.C., Chen, B.M., Chen, S.H., Ramakrishnan, V.: Development of a Web-Based Laboratory for Control Experiments on a Coupled Tank Apparatus. *IEEE Trans. on Education* 44(1) (February 2001)
6. Kim, D., et al.: A Web-based Virtual Laboratory for Basic Electrical Circuits. *Journal of Engineering Education Research* 5(1) (2002)
7. Kim, D., Choi, K., Lee, S.: Implementation of a web-based virtual laboratory for digital logic circuits using multimedia. *Korean Society for Engineering Education & Technology* 5(1) (2002)
8. Luigino, B., et al.: A Web-Based Distributed Virtual Educational Laboratory. *IEEE Trans. on Instrumentation and Measurement* 49(2), 349–356 (2000)

9. Gillet, D., Salzmann, C., Latchman, H.A., Crisalle, O.D.: Advances in Remote Experimentation. In: 19th American Control Conference, Chicago, Illinois, USA, pp. 2955–2956 (2000)
10. <http://www.lab-on-web.com/NET/WebApplication/LoW/Contact.aspx>
11. Ko, C.C., Chen, B.M., Chen, S.H., Ramarkrishnan, V.: Development of a Web-Based Laboratory for Control Experiments on a Coupled Tank Apparatus. *IEEE Trans. on Education* 44(1) (February 2001)
12. <http://dynamics.soe.stevens-tech.edu/>
13. Hercog, D., Gergič, B., Uran, S., Jezernik, K.: A DSP-based Remote Control Laboratory. *IEEE Transactions on Industrial Electronics* 54(6), 3057–3068 (2007)

LAPSE+ Static Analysis Security Software: Vulnerabilities Detection in Java EE Applications

Pablo Martín Pérez, Joanna Filipiak, and José María Sierra

Evalues Lab - Universidad Carlos III de Madrid
pmperez@di.uc3m.es, jfilipiak@gmail.com, sierra@inf.uc3m.es

Abstract. This paper presents the study and enhancement of LAPSE, a security software based on the static analysis of code for detecting security vulnerabilities in Java EE Applications. LAPSE was developed by the SUIF Compiler Group of Stanford University as a plugin for Eclipse Java IDE. The latest stable release of the plugin, LAPSE 2.5.6, dates from 2006, and it is obsolete in terms of the number of vulnerabilities detected and its integration with new versions of Eclipse. This paper focuses on introducing LAPSE+, an enhanced version of LAPSE 2.5.6. This new version of the plugin extends the functionality of the previous one, being updated to work with Eclipse Helios, providing a wider catalog of vulnerabilities and improvements for code analysis. In addition, the paper introduces a command-line version of LAPSE+ to make this tool independent of Eclipse Java IDE. This command-line version features the generation of XML reports of the potential vulnerabilities detected in the application.

1 Introduction

Nowadays, web applications play an important role in Information and Communication Technologies. This is motivated by the fact that the services leading this scope, such as e-commerce, e-learning, social networks and cloud computing, make use of this sort of applications. The mentioned services manage sensitive information that needs to be protected against attacks that can compromise its confidentiality, availability and integrity. Thus, the organizations offering these services can give confidence to their users, ensuring the continuity of business and being in compliance with the current security guidelines and legislation.

Ensuring web applications are not vulnerable to attacks is a hard task for developers and auditors. It is specially complicated when they have to deal with applications consisting of a complex structure and thousands of lines of code. To ease this task, they can make use of automatic tools to analyze the application in search of vulnerabilities. These tools allow them to identify the application points that are vulnerable to an attack. Depending on the analysis techniques used by these tools, we can mainly classify them in Dynamic or Static Analysis tools.

Dynamic Analysis of web applications involves the study of the program logic behaviour. Tools based on this technique can help to find flaws from the inconsistency of the outputs, in relation with the input data the program receives in execution time. These tools provide an initial approach to detect the vulnerabilities of the application.

The advantage of dynamic analysis lies on the fact that it is a technique that does not entail the study of the code to identify the vulnerabilities, inferring them from the behaviour of the application. The problem with this technique is that it cannot ensure the detection of all vulnerabilities, due to the complexity of covering all the possible execution scenarios that can lead to an attack.

Static Analysis entails the search of security flaws by reviewing the source code of the application. It is an static analysis since it does not comprise its execution. This technique allows to analyze the data flow, check the syntax and verify if the states of the application are finite.

The advantage of Static Analysis compared to Dynamic Analysis is that the first can be performed during development phase. On the other hand, Static Analysis is more complex than Dynamic Analysis since it implies a deep understanding of the applications behaviour and a broad knowledge of the code and the programming language. The effectiveness of Static Analysis to detect vulnerabilities makes of this technique an essential point in the process of ensuring the security of web applications. This statement, along with the difficulty that it involves, are two strong incentives to develop tools that help developers and auditors to carry out this analysis.

There are several tools designed to meet these objectives, i.e., ITS4[1], RATS[2]. However, most of the tools for code static analysis are intended for applications written in C. But the widespread development of Java EE Applications and the continuous emergence of frameworks for this purpose[3][4][5], make necessary to count on a tool for analyzing vulnerabilities in applications written in Java.

Hence, this paper presents LAPSE+, a security tool based on the static analysis of source code for the detection of vulnerabilities in Java EE Applications. LAPSE+ has been developed from the study of LAPSE, the security scanner by SUIF Compiler Group at Stanford University. Considering the advantages mentioned before about the completeness of the static analysis of code for detecting security flaws, since LAPSE+ is based on this technique, it represents an advance in the security process of Java EE Applications, both in the development and in the audit.

1.1 Paper Organization

The paper is structured in five sections to understand the functionality of LAPSE, the security scanner for the static analysis of Java EE Applications, and its evolution to LAPSE+, the enhanced version of this tool. The first section refers to the most common vulnerabilities in web applications in order to understand what kind of attacks is LAPSE focused on. In the second section we have an overview of LAPSE, referring to the history of this tool and its features, referring to the features of the latest stable release, LAPSE 2.5.6. The third section introduces LAPSE+, including a deep analysis of the improvements in this version, studying the new vulnerabilities detected, the interpretation of new method parameters and the integration of this tool in Eclipse Helios. The fourth section explains the command-line version of LAPSE+, mentioning the reasons of its development and introducing its features. The fifth section consists of the conclusions extracted in the study of the vulnerabilities in web Applications and the enhancements of LAPSE+ to become an important Free Open Source Software (FOSS) tool in the process of auditing Java EE Applications.

2 Common Vulnerabilities in Web Applications

OWASP Top 10 Release of 2010 ranks Injection Attacks in the first place of The Ten Most Critical Web Application Security Risks, rating it as a vulnerability of Easy Exploitability and Severe Impact. The document states that Injection flaws are often found in SQL, LDAP or XPath queries, OS commands and program arguments. This ranking can be complemented with the 2010 CWE/SANS Release about the Top 25 Most Dangerous Software Errors. It ranks Improper Neutralization of Input During Web Page Generation as the most dangerous weakness, being it related with Cross-site Scripting attack. Hence, we have that the most common vulnerabilities in web applications are mainly related with the input validation. Web applications that not sanitize their input are prone to this sort of attacks, which can manipulate the behaviour of the application, being able to endanger the availability, integrity and confidentiality of its information.

3 LAPSE Overview

LAPSE is a security software, based on the static analysis of code, for detecting vulnerabilities in Java EE Applications. It was developed by the SUIF Compiler Group at Stanford University, released in 2006. LAPSE stands for Lightweight Analysis for Program Security in Eclipse, and as its acronym states, it is a plugin for Eclipse, the well-known open-source Java IDE. The tool aims to find security flaws caused by the inadequate or non-existent validation of the user input data. These sort of vulnerabilities are known as the most common among web applications. The main idea of LAPSE is to help the developer or auditor to sanitize the input, problem based on the *tainted mode* of Perl[6]. LAPSE extends the tainted mode of Perl defining the problem of *tainted object propagation*[7]. LAPSE defines three concepts to determine the existence of a vulnerability in a Java EE Application: Vulnerability Sources, Vulnerability Sinks and Backward Propagation.

- **Vulnerability Sources** comprise the points of code that can be target for untrusted data injection, i.e., when getting HTML form parameters, Cookies parameters or HTML headers.
- **Vulnerability Sinks** refer to the manipulation of the web application once the malicious data have been injected. Expressions mostly related to accesses to databases or file systems, in order to get sensitive information or even gaining privileges to compromise its availability and integrity.
- **Backward Propagation** involves the construction of a syntax tree to determine if the untrusted data propagates through the web application and accomplish to manipulate its behaviour. The root of the syntax tree is a Vulnerability Sink. Hence, the tree is covered backwards, analyzing the value the Vulnerability Sink parameters take through the different assignments and method calls. When it is possible to reach a Vulnerability Source from a Vulnerability Sink, then the web application has a security vulnerability.

3.1 Vulnerabilities Detected by LAPSE

The vulnerabilities detected by LAPSE are classified by its implication as vulnerability sources or sinks and the categories of attacks they can be target of. The vulnerability sources refer to attacks consisting of Parameter Tampering, URL Tampering, Header Manipulation and Cookie Poisoning[7]. The vulnerability sinks have to do with attacks of SQL Injection, Cross-site Scripting, HTTP Response Splitting and Command Injection. Another kind of attack detected by LAPSE is that related to Information Leakage. Although it is not an injection data attack, it is considered as a possible source of attack. If the developer reveals sensitive information about the web application, the user can take advantage of this to carry out an attack. For example, if the name of the database tables are revealed and the application is also vulnerable to SQL Injection attack, the user can make unauthorized changes on the database, even deleting the tables.

4 LAPSE+: An Enhanced Version of LAPSE

Evalues ICT Security Lab at Universidad Carlos III de Madrid has developed an enhanced version of LAPSE, which has been given the name of LAPSE+. This new version of the plugin detects new categories of attacks and provides new features to analyze the code and determine the existence of vulnerabilities. The points that make up this section aim to explain the improvements of LAPSE+ in terms of the new vulnerabilities detected, the interpretation of new method parameters and the integration of LAPSE+ in Eclipse Helios. The version of LAPSE that has been taken as reference for the comparison study is LAPSE 2.5.6, the latest stable release of this tool.

4.1 New Vulnerabilities Detected

The new vulnerabilities that have been included in LAPSE+ are focused on the injection of XML code, queries on LDAP servers and the unauthorized access to files and directories stored outside the root folder of a web application server. Specifically, LAPSE+ includes the detection of vulnerabilities corresponding to Path Traversal, XPath Injection, XML Injection and LDAP Injection attacks.

4.1.1 Path Traversal

Path Traversal attack aims to access directories and files out of the root path of the web server application, by manipulating the variables that reference files, using “dot-dot-slash (../)” sequences and its variations. Although LAPSE 2.5.6 has defined several Path Traversal methods in its catalog of vulnerabilities, it never reaches to detect these vulnerabilities because it does not scan constructor methods. Due to the fact that the vast majority of Path Traversal attacks in Java come from constructor methods, the analysis of these has been included in LAPSE+.

4.1.2 XPath Injection

XPath is the query language for managing the information of a XML document. Web applications which use XML databases to store their data, communicate with them by XPath queries. XPath Injection attacks are based on embedding XPath code into the

input data of the web application, in order to exploit the parameters of the XPath queries. Thus, the attacker can extract sensitive information from the database or alter it. LAPSE+ includes in its catalog of vulnerabilities the Java methods that can propagate this attack. The catalog covers the methods that belong to the most common libraries of XPath processing, such as Xalan and JXPath.

4.1.3 XML Injection

XML Injection is an attack based on attaching malicious data into the fields of a XML message. The injected inputs can change the structure of the XML fields, not only its content, but also its tags. These malicious data can manipulate the logic of the application and hence compromise its behaviour. LAPSE+ detects this vulnerability by adding to its catalog the Java methods that can propagate this vulnerability. These methods belong to the XML:DB API. Specifically, it has been taken into account the XMLResource module.

4.1.4 LDAP Injection

LDAP Injection attacks are based on similar techniques to XPath Injection. Their aim is to take advantage on the parameters in the LDAP query to extract sensitive information. Also, the attacker can get permissions on the LDAP server to modify the content inside the LDAP tree. LAPSE+ includes as an improvement the detection of Java methods that can propagate LDAP Injection. Hence, this new version adds to its catalog the LDAP classes and libraries for Java, called JLDAP[8].

4.2 Interpretation of Method Parameters

One weakness found in LAPSE 2.5.6 is the fact of identifying only simple types as method parameters. However, it is very common when programming in Java to have as parameter an expression consisting of reference variables, e.g., string concatenations, accesses to an array and methods calls. Likewise, it is very common to have these variables in brackets or being modified by derived methods. The interpretation of this kind of expressions is included in LAPSE+, due to the fact that Java is an Object-Oriented Language and it is based on both simple and reference variables. The expressions that have been considered are those related to array calls, method calls, class instance creation, string concatenations, expressions in brackets and derived methods. In the next subsections we explain in more detail each of them.

4.2.1 Array Calls

LAPSE+ considers the interpretation of array calls as method parameters. For example, the following Java method call

```
obj.method(array[0]);
```

would not perform the backward propagation in LAPSE 2.5.6, because it contains an access to an array position as parameter. It is a significant weakness, because vulnerability sources in Java Web Applications can be arrays, as we can see in the value returned by the method *String[] getParameterValues(String name)* that belongs to Java ServletRequest Interface.

4.2.2 Method Calls

As well as array calls, it has been included method calls as parameter of a method. Therefore, backward propagation can be performed in LAPSE+ when we meet in the parameters a call to a method by an object reference, or to a method of the same class. The following examples would be performed by LAPSE+.

```
method(methodCall());  
method(obj.methodCall());
```

4.2.3 Class Instances Creation

Java interpreters provide the option of creating class instances as parameter of a method. Methods as

```
method(new Object(parameters));
```

are taken into account for the backward propagation in LAPSE+, because they can receive malicious data in its own parameters. Besides, the creation of a class instance can involve the call to another method when the constructor initializes the attributes of the object.

4.2.4 String Concatenations

Another improvement included in LAPSE+ is the interpretation of string concatenations as parameter of a method. For example, the method

```
method(rootPath+"/resources");
```

would give an error in LAPSE 2.5.6 when performing the backward propagation. Since this type of expressions are very common as method parameters, for example, when managing files, they have been considered in LAPSE+.

4.2.5 Expressions in Brackets

The interpretation of expressions in brackets as method parameters has been also included in LAPSE+. Expressions such as

```
method((var1+var2)+(var3+var4));
```

in which the backward propagation is performed recursively for each sub-expression. In the example given, we would have two propagations: one for the expression $(var1+var2)$ and other for $(var3+var4)$.

4.2.6 Derived Methods

Derivation methods for making transformations on strings, i.e., *toString* and *substring*, are considered to perform the backward propagation. The following examples perform the backward propagation in LAPSE+, doing it from *var*, the variable that call these methods, instead of carrying out the evaluation of the whole expression.

```
method(var.toString());  
method(var.substring(position));
```

4.3 LAPSE+ Integration in Eclipse Helios

As it has been mentioned before, LAPSE is a plugin for Eclipse, the software environment for developing Java applications. Specifically, the 2.5.6 version of LAPSE works with Eclipse Callisto, released on June 26, 2006. But we have to consider the evolution of this development environment until today. Since Eclipse Callisto release, it has been four releases more of this tool, comprising Eclipse Europa, Ganymede, Galileo and Helios. The latest one is Eclipse Helios, released on June 23, 2010. Considering that Eclipse is an open-source tool and its use is widespread among Java EE developers, it is significant to have LAPSE+ working with the latest version of this environment. For this reason, LAPSE+ has been developed to work with Eclipse Helios.

5 LAPSE+ Command-Line Version

One of the main features of Java is its portability. It means that an application developed in Java can run in several platforms, we just need a Java Run Environment (JRE) to run it. Since LAPSE+ has been developed in Java, specifically, using Java Development Kit (JDK) 1.6, it has been considered the property of portability of this language to take advantage of it. Not only to have an application that can run in different platforms. LAPSE+, as a plugin for Eclipse, can already run in different platforms, since Eclipse IDE is a portable application as well. The idea is to develop a version of the plugin that also can be independent of the Eclipse IDE, not completely leaving this environment, but using internally the features that it provides. Therefore, it has been developed a command-line version of LAPSE+ to analyse the source code of Java EE Applications without having them as a project in Eclipse IDE. This command-line version receives the source code to analyze and returns an XML report with the potential vulnerabilities found on the application. The command-line version is based on Equinox OSGi, the framework used internally by Eclipse to deploy and manage bundles, such as plugins and services. Among the features of this framework is the OSGi console, that we can be used to carry out these tasks. OSGi console consists of a Java JAR file that can be executed with a JRE. Hence, we have a console portable environment to deploy and manage LAPSE+ as a command-line application. This command-line version of LAPSE+ uses Ant to batch all the process of deployment and execution of the plugin inside OSGi framework. The Ant batch process receives as parameters the folder containing the source code to analyze and the folder where we want to store the XML report. LAPSE+ Command-line Version has been deployed as a service for VulneraNET, a collaborative platform for the detection, prediction and correction of vulnerabilities in web applications.

6 Conclusions

The static analysis of code is an essential process for detecting vulnerabilities in Java EE Applications. However, this sort of analysis needs a deep knowledge of the code, in terms of the language in which the application is written and the structure that it follows. The difficulty of this process increases when we face large applications,

consisting of thousands lines of code or having a complex structure with many Java classes. Therefore, it is important for the auditor or developer complementing the analysis of code by using tools that allow them to carry out this task in the most effective and efficient way. Thus, LAPSE+ is intended to provide this support to developers and auditors, but with the aim, above all, that security is considered from the development of the application, since this is the most important phase to correct all the possible vulnerabilities that can be presented.

Java EE development comprises a wide range of possibilities in the use of Java classes and libraries. This includes the large number of Java interfaces for communicating with other applications, such as SQL, XML or LDAP databases. Due to this heterogeneity, we need a tool that provides a complete catalog for detecting all the possible vulnerability sources and sinks that can be present in these applications. For this reason, LAPSE+ extends its catalog including the identification of vulnerability sources and sinks related to the management of XML and LDAP databases.

Another key point in the static analysis of code is the classification of the vulnerabilities. Using a tool that classifies the vulnerabilities by its nature is of great importance to apply the necessary security measures to fix them. It has to be considered that the tool must include an updated catalog with all the possible attacks that the application can be target of. Thus, LAPSE+ includes three categories of attack more than LAPSE 2.5.6, related to XPath Injection, XML Injection and LDAP Injection.

The vulnerabilities detected by LAPSE+ correspond to the injection of untrusted data in order to manipulate the behaviour of the application. Consequently, it is important to know how the malicious data propagate through the application and if they achieve to modify its normal operation. Hence, LAPSE+ enhances the performing of the backward propagation from a vulnerability sink to its source, including the identification of array accesses, method and constructor calls, string concatenations, expressions in brackets and derived methods.

The development of LAPSE+ with Java SE 6 entails a progress because of the performance improvements of this Java version compared to Java SE 5[9] with which LAPSE 2.5.6 runs. Furthermore, it allows the integration of LAPSE+ with Eclipse Helios, the latest release of this open-source Java development environment. Thus, the developer can use the features that this new version of Eclipse provides for Java EE Applications development.

The development of a command-line version of LAPSE+ means the independence of this tool as a plugin that works only for Eclipse IDE. Besides, this version provides the possibility of generating reports of all the potential vulnerabilities detected. These reports can be used by the developer as a historic database of the most common vulnerabilities detected on the code.

Finally, it is remarkable that LAPSE+ represents a progress on Free and Open Source Software (FOSS), being a GNU General Public License v3 software that is in constant development and can count on the collaboration of developers community.

Acknowledgements

LAPSE+ is part of VulneraNET Project, a collaborative platform for the detection, prediction and correction of vulnerabilities in web applications. The project has the support of Plan Avanza2, an initiative by the Spanish Ministry of Industry, Tourism

and Trade. LAPSE+ is also part of OWASP LAPSE Project. LAPSE+ provides OWASP with an updated security tool, enhancing the reliability on this prestigious open security project.

References

- [1] Viega, J., Bloch, J.T., Kohno, Y., McGraw, G.: ITS4: A static vulnerability scanner for C and C++ code. In: 16th Annual Conference on Computer Security Applications, ACSAC 2000, pp. 257–267 (2002)
- [2] McGraw, G.: Automated code review tools for security. *Computer* 41(12), 108–111 (2008)
- [3] Johnson, R.: J2EE development frameworks. *Computer* 38(1), 107–110 (2005)
- [4] Alur, D., Malks, D., Crupi, J.: *Core J2EE patterns: best practices and design strategies*. Prentice Hall PTR, Upper Saddle River (2001)
- [5] Kereki, F.: Web 2.0 development with the Google web toolkit. *Linux Journal* 2009(178), pages 2 (2009)
- [6] Tang, H., Huang, S., Li, Y., Bao, L.: Dynamic taint analysis for vulnerability exploits detection. In: 2010 2nd International Conference on Computer Engineering and Technology (ICCET), vol. 2, pages V2 (2010)
- [7] Livshits, V.B., Lam, M. S.: Finding security vulnerabilities in Java applications with static analysis. In: *Proceedings of the 14th conference on USENIX Security Symposium*, vol. 14, pages 18 (2005)
- [8] Barman, A.: LDAP application development using J2EE and .NET. In: *Proceedings of the First India Annual Conference, IEEE INDICON 2004*, pp. 494–497 (2005)
- [9] Kotzmann, T., Wimmer, C., Mössenböck, H., Rodriguez, T., Russell, K., Cox, D.: Design of the Java HotSpot client compiler for Java 6. *ACM Transactions on Architecture and Code Optimization (TACO)* 5(1), 1–32 (2008)

A Low-Power Wakeup-On-Demand Scheme for Wireless Sensor Networks^{*}

Byung-Bog Lee¹, Se-Jin Kim^{2,**}, and Choong-Ho Cho²

¹ RFID/USN Research Division, ETRI, Daejeon, Korea
bblee40@etri.re.kr

² Department of Computer and Information Science, Korea University, Korea
{kimsejin, chcho}@korea.ac.kr

Abstract. This paper proposes a wakeup-on-demand scheme based on the idea that a device should be woken just when it has to receive a packet from a neighboring device. To prove this scheme, this paper proposes a mathematical model based on the busy cycle of M/G/1 queuing systems to obtain a battery lifetime of one-hop cluster-topology shaped W-WSN. Also, we estimate the performance based on a more realistic log-normal shadowing channel model to prove the correctness of the mathematical analysis of the W-WSN with parameters such as the number of devices, the value of the wakeup period and communication distance variations.

Keywords: IEEE 802.15.4, Wakeup-On-Demand MAC Scheme, The unslotted CSMA/CA.

1 Introduction

Many wireless sensor network (WSN) applications must be designed to ensure that an event that happened is forwarded to the information sink in the order of hundreds of milliseconds. Even though WSNs have a wide range of useful applications they have the critical weakness of sleep-delay for the power saving gained by scheduled rendezvous schemes [1]. The drawback of the existing approaches [2] using a low-power radio for the wakeup channel is that the transmission range of the wakeup radio is significantly less than 10 meters. This may limit the applicability of such a technique as a device may not be able to wake up a neighboring device even if it is within its data transmission range.

Hence, this paper proposes a wakeup-on-demand scheme based on the idea that a device should be woken just when it has to receive a packet from a neighboring device. This scheme helps to reduce end-to-end delay while taking into account a dedicated low-power wakeup radio receiver with a very-short duty cycle clock. The duty cycle clock can be configured by software. A wakeup-on-demand device has multiple

^{*} This work was supported by the IT R&D program of MKE/KEIT (Project No. 10035310-2010-01 and 10038653-2010-01).

^{**} Corresponding author.

radios, namely, a main radio and a wakeup radio. The main radio can be switched between an IEEE 802.15.4 [3] compliant transceiver (IRT) and a wakeup radio transmitter (WRT) according to a configured modulation scheme. It is designed to operate over a range of 10 to 30 meters with ISM 2.4GHz carrier frequency. The IRT and WRT allows for the data rate of up to 250 Kbps with OQPSK modulation and 1 Kbps with OOK modulation, respectively. The other wakeup radio must consume very little power compared to the main radio because it remains idle-listening at all times. The proposed wakeup-on-demand scheme has an effect on WSNs which are located in remote areas and ideally are untouched for many years after installation.

The remaining sections of this paper are organized as follows: Section 2 describe the wakeup-on-demand radio and MAC scheme for the proposed W-WSN. Section 3 introduces models to estimate the performance of one-hop star-topology shaped W-WSN. Section 4 analyzes the results which are obtained by the proposed mathematical and the simulative models. Finally, Section 5 concludes the paper.

2 Wakeup-On-demand MAC (W-MAC) Scheme with the Unslotted CSMA/CA

A dedicated wakeup radio receiver (WRR) has two power modes as shown in Fig. 2. The doze mode operates during the toggle switch (SW) is connected directly to an envelope detector.

After the completion of this mode, if an 8-bit sequence collected from the OOK RF signal matches with SFD criteria, the SFD detection block wakes up the 2nd amplifier (AMP) and address decoder which is in deep-sleep. By switching the SW to the 2nd AMP, the above action completes the receive mode that collects and decodes addresses with Manchester coding in physical layer service data unit (PSDU). After receiving the complete address, the WRR transits from the receive mode to doze mode again.

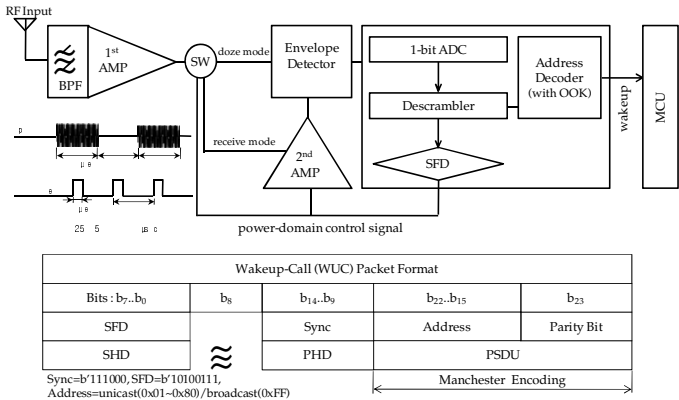


Fig. 1. The wakeup radio receiver and its sampling cycle clock

The device attempts to reduce power consumption by keeping the micro-controller unit (MCU) and the main radio powered down until an interesting event takes place. In charge of sensing this interesting event and awaking the other components is the WRR or sensor in the doze mode. The WRR and sensor generally use a simple hypothesis functions to test the existence of potential events in its input. Once the hypothesis is proven, the WRR or sensor wakes up the MCU to process information. Another hypothesis function for waking up main radio tests for the existence of events of interest in its input parameters. If the processed information turns out to be unimportant, it is discarded and the device returns to the doze mode. Otherwise, the data is stored in the memory, waiting to be transmitted to the other device once a predefined transmission buffer is full. Hence, when the buffer is just full, the woken MCU wakes up the main radio to perform the CCA, wakeup call, and data packet transmission as shown in Fig. 2 and Table 1.

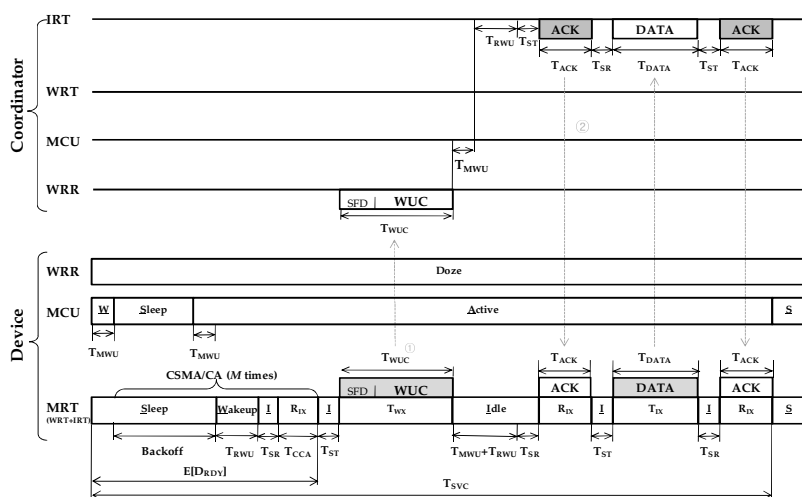


Fig. 2. The proposed wakeup-on-demand MAC (W-MAC) scheme model

Table 1. The energy and time characteristics for W-MAC

Symbol	Note	Value
E_{IS}	energy when IRT in the sleep mode	0.02mAh
E_{II}	energy when IRT in the idle mode	10mAh
E_{IT}	energy when transmitting with IRT	18mAh
E_{IR}	energy when receiving with IRT	20mAh
E_{WS}	energy when WRT in the sleep mode	0.426mAh
E_{WD}	energy when WRR in the doze mode	0.09mAh
E_{WT}	energy when transmitting with WRT	18mAh
E_{WR}	energy when receiving with WRR	1 mAh
E_{MS}	energy when MCU in the sleep mode	0.0007mAh
E_{MA}	energy when MCU in the active mode	0.27mAh
E_{MWU}	energy when MCU in the wakeup mode	0.26mAh
E_{RWU}	energy when MRT in the wakeup mode	9mAh
T_{MWU}	MCU wakeup time	0.006msec

Table 1. (Continued)

T_{RWU}	MRT wakeup time	1.5msec
T_{ST}, T_{SR}	turnaround time between the transmit and receive mode	0.192 msec
σ	backoff slot length	0.32msec
T_{CCA}	the time for performing a carrier sense ($T_{MWU} + T_{RWU} + T_{ST} + \Delta t$)	0.581msec
T_{WUC}	wakeup-command transmission time	6msec
T_{DATA}	data transmission time	0.8msec
T_{ACK}	acknowledgment transmission time	0.352msec
M	macMaxCSMABackoff, the number of preamble retransmission	4

3 Performance Analysis

3.1 Mathematical Model

We assume that: (i) $\{1 \leq n \leq 127\}$ devices are associated with a coordinator, (ii) each device generates packets according to the Poisson process with the rate λ for uplink traffic service, and (iii) the data packet size is constant, so that the transmission time of a data packet is fixed.

The stochastic process model $\{X(t), t \geq 0\}$ which describes stochastic behaviors of the tagged device is as follows:

$$X(t) = \begin{cases} \text{Doze,} & \text{when the device does is in waiting the SFD of wakeup call at } t \\ \text{Backoff,} & \text{when the device is in backoff process at } t \\ \text{CCA,} & \text{when the device is in CCA at } t \\ T_{WX}, & \text{when the device is in the wakeup call packet transmission at } t \\ T_{IX}, & \text{when the device is in IEEE802.15.4 packet transmission at } t \end{cases}$$

For analysis, the tagged device is modeled as the busy model of the M/G/1 queuing system where the service time is independent and identically distributed, the service time being the duration from the epoch which the data packet arrives at the head of the queue to the epoch where the data packet is transmitted successfully or discarded. Let t_k be the epoch terminated in the k^{th} busy period of the M/G/1 queuing system. Then, $\{X(t), t \geq 0\}$ is a regenerative process where a busy cycle of the M/G/1 queue is a regenerative cycle. In the regenerative process, the expected fraction of time that the system is in a given state is equal to the expected fraction of time during a single cycle that the system is in that state [4].

The tagged device senses that the channel is busy if it starts the CCA during another device's packet transmission period including the CCA. Since all the devices will have equal opportunity to transmit during one busy cycle of the tagged device, all other $n-1$ devices would statistically have one regenerative point so that the average successful transmission number of a device in one busy cycle will be calculated by $(1-P_{loss}) \cdot E[I]$ where P_{loss} is the packet loss probability. I is the number of packets served in a busy period of the M/G/1 queuing system and $E[I] = 1/(1-\rho)$ where traffic intensity $\rho = \lambda \cdot T_{SVC}$, T_{SVC} denotes the expectation of service time for the M/G/1 queuing system.

The service time is the duration from the epoch which the data packet arrives at the head of the queue to the epoch when the data packet is transmitted successfully or discarded. The service time T_{SVC} which denotes the expectation of service time for the M/G/1 queuing system can be obtained as

$$T_{SVC} = E[D_{RDY}] + T_{WUC} + T_{MWU} + T_{RWU} + 2 \cdot T_{ACK} + T_{DATA} + 2 \cdot T_{ST} + 2 \cdot T_{SR} \quad (1)$$

where $E[D_{RDY}]$ denotes the delay which is defined as the duration from the epoch when the data packet arrives at the head of the queue to the epoch just before the wakeup-call command packet transmission or discard. T_{WUC} , T_{DATA} and T_{ACK} are the transmission period of the wakeup-call command, data and acknowledgment packets respectively, and T_{ST} and T_{RT} are the receive-to-transmit or transmit-to-receive turn-around time.

The time period that the channel is occupied by $n - 1$ devices in a busy cycle is $(n - 1) \cdot (1 - P_{loss}) \cdot E[\Gamma] \cdot (T_{CCA} + T_{WUC} + T_{MWU} + T_{RWU} + 2 \cdot T_{ACK} + T_{DATA} + 2 \cdot T_{ST} + 2 \cdot T_{SR})$. Since the channel busy probability at the CCA is equal to the probability that the channel is busy given that the tagged device is not in the transmission state, α is calculated by Eq. (1) ($0 < \lambda < 1/T_{SVC}$). Eq. (2) is referred to as (1).

$$\alpha = \frac{(n-1) \cdot (1-P_{loss}) \cdot E[\Gamma] \cdot (T_{CCA} + T_{WUC} + T_{MWU} + T_{RWU} + 2 \cdot T_{ACK} + T_{DATA} + 2 \cdot T_{ST} + 2 \cdot T_{SR})}{\frac{1}{\lambda} + E[\Gamma] \cdot E[D_{RDY}]} \quad (2)$$

To compute $E[D_{RDY}]$, let us adopt the notation $W_i = \min\{2^i \cdot W_{min}, W_{max}\}$ where $W_{min} = 2^{BE_{min}}$, $W_{max} = 2^{BE_{max}}$. BE_{min} and BE_{max} denote the minimum backoff exponent (macMinBE) and the maximum backoff exponent (macMaxBE), respectively. Default values of these exponents are 3 and 5, respectively. Then, the expected delay $E[D_{RDY}]$ can be formulated as Eq. (3).

$$E[D_{RDY}] = T_{MWU} + N_{cca} \cdot (T_{MWU} + T_{RWU} + T_{SR}) + \sum_{v=0}^M \alpha^v \cdot (1 - \alpha)$$

$$\left\{ \sum_{i=0}^v \frac{W_i - 1}{2} \cdot \sigma + (v + 1) \cdot T_{CCA} \right\} + \alpha^{M+1} \left\{ \sum_{i=0}^M \frac{W_i - 1}{2} \cdot \sigma + (M + 1) \cdot T_{CCA} \right\} \quad (3)$$

$$N_{cca} = \sum_{i=0}^M \alpha^i \cdot (1 - \alpha) + \alpha^{(M+1)} \cdot (M + 1) \quad (4)$$

Here, σ is the length of a backoff slot, T_{MWU} and T_{RWU} denotes the time necessary to wake up the MCU and radio chip in the sleep mode, and N_{CCA} is the average number of CCAs until successful transmission or discard of the packet as Eq. (4). The first summation of Eq. (3) corresponds to a situation where a packet is successfully transmitted and the second term of Eq. (3) corresponds to a situation where a packet is discarded after $(M + 1)$ attempts at CCA have failed. So, packet loss probability can be obtained as

$$P_{loss} = \alpha^{M+1} \quad (5)$$

Note that α in Eq. (2) is expressed in terms of $E[D_{RDY}]$ and P_{loss} . Also, $E[D_{RDY}]$ in Eq. (3) and P_{loss} in Eq. (5) are expressed in terms of α . Therefore, by solving nonlinear Eqs. (2), (3) and (5), we obtain necessary value α . In addition, The expected delay $E[D_{RDY}]$ and packet loss probability P_{loss} are derived.

The expected delay $E[D]$ from the epoch where data arrives at the device to the epoch just before transmission of a wakeup command packet or discard is calculated as follows by the theory of M/G/1 queuing systems.

$$E[D] = \frac{\lambda \cdot (T_{svc})^2}{2(1-\rho)} + E[D_{RDY}] \quad (6)$$

This paper calculates the average current consumption E per millisecond ($mA/msec$), one of the primary performance metrics.

$$E = \frac{E[\Gamma] \cdot (E_{MCU} + E_{MRT} + E_{OHEAR})}{\left(\frac{1}{\lambda} + E[\Gamma] \cdot (T_{svc} + E[D] - E[D_{RDY}])\right)} + E_{IS} + E_{MS} + E_{WD} + E_{SA} \quad (7)$$

where E_{MCU} and E_{MRT} are computed by:

$$\begin{aligned} E_{MCU} &= (E_{MWU} - E_{MS}) \cdot T_{MWU} + (E_{MA} - E_{MS}) \cdot \\ &\quad \{N_{CCA} \cdot (T_{MWU} + T_{RWU} + T_{SR} + T_{CCA}) + T_{SVC} - E[D_{RDY}]\}, \\ E_{MRT} &= (E_{RWU} - E_{IS}) \cdot N_{CCA} \cdot T_{RWU} + (E_{IR} - E_{IS}) \cdot N_{CCA} \cdot T_{CCA} + (E_{II} - E_{IS}) \cdot \\ &\quad \{N_{CCA} \cdot T_{SR} + T_{MWU} + T_{RWU} + 2 \cdot T_{SR} + 2 \cdot T_{ST}\} + 2 \cdot (E_{IR} - E_{IS}) \cdot T_{ACK} \\ &\quad + (E_{IT} - E_{IS}) \cdot T_{DATA} + (E_{WT} - E_{IS}) \cdot T_{WUC} \cdot 250 / (2 \cdot B_W) \end{aligned}$$

By taking into consideration the overhearing of the wakeup-call packet, E_{OHEAR} is given as:

$$E_{OHEAR} = (n - 1) \cdot (1 - P_{loss}) \cdot T_{OHEAR} \cdot (E_{WR} - E_{WD})$$

where T_{OHEAR} is equal to the transmission time of the wakeup PPDU without SHD field.

As shown in Fig. 2 and Table 1, let the energy consumption in each power mode be the current consumption per millisecond. Note that time elements are lengths per millisecond. Finally, the lifetime of a battery is computed as follows:

$$L = \frac{E_{TOTAL}}{E} \quad (8)$$

where E_{TOTAL} is the initial current capacity of a battery's energy.

3.2 Simulative Model

The performance of the proposed W-MAC scheme has been analyzed by mathematical computation. However, the numerical results are based on optimistic assumptions: no collision, and no noise or interference. These are not the real working situation for a WSN.

Therefore, we estimate the performance of the proposed W-MAC scheme based on a more realistic channel model with the parameters limited as follows: (i) the IRT provides for a wide transmitter output power range, but the device must be capable of transmitting -3 dBm; (ii) the IRT and WRR must be capable of correctly decoding a signal with an input power of -85 dBm and -70 dBm, respectively, in the 2.4 GHz band. Better sensitivity is not prohibited.

In the case of free space, the ratio of the power at a receiving antenna to the power at a transmitting antenna is given by the Friis model. For simplicity, we treat the approximately 2.4 GHz so that the path loss can be expressed as:

$$L_B(d) \approx 100.04 + 20 \log_{10} d_{km}$$

For a transmitter rated at 0 dBm, the maximum free space ranges of the IRT and WRR must be approximately 177 m and 31 m, respectively.

To approximate the propagation behavior in RF channels in real environments, we add fading effects into the free space model, a so called “log-normal shadowing” model is given by

$$L_B(d) = L_B(d_0) + 10 n \log_{10} \left(\frac{d}{d_0} \right) + X_{dBm},$$

where d_0 is some reference distance, $L_B(d_0)$ is the path loss at this distance, n is the path loss coefficient from 0 to 4 , and X_{dBm} is a Gaussian random variable with zero mean and standard deviation from 0 to 10 . This log-normal shadowing model is used to simulate both the path loss and the fading effects.

4 Numerical and Simulative Results

The analytical results of an ideal scheduled rendezvous scheme will be compared to estimate the performance of the proposed W-MAC scheme. And, the simulation will be conducted to prove the correctness of the mathematical analysis of the proposed W-MAC scheme. These are based on one-hop cluster topology with traffic that obeys the Poisson process traffic model. And, the different parameters are limited to a reasonable range for a fair comparison as follows: (i) the sampling duration $T_s = T_{CCA}$ is $T_{MWU} + T_{RWU} + T_{ST} + \Delta t$; (ii) all the schemes take into account $T_{MWU} + T_{RWU} + T_{SR}$, the MCU and the radios’ transition time from the sleep to active mode; (iii) BEmin and BEmax denote the minimum backoff exponent (macMinBE) and the maximum back-off exponent (macMaxBE), respectively. The default values of these exponents are 3 and 5 , respectively; (iv) the maximum number of backoffs (M) which the CSMA-CA algorithm will attempt before declaring a channel access failure. The default value of M is 4 ; (v) the length of data and acknowledgment’s PPDU in octet assume 25 and 11 , respectively; (vi) the ideal X-MAC takes into account only both MCU and IRT with the OQPSK without the WRT and WRR; (vii) the initial current capacity of battery energy is assumed to be 2100 mAh. Note that the self-discharging effect of the battery is not considered here.

4.1 Numerical Results

As shown in Fig. 3, in the case of no traffic, the W-MAC consumes much less power than the ideal X-MAC scheme, especially when the length of a wakeup period (T_w) of the ideal X-MAC is short. In Fig. 3, the cross over point between the W-MAC and ideal X-MAC curves can be moved further to the right by increasing the value of the wakeup period, B_w . This will, of course, accompany an increase in delay.

Fig. 4 shows the battery lifetime of a device with uplink traffic. Thus, the W-MAC scheme is better in the sense of power consumption during the same packet arrival rate. And it consumes low-overhear-power than the ideal X-MAC when the traffic is very high.

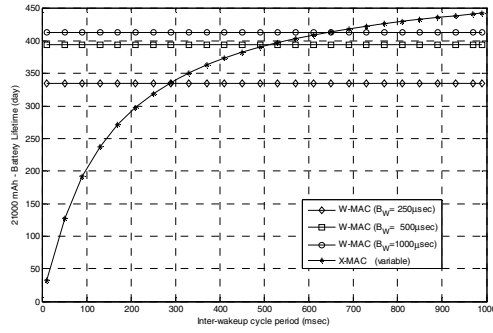


Fig. 3. Battery lifetime comparison with no traffic: proposed W-MAC vs. ideal scheduled rendezvous X-MAC

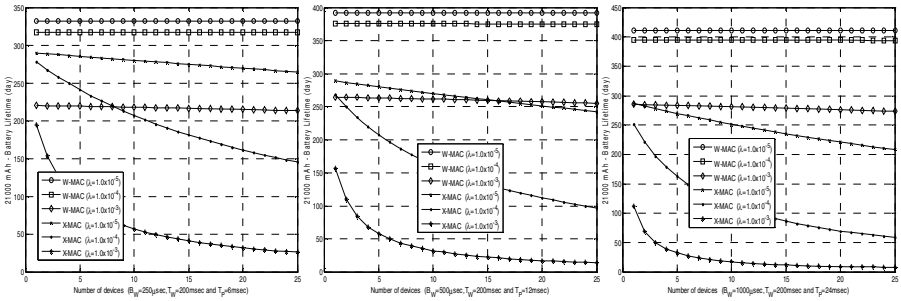


Fig. 4. Battery lifetime comparison with uplink traffic: proposed W-MAC vs. ideal scheduled rendezvous X-MAC

4.2 Simulative Results

Fig. 5 shows the mean path loss on 50 times of the log-normal shadowing model compared to the Friis free space model.

In log-normal shadowing model, the path loss coefficient can vary between $n = 0$ and $n = 4$. Furthermore, so can the variance in the received power ranges from $X_{dBm} = 0$ to $X_{dBm} = 10$. Hence, the mean path loss of the shadowing model shows a more fickle characteristic of the channel with distance than that of the free space model, which just indicates that the channel suffers from severe path loss when the distance increases. At a boundary of $20m$ for the communication range, the mean path loss damage of the shadowing becomes 9.29% compared to the free space model. This will, of course, accompany an increase in packet transmission delay and the average number of CCAs having an effect on the device's battery lifetime.

As shown in Fig. 5, transmission range of the wakeup call signal is shorter than that of the data radio. This may limit the applicability of such a technique as a device may not be able to wake up a neighboring device even if it is within its data transmission range, 74 meters.

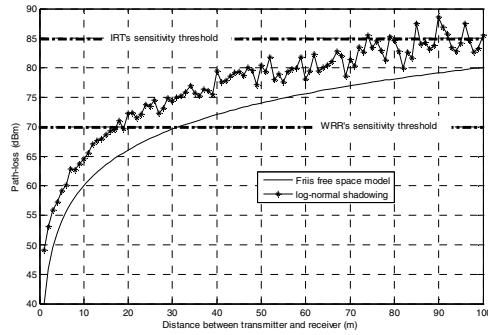


Fig. 5. The mean path loss of W-WSN

When the number of devices and traffic are further increased, the numerical and simulative results of the battery lifetime over X-MAC scheme are as shown in Fig. 6.

As shown in Fig. 6, the numerical results are larger than the simulative results in most cases, and are almost similar to the simulative results when the traffic is very small.

In case of a device with $\lambda = 1.0 \times 10^{-3}$ and $B_w = 250 \mu\text{sec}$, the battery lifetime gains of the numerical results increase to 7.6% , 8.0% and 9.3% compared to the simulative results at distances varying between transmitter and receiver, $10m$, $15m$ and $20m$, respectively. The distance accompanies an increase in the battery lifetime gain of the numerical against simulative results. When the value of B_w is further increased, the battery lifetime gain is similar to the gain of $B_w = 250 \mu\text{sec}$.

As a result, the difference between numerical and simulative results can be neglected. The main reason for neglect is that the traffic in many WSNs is very-low.

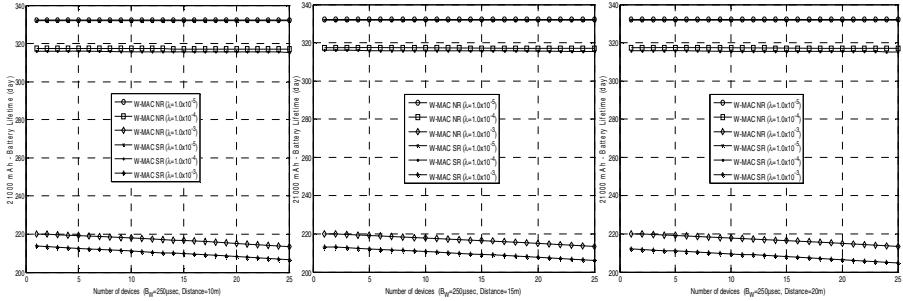


Fig. 6. Battery lifetime comparison of W-MAC scheme with traffic and distance varies : numerical (NR) vs. simulative (SR) results

5 Conclusions

In this paper, we proposed a wakeup-on-demand scheme for WSNs which employs a dedicated low-power wakeup radio receiver. The power consumption of the proposed scheme was obtained using an M/G/1 busy cycle analysis.

We compared the performance of the proposed scheme in terms of power consumption and average delay with that of X-MAC, the best of the best of scheduled rendezvous MAC schemes. By numerical and simulative examples, we demonstrated that the proposed scheme allows a substantial decrease in power consumption while achieving a relatively lower average delay compared with the existing schemes.

For flexible application of the proposed scheme, the tradeoff between power consumption and end-to-end delay can be made by adjusting the value of the wakeup period Bw .

References

1. Buettner, M., Yee, G., Anderson, E., Han, R.: X-MAC: A Short Preamble MAC Protocol For Duty-Cycled Wireless Sensor Networks. In Technical Report CU-CS-1008-06 (May 2006)
2. Pletcher, N., Gambini, S., Rabaey, J.: A 65 μ W, 1.9 GHz RF to digital baseband wakeup receiver for wireless sensor nodes (2007), doi:10.1109/CICC
3. IEEE Recommendation, Wireless LAN medium access control(MAC) and physical layer (PHY) specifications for low-rate wireless personal area network (LR-WPANs), IEEE 802.15.4 (2006)
4. Kim, T.O., Park, J.S., Chong, H.J., Kim, K.J., Choi, B.D.: Performance analysis of the IEEE 802.15.4 Non-beacon Mode with the Unslotted CSMA/CA. IEEE Commun. Letter 12(4) (April 2008)

The Improved Space-Time Trellis Codes with Proportional Mapping on Fast Fading Channels

Ik Soo Jin

Dept. of Information & Communication Engineering, Kyungnam University,
449 Woryeong-dong, Masanhappo-gu, Changwon-si, Gyeongsangnam-do, 631-701, Korea
isjin@kyungnam.ac.kr

Abstract. The use of multiple antennas in most future mobile communication systems seems to be inevitable. Today, the main question is how to include multiple antennas and what are the appropriate methods for specific applications. In this paper, a space-time trellis code (STTC) with proportional mapping is proposed in order to improve the bit error rate (BER) performance. The core of proportional mapping assigns information bits with a Hamming distance in proportion to the sum of the Euclidean distance to each trellis branch of STTC. Following these rules, we can construct 8-state 8-PSK STTC with proportional mapping. To the best knowledge, this combination has not been considered yet. The BER performance is examined with simulations and the performance of proportional mapping is compared to that of well known natural mapping and Gray mapping on both fast Rayleigh as well as fast Rician fading channels. It is shown that in case of 8-state 8-PSK STTC, proportional mapping outperforms Gray mapping and natural mapping by about 0.5 dB and about 0.7 dB at a BER of 10^{-5} , respectively, over fast Rician fading channels, without any additional loss.

Keywords: STTC, proportional mapping, Gray mapping, natural mapping.

1 Introduction

Mobile communication comprises a wide range of technologies, services, and applications that have come into existence to meet the particular needs of users in different deployment scenarios. Wireless multimedia traffic is increasing far more rapidly than voice, and will increasingly dominate traffic flows. The use of multiple antennas at transmitter and/or receiver, commonly known as the multiple-input multiple-output (MIMO) system, has become a pragmatic and cost-effective approach that offers substantial gain in mobile communication. Today, the main question is how to include multiple antennas and what are the appropriate methods for specific applications.

Space-time trellis codes (STTCs), which are able to combat effects of fading, incorporate jointly designed channel coding, modulation, transmit diversity and optional receiver diversity. The STTCs have provided the best trade-off between data rate, diversity advantage and trellis complexity. In [1] Tarokh *et al.* introduced the concept of STTC as an extension to conventional trellis coding, and derived analytical bounds and design criteria to obtain codes for slow and fast fading channels. These codes may

be designed to extract diversity gain and coding gain using the design criteria. STTCs have attracted considerable attention mainly due to the significant performance gains they can provide. A numerous of investigations have shown the great promise of STTC [2]-[5].

In general, the natural mapping is used for mapping between information bits and M -ary channel symbols in STTCs. However, it is possible to gain a further performance improvement using proper bit-to-symbol mapping. In [6]-[8], Gray mapping is used to improve the bit error rate (BER) performance. However, one can say that in Gray mapping, the Hamming distances between information bits are imperfectly proportional to the Euclidean distance between M -ary channel symbols.

The motivation of this work stems from the application of new mapping to STTC to obtain a further performance improvement. In new mapping, the Hamming distances between information bits are perfectly proportional to the Euclidean distance between M -ary channel symbols. The core of new mapping assigns information bits with a Hamming distance in proportion to the sum of the Euclidean distance to each trellis branch of STTCs. We refer the new mapping as proportional mapping. Following these rules, we can construct 8-state 8-PSK STTC with proportional mapping. In addition, little has so far been reported on the application of proportional mapping to STTC over fading channels. We have not attempted to modify the channel signal set obtained by Tarokh/Seshadri/Calderbank (TSC) [1] because it is beyond the scope of this paper.

In this paper, The BER performance of STTC with proportional mapping is examined with simulations and the performance of proportional mapping is compared to that of well known natural mapping and Gray mapping on both fast Rayleigh as well as fast Rician fading channels. The 8-state 8-PSK/4-PSK TSC STTC with two transmit antennas and one receive antenna in TSC code [1] is considered.

2 System Model

Fig. 1 illustrates the block diagram of the considered baseband space-time coded system with $n_T = 2$ transmit antennas and $n_R = 1$ receive antenna. The signals on the matrix channel, i.e., the $n_T \cdot n_R$ transmission paths between transmitter and receiver,

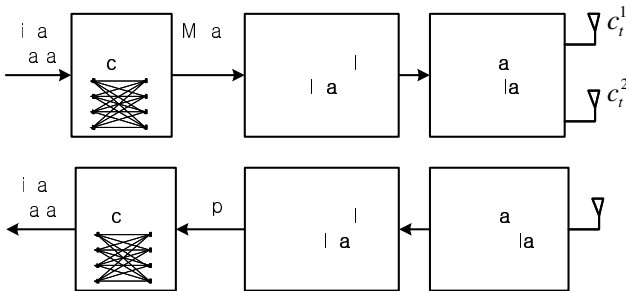


Fig. 1. General STTC system model

are modeled as independent samples of a complex Gaussian random variable with mean zero and variance 0.5 per dimension. This assumption corresponds to fast Rayleigh fading but the generation to Rician fading is straightforward. In fast fading channel, it is assumed that the fading coefficients are constant within each symbol period and vary from one symbol to another. It is also assumed that channel state information is known to receiver perfectly. The block size of symbol interleaver per antenna is 10×10 . The received signal r_t at time t is given by

$$r_t = \sum_{i=1}^{n_T} h_t^i c_t^i \sqrt{E_s} + N_t \quad (1)$$

where h_t^i is the complex Gaussian channel path gain from transmit antenna i ; E_s is the energy per symbol; c_t^i is space-time trellis coded symbol transmitted via transmit antenna i at time t ; N_t is the additive complex white Gaussian noise at time t with zero mean and variance 0.5 per dimension.

The amplitude of the envelope of the received signal is a normalized random variable with a Rician probability density function given by

$$p(h) = 2h(1 + K_f) e^{-K_f - h^2(1 + K_f)} I_0(2h\sqrt{K_f(1 + K_f)}), \quad h \geq 0 \quad (2)$$

where the Rician fading parameter K_f represents the ratio of the direct and specular signal components to the diffuse component and $I_0(\cdot)$ is the zero order modified Bessel function of the first kind. As a special case, $K_f = 0$ yields Rayleigh fading, and $K_f = \infty$ describes the AWGN channel. The outputs of symbol deinterleaver are sent to STTC decoder using the conventional Viterbi algorithm with no quantization.

3 Mapping between Information Bits to M-ary Channel Symbols

3.1 Proportional Mapping

The mapping or assignment of k information bits to the $M = 2^k$ possible signal may be done in a number of ways. Proper bit-to-symbol mapping can reduce the number of erroneous bits. The preferred assignment of k information bits to the M -ary signal is natural mapping in STTCs. However, it is possible to gain a further performance improvement using proper bit-to-symbol mapping.

It is well known that Gray mapping can also be a method for performance improvements in STTC [6]-[8]. Fig. 2 shows the three kinds of mapping for 8-PSK constellation. Gray mapping is one in which two successive values differ in only one bit. In Fig. 2(b), one can find that the Hamming distance between the 8-PSK symbol **0** and **3**, is $d_H = 1$, and the Hamming distance between the 8-PSK symbol **0** and **4**, is $d_H = 2$, respectively. But the Euclidean distance between the 8-PSK symbol **0** and **4**, is much larger than the Euclidean distance between the 8-PSK symbol **0** and **3**.

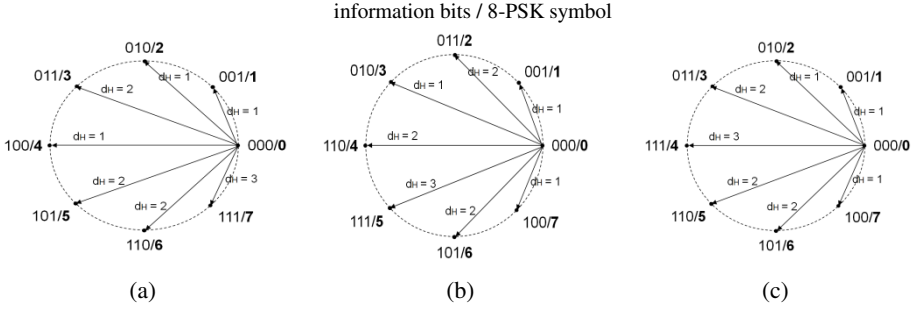


Fig. 2. The three kinds of mapping for 8-PSK constellation. (a) Natural mapping. (b) Gray mapping. (c) Proportional mapping.

Therefore, one can say that in Gray mapping, the Hamming distances between information bits are imperfectly proportional to the Euclidean distance between M-ary channel symbols. Fig. 2(c) shows proportional mapping for 8-PSK constellation. One can find that the Hamming distances between information bits are perfectly proportional to the Euclidean distance between M-ary channel symbols in proportional mapping. For special case, proportional mapping is exactly the same as Gray mapping in 4-PSK constellation due to the small signal set size.

3.2 The 8-State 8-PSK TSC Code with Proportional Mapping

This paper studies the performance of STTC with proportional mapping in order to improve the BER performance. When we try to apply the proportional mapping concept to STTC, it is necessary to modify the proportional mapping slightly because there is more than one M-ary symbol in the STTC trellis branch.

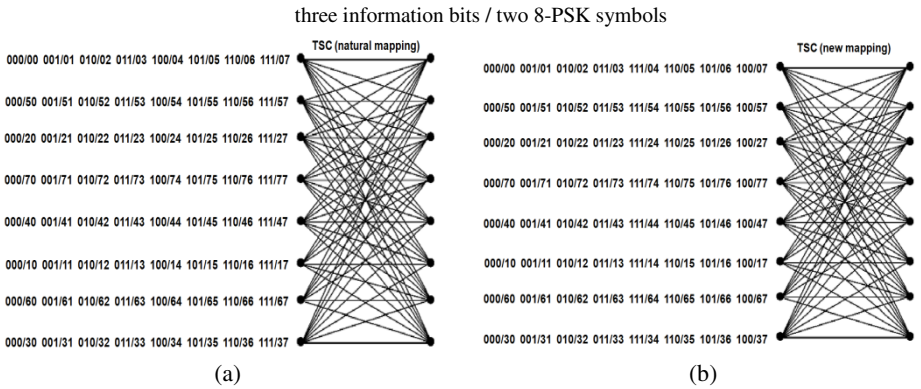


Fig. 3. Trellis diagram of 8-state 8-PSK TSC code with 2 transmit antennas. (a) TSC code with natural mapping. (b) TSC code with proportional mapping.

The core of modification assigns information bits with a Hamming distance in proportion to the sum of the Euclidean distance to each trellis branch of STTC. Following this rule, we can construct 8-state 8-PSK STTC with proportional mapping.

The trellis diagram of 8-state 8-PSK TSC code with 2 transmit antennas is shown in Fig. 3. In Fig. 3, the trellis diagram of 8-state 8-PSK TSC code with Gray mapping is skipped because it can be easily obtained from the similar way. The comparisons of Euclidean distance d_E and Hamming distance d_H are also shown in Fig. 4. The two 8-PSK symbols **00** and 3 information bits 000 are used as a reference when d_E and d_H are calculated. It is worthwhile to note that the Hamming distances of information bits are perfectly proportional to the sum of the Euclidean distance to each trellis branch of STTC in Fig. 3(b) and Fig. 4(b).

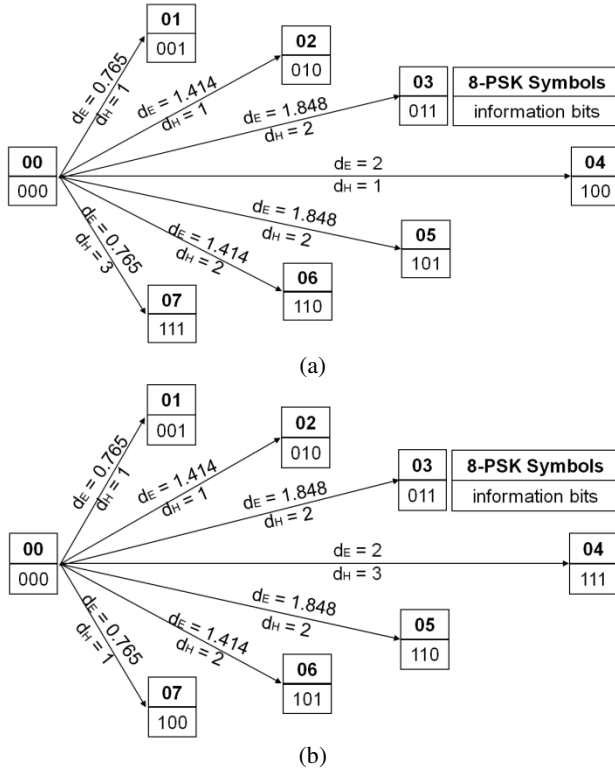


Fig. 4. Comparisons of Euclidean distance and Hamming distance. (a) Natural mapping. (b) Proportional mapping.

4 Simulation Results

The bit error rate (BER) performance is evaluated by simulation. In the simulations, each frame consisted of 100 symbols transmitted from each antenna. The number of transmit antenna is two and the number of receive antenna is one. A maximum

likelihood Viterbi decoder with perfect channel state information (CSI) is employed at the receiver. Jakes' model is used for Rayleigh fading based on summing sinusoids [9]. C-language is used to conduct the simulation. The simulation conditions are listed in Table 1.

Table 1. Simulation conditions

Parameters	Description
Frame Length	20 ms
Carrier Frequency	2.0 GHz
bits/frame	200
Modulation	4-PSK, 8-PSK
State	8
Transmit Antenna	2
Receive Antenna	1
Interleaver	10× 10 block interleaver per antenna. Symbol by symbol interleaver
Channel Model	Rayleigh fading or Rician fading with Rician parameter K_f . Fast fading. Jakes model for Rayleigh fading
Mobile Speed	120km/h
Channel Estimation	Ideal channel estimation (Channel information is known to receiver perfectly)
Decoder	Viterbi decoder with no quantization

Fig. 5 shows the BER performance of the 8-state 8-PSK/4-PSK TSC code over fast Rayleigh fading channels ($K_f = 0$). As stated previously, proportional mapping is exactly the same as Gray mapping in 4-PSK constellation. From Fig. 5, it can be observed that in case of 8-state 8-PSK TSC code, TSC code with proportional mapping is superior to TSC code with natural mapping by about 0.7 dB at a BER of 10^{-4} , while the performance of TSC code with proportional mapping is almost the same as that of TSC code with Gray mapping. The imperceptible difference between proportional mapping and Gray mapping is due to the severe channel conditions such as fast Rayleigh fading channels. It can also be observed that in case of 8-state 4-PSK TSC code, proportional mapping outperforms the natural mapping by about 1.2 dB at a BER of 10^{-5} .

Fig. 6 illustrates the BER performance of the 8-state 8-PSK/4-PSK TSC code over fast Rician fading channels ($K_f = 10$). It is worthwhile to mention that if the channel tends toward Rician, the performance differences between proportional mapping and Gray mapping can be clearly seen. From Fig. 6, we can observe that in case of 8-state 8-PSK TSC code, proportional mapping outperforms Gray mapping and natural mapping by about 0.5 dB and about 0.7 dB at a BER of 10^{-5} , respectively. The performance improvements are mainly resulting from the good distance properties that may be achieved through proportional mapping. As a further point, it is also observed that in case of 8-state 4-PSK TSC code, proportional mapping outperforms the natural mapping by about 0.1 dB at a BER of 10^{-5} .

Finally, notice that the cost that has to be paid for this improvement can be negligible because the hardware complexity of proportional mapping is almost the same as that of natural/Gray mapping.

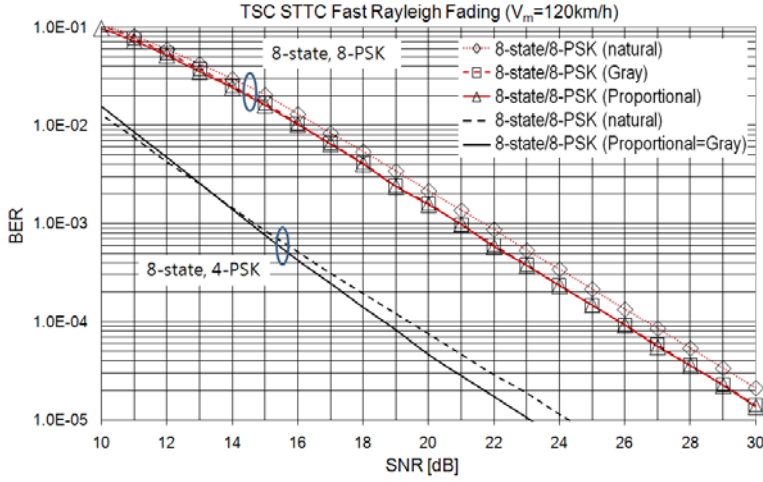


Fig. 5. BER comparisons of 8-state TSC codes on fast Rayleigh fading channels ($K_f = 0, n_T = 2, n_R = 1$)

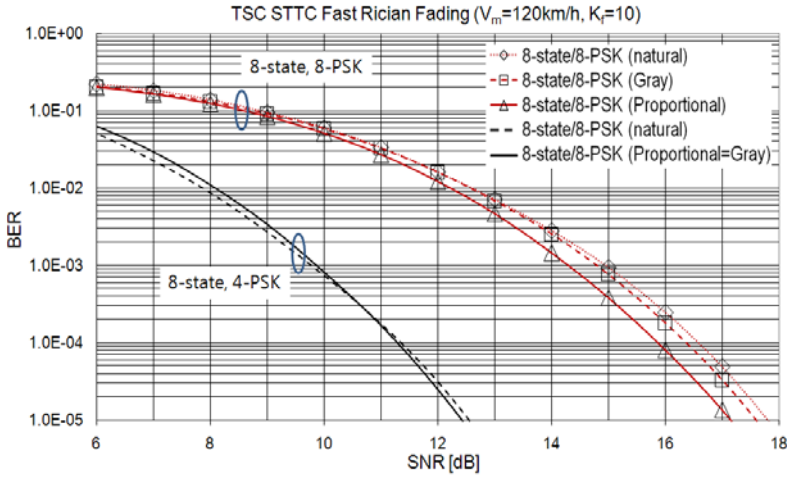


Fig. 6. BER comparisons of 8-state TSC codes on fast Rician fading channels ($K_f = 10, n_T = 2, n_R = 1$)

5 Conclusions

We presented a STTC with proportional mapping to improve the BER performance. In proportional mapping, the Hamming distances between information bits are perfectly proportional to the Euclidean distance between M-ary channel symbols. The core of proportional mapping assigns information bits with a Hamming distance in proportion to the sum of the Euclidean distance to each trellis branch of STTC.

Following these rules, we can construct 8-state 8-PSK STTC with proportional mapping. The BER performance of STTC with proportional mapping is examined with simulations, and the performance of proportional mapping is compared to that of well known natural mapping and Gray mapping on both fast Rayleigh as well as fast Rician fading channels. From the simulation results, it is shown that in case of 8-state 8-PSK TSC code, proportional mapping outperforms Gray mapping and natural mapping by about 0.5 dB and about 0.7 dB at a BER of 10^{-5} , respectively. It is worthwhile to mention that if the channel tends toward Rician, the performance differences between proportional mapping and Gray mapping can be clearly seen. As a further point, it is also showed that the cost that has to paid for this improvement can be negligible because the hardware complexity of proportional mapping is almost the same as that of natural/Gray mapping. Although this paper deals with 8-state TSC code exclusively, the ideas can be applied to any other STTC code.

Acknowledgments. This work was supported by Kyungnam University Foundation Grant, 2011.

References

1. Tarokh, V., Seshadri, N., Calderbank, A.R.: Space-Time Codes for High Data Rate Wireless Communication - Performance Criterion and Code Construction. *IEEE Trans. Inform. Theory* 44, 744–765 (1998)
2. Sibille, A., Oestges, C., Zanella, A.: MIMO From Theory to Implementation. Academic Press, San Diego (2011)
3. Shr, K.T., Taylor, D.: A Low-Complexity Viterbi Decoder for Space-Time Trellis Codes. *IEEE Trans. Circuits and Systems I* 57, 873–885 (2010)
4. Turner, J., Chen, H.D., Huang, Y.H.: Reduced Complexity Decoding of Space Time Trellis Codes in the Frequency Selective Channel. *IEEE Trans. Commun.* 57, 635–640 (2009)
5. Flores, J., Chen, S.J., Jafarkhani, H.: Quasi-Orthogonal Space-Time-Frequency Trellis Codes for Two Transmit Antennas. *IEEE Trans. Wireless Communications* 9, 2125–2129 (2010)
6. Tran, N.H., Nguyen, H.H., Le-Ngoc, T.: Coded unitary space-time modulation with iterative decoding: error performance and mapping design. *IEEE Trans. Commun.* 55, 703–716 (2007)
7. Panagos, A., Kosbar, K.: A Gray-Code Type Bit Assignment Algorithm for Unitary Space-Time Constellations. In: *Proc. IEEE Global Telecommun. Conf (GLOBECOM)*, pp. 4005–4009 (2007)
8. Villardi, G.O., Freitas de Abreu, G.T., Kohn, R.: Performance of STBCs with linear maximum likelihood decoder in time-selective channels. In: *Proc. the 11th IEEE Singapore Int'l. Conf on Communication Systems (ICCS)*, pp. 679–683 (2008)
9. Jakes Jr., W.C.: *Microwave Mobile Communications*. John Wiley & Sons, New York (1974)

Strategies for IT Convergence Services in Rural Areas

Hyongsoun Kim¹, Kwang-Taek Ryu¹, Sang-Yong Ha¹, and Eunyoung Lee²

¹ Digital Infrastructure Division, IT Convergence Service Department,
National Information Society Agency, Seoul, Korea
{khs, ryukt, yong}@nia.or.kr

² Dept. of Computer Science, Dongduk Women's University, Seoul, Korea
elee@dongduk.ac.kr

Abstract. The digital divide refers to the gap between people with effective access to digital and information technology and those with very limited or no access at all. The inequality of accessing information could cause the inequality of opportunity between different social groups. This inevitably results in social problems. The Korea government has been conducting a project resolving the digital divide between urban areas and rural areas since 2010. In this paper, we introduce the rural BcN project of the Korea government with the motivation and development plans. After that, we propose the strategies for boosting broadcast-communication services and the strategies for enhancing user experience.

Keywords: digital divide, BcN, convergence services, policies.

1 Introduction

The digital divide usually refers to the gap between people with effective access to digital and information technology and those with very limited or no access at all. It includes the imbalance both in physical access to technology and the resources and skills needed to effectively participate as a digital citizen [1],[2]. The digital divide is typically due to the imbalanced availability to IT technology by gender, income, race and location. The term *global digital divide* is much familiar to the public, which refers to differences in access between countries. However, there also exists the digital divide in one country, reflecting the gap between different social groups. The inequality of accessing information, i.e., the failure of obtaining the information timely, could cause the inequality of opportunity between different social groups. This inevitably results in social problems.

The Korea government has conducted a series of projects to build the network infrastructure of high speed, to provide various converged services, and to encourage the Internet service providers in Korea to invest IT-related researches [3]. The result of the series of government-driven projects is quite successful, so Korea is considered as one of the top IT countries in the world [4]. The recent project has concentrated on laying a foundation for universal broadcast and communication convergence services by expanding broadband broadcast communication networks [5],[6], but it results in the situation in which broadband networks are concentrated in cities, not in rural areas, according to the density of population.

The low speed of the network infrastructure and the low profitability of rural areas made a gap of information between urban areas and rural areas in Korea. In order to solve rural areas' economic, social and cultural issues and to vitalize the economy and industry, it is necessary to diffuse and spread specialized services based on broadband broadcast-communication networks in rural areas.

Constructing broadband broadcast-communication networks is the key to solve those problems; it is also needed to establish a nation-wide and inter-ministerial co-operation system. Local governments and carriers should be involved in the co-operation system, too.

Though the previous high-speed network development had ended quite successfully, the change of service model of ICT requires more network bandwidth than ever. To satisfy the request in rural areas, the Korea government began to build broadband networks in rural areas since 2007, and one dedicated plan called *Long-term Plan for Constructing Broadband Network in Rural Areas* was made and announced by KCC (Korean Communications Commission) in 2009 [7]. The elaborated version was announced by the Korea government in June 2010 [8],[9]. In these reports, more goals were identified and tasks for achieving the goals were clarified.

In this paper, we introduce the rural BcN project of the Korea government. After that, we propose the strategies for boosting broadcast-communication services and strategies for enhancing user experience. In Section 2, we analyze the status of digital divide of Korea in terms of the network infrastructure and the available IT services. In Section 3, we introduce the rural BcN project of the Korea government with its motivation and development objectives. After the brief introduction of rural BcN plan, we propose the strategies for boosting broadcast-communication services and strategies for enhance user experience based on our survey on rural residents in Section 4. We draw a conclusion in Section 5.

2 Analysis of Current Status

In this section, we analyze the status of digital divide of Korea in terms of network infrastructure and available IT services.

As of December 2009, 3,925,866 households, that is, 20.4 percent of the whole 19,261,292 households in Korea lived in rural areas, and 450,061 households, 11.5 percent of the rural households lived in small villages of less than 50 households.

The population and generations of rural communities have been diversified largely due to accelerating reduction and aging of population and increasing multi-cultural families. The population over 65 years old in rural areas is expected to increase continuously. While the rate of senior citizens in rural areas was 18.6% in 2005, the rate is estimated to be 22.9% in 2014. The number of foreign residents in rural areas is expected to increase, too. The number of foreign residents was 45,000 in 2005, and increased to 70,000 in 2008. The number will be expected to be 115,000 in 2014.

From 2002 to 2008, the construction of high-speed networks, including broadband networks, was completed for 3.76 million households (99.7%) out of 3.77 million rural households in the country. Table 1 shows the number of subscribers of broadband network (below 50 Mbps) and BcN (above 50 Mbps) in urban areas and rural

areas [12]. The data in Table 1 depicts that the subscription rate of rural areas is quite lower than the rate of urban areas. The gap between two subscription rates could be even wider than the gap shown in the table, because the urban areas of the survey do not include the 6 metropolitan cities such as Seoul in Korea. The survey also identifies that the rate of BcN subscription in small rural towns (of less than 50 households) is almost zero [12].

Table 1. The subscription status of network services in Korea [12] (unit: thousand people)

	BcN (above 50Mbps)		Below 50Mbps		Total	
	Subscribers	Household	Subscribers	Household	Subscribers	Household
Urban	4,722	8,479	598	1,012	5,320	9,491
Rural	1,020	1,805	613	919	1,633	2,724
Total	5,742	10,284	1,211	1,931	6,953	12,215

Carriers are actively promoting the construction of broadband networks in rural areas, especially in apartment complexes of high residential density, but they are reluctant to invest in areas of small size or medium size where residential densities are too low to regain the investment costs quickly. In particular, broadband networks do not exist in areas of less than 50 households, in which carriers avoid investing due to lack of economic feasibility. Only 64 villages out of villages of less than 50 households have been equipped by broadband networks in 2010. All of those villages are in Busan Metro-City and Daegu Metro-City.

The construction of high-speed networks has contributed to narrowing the information access gap between rural residents and the general public, but the gap of skills for using Internet services and the Internet usage gap have rarely narrowed.

The digital divide index (DDI) is a measurement of information access gap, which is calculated by subtracting the informatization score (IS) of rural residents from that of the general public (supposed as 100). Table 2 shows the digital divide indices of various social groups [12]. The senior citizens refer to the people over 50 years old, and the average is a weighted average based on the size of each social group. Considering the population distribution of rural areas, the survey confirms that the digital divide gap is as serious as we conjecture.

3 Rural BcN Development Plan

When the early version of the rural BcN project was made in 2009, a feasibility study on the project was conducted. The result of the feasibility study was reflected in the plan called “Mid- and Long-term Development Plan for Broadcast & Communication Networks” in 2009 [7]. In 2010, the panel on the construction of smart infrastructure in rural areas has been formed, and the project has begun.

The broadcast-communication convergence network in rural areas is needed to lay a foundation for universal broadcast-communication convergence services by expanding broadband broadcast-communication networks, which are concentrated in cities, into rural areas. Constructing BcN in rural areas will help solve rural areas’ economic,

social and cultural issues and vitalize the economy and industry through the diffusion and spread of specialized services based on broadband broadcast-communication networks in the areas. It is also required to establish a national cooperation system integrating ministries, local governments and carriers in order to construct broadband broadcast-communication networks and expand broadcast-communication convergence services.

Table 2. The Digital Divide Indices of different social groups [12]

	2004	2005	2006	2007	2008	2009
Handicapped	58.9	50.0	39.0	36.6	34.0	27.8
Low income	50.1	41.7	32.9	32.4	30.9	25.3
Rural	81.7	75.0	70.9	69.5	66.8	64.8
Senior	82.3	76.7	67.6	66.3	65.5	63.0
Average	72.5	65.8	57.1	55.5	54.3	51.1

In order to construct broadband networks and create an environment for increasing the use of broadcast-communication convergence services, executive plans for rural BcN have four sub-goals: network infrastructure, service accessibility, customized services, and political and economical support.

Broadband networks in small rural villages (less than 50 households) will be built by 2016. The financial support will be given by the government, and the central government, local governments, and carriers share the costs in the ratio of 1:1:2. For the broadband network construction in medium-sized rural villages (more than 50 households), the efficiency of the broadband network will be increased. The construction of broadband networks in medium-sized areas (50-240 households) will be completed by 2012. This construction will be supported through efficient policies, for example, inducements to carriers to divide construction projects among themselves or construct networks jointly. Broadband networks in remote islands and isolated areas will be built, too. Remote islands and isolated areas are considered as places where new wired networks cannot be constructed. For these areas, a foundation to provide broadcast-communication convergence services will be made by considering and applying various alternatives.

The use of wireless networks such as 4G will be tested for rural areas. Building Test beds for wireless ultra wide-band networks in small rural areas that have no broadband networks will be built, and the possibility that they will replace fixed broadband networks will be verified. The government is trying to rent the optic fiber cables of Korea Electric Power Corporation in rural and isolated areas can be considered as another option, but it requires the governmental involvement in the process of mediation and negotiation. Pilot projects will be promoted to verify other wireless technologies such as Wi-Fi Mesh, WiBro. Services through a two-way satellite will be provided for areas where other alternatives cannot be applied.

In order to encourage service carriers to construct broadband networks efficiently, a supportive legislation for carriers will be enacted. Carriers will be encouraged to voluntarily invest in building broadband networks in non-competitive rural areas by

reforming laws and regulations that deal with the construction of broadcast- communication infrastructure and developing a support system.

Policy and institutional supports for facilitating the use of services will be provided. The use of services will be facilitated by organizing and operating an effective body that promotes and supports the projects, and developing a reasonable and affordable tariff plan for specialized services for rural areas. The broadcast communication industry's ecology to boost the local industry in rural areas will be established. A basis for sustainable development of rural areas will be founded, and the opportunity to grow together will be offered to the related industries by establishing an ecological system that promotes a virtuous cycle.

4 Strategies for Service Applications

We surveyed the local residents of target rural areas on the IT services they want most [13]. Table 3 shows the survey result of different social and gender groups. The feasible IT services are divided in 5 categories: welfare, safety, education, commerce, and culture. The survey result shows that the welfare services are requested most by the rural residents. The welfare services are asked most by the female group, the over-seventy senior citizen group, the one-man household group, and the low-income group. The safety services are requested most by the over-seventy senior citizen group. The survey shows that the younger, the households with students and the higher-income households want to get good education services. Based on the survey result, we propose the strategies for boosting the IT services of each category and the strategies for enhancing the user experience.

Boosting IT Services. We propose to supply multimedia content services with high speed and real time for commercial services. Owing to the high speed of BcN infrastructure, it is possible to provide high-quality multimedia services on N-screens. The high quality multimedia services will maximize the advertising effect of local tourist industry, and prompt the e-commerce in rural areas.

The proposed services will be as follows: vitalizing the ecommerce trading of regional products in rural areas, discovering a new business model of already informatized villages, expanding the experiential and refresh tourist industry in rural areas, and increasing the advertisement of rural life and areas. The proposed services will be as follows: vitalizing the ecommerce trading of regional products in rural areas, discovering a new business model of already informatized villages, expanding the experiential and refresh tourist industry in rural areas, and increasing the advertisement of rural life and areas.

We also propose to provide multimedia educational services on N-screens with high quality and fast delivery. We think that the educational services provided by BcN infrastructure will increase the opportunity of education and the efficiency of education in rural areas. The educational services can be categorized by the target groups: supplementary educational contents for the students in rural areas, language training courses (especially Korean language courses) for multi-cultural households, continuing education services for senior citizens, and new agricultural technology courses for farm owners.

For community union, we propose to invigorate the online communities of rural areas. We call it e-Hometown plan. Every village is encouraged to open its own online community at a popular portal site such as Daum in Korea. The online community can be used in order to unify the members of the village, and to advertise the village to the general public. We insist that the online communities of rural villages will play the role of gate between urban areas and rural villages. The online community will be the place of information exchange which is essential to whom are planning to go and settle down in rural areas.

Table 3. Preferred IT Services of Different Social Groups [13]

		Cases	Welfare	Safety	Education	Industry	Culture
Total		(1000)	30.4	20.7	18.2	18.0	12.7
Sex	Male	(498)	27.7	19.9	18.9	19.9	13.7
	Female	(500)	33.1	21.5	17.5	16.1	11.8
Age	< 50	(401)	19.2	19.0	31.4	18.0	12.5
	50 – 59	(198)	27.3	17.7	14.1	24.7	16.2
	60 – 69	(201)	40.8	22.9	7.0	14.9	14.4
	> 70	(200)	45.5	25.0	7.0	14.5	8.0
Job	Agricultural	(705)	31.8	20.0	14.8	20.6	12.9
	Students/ Housewives/None	(144)	39.5	16.7	18.4	12.3	13.2
	Etc.	(181)	19.3	26.0	31.5	11.6	11.6
Number of Housholds	1	(97)	54.6	21.6	5.2	13.4	5.2
	2	(419)	33.4	23.9	11.2	17.2	14.3
	3	(178)	27.0	20.2	16.3	19.7	16.9
	4	(182)	20.9	16.5	33.0	20.9	8.8
	> 5	(124)	20.2	16.1	33.1	17.7	12.9
Family Type	Living Alone	(105)	53.3	20.0	5.7	15.2	5.7
	With Students ⁽¹⁾	(288)	24.0	16.7	29.2	18.8	11.5
	Multi-Cultural ⁽²⁾	(20)	25.0	10.0	35.0	10.0	20.0
	(1)+(2)	(3)	33.3	0	66.7	0	0
	Etc.	(584)	29.6	23.3	14.2	18.5	14.4
Income	< 100 KW	(315)	42.9	23.5	8.6	13.3	11.7
	< 200 KW	(392)	27.0	19.6	19.4	20.2	13.8
	> 200 KW	(293)	21.5	19.1	27.0	20.1	12.3
Family in long distance	O	(842)	32.8	20.4	16.0	18.6	12.1
	X	(158)	17.7	22.2	29.7	14.6	15.8
Size of Village	> 50 households	(230)	30.9	20.9	25.2	12.2	10.9
	< 50 households	(770)	30.3	20.6	16.1	19.7	13.2

We propose two feasible welfare services to rural residents; for senior citizens, u-Health services should be provided as soon as possible. To make u-Health services work, the related laws and regulations should be amended first. The survey showed that the welfare services are requested most by rural residents. The other service is the

day-care system for infants and children. With the BcN infrastructure of high speed and fast delivery, the day-care services with remote monitoring can be supported in rural areas. Or more educational multimedia contents can be sent to day-care centers through IPTV services.

For the safety services, we propose the disaster monitoring service and the history recording service of agricultural products. It is possible to monitor the farms, local roads, and chronicle disaster areas through CCTV in real time, owing to the BcN infrastructure with high speed and fast delivery. The history recording service of agricultural products will be used to increase the reliability of regional products and make the products more profitable. We think that the support of the government to build the history recording system will boost the local economy evidently.

Enhancing User Experience. With the recent transition from IPTV to Smart TV, the user interface (UI) of services is considered to be important than ever. So, the design of user interfaces becomes the key to the success of the services. The user interfaces of current services are provider-oriented: menus are complicated and not easy to use. Remote controllers are not convenient because they have too many functions and buttons.

Considering that the population of rural areas is getting old faster than before, we propose to develop a new user interface model for rural residents. For the visual user interfaces, we propose to simplify them and to enhance the accessibility. The control user interfaces must be also made simple and easy to access. The consistency and the uniformity of user interfaces should be enhanced. For this purpose, we suggest that the governmental support for developing a design model must follow because of the need of UI standards.

We think the visual user interface must be designed with intuitive menus, reflecting the cognitive and physical ability of the users and usage patterns. We propose that the user interface of controller must have less buttons and recognizable button labels. The controllers must provide tactile responses, less fatigue and familiar appearance. The user interface of provided services should be amended. The recent field research shows that senior citizens have trouble in browsing and selecting IPTV contents because of the complicated user interface of IPTV. We recommend developing a system of recommending contents to the audience according to their age, sex and social groups.

We also propose a service interface which enables family members or friends to recommend service contents using social networking services (SNSs). With this service interface, family members or friends can recommend or reserve some contents in advance, and share their feelings after watching the contents. The power of social networking services is expected to be powerful especially to senior citizens. Sharing the TV contents with their family members or friends will let senior citizens feel less isolated. SNS will help senior citizens get information for daily lives, too. At the same time, it will help senior citizens response emergencies more quickly.

5 Conclusion

The digital divide usually refers to the gap between people with effective access to digital and information technology and those with very limited or no access at all. It

includes the imbalance both in physical access to technology and the resources and skills needed to effectively participate as a digital citizen. The low speed of network makes broadcast-communication convergence services in rural areas difficult, and it leads to the digital divide between cities and rural areas, which will emerge as a social issue. The Korea government has set up the BcN establishment project to provide high-speed network convergence services such as IPTV and VoIP to sparsely-populated rural areas, and conducted the project since 2010.

In the paper, we introduced the rural BcN project of the Korea government, which aims at accelerating the network infrastructure of rural areas. We proposed the strategies for boosting broadcast-communication services and the strategies for enhancing user experience in rural areas. The suggested boosting strategies direct to two directions: enhancing user accessibility to the services and developing specialized services.

The rural BcN project of the Korea government is under progress now, and it will become complete by 2016. We believe that developing a good assessment model for the development plan and its achievement is very important, and leave it as a future work.

References

1. Bargh, J.A., McKenna, K.Y.A.: The Internet and Social Life. *Annual Review of Psychology* 55, 573–590 (2004)
2. Compaine, B.M. (ed.): *The Digital Divide: Facing a Crisis or Creating a Myth?* MIT Press, Cambridge (2001) ISBN 0262531933
3. Ministry of Information Communication of Korea: Basic blueprint for building the Broadband convergence Network, MIC, Tech. Rep. (2004)
4. Ministry of Information Communication of Korea: Second phase plan for establishing BcN, MIC, Tech. Rep. (2006)
5. Korea Communication Commission: Third phase plan for establishing BcN, Korea Communication Commission, Tech. Rep. (2008)
6. Lee, E.: Plans and Strategies for UBcN Networks and Services. *Journal of Information Processing Systems* 6(3) (2010)
7. Korea Communication Commission: Directions for BcN infrastructure of rural areas, Korea Communication Commission, Tech. Rep. (2009)
8. Korea Communication Commission: Project plans for promoting BcN construction in rural areas, Korea Communication Commission, Tech. Rep. (2010)
9. Korea Communication Commission: Long-term plan for constructing broadband networks in rural areas, Korea Communication Commission, Tech. Rep. (2010)
10. National Information Society Agency of Korea: Broadband convergence Network annual report 2007, MIC, Tech. Rep. (2007)
11. Korea Communication Commission: UBcN deployment policy, Korea Communication Commission, Tech. Rep. (January 2009)
12. National Information Society Agency of Korea: Mid- and Long-term Plan for Constructing Broadband Networks in Rural Areas (2010)
13. National Information Society Agency of Korea: A Research Study on Broadband Networks in Rural Areas (2010)

Enlarging Instruction Window through Separated Reorder Buffers for High Performance Computing^{*}

Min Choi¹, Jonghyuk Park², and Young-Sik Jeong^{1,**}

¹ Wonkwang University, Iksan, Republic of Korea

{mchoi, ysjeong}@wku.ac.kr

² Seoul National University of Science and Technology, Seoul, Republic of Korea

parkjonghyuk1@hotmail.com

Abstract. Modern microprocessors achieve high application performance at the acceptable level of power dissipation. In terms of power to performance trade-off, the instruction window is particularly important. This is because enlarging the window size achieves high performance but naive scaling of the conventional instruction window can severely increase the complexity and power consumption. In this paper, we propose low-power instruction window techniques for contemporary microprocessors. First, the separated reorder buffer (SROB) reduces power dissipation by deferred allocation and early release. The deferred allocation delays the SROB allocation of instructions until their all data dependencies are resolved. Then, the instructions are executed in program order and they are released faster from the SROB. This results in higher resource utilization and low power consumption.

Keywords: Low power, Reorder buffer, Instruction Window, High performance computing, Separated reorder buffer.

1 Introduction

Enlarging the size of instruction windows can lead to performance improvement. However, naive scaling of the conventional reorder buffer severely affects the complexity and power consumption. In fact, Folegnani and Gonzalez [1] showed that the reorder buffer is the most complex and power-dense parts in dynamically scheduled processors. Thus, much research has been conducted to increase the size of the reorder buffer without negatively impacting power consumption.

In this context, we propose a novel technique for reducing power dissipation and improving performance of the reorder buffer (ROB). Our proposed method, called separated reorder buffer (SROB), is distinct from other approaches in that we achieve early release without depending on any checkpointing. This feature gives us good performance with relatively low power dissipation. In this paper, we introduce a sepa-

^{*} This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0025748 and 2010-0022589).

^{**} Corresponding author, ysjeong2k@gmail.com

rated architecture for reorder buffer. First, we focus on the fact that many instructions waste ROB resources without doing any useful work during data dependency resolution. To reduce such wasteful resource usage, we introduce the novel concept of reorder buffer, named as separated reorder buffer (SROB). The SROB structure consists of two parts which are in charge of dependent instructions and independent instructions, respectively. For the dependent instructions, our SROB architecture executes the instructions in program order and releases the instructions faster. This results in higher resource utilization and low power consumption. The power reduction stems from deferred allocation and early release. The deferred allocation technique inserts instructions into the SROB only after fulfilling the data dependency. The SROB releases instructions earlier immediately after the execution completes, because precise exception is trivial under in-order execution. Second, in order to deal with the power problem on issue queue, we focus on a well known fact that the vast majority of instruction dependency exists within a basic block. In practice, a basic block is comprised of about 6 instructions on average [22].

The rest of this paper is organized as follows. Section 2 presents a brief review of the existing approaches. Section 3 describes our modified reorder buffer architecture, the SROB, and the concept of deferred allocation and early release. We evaluate its performance and power consumption in Section 4. Finally, we conclude by summarizing our results in Section 5.

2 The Separated Reorder Buffer

In general, the function of the reorder buffer (ROB) is to put the instructions back into the original program order after the instructions have finished execution possibly out of order. The ROB maintains an ordered list of the instructions and takes into account recovery and precise exception. Conventionally, the instructions are inserted into both the issue queue and the ROB. The instructions stay in the ROB until the instruction commits. As soon as the dependency of an instruction is fulfilled, the processor executes the instruction sequentially in program order. In addition to the ordinary ROB, we propose the concept of the separated ROB (SROB), as depicted in Figure 1.

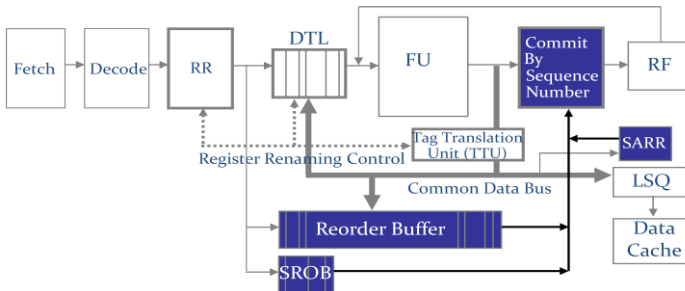
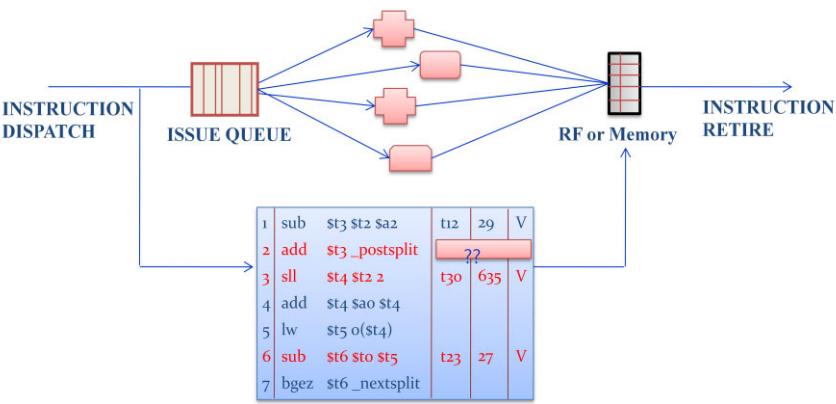


Fig. 1. Pipeline architecture model

Figure 1 shows the overall pipeline architecture in which the colored components represent the modified (or newly added) parts in this work. For load and store in

structions, they are assigned to entries in load-store queues (LSQ). Instructions leave the instruction queue when they are issued, and free their reorder buffer entries when they commit. Reorder buffer holds the result of an instruction between the time the operation associated with the instruction completes and the time the instruction commits.



The functional units (FU) can execute an operation of a certain type. The system retrieves the operands from register file (RF), and stores the operands into the register file. The stand-alone rename registers (SARR) are split register file to implement the rename buffers.

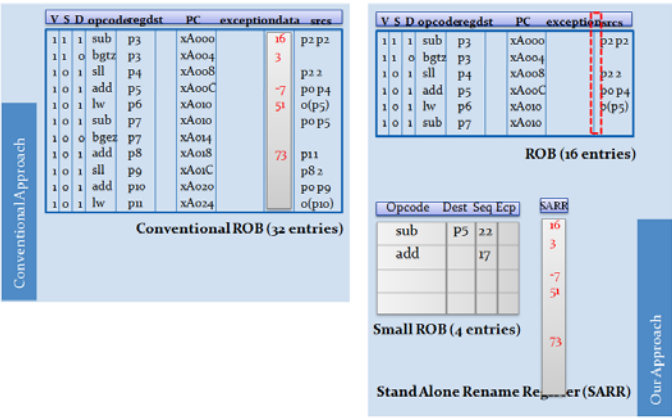


Fig. 2. Structural difference between ordinary ROB and SROB

Usually in conventional reorder buffer, the rename registers are integrated into re-order buffer. Each entry in the SROB has the same structure as the ordinary ROB, but in the SROB architecture the rename registers are stand-alone, so we named that as SARR. The structural difference of separated ROB is as shown in Figure 2. Each separated part of the SROB manages dependent and independent instructions, respectively. One part of SROB processes control instructions, independent instructions, and

load/store instructions. Another part of SROB is in charge of dependent instructions. The execution of dependent instructions is serialized inherently by true dependency. The dependent instructions will wait for a long time to resolve their data dependencies, even if we put the dependent instructions into the general ROB.

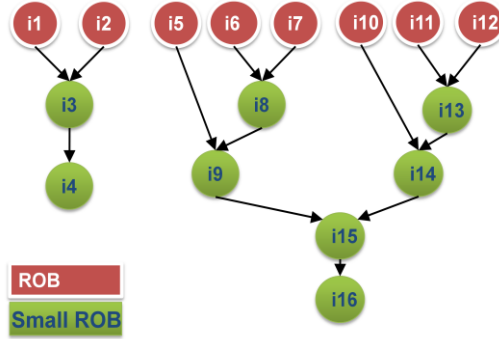


Fig. 3. Instruction allocation on separated ROB

Figure 3 shows the example of instruction allocation in the SROB architecture. The instructions waiting in the ROB do not any useful work and severely affect the power consumption and the instruction level parallelism (ILP). This is because the ROB is a complex multi-ported structure and represents a significant source of power dissipation. Moreover, if the dependent instructions are in a long dependency chain, power and performance problem gets worse.

2.1 Deferred Allocation and Early Release

In order to resolve the power and performance problems, we prevent dependent instructions from moving through the ROB at dispatch time. The instructions wait for issue on the instruction queue, not on the ROB. After the instruction dependency is fulfilled, the instructions can go to the SROB. As a result, one instruction of a dependency chain executes in the SROB at a time naturally as shown in Figure 4.

We call this the deferred allocation feature of the SROB. Moreover, the instructions in the SROB are released earlier and the result of the instruction is written into rename buffers immediately after the execution completes. Then, the result values in the rename buffer are written into the architectural register file at the commit state. Since the instructions in the SROB are executed in program order, we need not maintain the order of instructions and thus we have to take the results only. For implementation of the deferred allocation, we need to check whether an instruction is in a certain dependency chain or not. However, facilitating such a hardware checker causes complexity at the front-end. So, we take a straightforward approach to realize a simple instruction classification at the decoding stage. Our classifier checks only the operand availability of each instruction. If operands are available, the instruction is independent. Otherwise, the instruction is dependent and it thus goes to the SROB. This classification mechanism is very simple, yet able to capture in a uniform way all the dependency chains through a given microarchitectural execution of a program.

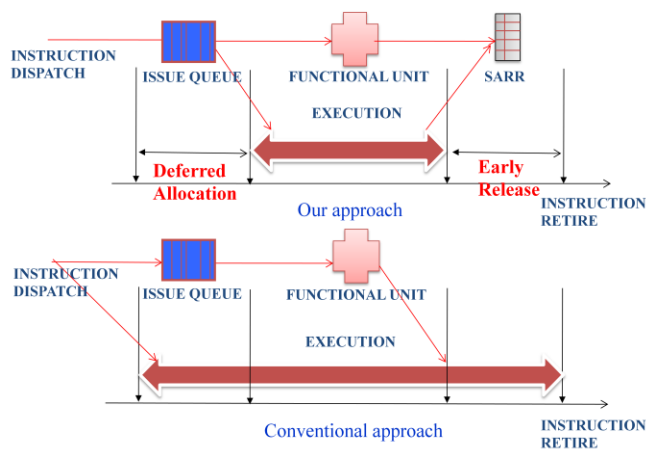


Fig. 4. Resource usage between SROB and conventional ROB

3 Experimental Results

All tests and evaluations were performed with programs from the SPEC2000 CPU benchmark suite on Sim-Panalyzer [6]. The Sim-Panalyzer is a cycle accurate and architecture level power simulator which is built on the SimpleScalar simulator. The Sim-Panalyzer lumps the issue queue, the reorder buffer, and the physical register file into a register update unit(RUU). In order to better model the power consumption of contemporary microprocessor architecture, we split the RUU into the reorder buffer and the issue queues.

Table 1. Simulation parameters

	Baseline	SROB Method
fetch:speed	4/16	4/16
decode:width	4/16	4/16
issue:width	4/16	4/16
commit:width	4/16	4/16
rn:size	32	32
rob:size	8/16/32/64	8/16/32/64
srob:size	-	4/16
ialu, imult, fpalu, fpmult	8, 2, 2, 2/8, 8, 8, 8	8, 2, 2, 2/8, 8, 8, 8
Branch Prediction		
bpredictor	2lev. Bimodal, 2K	2lev. Bimodal, 2K
btb	128, 4-way	128, 4-way
return address stack	8 entry	16 entry

To evaluate the performance of the SROB architecture, we use the Alpha 21264 architecture as the baseline platform. The Alpha is an out-of-order-issue microprocessor that can fetch and execute up to four instructions per cycle. It also features dynamic scheduling and speculative execution to maximize performance. The Alpha

pipeline contains four integer execution units. The two of the integer execution units can perform memory address calculations for load and store operations. The 21264 pipeline also contains two floating-point execution units to perform add, divide, square root, and multiply functions. The 21264 pipeline has 7 stages which consist of instruction fetch, branch prediction, register renaming, instruction issue, register access, execution and writeback. The architecture parameters used in our Sim-Panalyzer simulations are listed in Table 1.

The `rn:size` is for adjusting the range of register renaming. It indicates how many physical registers are mapped to logical register names. Without register renaming, running a binary executable compiled for 32 registers on 64 register machine will repetitively make use of first 32 registers only. This is because the renamed register tag is used as an index to lookup the IW in DLT architecture. This technique avoids recompilation overhead when a binary executes on different architecture in terms of physical register size. The `srob:size` configures the size of the SROB buffer. The reason we set this parameter as 4 is to make the SROB size equal to the issue/commit bandwidth. If the size is more or less than the bandwidth, it may result in performance bottleneck or resource waste.

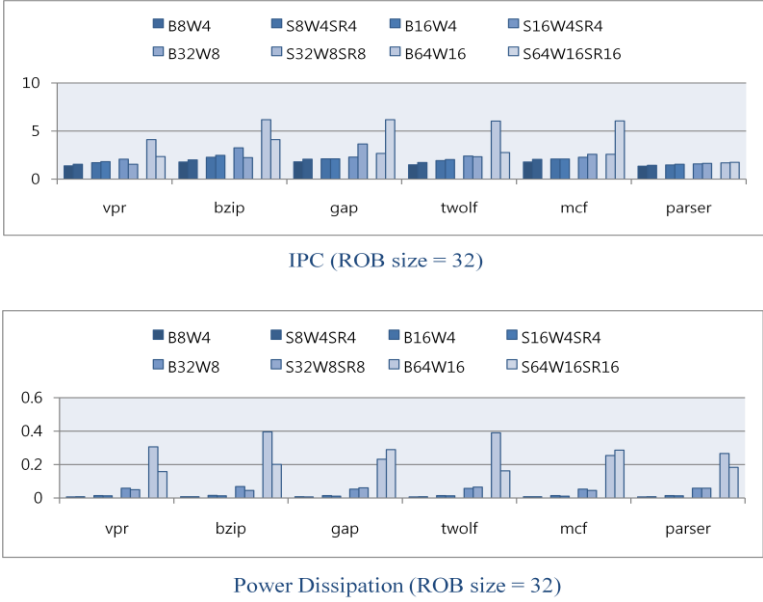


Fig. 5. Evaluation on 32, 64, 128 entry ROB configuration

The top half of Figure 5 shows an average of IPC attained by SpecInt applications in simulations. The results are normalized to the baseline values. The performance degradation is due to the SROB contention. The exception is that apsi delivers even better performance while maintaining an effective power consumption level (4.9% less than the baseline power). The bottom half of Figure 5 represents the evaluated power dissipation. The SROB method achieved power reduction to 11.2% of baseline power. The power reduction stems from deferred allocation and early release in the

SROB. The power savings come with a performance penalty of only 3.7% on average. We note that power saving of the 11.2% is not total system savings, but a portion of the total system savings. The savings only applies to the power saving in the ROB unit. However, the overall power savings in the perspective of total system are not negligible. This is because the ROB consumes the most significant amount of energy among all structures. In fact, it takes 27.1% of total system power dissipation. At the same time, we achieved power reduction of the ROB unit to 11.2%. Therefore, the overall power savings in the perspective of total system are 3.04%.

4 Concluding Remarks

The separated reorder buffer (SROB) reduces power dissipation by deferred allocation and early release. These two techniques result in higher resource utilization and low power consumption. Therefore, up to 3.04% of power saving comes with an average of only 3.7% performance penalty. In current version of implementation, we limited the role of the SROB to process only dependent instructions. The power saving will be much increased if the SROB approach is extended to all types of instructions as future work. Even though there is a little performance penalty, our SROB technique for reducing the power dissipation is still meaningful, especially on the embedded computing. In the embedded environment, the energy saving is the most critical due to the limited battery capacity.

References

1. Folegnani, D., Gonzalez, A.: Energy-Effective Issue Logic. In: The Proceedings of the IEEE International Symposium on Computer Architecture, ISCA (2001)
2. Cristal, A., Santana, O., Cazorla, F., Galluzzi, M., Ramirez, T., Pericas, M., Valero, M.: Kilo-Instruction Processors: Overcoming the Memory Wall. *IEEE Micro* (2005)
3. Kirman, N., Kirman, M., Chaudhuri, M., Martinez, J.: Checkpointed Early Load Retirement. In: Proceedings of the International Symposium on High-Performance Computer Architecture, HPCA (2005)
4. Martinez, J., Renau, J., Huang, M., Prvulovic, M., Torrellas, J.: Cherry: Checkpointed Early Resource Recycling in Out-of-Order Microprocessors. In: Proceedings of the IEEE International Symposium on Microarchitecture, MICRO (2002)
5. Dundas, J., Mudge, T.: Improving Data Cache Performance by Pre-executing Instructions under a Cache Miss. In: Proceedings of the ACM International Conference on Supercomputing (ICS) (July 1997)
6. Mutlu, O., Stark, J., Wilkerson, C., Patt, Y.N.: Runahead Execution: An Alternative to Very Large Instruction Windows for Out-of-order Processors. In: Proceedings of the IEEE International Symposium on High Performance Computer Architecture (HPCA), pp. 129–140 (February 2003)
7. Sima, D.: The Design Space of Register Renaming Techniques. *IEEE Micro* (2000)
8. Obaidat, M.S., Dhurandher, S.K., Gupta, D., Gupta, N., Deesr, A.A.: Dynamic Energy Efficient and Secure Routing Protocol for Wireless Sensor Networks in Urban Environments. *Journal of Information Processing* 6(3) (September 2010)
9. Nan, H., Kim, K.K., Wang, W., Choi, K.: Dynamic Voltage and Frequency Scaling for Power-Constrained Design using Process Voltage and Temperature Sensor Circuits. *Journal of Information Processing Systems* 7(1) (March 2011)

Smart Mobile Banking and Its Security Issues: From the Perspectives of the Legal Liability and Security Investment

Se-Hak Chun

Department of Business Administration, Seoul National University of Science and Technology,
Kongneung-gil 138, Nowon-gu, Seoul 139-743, Republic of Korea
{shchun}@seoultech.ac.kr

Abstract. As the Internet prevails, financial transactions using mobile handsets increase with electronic fraudulent transactions at the same time. To prevent fraudulent transactions, security liability rules are needed as well as more security protection techniques. In comparison with the amount of research on technical approach, little research on legal approach has been done. This paper analyzes how a legal approach can contribute to fraudulent transaction protection and investigates how this is related to banks' investment in security.

Keywords: Mobile banking, Security technologies, Security investment, Legal liability, Fraudulent transaction.

1 Introduction

With widespread use of mobile phones, mobile payment and banking has recently spread with growth projected at 50% per year [1]. However, most customers remain skittish about mobile banking because a majority of customers do not believe that mobile banking is safe and secure [2]. According to the Online Fraud Report (2010) from 2006 to 2008 the percent of online revenues lost to payment fraud was stable and online merchants consistently reported an average loss of 1.4% of revenues to payment fraud [3]. In particular, the numerous viruses, network assaults and mobile phone thefts are threatening mobile security and increasing the amount of fraudulent transactions. Since the birth of GSM technology, security and trust have been key parts for handset manufacturers, making it easy to use for customers and creating trust in banking services. So far, the technology enablers available in most low-end handsets are SMS, typically used with SIM ATK (application toolkit), USSD (unstructured supplementary service data) and voice calls with DTMF (dual-tone multifrequency interaction). However, as smart phones and mobile banking become more widespread, existing security solutions have become quite fragmented [4]. Thus, a perspective of technology standardization would be needed to avoid further fragmentation as mobile banking proceeds to grow. With the development of smart security technologies, the governing rule on security disaster becomes a practical and key issue for mobile financial transactions. Even though mobile banking has become an enormous market,

so far there has been little research on security problems encompassing legal issues in mobile banking [2,5]. We focus on liability between service providers such as financial and wireless carriers and customers (or subscribers) from the perspective of burden of proof when fraudulent transactions occur.

2 The Model

The level of security investment may depend on how government authorities regulate. If the law is more favorable to a bank when fraudulent transactions are disputed, the bank has less incentive to invest in security. Vise versa, if the law is less favorable, the bank has stronger incentive. We focus on a bank (or financial institution)'s decision on investments in security as a decision variable because it is one of the most fundamental decisions when the bank determines the level of its security. In the U.S., when a dispute regarding financial transactions arises between the bank and a customer, providing burden of proof is on the bank. If the bank cannot prove the transaction is correct, it has to accept the customer's argument, and it has to refund money to the customer. But, in the U.K., Norway, and the Netherlands, the burden of proof is on customers [6,7]. We analyze two different legal systems regarding whether the burden of proof should lie on banks or customers when fraudulent transactions are disputed.

2.1 Case 1: The Bank Has the Burden of Proof

We assume that customers' service utilities are uniformly distributed along $[0, V]$ according to their reservation service fee, p . A customer, $v \in [0, V]$ will obtain the surplus, $U(v) = v - p$, in using the financial service and will use the service if $v - p > 0$. Thus, the demand, Q , will be $V - p$. There are two types of costs when a security accident occurs. The first type is direct security costs related to security accident or disaster, which is denoted by L . The second type is indirect costs related to procedural burden of proof costs when a bank spends to prove that the accident is not its responsibility, which is denoted by B . We assume that the direct security loss of the bank has a positive relationship with the size of the financial transaction, $p * Q$, which assumes that the potential damages or losses from security breaches are likely to increase as the size of the transaction services increase. The security loss based on the model of Gordon and Loeb [8] is determined by the probability of a security breach and the size of the financial transaction. The probability of a security breach per unit of transaction can be represented by the expected probability of an incident, $\frac{s}{1+kl}$, where I is the monetary investment of the bank in security, s is the vulnerability to exposure to a security incident when the bank does not invest in any kinds of security, and k is the efficiency of the security investment. We also assume that the proof cost, B , is a portion of the L denoted by δ . Both costs are reduced when the bank invests more in security and is assumed to have negative relationships with the amount of the investment. Thus, $\partial L / \partial I < 0$ and $\partial B / \partial I < 0$. Also, these are slowly reduced as the bank increases the investment level, thus $\partial^2 L / \partial I^2 < 0$ and $\partial^2 B / \partial I^2 < 0$. These two types of costs can be affected by other external security factors such as the initial probability of security accidents, s , and the efficiency of the security investment (or an elasticity

of investment in security level), k that can represent overall unit investment effect on the security risk, a measure of the morality or attitude of the society and security infrastructure rather than an individual bank's security level. Thus, low k means the unit effect of the investment is low, and it can represent that the social security system matures or the society may have a higher IT infrastructure level. While high k means that the unit effect of the investment is high and represents that social security system does not mature or the society has a lower IT infrastructure level because a little increase in investment can make the security risk lower. The parameter k can be a measure of the efficiency of a security investment that refers to how much the bank can reduce its security incidents with 1 unit of investment in security. Then, the bank in this case will find an optimal security investment level to maximize its profit and the profit function can be represented as below:

$$\begin{aligned}\Pi_1 &= p_1 Q_1 - L(I_1; s_1, k_1) - B(I_1; s_1, k_1) - I \\ &= p_1(V - p_1) - \frac{s_1}{1 + k_1 I_1} p_1(V - p_1) - \delta \frac{s_1}{1 + k_1 I_1} p_1(V - p_1) - I_1.\end{aligned}\quad (1)$$

From the first order condition we find an optimal service fee and a level of investment in security that maximizes the bank's profit as follows:

$$p_1^* = \frac{V}{2}, I_1^* = -\frac{1}{k_1} + \frac{V}{2} \sqrt{\frac{(1 + \delta)s_1}{k_1}}. \quad (2)$$

2.2 Case 2: The Customer Has the Burden of Proof

When the burden of proof for security accidents lies on the customer side, customers have less incentive to use online financial services. The value which customers feel when they use financial service decreases as security disaster risk increases, thus, net customer utility, U , can be written as follows:

$$U = v - p_2 - \frac{\delta s_2}{1 + k_2 I_2} p_2. \quad (3)$$

Thus, a customer whose net utility is nonnegative will use financial services and the market demand will be

$$Q_2 = V - p_2 - \frac{\delta s_2}{1 + k_2 I_2} p_2. \quad (4)$$

Thus, the bank does not consider the proof cost as follows:

$$\Pi_2 = p_2 Q_2 - L(I_2; s_2, k_2) - I_2 = p_2 Q_2 - \frac{s_2}{1 + k_2 I_2} p_2 Q_2 - I_2. \quad (5)$$

From the first order condition, we find an optimal service fee and a level of investment in security that maximizes the bank's profit as follows:

$$p_2^* = \frac{V}{2} - \frac{\delta \sqrt{s_2}}{\sqrt{(1 + \delta)k_2}}, I_2^* = -\frac{1 + \delta s_2}{k_2} + \frac{V}{2} \sqrt{\frac{(1 + \delta)s_2}{k_2}}. \quad (6)$$

3 The Optimal Security Investment and Implications

In this section, we discuss the question of why the allocation of the burden of proof for fraudulent transactions can be different across countries, cultures or IT infrastructure levels, and we discuss some implications on concerning the financial transaction law.

3.1 Why Do Different Regulations Exist?

We can obtain the banks' profits from the optimal service fee and security investment level, and then compare two different regulation regimes. For comparing two regulation effects, we set $s_1=s_2=s$ and $k_1=k_2=k$. Then, we obtained the following proposition:

Proposition 1. $\Pi_1^* \geq \Pi_2^*$ if $\sqrt{k^*} \geq \frac{\delta(\sqrt{\delta+1} + \sqrt{\delta})\sqrt{s}}{V}$.
 $\Pi_1^* < \Pi_2^*$ if $\sqrt{k^*} < \frac{\delta(\sqrt{\delta+1} + \sqrt{\delta})\sqrt{s}}{V}$.

Proof. Omitted because of simple calculation. □

Proposition 1 means that if k is high enough, the profit of the bank in case 1 is greater than in case 2 while if k is not high enough, the profit of the bank in case 1 is less than in case 2. Proposition 1 implies that if k is high, the bank may have greater profits even though the government imposes the burden of proof on the bank rather than on customers.

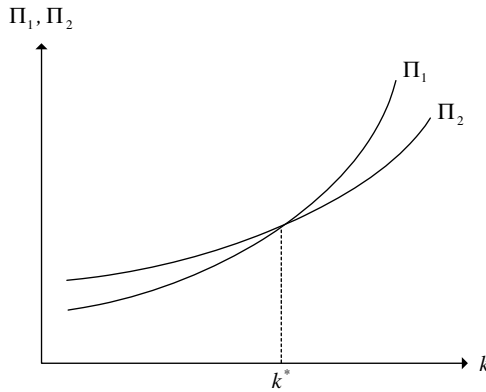


Fig. 1. Profits and k^*

Corollary 1. $\partial k^* / \partial \delta > 0$, $\partial k^* / \partial s > 0$, $\partial k^* / \partial V < 0$.

Corollary 1 means that the profit in case 2 is likely to greater than in case 1 as δ and s increase and V decreases. This implies that the government tends to impose the burden of proof on the bank when the burden of proof cost is small, the initial vulnerability is small and customer's maximum reservation fee (or potential market size), V , is large.

3.2 Security Investment Level

Security level can influence the bank's performance in various ways. For example, insecure systems may result in decrease of revenue by discouraging potential customers from using services or may cause security accidents that could lead to a huge financial loss on the part of the bank. Thus, the bank has an incentive to invest in security. To enhance the security systems, the bank needs to approach the issue from both a managerial and a technical perspective. In addition, the bank needs to understand regulations and laws. In this paper, we focus on the level of investment in security in a perspective of the electronic financial law. From the optimal security level in the previous section, we obtain the following proposition:

Proposition 2. $I_1^* \leq I_2^*$ if $\sqrt{k^{**}} \geq \frac{2\delta(\sqrt{1+\delta} + \sqrt{\delta})\sqrt{s}}{V}$.

Proof. Omitted because of simple calculation. □

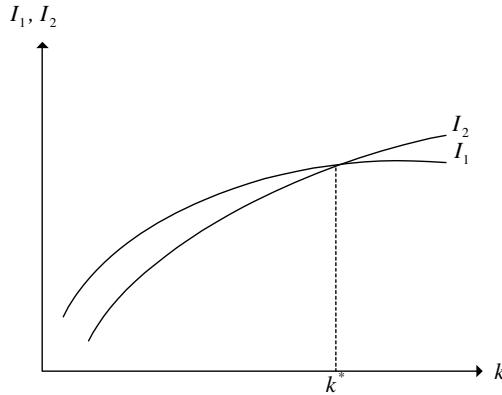


Fig. 2. Optimal Investment and k^{**}

Proposition 2 is closely related to the first proposition. This means that the bank needs to increase investment in security in a situation that the government enforces the burden of proof on the bank if k is not low because $k^* < k^{**}$. It is obvious that if the efficiency of the security investment is not low enough, the bank should invest more in the security level even if the burden of proof is on the bank. However, the bank hesitates to invest in security even in a situation that k is high when the government enforces the burden of proof on the bank side. Corollary 1 and proposition 2 imply that the bank needs to invest more in security when the burden of proof cost and the initial vulnerability are small and the maximum reservation fee, V , is large. Also, it implies that within some areas of k ($k^* < k < k^{**}$), the bank has an incentive to invest more in security to attain better profitability even though the government imposes the burden of proof on the bank. This is in line with the fact that recently, the U.S. government tends to force banks to spend more on security and privacy levels.

4 Conclusion

There are many studies that address banks' security issues qualitatively, either with a technical approach or with a managerial approach. However, there are few studies that provide the optimal level of banks' investment in security using analytical models. This study contributes to an understanding of the situation of lawmakers and banks by investigating the emerging issue of fraudulent transactions. Few studies have investigated the strategic investment of banks in security by considering the legislative conditions of countries. The results of this study can help lawmakers to understand the consequences of the two alternative regulations and can help banks to understand the impact of the financial transaction regulations.

Acknowledgments. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-220-B00006).

References

1. Vyas, C.: From Niche Play to Mainstream Delivery Channel: US Mobile Banking Forecast, 2008-13. Tower Group (May 2009)
2. Joyce, F.M.: Mobile Banking Liability: The Elephant in the Parlor. *The Innovator* 3(3), 29–32 (2010)
3. <http://www.ecommercetimes.com/story/2771.html?wlc=1279873107>
4. Nokia and Nokia Siemens Networks: Mobile phones can bring banking within everyone's reach. *Expanding Horizons* 1 (2008)
5. Chun, Se-Hak, Kim, J.-C., Cho, W.: Who is responsible for the onus of proof on online fraud transaction? Korean Conference of Management Information System, Seoul Korea Fall (2007)
6. Anderson, R.: Why Cryptosystems Fail. *Communications of the ACM* 37(11), 32–40 (1994)
7. Anderson, R.: Why Information Security is Hard. University of Cambridge, working paper (2002)
8. Gordon, L.A., Loeb, M.P.: The economics of information security investment. *ACM Trans. Inform. System Security* 5(4), 438–457 (2002)

Future Green Technologies and Performances: A Case Study of the Korean Stock Market

Se-Hak Chun¹ and Seonmin Kim^{2,*}

¹ Department of Business Administration,
Seoul National University of Science and Technology,
Kongneung-gil 138 Nowon-gu, Seoul 139-743, Republic of Korea
shchun@seoultech.ac.kr

² Department of Business Administration,
Seoul National University of Science and Technology,
Kongneung-gil 138 Nowon-gu, Seoul 139-743, Republic of Korea
skim@seoultech.ac.kr

Abstract. Nowadays, for sustainable economic development, many countries or companies are investing in green technologies to reduce CO₂ and create new revenue streams. This paper evaluates the performances of the green technologies comparing them with those of other technologies using the Korean stock market. Using a regression framework, we also investigate how the performances of green technologies are related to some factors such as types of technology, price competitiveness and investment timing.

Keywords: Sustainable economy, Green technology, Green computing, Performance measure, Investment timing.

1 Introduction

Global warming makes the public aware of problems such as climate change, peak-oil, high volume of greenhouse gas emissions, loss of biodiversity, potable water pollution, growing landfill areas, and population pressure [1]. The definition of Green contains the notion of “helping to sustain the environment for future generations.” Many leaders understand that sustainability is now a critical part of the core value of the company because returns from launching green projects like positive cash flow, reduced energy, material, and operating costs can make or break a company today [2]. Governments and transnational institutions, such as the European Union and the United Nations, have stated the need of concerted actions worldwide to tackle the problem of environmentally sustainable development and have also emphasized the role that scientific research has in advancing knowledge for this goal [2]. Also, many countries are investing in green technologies to cope with environment problems and to seek future revenue streams towards economic growth. For example, the Korean government plans to expand its R&D investment in green technologies from 2 trillion won in 2009 to 3.5 trillion won by 2013, making a cumulative amount of 13 trillion

* Corresponding author.

won as Korea’s greenhouse gas emissions almost becomes doubled between 1990 and 2005, the highest growth rate in the OECD area [3].

This paper evaluates the performances of the green technologies comparing them with those of other technologies using the Korean stock market. Using a regression framework, we also investigate how the performances of green technologies are related to some factors such as types of technology, price competitiveness and investment timing. The paper is organized as follows. In Section 2, we describe the data used in the study. In Section 3, we show the results of our analyses. Section 4 concludes this study.

2 Data

We collected data on stock prices from the largest portal site in Korea (<http://finance.naver.com>). We chose the starting date as January 2, 2009 because 27 core technologies (Table 1) were announced in January 2009 as new growth engines for Korea, and the Korean government planned to focus on them as R&D investment areas [3].

Table 1. Selected sample technologies among 27 core green technologies (January 2, 2009 – March 30, 2011)

Sector	Selected green technologies	Sample companies from KOSPI	Competitiveness	Invest Timing
Energy source technology (input)	Non-silicon based solar cells	KCC/WoongJin H/STX engine/KC Green H /TaiHan Elec.Wire/ OCI/KISCO H/Sinsung H	High	Long-term
	Bio-energy	Daewoo Construction	Medium	Long
Technologies to improve Efficiency (Green IT)	LED	Samsung E/LG Innotek /SamYoung E/Kwang E	Low	Short
	RFID	IIJin E/LG Sanjun/Sinsekye I&C/DongYang Sys	Low	Short
	Cloud Computing	DawooTech	Low	Short
	Smart grid	IIJin E/LG Sanjun/ SamHwa C/SunDo E/IIJin H /KEPCO/LS	Medium	Long
Technologies to improve Efficiency (Logistics)	Green Cars (Hybrid Cars)	SamHwa E /SamHwa C /SamHwa E/DaiHo AL	High	Medium
	Green Cars (Electronic)	SK Inovation/LG Chem. /LS Sanjun/Kia Motors/Hyudai Mobis/Hyudai Motors/Samsung SDI	High	Medium
	Green Buildings	LSSanjun/KyunDong /KCC	High	Long
End-of-pipe technology	Water quality/ management	Hankook JCK/WoonJin Chem./DongYang CK	Low	Medium
Final sample size		42		

Table 1 shows selected sample technologies among 27 core green technologies with its price competitiveness against other countries and the government’s R&D investment timing. We analyze how related factors can affect the performance of green technologies. For this, we categorized 4 types of green technologies as stages of product life cycle: input, process for direct energy efficiency, process for macro level and end-of pipe stages. According to the Korean government’s plan of investment timing we categorized three types of investment time: short term, medium term and long term. We also used price competitiveness information from Korean government’s reports [4,5]. Table 2 shows average rates of returns for each categories of green technologies from Jan 2, 2009 to March 30, 2011.

Table 2. Green Technologies’ Rate of Returns (Jan 2, 2009 ~ March 30, 2011)

Sector	Selected green technologies	No. of companies	Rate of Return
Energy source technology (input)	Non-silicon based solar cells	8	115%
	Bio-energy	1	121%
Technologies to improve Efficiency (Green IT)	LED	4	279%
	RFID	4	184%
	Cloud computing	1	189%
	Smart grid	7	155%
Technologies to improve Efficiency (Logistics)	Green Cars(Hybrid Cars)	4	247%
	Green Cars(Electronic)	7	455%
	Green Buildings	3	143%
End-of-pipe technology	Water quality/management	3	98%
Total		42	216%

3 Results

Table 3 shows the rate of return on core green technologies, direct energy related technologies and other general companies and compares the rate of return on core green technologies with those on others.

Table 3. Green Technologies’ Rate of Returns (Jan 2, 2009 ~ March 28, 2011)

Sector	Green technologies vs. General Energy	Green technologies vs. Total companies(KOSPI)
Rate of return	216% vs.293%	216% vs.177%
# of companies	42 vs. 13	42 vs. 829
t-value	-2.701	1.374
Decison	Reject H_0	Accept H_0

Table 3 show that the core green technologies does not significantly outperform the general companies while the direct energy related technology outperformed the green core technologies and other general companies.

Next, we investigate how performances of green technologies can be affected by other factors such as stage of technology, price competitiveness and investment timing. For this, we assume a regression framework as follows:

$$\text{Green Performance} = \alpha + \beta_1 * \text{Stage of Tech} + \beta_2 * \text{Competitiveness} + \beta_3 * \text{Timing} + \varepsilon$$

α : constant β : standardized coefficient ε : residuals

Table 4 shows the results of the regression of green performance on stock prices and three variables. The dependent variable is the green performance. Table 4 indicates that two factors such as price competitiveness and investment timing are statistically significant at level $p < 0.01$ and the stage of technology is not significant.

Table 4. Regression Results of Green Technology Performance

	Coefficient	Standard error	t-statistic	P-value
Constant	1.958	1.204	1.627	.112
Stage of Technology	.299	.307	.973	.336
Price Competitiveness	1.057	.376	<u>2.815</u>	<u>.008</u>
Investment Timing	-1.269	.436	<u>-2.910</u>	<u>.006</u>
$R^2=0.266$ $F=4.579$ Signif $F=0.008$				

4 Conclusion

There is a growing concern worldwide about sustainable development. Moreover, there are many open challenges for research in this area that deserve the attention of the research community. In this study, we examined the performances of the green technologies comparing them with those of other technologies using the Korean stock market and investigate how they are related to some factors such as types of technology, price competitiveness and investment timing. A promising direction for the future is to investigate these arguments in greater detail and find other factors affects on green performances.

References

1. Wills, B.: The Business Case for Environmental Sustainability (Green). A 2009 HPS White Paper (2009)
2. OECD. Towards Green ICT Strategies: Assessing Policies and Programmes on ICT and the Environment (2009)
3. Jones, R.S., Yoo, B.: Korea’s Green Growth Strategy: Mitigating Climate Change and Developing New Growth Engines. OECD Economics Department Working Papers. No. 798, OECD Publishing (2010)
4. <http://www.mest.go.kr>
5. <http://www.greengrowth.go.kr/www/policy/skill/skill1.cms>

Corporate Social Responsibility and Its Performances: Application to SRI (Socially Responsible Investing) Mutual Funds

Joon Ho Hwang¹, Dong Han Kim¹, and Se-Hak Chun^{2,*}

¹ Korea University Business School, Anam-dong, Seongbuk-gu, Seoul, Korea 136-701
joonhwang@korea.ac.kr, openletters@naver.com

² Department of Business Administration,
Seoul National University of Science and Technology,
172 Gongreung 2-dong, Nowon-gu, Seoul, 139-743
shchun@seoultech.ac.kr

Abstract. We compare the performance of the SRI (Socially Responsible Investing) mutual funds and that of regular mutual funds in Korea. We find that the SRI funds show higher Sharpe ratios and Treynor's measure compared to regular mutual funds. We also show that return measures are sensitive to weighting scheme; since SRI funds outperform regular funds by 1.2 times in terms of equal-weighted portfolio, but outperform by 62 times in terms of value-weighted portfolio. Finally, using a regression framework controlling for factors such as fund age, risk, and size, we confirm that SRI funds perform better than regular mutual funds.

Keywords: Socially Responsible Investing, Mutual Fund, Performance Measure.

1 Introduction

Whereas regular investment typically focuses on the financial aspect of investment, socially responsible investing refers to investment which also considers non-financial aspect of investment, such as environment and society as a whole. Recent literature started to analyze the financial performance of these socially responsible investing [1, 2, 3, 4]. Our study adds to the literature by examining whether there is a significant difference in the performance between the SRI mutual funds and regular mutual funds. The measurement of performance is done by using Sharpe ratio and Treynor's measure [5]. We carry out the task using both equal-weighted portfolio and value-weighted portfolio. The reason for doing so is to see if the result can be driven by a few large funds which dominate the market. Finally, we examine if the results hold even after controlling for factors such as the age of the fund, fund size, riskiness of the fund measure by the standard deviation of return, and the cost of investment. The paper is organized as follows. In Section 2, we describe the data used in the study. In Section 3, we show the results of our analyses. Section 4 concludes this study.

* Corresponding author.

2 Data

We collect data on fund performance from Korea's Fund Doctor (www.funddoctor.co.kr). Specifically, for each fund, we collect information on fund name, founding date, asset size, return, standard deviation of return, and beta, for a one-year period until the most recent date of March 22, 2011. To calculate the Sharpe ratio and Treynor's measure, we use 4.64% as the 12-month risk-free rate. We only include in our sample those funds with asset size greater than 5 billion Korean Won. Table 1 and Table 2 show our sample size.

Table 1. Sample size of regular mutual funds in Korea

Funds	Sample size
Regular funds in existence in March 22, 2011	411
Funds with asset size less than 5 billion Korean Won	152
Final sample size	259

Table 2. Sample size of SRI mutual funds in Korea

Funds	Sample size
SRI funds in existence in March 22, 2011	11
Funds with asset size less than 5 billion Korean Won	5
Final sample size	6

3 Results

Sharpe ratio is defined as the abnormal return per unit of risk measured by the standard deviation of return. The higher the Sharpe ratio, the better the performance relative to risk. Treynor's measure is similar to Sharpe ratio, except that the measure of risk is the beta of the portfolio rather than the standard deviation of return. That is, the risk in Treynor's measure is the systematic risk, whereas the risk used in the Sharpe ratio is the total risk of the portfolio. Table 3 and Table 4 show that SRI funds perform better than regular mutual funds, in terms of both the Sharpe ratio and the Treynor's measure.

Table 3. Sharpe ratio

	Regular funds	SRI funds
Average of Sharpe ratio	1.16	1.46
Variance of Sharpe ratio	0.13	0.15
T-statistic in the difference of Sharpe ratio	-1.87	

Table 4. Treynor's measure

	Regular funds	SRI funds
Average of Treynor's measure	20.39	25.25
Variance of Treynor's measure	54.03	53.97
T-statistic in the difference of Treynor's measure		-1.60

Table 5 and Table 6 show the average performances of the fund using measures of equal-weighted and value-weighted. In both tables, results confirm that SRI funds outperform regular funds, although the size of significance is reduced in the case of value-weighted funds. Specifically, using equal-weighted portfolio, SRI funds outperform regular funds by a magnitude of 1.2 times. More drastically, using value-weighted portfolio, SRI funds outperform regular funds by 62 times. Also, results show that fund performance is reduced if the performance is measured using value-weighted scheme, relative to equally-weighted scheme.

Table 5. Performance measure using equal-weighted portfolio

	Regular funds	SRI funds
Average return	23.38	28.31
Variance of return	35.57	35.65
T-statistic in the difference		-2.00

Table 6. Performance measure using value-weighted portfolio

	Regular funds	SRI funds
Average return	0.09	5.49
Variance of return	0.03	66.09
T-statistic in the difference		-1.63

Table 7 shows the results of the regression of fund performance on SRI fund dummy and various control variables. The dependent variable is the Sharpe ratio. Results show that even after controlling for factors such as fund age, asset size, standard deviation of return, the SRI dummy variable is positive and statistically significant at the 90% confidence level. Results also show that younger funds perform better, and riskier funds perform better. The asset size of the funds is not a significant determinant of the fund performance.

In Table 8, we repeat the regression using the Treynor's measure as our dependent variable. The results are qualitatively similar to Table 7, except that the coefficient of the standard deviation variable is no longer significant in Table 8.

Table 7. Regression of fund performance (Dependent variable: Sharpe ratio)

	Coefficient	Standard error	t-statistic	P-value
Constant	0.78	0.25	3.12	0.00
Fund age	-5.60E-05	1.73E-05	-3.23	0.00
Asset size	-6.32E-06	1.09E-05	-0.58	0.56
Standard deviation	0.03	0.015	1.96	0.05
SRI dummy	0.27	0.15	1.89	0.06

Table 8. Regression of fund performance (Dependent variable: Treynor's measure)

	Coefficient	Standard error	t-statistic	P-value
Constant	19.97	5.16	3.87	0.00
Fund age	-0.00	0.00	-3.41	0.00
Asset size	-2.79E-05	0.00	-0.12	0.90
Standard deviation	0.17	0.32	0.52	0.60
SRI dummy	4.50	2.99	1.51	0.13

4 Conclusion

In this study, we examined the performance of SRI funds against the regular mutual funds in the Korean market. We show that SRI funds outperform regular mutual funds for various measures of performance. Specifically, SRI funds perform better in terms of both Sharpe ratio and Treynor's measure. The results are robust to different types of averages; equal-weighted and value-weighted. Also, we confirm the results using regression analysis while controlling for other possible factors which can affect firm performance.

References

1. Geczy, C., Stambaugh, R.F., Levin, D.: Investing in Socially Responsible Mutual Funds. Working Paper (2005)
2. Mohr, L.A., Webb, D.J., Harris, K.E.: Do Consumers Expect Companies to be Socially Responsible? The Impact of Corporate Social Responsibility on Buying Behavior. *J. of Consumer Affairs*. 35, 45–72 (2005)
3. Bauer, R., Derwall, J., Otten, R.: The Ethical Mutual Fund Performance Debate: New Evidence from Canada. *J. of Business Ethics* 70, 111–124 (2007)
4. Hamilton, S., Jo, H., Statman, M.: Doing Well while Doing Good? The Investment Performance of Socially Responsible Mutual Funds. *Financial Analysts J.* 49, 62–66 (1993)
5. Sharpe, W.F.: Mutual Fund Performance. *Journal of Business* 39, 119–138 (1966)

Mobile Cloud e-Gov Design and Implementation Using WebSockets API

Young-Hyuk Kim, Il-Kown Lim, Seung-Gu Kang, and Jae-Kwang Lee*

Department of Computer Engineering, Hannam University, Daejeon, Korea
{yhkim, iklim, sgkang, jklee}@netwk.hannam.ac.kr

Abstract. As the pan-government computational resources integrated building project of Korea in 2010, computational resources integrated building and use plan for 139 systems of 34 government departments using Cloud Computing are under way. Therefore, in this paper, we design Mobile Cloud Computing e-Gov using central government Cloud cluster by using WebSockets API newly added to HTML5. We build UCS performing Mobile Cloud Computing service directly and communicate Full Duplex with each organization WAS in Mobile Web of UCS directly. As a result, the model presented in this paper enables data push from the server of true meaning that was not fully realized in HTTP until now and materializes interactive communication in which real time performance of Web Application improves greatly.

Keywords: Mobile Cloud Computing, HTML5, WebSockets, WAS, m-Gov.

1 Introduction

It is known that Christophe Bisciglia of Google proposed ‘Cloud Computing’ to Eric Schmidt, CEO in 2006 for the first time [1].

In Gartner, Cloud Computing is defined as ‘one form providing resources with a high level of scalability to many customers as the service by using Internet technology’ and in IBM, it is defined as ‘the environment where large capacity database is distributed and processed in the Internet virtual space by using Web based Application and this data can be brought or processed in various terminals such as PC, notebook, smart phone etc. And in the Forrester Research, it is defined as ‘Computing that provides standardized IT based functions through IP and that is always accessible and variable according to change of acceptance and that is provided through web or programmatic Interface.’ [2-5].

According to the pan-government computational resources integrated building plan in 2010, Korea’s central government enhances flexible resources deployment and utilization rate in common use and overlapping · excessive investment prevention by integrating and building computer equipment that has been built by department previously by Cloud Computing technology and the project to integrate and use computational resources of 139 systems of 34 government departments is under way [6].

* Corresponding author.

Therefore, in this study, we design Mobile Cloud e-Gov(Electronic Government) responding to the Mobile Environment which is rapidly spreading lately by using central government Cloud cluster.

Mobile Cloud e-Gov is ‘the technology providing high level of integrated e-Gov service to many users by bringing Web service of each government organization distributed into virtualized Mobile service space by using the Internet technology’ and can be expressed as [Figure 1].



Fig. 1. The concept of Mobile Cloud e-Gov

In a broad sense, virtualization means abstraction on computer resources and there are various types of virtualization.

In Cloud Computing, virtualization means resource virtualization and resource virtualization means virtualization on a specific system resource such as storage volume, name space, network resource etc. [7 - 8].

The method proposed as a way of Mobile Cloud e-Gov realization like [Figure 1] is user request data processing using WebSockets API of HTML5, next-generation Web standard language. Without giving any special changes to the organization systems providing unique service individually, virtualized integrated space UCS(Ubiquitous Cloud Space) is composed ‘so that a number of systems divided physically can be seen as the one logical system like integrated one’ through Virtualization technology of Cloud Computing between each organization and users.

As the characteristic of UCS, while the Cloud Computing project performed by the government is to integrate each organizations into one, the objectives of this paper are to take care of users' request through TCP communication between WAS(Web Application Server) and Web in which WebSockets code is inserted and to deliver a feeling that all government organization services are done at one Mobile Web to users by applying Swipe method and Stack method, trends of recent Mobile Web.

For this study, we check Mobile e-Gov trend in chapter 2 and analyze the technology to implement it substantially. In chapter 4, we design and implement Mobile Cloud e-Gov system for practical service based on previously researched technology and finish this paper by setting future research goals.

2 Mobile Cloud e-Gov Trend

All over the world, the countries conducting Mobile Cloud e-Gov do not exist and in case of the U.S, general Cloud Computing e-Gov pilot project except for Mobile service is under way.

2.1 Overseas Trend

1) U.S.

Mobile Web was developed to easily see USA.gov, GSA(General Service Administrator) federal e-Gov portal even in a small screen of Mobile devices. Currently, 'The White House App.' Where the contents of the White House government can be seen is distributed freely and Mobile only homepage was opened [Figure 2].



Fig. 2. Mobile Web service screen of the U.S. White House

Also, as part of low carbon green IT technology introduction of Obama Government, the U.S. Government announced Apps [Figure 3], public service site as a pilot project aiming at e-Gov one stop service provision through Cloud Computing.

The U.S government is hitting its stride in Cloud Computing building for environment protection and IT infrastructure building cost reduction and regards moving light document management (work flow) system into Cloud Computing environment as the top priority. In fiscal year of 2011, they will present Cloud Computing guide lines targeting each level of authority.



Fig. 3. The U.S. public service site (Apps)

App, the current public service site, is a pilot project and though it is small Cloud Computing, it is evaluated as epoch making public service site having totally new civil service distribution center role model that did not exist previously.

2) England

Directgov[Table 1] of England has provided public service by using mobiles since 2007 and service can be used in all Internet enabled mobile devices.

For use of the service, first, enter 'MOBILE' in a mobile in text for text service or enter '83377'. Second, for Browsing service, press 'm.direct.gov.uk' in Internet browser of a mobile.

Table 1. The main services provided in Directgov

Classification	Main contents
JOBS	Providing information about 400,000 jobs through Jobcenter Plus
Journey Planner	Providing departure time and arrival time of trains, public transportation information and road information
Blue Badge Service	Providing information on police stations and public bathrooms, parking information for disabled people
Public announcements	Providing terror warning, government news and emergency information in case of emergency etc.
Local Health Services	Providing medical information such as drugstores, dentists, hospitals etc. near the area
Information	Providing information on banks, house information, real estate, schools, trip, passport and legal issues etc.
Local Search	Location information search on local council, post offices, passport issue window etc. is available
Travel News App	Providing trains, roads, subway information etc. by iPhone and Android

2.2 Korean Trend

e-Gov defined in Korea is 'government efficiently performing administrative tasks for the people or mutual administrative agencies by digitalizing work of administrative organizations by using information technology (e-Government law)'. Therefore, Ministry of Public Administration and Security is making Mobile e-Gov service vitalization strategy to implement 'Smart government, advanced government'.

M-Gov(Mobile Government) is one form of Electronic Government and means government performing efficient administrative work by providing service and information to the people and companies or performing government affairs by using Mobile wireless communication technology and domestic Mobile e-Gov status being carried out currently is limited to information provision in the form of a message.

3 HTML5 WebSockets API Analysis

WebSockets is the technology to realize interactive Full Duplex between browser and users. Since both clients and the server should be able to use this function in order to

use WebSockets, WAS on the server side is required. Therefore, if WebSocket connection is set, either side of a client and the server can send messages at any time and it means data Push from the server of true meaning that has not been fully realized in HTTP until now is possible. It means that not only real time performance of Web Application is greatly improves but totally new Application and service based on interactive communication are possible through this.

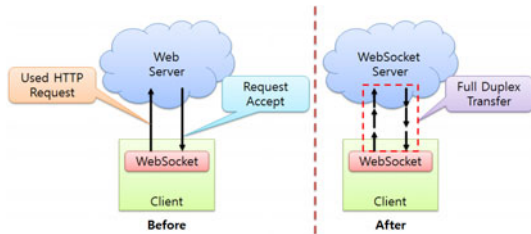


Fig. 4. WebSockets Communication Flow

3.1 WebSockets API (W3C: World Wide Web Consortium)

As a way to communicate with the server by using JavaScript, there has been XMLHttpRequest(Ajax) until now but it only could implement communication of a form that the server responds when a browser requests. Therefore, real time interactive communication cannot be exchanged and data cannot be sent from the server to a browser at a certain point. In contrast, WebSockets API achieves Full Duplex with the server and real time interactive communication is possible [Figure 5].

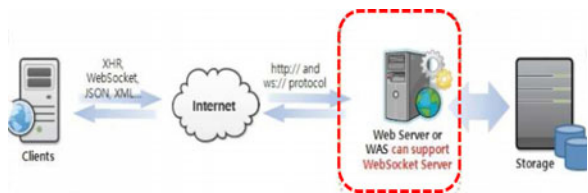


Fig. 5. WAS-WebSocket Communication

3.2 WebSockets Protocol (IETF: Internet Engineering Task Force)

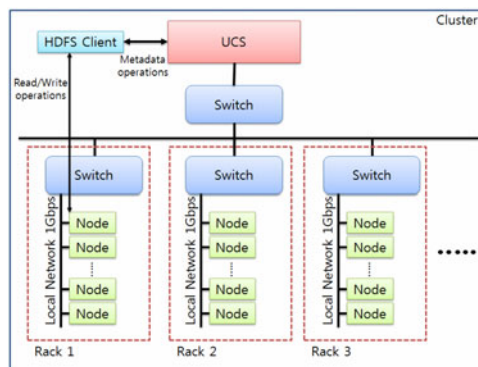
WebSockets Protocol uses WebSockets only protocol without using HTTP as communication protocol. In Web, communication is possible through TCP Sockets and one or more TCP ports are necessary. In this protocol, IETF is responsible for WebSockets Protocol.

Table 2. WebSocket Server Opening Handshake

Client Handshake Request	Server Handshake Response
GET /demo HTTP/1.1	HTTP/1.1 101 WebSocket Protocol Handshake
Upgrade: WebSocket	Upgrade: WebSocket
Connection: Upgrade	Connection: Upgrade
Host: example.com	Sec-WebSocket-Origin: http://example.com
Sec-WebSocket-Key1: 4 @1 465 46XW01 1 5	Sec-WebSocket-Location: ws://example.com/demo
Sec-WebSocket-Key2: 12998 5 Y3 1 .P00	Sec-WebSocket-Protocol: sample
Sec-WebSocket-Protocol: sample	
Origin: http://example.com	

4 Mobile Cloud Government Design

Hadoop Framework is used to bind distributed each organization system and resources into Cloud Computing environment and the structure is as shown in [Figure 6]. Hadoop Framework [9] is one of Apache Luccen projects including MapReduce and HDFS(Hadoop Distributed File System). HDFS can be installed on low specification computers and provides high Fault-Tolerant and is the file system suitable for processing large capacity data. [10]. MapReduce operates based on HDFS, which was developed for the purpose of efficient resources management of distribution and parallel system dealing with large quantity of resources. [11].

**Fig. 6.** UCS Hadoop Architecture

Overall Mobile Cloud e-Gov service is UCS based on HTML5 and is done by access of users.

UCS is divided by Private Cloud Service and Public Cloud Service and provide different services to classified users [Figure 7].

UCS, the service subject, is the Web based system made by HTML5 language and uses Stack structure system to provide convenience of service use and efficient menu access and serviceability to users in Mobile Web environment.

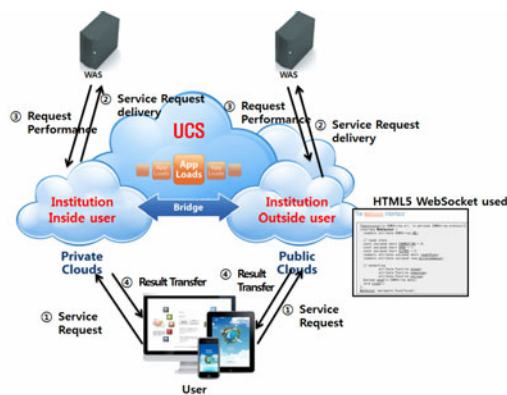


Fig. 7. UCS Service flow chart

Each layer service configured on screen by Stack method communicates directly to actual service corresponding internally through WebSockets code and shows the result to users [Figure 8] is the screen implemented through UCS after defining the actual service running independently as the single service at present as ‘Mobile Seoul’, ‘The National Library of Korea’, ‘LH Corporation’, ‘Korea Meteorological Administration’ temporarily through Storage virtualization within the server.

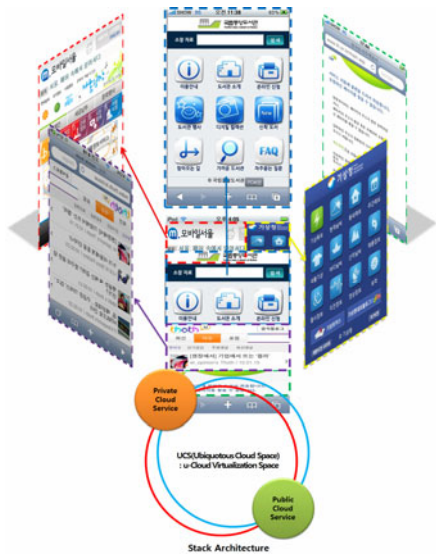


Fig. 8. Mobile Cloud e-Gov implementation screen through UCS

5 Conclusion

In this paper, we designed and implemented Mobile Cloud e-Gov system to which Mobile, global IT trend is combined based on Cloud Computing based e-Gov cluster that Korea is promoting. UCS, the service subject to service Mobile Cloud e-Gov, implemented it through HTML5 language based on Web. Since WebSockets API is newly added to HTML5, Full Duplex real time interactive communication can be implemented. Therefore, the systems of each independent organization were built into Cloud cluster by using Hadoop Framework, one of Apache Luccen projects and access and access path of all services were unified by putting UCS on Root.

UCS is divided into Private Cloud Service for internal organization users and Public Cloud Service for general users. Mobile Web where the actual service is done configured the screen by using Stack structure Layer method spread recently. For actual implementation, we made 5 spaces through Storage virtualization within the server and materialized so that communication through WebSockets between UCS and Mobile Web can be made with different data. Current design system implemented only communication simply using HTML5 WebSockets and security technology and encryption are not applied. Therefore, studies on future security vulnerability and security technology application measures are necessary.

References

1. Min, O.G., Kim, H.Y., Nam, G.H.: Trends in Technology of Cloud Computing. ETRI Journal 24(4), 1–13 (2009)
2. Wikipedia, http://en.wikipedia.org/wiki/Cloud_computing
3. Vision, Hype, and Reality for Delivering IT Services as Computing Utilities, HPCC, Keynote (2008)
4. IBM Cloud Computing Strategy: Blue Cloud computing paradigm, Microsoft (2009)
5. Valdes, R.: Google App Engine Goes Up Against Amazon Web Services. Gartner (April 2008)
6. Shin, S.-Y.: Pan-government Cloud Computing Activation Master Plan. Korea Local Information Research & Development Institute 61, 46–51 (2010)
7. IBM Virtualization technology, <http://www.ibm.com>
8. Kim, J.-M.: Virtualization technology for next generation Computing. ETRI Journal 23(4), 102–114 (2008)
9. Hadoop, <http://www.hadoop.apache.org>
10. Ghemawat, S., Gabioff, H., Leung, S.-T.: The Google File System. In: Proceeding of ACM Symposium on Operating System Principle
11. Jeffrey Dean, S.G.: MapReduce Simplified Data Processing on Large Clusters. In: Proceedings of the 6th Conference on Symposium on Operating Systems Design & Implementation, San Francisco, pp. 1–13 (2004)

Noise Reduction in Image Using Directional Modified Sigma Filter

Hye-Youn Lim, Mi-Ran Gu, and Dae-Seong Kang

Department of Electronics Engineering Dong-A University, Busan, Korea
dskang@dau.ac.kr

Abstract. In this paper, we propose a new method using a modified sigma filter. The sigma filter among many algorithms is one of the simplest de-noising methods. The threshold of the sigma filter uses the estimated standard deviation of the noise by block-based noise estimation using the adaptive Gaussian filtering. In the proposed method, an input image is first decomposed into two components according to direction features. Then, two components are applied; HPF and LPF. By applying a conventional sigma filter separately on each of them, the output image is reconstructed from the filtered components. Comparative results from experiments show that the proposed algorithm achieves higher gains than the sigma filter and modified sigma filter, which are 2.6 dB PSNR on average and 0.5 dB PSNR, respectively. When relatively high levels of noise are added, the proposed algorithm shows better performance than the two conventional filters.

Keywords: Adaptive Gaussian filtering, image de-noising, noise estimation, noise removal, sigma filter.

1 Introduction

Noise reduction is a very important processing step in all digital imaging applications. Moreover, the noise level is still high despite the latest manufacturing advances in the camera sensors. Consequently, image de-noising is always an important research field.

Existed noise reduction methods can be categorized into spatial filters and frequency domain filters. The spatial filters are linear filters (average filter and Gaussian filter) and non-linear filters (median filter and sigma filter). The frequency domain filters are band-pass filter and notch filter [1] [2] [3] [4].

In this paper, we propose a new method using a directional modified sigma filter [5]. The threshold of the sigma filter uses the estimated standard deviation of the noise by block-based noise estimation using the adaptive Gaussian filtering [6] [7]. In the proposed method, the input image is first decomposed in two components that have features of horizontal, vertical, and diagonal direction. Then, two components are applied, which are HPF and LPF. By applying a conventional sigma filter separately on each of them, the output image is reconstructed from the filtered components.

2 Sigma Filter

2.1 Introduction of Sigma Filter

The sigma filter proposed by Lee is based on the sigma probability of Gaussian distribution. The basic principle of this filter is to replace the current pixel by the averaging of the pixels which have their intensity in a $(2m + 1) \times (2n + 1)$ sliding window [4]. This algorithm is defined as follow:

$$y(i, j) = x(i, j) + n(i, j), \quad (1)$$

$$x^*(i, j) = \frac{\sum_{k=i-n}^{n+i} \sum_{l=j-m}^{m+j} \delta_{k,l} y(k, l)}{\sum_{k=i-n}^{n+i} \sum_{l=j-m}^{m+j} \delta_{k,l}}, \quad (2)$$

$$\delta_{k,l} = \begin{cases} 1 & |y(k, l) - y(i, j)| \leq \Delta \\ 0 & |y(k, l) - y(i, j)| > \Delta \end{cases} \quad (3)$$

where $y(i, j)$ is a noisy image, $x(i, j)$ is its original image, $n(i, j)$ is assumed to be a binary symmetric channel (BSC) noise, $x^*(i, j)$ is called the sigma averaging and the sigma range Δ is fixed to a certain value.

2.2 The Modified Sigma Filter

The modified sigma filter [5], which is the transformation of the conventional sigma filter, is decomposed in four components for the input image. The sigma filter is applied independently on each of them. Then the output image is reconstructed from the four components filtered by the modified sigma filter.

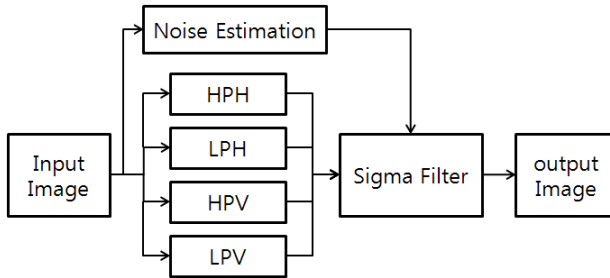


Fig. 1. The block diagram for modified sigma filter with noise estimation

The block diagram for modified sigma filter with noise estimation is shown in Fig. 1. The blocks denoted as HPH, HPV, LPH, LPV mean computing a high-pass filtering and low-pass filtering on the vertical and horizontal directions respectively.

3 The Proposed Approach

3.1 Noise Estimation

Generally, noise estimation algorithms in the spatial domain are classified into two approaches: block-based and filtering-based (smoothing-based).

Block-based noise estimation using adaptive Gaussian filtering [7] is based on block-based noise estimation. The input image is the image with the white Gaussian noise and a filtering process is performed by an adaptive Gaussian filter. To make small block size, the gradient of blocks using local gradient derivative pattern (LGDP) [8] are gotten and an orientation histogram is formed from the local gradient. The regions of noise estimation are set by the region of high orientation histogram. Fig. 2 shows the regions of noise estimation are black blocks ($= B^*$) when the standard deviation of the Gaussian noise is 5.0.

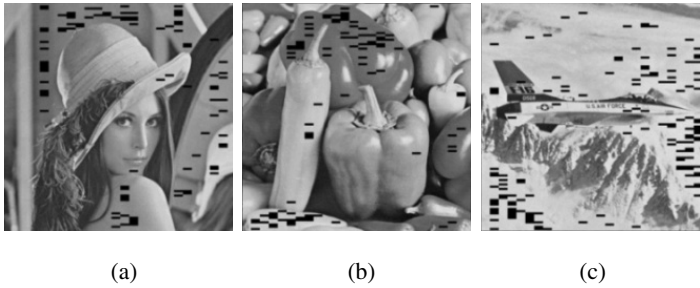


Fig. 2. The selected blocks B^* ($\sigma_n=5$). (a) “Lenna” (512×512), (b) “Pepper” (512×512), (c) “Airplane” (512×512)

3.2 The Proposed Method for Noise Reduction

In this section, the modified de-noising method based on the modified sigma filter [5] is proposed. The input image is first decomposed in two components that have features of horizontal, vertical, and diagonal direction. Then, two components are applied, which are HPF and LPF. By applying the conventional sigma filter separately on each of them, the output image is reconstructed from the filtered components. The added noise is removed and the proposed method preserves the edges from the image.

Fig. 3 shows the block diagram for the proposed method. The blocks denoted as HPF_comp1 and HPF_comp2 perform a high-pass filtering on two components. The blocks denoted as LPF_comp1 and LPF_comp2 perform a low-pass filtering on two components, where two components have features of horizontal, vertical, and diagonal direction. Block-based noise estimation using adaptive Gaussian filtering is performed by the block denoted as BNE using AGF and LGDP. The block denoted as sigma filter performs the sigma filtering and recombines the filtered components to obtain the restored image. The proposed algorithm is shown in Table 1.

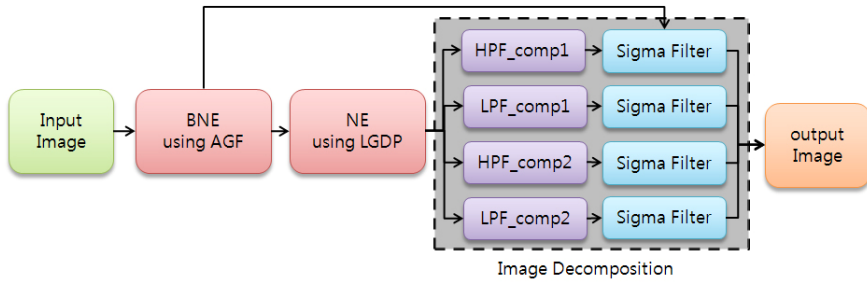


Fig. 3. The block diagram for proposed method

Table 1. The proposed algorithm for noise reduction

The proposed algorithm
(1) First, the input image is decomposed in two components that have features of horizontal, vertical, and diagonal direction. $comp1 = y(i, j - 1) + y(i, j + 1) + y(i - 1, j) + y(i + 1, j)$ $comp2 = y(i - 1, j - 1) + y(i - 1, j + 1) + y(i + 1, j - 1) + y(i + 1, j + 1)$ (2) compute the differences and sum between the $comp1$ and the center pixel of the input image: $y_{HPF\_comp1}(i, j) = \frac{1}{8}(4y(i, j) - comp1), \quad y_{LPF\_comp1}(i, j) = \frac{1}{8}(4y(i, j) + comp1)$ (3) compute the differences and sums between the $comp2$ and the center pixel similar to $y_{HPF\_comp2}(i, j) = \frac{1}{8}(4y(i, j) - comp2), \quad y_{LPF\_comp2}(i, j) = \frac{1}{8}(4y(i, j) + comp2)$ (4) Apply a sigma filter separately on the four computed components $y_{HPF\_comp1}(i, j)$, $y_{HPV\_comp1}(i, j)$, $y_{LPF\_comp2}(i, j)$ and $y_{LPV\_comp2}(i, j)$ to obtain f_{HPF_comp1}, f_{HPV_comp2}, f_{LPF_comp1} and f_{LPV_comp2} respectively. This is done by the blocks denoted as sigma filter in Fig. 3.

4 Simulation Results and Discussions

In this section, we show the comparative performances of our proposed method and the standard sigma filter and the modified sigma filter. To this end, we selected three images, “Lenna,” “Pepper,” and “Airplane,” and we added zero mean Gaussian noise with different variances to them. The original images were represented on 8 bits (values in the range [0, 255]). In experiments, we use a 5×5 sigma filter.

To visually compare the performances of the three methods, in the left images of Fig. 4, we show parts of the processed images. In Fig. 4, shown the three methods result on a noisy grayscale image “Lenna” (512×512); (a) is the original image; (b) is the pre-processed image which was applied the Gaussian noise of 25 standard

deviations; (c), (d), and (e) show the obtained image using the standard sigma filter, the modified sigma filter, and the proposed method respectively. The effect of image smoothing and edge preservation can be seen in the left images of Fig. 4. The added noise and fine texture in Fig. 4 (e) have been mostly removed. The obtained graph using the proposed method is the most similar original graph. We note that our proposed method better preserve small details.

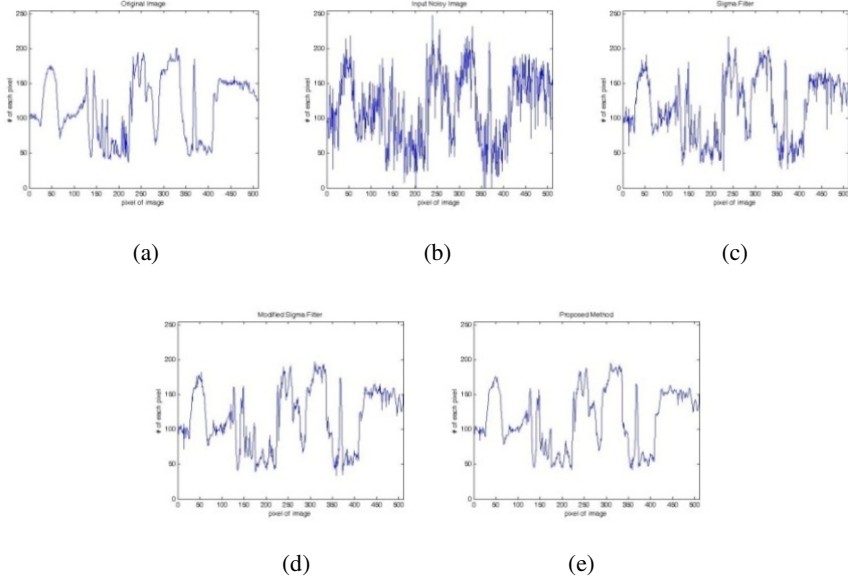


Fig. 4. The noise reduction results of each algorithm (a) Original image “Lenna”, (b) Noisy image, (c) Sigma filter, (d) Modified sigma filter, (e) Proposed method

Evaluation factors for comparing the performance are mean square error (MSE) and peak signal to noise ratio (PSNR).

$$MSE = \frac{1}{MN} \sum_{i=0}^M \sum_{j=0}^N [x(i, j) - \hat{x}(i, j)]^2 \quad (4)$$

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} [dB] \quad (5)$$

Where $x(i, j)$ is the original image, $\hat{x}(i, j)$ is the filtered image, and M, N are the horizontal and vertical size of the image.

Table 2 shows the PSNR comparison of each algorithm according to the standard deviation of added Gaussian noise. Comparative results from experiments show that the proposed algorithm achieves higher gains, which is 2.6 dB PSNR on average, than the sigma filter, and 0.5 dB PSNR than the modified sigma filter. When relatively high levels of noise are added, the proposed algorithm shows better performance than two conventional filters.

Table 2. The PSNR comparison of each algorithm according to the standard deviation of added Gaussian noise

Standard deviation of noise	Gaussian filter	Sigma filter	Modified sigma filter	Proposed method
5	32.238	36.747	37.479	37.290
10	31.616	31.382	33.451	33.488
15	30.740	29.349	31.718	32.353
20	29.725	28.028	29.694	30.331
25	28.758	25.579	28.618	29.739
Average value	30.558	30.558	32.605	33.127

5 Conclusions

This paper presents a new method using a modified sigma filter for image de-noising. The new algorithm has improved performances in terms of MSE and also preserves better the fine details of the processed image as opposed with the standard sigma filter and the modified sigma filter. Comparative results from experiments show that the proposed algorithm achieves higher gains, 2.6 dB PSNR on average, than the sigma filter, and 0.5 dB PSNR than the modified sigma filter. When relatively high levels of noise are added, the proposed algorithm shows better performance than the two conventional filters.

Acknowledgments. This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the Core Technology Development for Breakthrough of Robot Vision Research support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2010-C7000-1001-0006).

References

1. Lim, J.S.: Two-Dimensional Signal and Image Processing. Prentice Hall, Englewood Cliffs (1990)
2. Gonzales, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, Englewood Cliffs (2002)
3. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis, and Machine Vision, ITP (1999)
4. Lee, J.S.: Digital Image Smoothing and the Sigma Filter. Computer Vision, Graphics, and Image Processing 24(2), 255–269 (1983)
5. Bilcu, R.C., Vehvilainen, M.: A Modified Sigma Filter for Noise Reduction in Images. In: Proceedings of the 9th WSEAS Circuits, Systems, Communications and Computers multiconference, WSEAS/CSCC 2005, vol. 15 (July 2005)
6. Olsen, S.I.: Estimation of noise in images: An evaluation. Graphical Models and Image Process 55, 319–323 (1993)
7. Shin, D.-H., Park, R.-H., Yang, S., Jung, J.-H.: Block-Based Noise Estimation Using Adaptive Gaussian Filtering. IEEE Transactions on Consumer Electronics 51(1), 218–226 (2005)
8. Zheng, X., Kamata, S., Yu, L.: Face Recognition with Local Gradient Derivative Patterns. In: TEMCON 2010 – 2010 IEEE Region 10 Conference, pp. 667–670 (November 2010)

An Approach to Real-Time Region Detection Algorithm Using Background Modeling and Covariance Descriptor

Jong-Dae Park, Hye-Youn Lim, and Dae-Seong Kang

Department of Electronics Engineering DongA University, Busan, Korea
dskang@dau.ac.kr

Abstract. In this paper, we present a region detection algorithm using robust image descriptor. It abstracts objects from moving picture using Gaussian mixture background modeling on a moving object. And feature image is comprised of spatial properties and statistical properties on abstracted objects, and covariance matrices are formatted using region variance magnitudes. By using it to detection region, this paper puts a proposal for outstanding real-time region detection algorithm to diverse situations with strong noise, rotation and illumination. To estimate performance evaluation of proposed algorithm, we conduct an experiment on region detection in diverse traffic circumstance moving picture. And then we get a great performance result on interested region detection

Keywords: Covariance matrix; region detection; Gaussian mixture background modeling.

1 Introduction

As an applied part for mobile, the needs of pattern recognition have been increased recently. It is being used for generic technologies in a variety of applied parts such as recognitions for face, trademarks, and license plates in real-time images, besides, it will need to keep studying in the future [1][2].

The study about traditional area recognition is difficult to do real-time processing due to complexity of realization in the statistical approaches, Gaussian Markov random field (GMRF) [3], gray-level concurrence matrix, local linear transformation and structural approaches when we examine the way for approaching by using wavelet transform. Additionally, it can be confirmed performance decrease by intensity of illumination and noise in being allowed situation.

In this paper, therefore, distinguishing images included external features of image territory which try to find and statistical features in real-time video will be formed in advance, further, covariance matrix is going to be suggested to use for detecting the wanted area in the complicated video by various pictures. The experiments about the suggested performance are going to be modified by detecting the license plates in vehicles.

2 Gaussian Mixture Background Modeling

2.1 Previous Study on Background Modeling

Previous background modeling methods could have to need a manual initialization. And then, most researchers have given up non-adaptive methods of background modeling. Without doing initialization again, errors in the background accumulate over time, making this method useful only in highly-supervised, short-term tracking applications without significant changes in the scene [4].

A general method of adaptive background modeling is averaging the images over time, making a background approximation which is similar to the current static scene except for motion region.

While this is useful in positions where objects move constantly and the background is visible a serious portion of the time, it is not robust to scenes with many moving objects particularly in case they move slowly. And it also cannot handle bimodal backgrounds, recovers slowly when the background is uncovered, and has a single, predetermined threshold for the entire scene. Variations in bright scene can cause problems for many background modeling methods.

2.2 Effective Background Modeling

Friedman and Russel proposed to model each background pixel using a mixture of three Gaussians corresponding to road, vehicle and shadows in the context of a traffic surveillance system [5]. It is initialized by using expectation maximization (EM) algorithm. Then, the Gaussians are manually labeled in a heuristic manner as follows: the darkest component is labeled as shadow; in the remaining two components, the one with the largest variance is labeled as vehicle and the other one as road. This remains fixed for all the process giving lack of adaptation to changes over time. For the foreground detection, each pixel is compared with each Gaussian and is classified according to it corresponding Gaussian. The maintenance is made using an incremental EM algorithm for real time consideration. Stauffer and Grimson generalized this idea by modeling the recent history of the color features of each pixel $X_1, \dots, X_t$ by a mixture of K Gaussians.

2.3 Gaussian Mixture Modeling

After conducting preprocessing using Gaussian filtering to remove noises of outside environment, background is modeled by using adaptive Gaussian mixture model. At any time, t , probability observed as value of X_t on an optional pixel is represented as a formula (1) [6].

$$\rho(X_i) = \sum_{i=1}^k w_{i,t} \cdot \eta(X_t, \mu_{i,t}, \sum_{i,t}) \quad (1)$$

where K is the number of distributions, $W_{i,t}$ is an estimate of the weight (what portion of the data is accounted for by this Gaussian) of the i^{th} Gaussian in the mixture at time t , μ , t is the mean value of the i^{th} Gaussian in the mixture at time t , $\sum_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the mixture at time t , and η is a Gaussian probability density function.

2.4 Background Estimation

Parameters of background modeling are estimated by using K-means. And then, we exam a adjustment in Gaussian mixture distribution. And variable on each distribution is lined up using a formula (2).

$$B = \arg \min_b \left(\sum_{k=1}^b w_k > T_{BG} \right) \quad (2)$$

where T is a measure of minimum portion of data that should be accounted for by background. This takes the good distributions until a certain portion, T , of the recent data has been accounted for. When a small value for T is chosen, the background model is usually uni-modal. If this is the case, using only the probable distribution will save processing.

A flowchart of the background method is shown in Fig. 1.

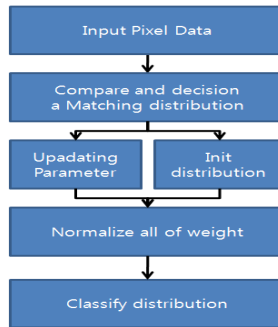


Fig. 1. The block diagram for adaptive background modeling

If value of T_{BG} is very large, a part of object is detected as background image. On the contrary, if value of T_{BG} is very small, a part of noise is detected as foreground image. A result of adaptive background modeling is shown in Fig. 2 Foreground image is abstracted using adaptive background modeling. And post-processing step eliminates noise from abstracted image.



(a) Original image (b) Foreground image (c) Post-processing

Fig. 2. Foreground image abstracted from original image

3 Covariance Descriptor

Basically, covariance is the measure of how much two variable vary together. According to relations between two variables, covariance values have diverse value in a Fig. 3 If two character values are increased, covariance values have a natural number. But when one character value is increased and another character value is decreased, covariance values have a minus value. Finally, if relation between covariance is independence, covariance value is zero.

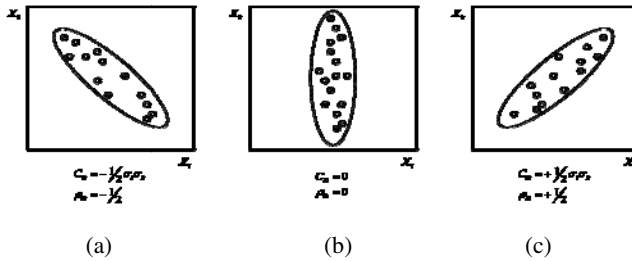


Fig. 3. Covariance matrix and correlation coefficient

In this paper, first derivative and second derivative toward x and y on space character is consist of following formula (3). Fig. 4 is made from space characters.

$$Z_k = \left[g(x', y') | r(x', y') I(x, y) I_x(x, y) I_y(x, y) I_{xx}(x, y) I_{yy}(x, y) \right] \quad (3)$$



Fig. 4. Covariance matrices generated for features

4 The Proposed Method and Experimental Result

The input frame is divided into two images – the first for background image, and the second for foreground image using Gaussian mixture background modeling. And then, we conduct a processing for noise elimination on abstracted foreground image. So, the moving object detected using background modeling is set as the region of interest (ROI). A license plate in determined ROI is abstracted by using Covariance Matrix. A license plate region abstraction is scanned by using fixed size. A flowchart of the proposed entire system is shown in Fig. 5.

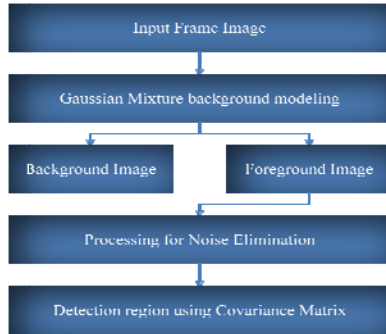


Fig. 5. The block diagram of a proposed method

Proposed distance measure formula (4) is used for measuring the dissimilarity of two covariance matrices [7]. The metric for comparison between two covariance matrices is

$$\rho(X_1, X_2) = \sqrt{\sum_{i=1}^n \ln^2 \lambda_i(X_1, X_2)} \quad (4)$$

where $\lambda_i(X_1, X_2)$ are generalized Eigen-values for X_1 and X_2 . The metric is a scalar value, which indicates the dissimilarity between two images with covariance matrices. In other words, the smaller this number is, the greater the similarity between the two images. If $X_1 = X_2$, then the result of the metric will be $\rho=0$.



Fig. 6. The results of the experiment on a license plate region detection

We experiment in car images over a road with diverse illumination change. So, this paper presented the covariance features and related algorithm for license plate region detection. Superior performance of the proposed methods was demonstrated on experiment. Also background modeling is used for fast processing. Fig. 6 shows the result images of license plate detection.

5 Conclusions

This paper presents an effective ROI detection method using Gaussian mixture background modeling and covariance descriptor. To estimate a performance of proposed method, we experiment on license plate detection in diverse environment which has many illumination changes. And we gain the better detection rating. But proposed method is a bit slow because of complexity on covariance descriptor calculation. Additionally, we can combine with neural networks for fast covariance matrix to improve a proposed method.

Acknowledgments. This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the Core Technology Development for Breakthrough of Robot Vision Research support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2010-C7000-1001-0006).

References

1. Hu, W., Tan, T., Wang, L., Maybank, S.: A Survey on visual Surveillance of Object Motion and Behaviors. *Systems, Man and Cybernetics*, Park C 34(3) (August 2004)
2. Tilmaz, A., Javed, O., Shah, M.: Object tracking; A survey. *ACM Journal of Computing Surveys* 39(4) (2006)
3. Krishnamachari, S., Chellappa, R.: Multiresolution GMRF models for texture segmentation. In: *Proc. of IEEE International Conf. on Acoustics, Speech, and Signal Processing*, pp. 2407–2410 (May 1995); also Tech. Rep. CS-TR-3393, University of Maryland (1995)
4. Lipton, A., Fujiyoshi, H., Patil, R.S.: Moving Target Classification and Tracking from Real-Time Video. In: *Proc. IEEE Workshop Applications of Computer Vision (WACV)*, pp. 8–14 (October 1998)
5. Friedman N., Russell S.: Image Segmentation in Video Sequences: A Probabilistic Approach. In: *Proceedings Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI 1997)*, pp. 175–181 (1997)
6. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: *IEEE Computer Vision and Pattern Recognition* (1999)
7. Forstner, W., Moonen, B.: A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart university (1999)

Optimization and Generation of Knowledge Model for Supporting Technology Innovation

Seungsu Jun, Seungryong Lee, and Munbong Shin

Korea Institute of Science & Technology Evaluation and Planning, Knowledge&Information division, 11F dongwon Industry Bldg, 68 Mabang-gil Seocho-gu, Seoul, South Korea
{dabins, leesr7376, mbshin}@kistep.re.kr

Abstract. In information society, there are now worldwide economic network that are expanding prodigiously. There could be argued that social structure is in dire need of supporting technology innovation. This paper seeks to analysis and how to generate of knowledge model that supporting policy making of technology innovation. We let the user focus on some interested parts a whole decision making and to analysis using case and how to generate knowledge model and network through advanced semantic mining and formal language. Knowledge model is created to topic network that compositing driven keyword through text mining from natural language in document. And we show that the way of analyzing knowledge model and automatically generating feature keyword and relation properties into topic networks. In order that we analyze the semantic relationship and knowledge properties using CTL(Computational Tree Logic) on organized topic map and Kripke structure.

Keywords: Knowledge Model, Semantic Analysis, Technology innovation.

1 Introduction

Knowledge management and analysis has been recognized important issue and to be most essential process on technology innovation and using resource in knowledge based economic environment. But how to organize and inter-operation of information and knowledge integration is to solve complex and very difficult problems[1]. In this study, we are ongoing integrating knowledge resources and identification of business activities prior to implementation of knowledge model, focused on verification and optimization of knowledge model for supporting technology innovation. Some change in environment that knowledge based industries and economic development can produce a variety of complexity to the problem. Therefore, business organization for innovation is needed more rapid decision making and strategic analysis[2].

In particular, rapid change of external environment and increase of complex systems require investment strategies and more intelligent decision making process and utilization system. This paper reflects to design knowledge model of supporting innovation and validating structured process models. And we have been derived the topic keyword of the process model from task information, generated hierarchical structure model by relation between topic keyword. This topic mapping for process models to model the knowledge network configuration and it is converted into the

equivalent relationship. This is a topic of unstructured information to extract meaning from the process model is mapped to the abstract model is transformed to the equivalent relationship network. This analysis and generation of knowledge model makes possible the design of knowledge maps and analysis of large scale data in the process[3]. Thus, procedure of semantic analysis and relation between processes base on path of tracking evidence of decision-making. As shown in fig.1, it is to express the logical structure for creating of new knowledge model.

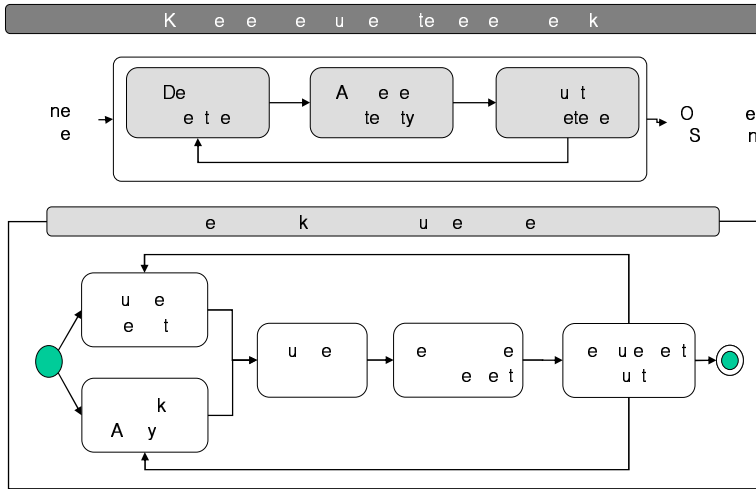


Fig. 1. The conceptual process model and information relation, decision making flow of strategies task for technology innovation

2 Optimization and Abstraction of Knowledge Model

In this study, we were divided and pre-defined a type for classification of knowledge and information on process of technology innovation. In recent research for business process design and validation, and have been investigated actively standardized graphical notation. Graphical notation for the specification of the model is easy to extend[4]. As well as it is also compatible with derivable logical language. Traditional business model was designed using the BPML(Business Process Modeling Notation), Activity Diagram, UML EDOC, IDEF, ebXML, BPSS, ADF(Activity Decision Flow), RossettaNet, LOVeM, EPCs(Event Process Chains), Petri Nets etc. In this study, be in a language that was utilized kripke structure. As shown in Fig.2, Here we denoted variables of the process Model(M) and Process(P) and Relations(R), Label(L). Kripke structure for the formal specification is the most easily graphic notation, and can be verified properties of the logic model[5]. In this paper, as shown in fig.2, on business activities in technology innovation was defined as abstracted process model. The activities in process model is labeled as a step in the process of the elements that represented by one state.

The relationship between the process(p) on abstracted process model(M) are denoted through transition and condition, interaction, reference etc.

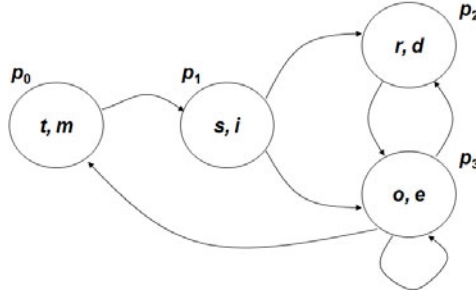


Fig. 2. Abstract knowledge model of Kripke structure derived from process model and activities for technology innovation

A process(p) has element(E) and element(E) has proposition(X). Each process(P) in the model is defined as follows : each transition(p, p') and has relation(R). Kripke structure's process model is defined as follows.

$$\forall p \in P \cdot \exists p' \in P \cdot (p, p') \in R$$

$$M : \text{Process Model} = \langle P, P_0, R, X, L \rangle$$

$$P : \text{Process} = \text{Set of finite process}(p)$$

$$P_0 : P_0 \subseteq P : \text{Set of initial process}$$

$$R : (p, p') \in R \subseteq P \times P = \text{State transition} (\forall p \in P, \exists p' \in P)$$

$$X : \text{True} \mid \text{False} = \text{Set of AP(Atomic Proposition)}$$

$$L : P \rightarrow 2^X = \text{assigns to each process state the set of atomic propositions that are true in that process state.}$$

$$P_0 : \text{Business Foresight} = \{\text{Trend } t, \text{Road Map } m\}, P_1 : \text{Business Planning} = \{\text{Strategy } s, \text{Invest } i\}, P_2 : \text{Performance Management} = \{\text{Research } r, \text{Development } d\}, P_3 : \text{Measurement Evaluation} = \{\text{Outcome } o, \text{Evaluation } e\}$$

Heterogeneous model checking in this study is easy to use for conceptual learning and analysis of knowledge model. And model checking to satisfy in a given model, the path of the property violated the state are presented. In this way, knowledge model to analyze can be continuously refining[6]. The model properties are expressed in CTL(Computation Tree Logic). CTL according to the passage of time representation of the property and it is easy to define a specific scope in model. CTL's syntax and semantics of temporal quarter is defined by the BNF(Backus Normal Form) as follow.

$$\Phi ::= \text{true} \mid x \mid \neg \Phi \mid \Phi_1 \vee \Phi_2 \mid AG \Phi \mid A(\Phi_1 W \Phi_2)$$

As existing in CTL, 'A' is universal operator and 'G' is global operator. And 'W' is until operator. Intuitively 'AG Φ ' is always true in all paths(true) means and 'A($\Phi_1 W \Phi_2$)' means that Φ_1 is forever true until Φ_2 to be true for all paths in model. Assuming a finite model M and CTL formula is true in the process P_0 is written as $p_0 \models \Phi$. A transition(p, p') in the model(M) indicated reachable from current process(P_n) to next process(P_{n+1}). Abstract process model(M) has the set of the transition(R). Transition relation (p_0, p_1) $\in R$ and by a single transition from p_0 to p_1 is reachable. On reaching it

from current process's forward and backward as the successor the predecessor said. As above, the current process(P_n) can be reached through a transition from one of the hollowing conditions to obtain the 'image computation' is called a calculation or reachability analysis. In the model, as the process reached the bottom of the 'image computation' algorithm and the same calculation when to stop is called fix-point. And the 'reachable' meaning in CTL formula is defined as follows.

$$R = p \in P ((p \rightarrow p') \vee (p' \rightarrow p))$$

$$R = \{(p_0, p_1), (p_1, p_2), (p_1, p_3), (p_2, p_3), (p_3, p_2), (p_3, p_3), (p_3, p_0)\}$$

$$successor(P_{\{i\}}, R) = \{p' | \exists p \in P \wedge (p, p') \in R\} = pre \forall (P)$$

$$predecessor(P_{\{i\}}, R) = \{p | \exists p' \in P \wedge (p, p') \in R\} = post \exists (P)$$

The Symbol μ is operator of least fix-point, and ν is operator of greatest fix-point.

$$R\Phi(P) = \mu Z. ((P \cup post \exists (Z)) \cap [\Phi])$$

$$Find(AG(true), P) = R \text{ true}(P), \quad Find(A(true \vee AG(\Phi)) W \Phi, P) = R^{\Phi}(P)$$

$$Find(AG(\Phi \Rightarrow AG \text{ true}), P) = R_{true}([\Phi] \cap post \exists (R^{\Phi}(P))),$$

$$Find(AG(\Phi_1 \wedge \Phi_2 \Rightarrow A[\text{true} \vee AG(\Phi_2) W \Phi_2]), P) = R^{\Phi_2}([\Phi_1 \wedge \Phi_2] \cap R_{true}(P))$$

$$AG \Phi = A(\Phi W \Phi \wedge false) = A(\Phi U \Phi \wedge false) \vee AG \Phi$$

$$= \nu Z. (\Phi \cap prev(Z)) = P_{init} \cup R_{\Phi} \cup post \exists (R_{\Phi})$$

$$Find(AG(true), P) = R \text{ true}(P), \text{ Init Process} = \{p_0\}, P_{\{i\}} = \{p_0\}$$

$$\text{Until} (Step_{i+1\{i\}} = Step_{i\{i\}}), \text{ Fixed Point} = R_{\Phi} = \{p_0, p_1, p_2, p_3\}$$

$$R_{\{p_0, p_1, p_2, p_3\}} = \text{Invariant} = (t \wedge m) \vee (s \wedge i) \vee (r \wedge d) \vee (o \wedge e)$$

3 Generation and Analysis of Knowledge Model

This chapter, we automatically generate that knowledge model is to explain. Knowledge model(M) depending on the definition of the unit is supposed that every proposition is the topic. Some words in the document representation with a significant suppose a topic. If we represent a particular set of documents from the relationship between the feature keyword and topics to be able to abstract, this is a very simple but useful model is to gain knowledge. Through natural language processing, we work activities that exist within specific processes and document analysis and text mining techniques to represent the document is extracted to terms with the characteristics of knowledge model(M) is mapped to an item. Document there is a repository of the address the process(P) is a detailed the task(T) name of directory to define that term for the extraction 'Weight Entropy' is shown below.

$$G_i = 1 + \sum_j \frac{P_{ij} \log_2(P_{ij})}{\log_2(n)} \quad \begin{array}{l} a_{ij} = \text{Frequency of terms is presented in the doc., } g_i = \\ \text{Collected in the doc., indicated the frequency of term, } n = \text{The} \\ \text{number of collected doc., } d_i = \text{number of documents with the} \\ \text{term, } p_{i,j} = a_{i,j} / g_i \end{array}$$

Automatically extracted terms to find the usefulness of confidence intervals for the topic at creation of knowledge models in terms(*Document Term*) of the whole document, we calculated from the extracted terms (*Detected Term*) through analysis of the most appropriate derived values (*Roll up Point*).

$AK : Author Keyword [\{Term_n\} \times X] == TM : Text Mining [\{Term_m\} \times (X+i)]$
 while $Author Keyword : AK [\{Term_n\} \times X] == Text Mining : TM [\{Term_m\} \times (X+i)]$

if $AK \neq TM$, $i = 100$, $m : Set of term in document$
 than $(i = + 100)$, Result Return(Roll up rate) for $(i) : Roll up$

$$rate(\%) = \frac{Detected\ Term\ Point : [i \times term]}{Document\ Total\ Term\ point : [m \times Document\ term]}$$

else $(AK = TM \mid Roll\ up\ rate \leq 60)$, Step(end)

In our experiment as shown in fig.3., abstract words and keywords automatically extracted from the concordance between the level should be approximately 80-90% effective when seen as an extract of the entire vocabulary of about 15% of the number of its induction into the most appropriate range. In addition, more than 90% of the keyword automatically in order to induce more than 30% of the full term should be induced to have confirmed that. The performance of general algorithm as Bayesian and weighted formula(*Weight Entropy*) algorithm do not differ. Therefore, how to extract the keywords from documents is very effective for summarizing and defining of complexity document and knowledge model.

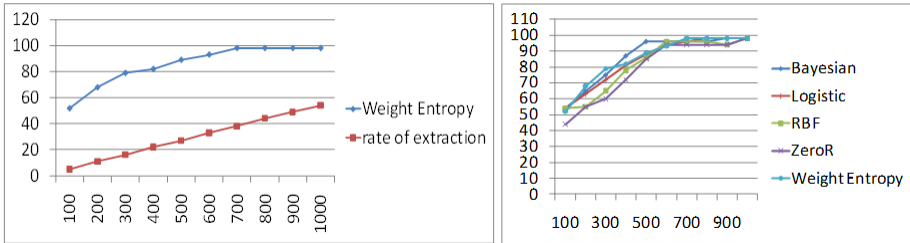


Fig. 3. Weight entropy based algorithm compared with the general mining algorithms, similar results were obtained

In this study, we used effective index that it calculate relationships between topics on knowledge model. Topics for the relationship between knowledge model extracted from the topics on document for the Closeness and Centrality of topic can be obtained through calculation. In this way, articles, patents, etc. The extracted social, economic, and technological knowledge of the topics discussed in the target generation and semantic analysis of the model can be applied. These methods for generating knowledge model can be used effective understanding for useful meaning that the relationship and quickly discovery of recent technical topics[7].

Especially if we could be use to understand for the impact of the various topics and can get a variety of implications. In addition, as shown in Figure 4, we created a visual analysis tool using the knowledge model structural analysis and hierarchical relationships between topics and took advantage of understanding.

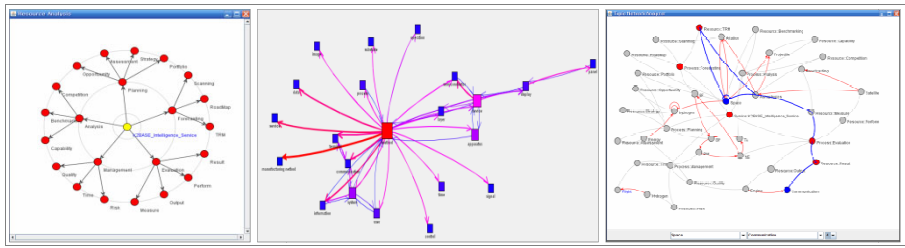


Fig. 4. Keyword's hierarchy and relationship of extracted knowledge model from document can be analyzed using visualization tools

4 Conclusion and Future Work

In this paper, we hoped to generate simple knowledge model for easy interpret and understanding of technology analysts. In this regard, we have suggested extraction algorithm of semantic keyword from complex technical document. To do this, we have defined process model of technology innovation, extracted semantic keyword and relationship property from technical document of each process. The extracted keywords and property is utilized to create and analyze knowledge model. The keyword in knowledge model is a property and semantic keyword that holds in every reachable process on knowledge model. Not only can be used in understanding and analysis of complex knowledge's relationships. For doing so, we let the user focus on some interested parts rather than a whole process state in a semantic model. Obviously, a set of process calculated in this way is a subset of every reachable process. Therefore, generated process and keyword is used to optimize and translate knowledge model. In future research, effective integration method of the extracted knowledge model and it based innovation trends will be applied actual system.

References

1. Barkly Rosser Jr., J.: On the Complexities of Complex Economic Dynamics. *Journal of Economic Perspectives*, 169–192 (1999)
2. Ruan, D., Hardeman, F.: *Intelligent Decision and Policy Making Support Systems*. Springer, Heidelberg (2008)
3. Thierauf, R.J.: *Optimal Knowledge Management*. IDEA Group Publishing, USA (2006) ISBN 1-59904-016-6
4. Devedzic, V.: Knowledge Modeling - State of the art. *ACM Integrated Computer Aided Engineering* 8(3), 25–281 (2001)
5. Clarke, E.M.: *Model Checking, foundations of software technology and theoretical computer science*. LNCS, vol. 136, pp. 54–56 (1997)
6. Jun, S.: Generation of place invariant in knowledge map and networks of business intelligence frameworks. In: *IEEE Proceedings of ALPIT*, pp. 308–313 (2008)
7. Sorenson, O.: Social networks and industrial geography. *Journal of Evolutionary Economics*, 513–727 (2003)

An Application of SCORM Based CAT for E-Learning

HwaYoung Jeong¹ and BongHwa Hong²

¹ Humanitas College of Kyunghee University, Hoegi-dong, Seoul, 130-701, Korea

² Dept. of Information and Communication, Kyunghee Cyber University,

Hoegi-dong, Seoul, 130-701, Korea

hyjeong@khu.ac.kr, bhhong@khcu.ac.kr

Abstract. Many CATs(Computerized Adaptive Testing) focus on analysis and evaluation of the results of learning. Because tests are based on learning materials, however, it is required to develop the learning process in which the learning materials are interlinked with the questions. For this purpose, this study suggests the CAT system which interfaces with SCORM. In this system, LMS processes and operates the learning materials, and LCMS manages the learning contents. The CAT system interlinked with SCORM provided the learners with the learning contents and items through a separate e-learning server.

Keywords: E-Learning, SCORM, CAT, E-Learning Contents, Learning materials, Learning multimedia resources.

1 Introduction

Using the Internet has become part of the everyday experience of millions of people throughout the world [1]. With the rapid development of technologies, the computer has evolved into a tool that can improve the accuracy, efficiency, interface, and feedback mechanism of tests. These developments, although resulting in significant improvements in the management of testing, remained in the format of traditional tests until the rise of computerized adaptive testing (CAT) [2]. CAT uses existing data to streamline and individualize the measurement process. To support selecting items to students in CAT, it has used IRT (Item response theory). IRT is the study of scoring tests and questions based on assumptions concerning the mathematical relationship between the examinee's ability (or other hypothesized traits) and the questions' responses.

Generally, CAT is aimed student's learning level oriented testing. In other words, e-learning for learning contents-based study has been studied separately so far from CAT for e-learning and testing [8, 9]. It is more efficient in e-learning, however, to handle testing together with the learning contents for improvement of students' learning effect. It is because that e-learning is intended for learning through rich web-based contents, and that questions for testing are set in relation with these learning contents.

This study suggests the workflow on application of SCORM-based CAT for e-learning. For this work, the research has to consider the connection between the structure of CAT and learning contents. CAT adopts question banks to store and

manage items, while learning contents use LMS (Learning Management System) and LCMS (Learning Contents Management System) to handle and manage learning materials. It adopts SCORM, an international standard for learning objects, as the system for operation and management of learning contents.

2 Environment of E-Learning System with SCORM, IRT and CAT

E-learning has played an important role in teaching and learning, which has become more and more popular not only in different levels of schools but also in various commercial or industrial companies [3]. One of the most significant developments in the use of e-learning has been the adoption of learning management systems (LMSs) to support the teaching and learning process. LMSs typically offer affordances beyond simple document repositories, by featuring discussion forums, online chat rooms, grade books, and the ability to give automatically marked tests such as multiple choice questionnaires and to track students' performance [4]. Consequently, it needs to manage multimedia based learning contents. LCMS (Learning Content Management System) deals with the process for e-learning system. Moreover, concerning the learning content adaptation to the system, the SCORM (Shareable Content Object Reference Model) standard was used. SCORM references specifications, standards and guidelines developed by other organisations that are adapted and integrated with one another to form a more complete and easier-to-implement model [5].

Moreover, IRT (Item Response Theory) was used to support learning items with difficulty for student. IRT (Item Response Theory) is a widely used theory in education measurement to select the most suitable items for examinees based on individual abilities [6]. In IRT, the 2-parameter logistic which considers the item difficulty and the item discrimination is as follows equation 1 and 2:

$$p(\theta) = \frac{1}{1 + e^{-a(\theta-b)}} \quad (1)$$

where,

- P(θ) = the probability of correct answer for an item,
- a = the parameter of degree of discrimination power,
- b = the parameter of degree of difficulty, and
- θ = the ability of a learner.

The item information function shows how accurate the test items estimate the abilities of examinees across the ability range. The higher item information value indicates more accurately estimated the abilities of examinees.

$$I_i(\theta) = a_i^2 P_i(\theta)1 - P_i(\theta) \quad (2)$$

Different types of computer-based exams have been devised where students use computer systems to complete tasks or respond to questions. The use of computer-based exams to assess IT courses has been used in many places for many years

although rarely for high-stakes purposes [7]. Computer-adaptive testing (CAT) is a method of test administration where the candidate responds directly to questions presented by computer.

3 E-Learning System with SCORM and CAT

3.1 CAT Based Learning Model

The CAT-based learning system suggested in this study interfaces with SCORM. Therefore, learning materials and items should be connected with the learning objects of SCORM. The 2-parameter logistic model was applied to individual adaptive test for learning items. This model gives the questions that are customized to the ability of a learner. If the learner fails to give a correct answer to a question, the system gives the learner with the feedback.

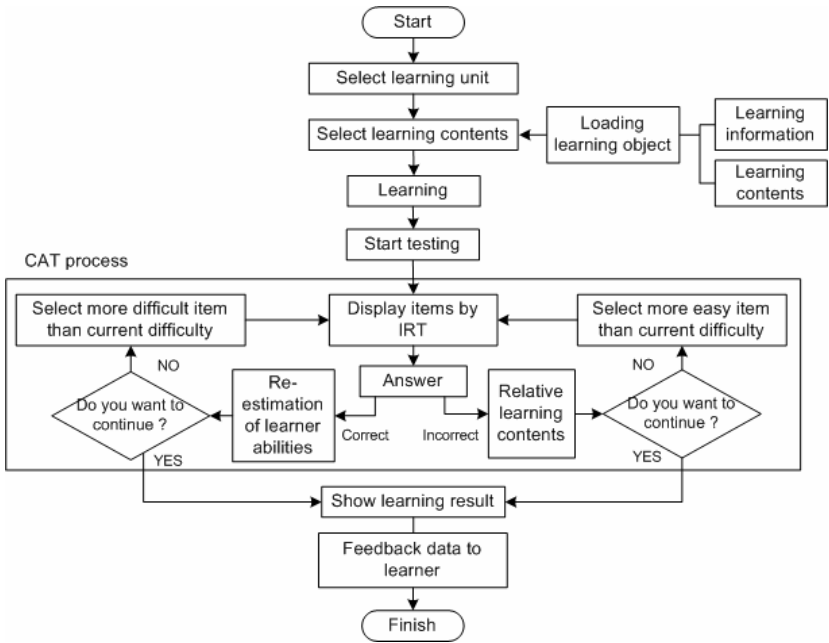


Fig. 1. CAT based learning process

Figure 1 illustrates the study and evaluation process flow. A learner should select a learning unit to start the learning process. Once the learning unit is selected, he or she selects a learning content out of the learning objects in the selected unit. The learning information and contents are loaded from the learning objects in SCORM through LMS and LCMS. As the learning is completed and the questions are made, the learner selects the items according to his or her level of learning. The items are displayed reflecting the ability data estimated with IRT based on the previous learning information of the learner. As the learner makes the correct answer for a question, the

system re-estimates the learner ability, and if he or she wants to continue, displays more difficult item than the current one. If the learner fails to give the correct answer, the system displays the relative learning contents, helping him to understand the question, and selects easier item than the current one. As the learner finishes all the questions, the system provides the learner with the feedback, and finishes the process.

3.2 Adaptation in SCORM

The system environment adopting LMS and LCMS is the required element for e-learning system which must provide large amount of learning information and contents in various categories. Figure 2 illustrates the block diagram of the learning system with the CAT learning process which interfaces with LMS, LCMS and SCORM. The teacher accesses the e-learning server over Internet, or may access and operate LMS directly to manage learning information. LMS has the data storages (item bank, learning information and student information database). LMS also interfaces with API Instance through the SCORM run-time environment data in order to provide the SCORM-based learning information. Learning contents are processed and managed through LCMS.

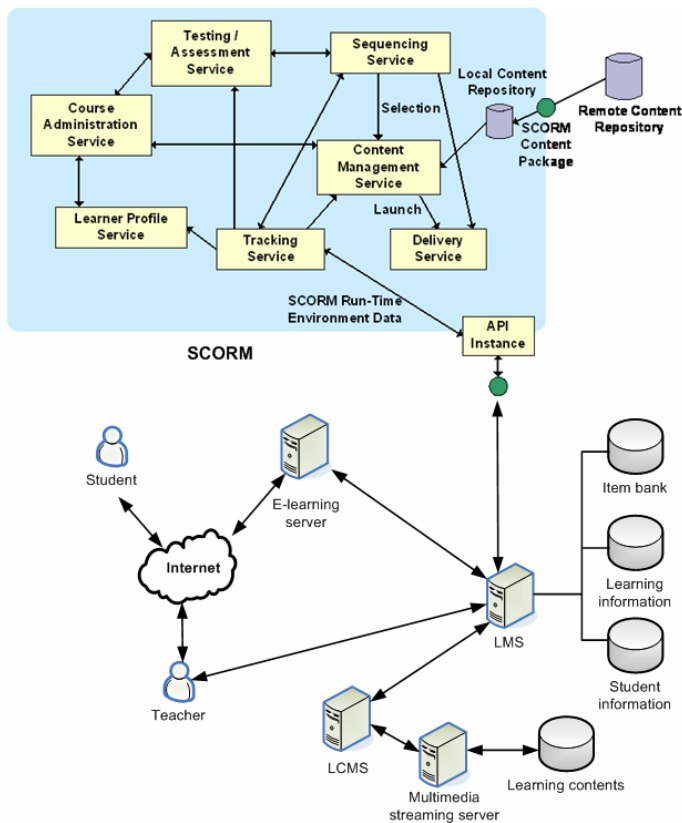


Fig. 2. The environment of e-learning system with LMS and LCMS

4 Experiment and Result

To assess the learning performance of the proposed e-learning system, this research recruited 40 university students who were studying a course in the School of General Education at KyungHee University to take part in this experiment as volunteers, ranging in age from 21 to 23 years old. And the sample system in this research was applied TOEIC simulation test. Among 40 students, there are 20 students who were served as the control group to perform the existent mode without supporting the proposed CAT process, and the remaining 20 students were served as the treatment group to perform the proposed learning mode.

SPSS statistical software was used to analyze the pre-test results for the control group and the treatment group. The pre-test results of independent samples t-tests are provided in Table 1.

Table 1. Group statistics of pre-test results in the control group and treatment group

Group	N	Mean	Std. deviation	Std. Error Mean
Control	20	819.25	35.275	7.888
Treatment	20	806.70	32.092	7.176

Table 2. Independent samples t-test of pre-test results in the control group and treatment group

	Levene's Test for Equality of Variance		t-test for Equality of Means						
	F	Sig.	T	df	Sig. (2- taile d)	Mean differe nce	Std.Er ror Differ ence	95% Confidence Interval of the Mean	
								Lower	Upper
Equ al vari ance assu med	.002	.696	-.276	38.000	.784	-3.100	11.244	-25.863	19.663
Equ al vari ance not assu med			-.276	37.990	.784	-3.100	11.244	-25.863	19.663

Table 2 shows the mean score by independent samples t-test of pre-test result. The result shows that $F=0.002$, the value is less than 0.05. So we are able to select "Equal variance assumed" and $t=-0.276$, $df = 819.25 - 806.70 - 2 = 38$. We can see that the null hypothesis is satisfied under 0.05 of the significant level and 0.784 of the P-value; that is, there is no difference between the control group and the treatment group under the significant level of 0.05 and $P=0.784>0.05$.

5 Conclusion

CAT is widely used for testing based on individual abilities. Since the existing CAT methods focus on analysis of result only, they have been utilized as a part of the learning system. This study includes the CAT-based learning process in the e-learning system, interfacing the system with the entire learning processes, learning information and contents. It is also designed to operate learning information and questions in a group. For this purpose the suggested system contains LMS which handles learning information, and LCMS which handles and manages learning contents, and SCORM which operates and manages learning information in an efficient manner. By operating learning information in the form of SCORM, the system handles learning information, fulfills the operation and management rules, and easily adds, edits, and deletes other learning information in compliance with the rules. In order to verify the suggested system, this study has sampled examination groups with 40 voluntary students. 20 of the students were classified into the control group in which the existing method is adopted, and the remaining 20 students into the treatment group in which the suggested method is adopted. The testing application was the TOEIC simulation test, and was divided into pre-test and post-test. The test result showed that the treatment group acquired higher increase of result than the control group, evidencing that the suggested method was effective for the students.

References

1. Kirkwood, A.: Getting it from the Web: why and how online resources are used by independent undergraduate learners. *Journal of Computer Assisted Learning* 24, 372–382 (2008)
2. Ho, R.-G., Yen, Y.-C.: Design and Evaluation of an XML-Based Platform-Independent Computerized Adaptive Testing System. *IEEE Transactions on Education* 48(2), 230–237 (2005)
3. Chao, R.-J., Chen, Y.-H.: Evaluation of the criteria and effectiveness of distance e-learning with consistent fuzzy preference relations. *Expert Systems with Applications* 36, 10657–10662 (2009)
4. Yasara, O., Adiguzel, T.: A working successor of learning management systems: SLOODLE. *Procedia Social and Behavioral Sciences* 2, 5682–5685 (2010)
5. Rey-López, M., Díaz-Redondo, R.P., Fernández-Vilas, A., Pazos-Arias, J.J., García-Duque, J., Gil-Solla, A., Ramos-Cabrer, M.: An extension to the ADL SCORM standard to support adaptivity: The t-learning case-study. *Computer Standards & Interfaces* 31, 309–318 (2009)

6. Chen, C.-M., Chung, C.-J.: Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education* 51, 624–645 (2008)
7. Paul Newhouse, C.: Using IT to assess IT: Towards greater authenticity in summative performance assessment. *Computers & Education* 56, 388–402 (2011)
8. Computerized Adaptive Testing (CAT) Overview, National Council of State Boards of Nursing (2008)
9. Armstrong, R.D., Jones, D.H., Koppel, N.B., Pashley, P.J.: Computerized Adaptive Testing with Multiple Form Structures, Computerized Testing Report 99-14, A Publication of the Law School Admission Council (May 2006)

Research of the Solar Orbit Compute Algorithm for Improving the Efficiency of Solar Cell

Sang-Bong Byun, Eun-Kyung Kim, and Yong-Hwan Lee

School of Electronic Engineering, Kumoh National Institute of Technology,
1, Yangho-dong, Gumi, 730-701 Gyungbuk, Korea
sb5816@kumoh.ac.kr

Abstract. In the solar industry, the development of tracking photovoltaic power generation devices is underway. So accurate and low-priced solar position tracking devices are very important to improve the economical efficiency and lower the invest price. This research is concerned with solar position algorithm.

Keywords: Solar Cell, photovoltaic, Solar Orbit Compute Algorithm.

1 Introduction

Our primary goal in the 21th century is global warming.[1] However most of the energies are from fossil fuel energy. Fossil fuel is a major contributor to global warming.[2] The solar photovoltaic does not greenhouse gases emissions. Solar industry is in the limelight. Generation efficiency is the most important issues in the solar photovoltaic. Solar cell is perpendicular to sun are most efficient at photovoltaic power generation. The Solar Orbit Compute Algorithm is more efficient than the sensor tracking.[3] In this paper, design of the solar orbit compute algorithm for improving the efficiency of solar cell and verification.

2 Compute Position of the Sun

The position of the Sun is computed just like the position of any other planet, but since the Sun always is moving in the ecliptic, and since the eccentricity of the orbit is quite small, a few simplifications can be made. Therefore, a separate presentation for the Sun is given.

The primary orbital elements are here denoted as:

e = eccentricity (0=circle, 0-1=ellipse, 1=parabola)

M = mean anomaly (0 at perihelion; increases uniformly with time)

E = eccentric anomaly

d =day

y =year

m =month

Orbital elements of the Son:

$M = 356.0470 + 0.9856002585 \cdot d$

$d = 367 \cdot y - 7 \cdot (y + (m+9)/12) / 4 + 275 \cdot m/9 + D - 730530$

Compute the eccentric anomaly E from the mean anomaly M and from the eccentricity.[4][5]

$$E = M + e(180/\text{deg}) \cdot \sin(M) \cdot (1.0 + e \cos(M)) \quad (1)$$

Then compute the Sun's distance r and its true anomaly v .

$$xv = r \cdot \cos(v) = \cos(E) - e \quad (2)$$

$$yv = r \cdot \sin(v) = (1.0 - e^2)^{1/2} \cdot \sin(E) \quad (3)$$

$$v = \text{atan2}(yv, xv) \quad (4)$$

$$r = (xv^2 + yv^2)^{1/2} \quad (5)$$

Formulas (5),(6),(7) is a function that converts an x, y coordinate pair to the correct angle in all four quadrants.

$$\text{atan2}(y, x) = \text{atan}(y/x) \quad \text{if } x \text{ positive} \quad (6)$$

$$\text{atan2}(y, x) = \text{atan}(y/x) \pm 180 \text{ degrees} \quad \text{if } x \text{ negative} \quad (7)$$

$$\text{atan2}(y, x) = \text{sign}(y) \cdot 90 \text{ degrees} \quad \text{if } x \text{ zero} \quad (8)$$

Convert $\text{lonsun}(v + w)$, r to ecliptic rectangular geocentric coordinates x_s, y_s .

$$x_s = x_e = r \cdot \cos(v + w) \quad (9)$$

$$y_s = r \cdot \sin(v + w) \quad (10)$$

Sun always in the ecliptic plane, z_s is of course zero. x_s, y_s is the Sun's position in a coordinate system in the plane of the ecliptic. To convert this to equatorial, rectangular, geocentric coordinates.

$$y_e = y_s \cdot \cos(\text{ecl}) \quad (11)$$

$$z_e = y_s \cdot \sin(\text{ecl}) \quad (12)$$

Finally, compute the Sun's Right Ascension (RA) and Declination (Dec).

$$\text{RA} = \text{atan2}(y_e, x_e) \quad (13)$$

$$\text{Dec} = \text{atan2}(z_e, (x_e^2 + y_e^2)^{1/2}) \quad (14)$$

3 Verification

Perform the experiment for efficiency test of the solar orbit compute algorithm. Used for experiment of solar altitude/azimuth compute tool in Korea Astronomy and Space Science Institute.(KASI) Result of experiment, maximum error has 0.008° and 0.006° . I confirmed that it is good at real-time solar tracking system because it can improve economical efficiency and it has low time complexity and low spatial complexity. Table 1 shows result of experiment.

Table 1. Result of experiment

Heading level	This Paper	KASI
Time	2011/03/20 12:00:00	2011/03/20 12:00:00
Latitude	37 00 00	37 00 00
Longitude	127 00 00	127 00 00
Right Ascension	023 56 54.47	023 56 54.76
Declination	163 54 56.78	163 54 56.00

4 Conclusion

Until now, mankind has depended on fossil fuels. It is mainly due to the increase in **carbon dioxide** levels. Solar is an infinite resource. The solar photovoltaic used the solar. They also cut fossil fuel consumption and greenhouse gases. The solar photovoltaic is green energy because it does not greenhouse gases emissions. Solar cell is perpendicular to sun are most efficient at photovoltaic power generation. Sensor-based solar tracking system is inefficient because it will probably have malfunction in a cloudy weather. Solar orbit compute system to calculate the exact orbit the sun. So it tracking the solar orbit in all weathers. Research and verification for the solar orbit compute algorithm.

References

1. Kim, D.-H.: Global Warming Effect on Marine Environments and Measure Practices against Global Warming. *Kosomes* 16(4), 421–425 (2010)
2. Rogner, H.-H., Sharma, D., Jalal, A.I.: Nuclear power versus fossil-fuel power with CO₂ capture and storage: a comparative analysis. *International Journal of Energy Sector Management* 2(2), 181–196 (2008)
3. Choi, J.-S.: Development of Automatic Tracking Control Algorithm for Efficiency Improvement of PV Generation. *KEA* 59(10), 1823–1831 (2010)
4. How to compute planetary positions,
<http://www.stjarnhimlen.se/comp/ppcomp.html#5>
5. Seo, M.-H.: A Development of the Solar Position Algorithm for Improving the Efficiency of Photovoltaic Power Generation. *KIIT* 6, 46–51 (2009)
6. Korea Astronomy and Space Science Institute (KASI),
<http://astro.kasi.re.kr/Life/SolarHeightForm.aspx?MenuID=108>

A Learning System Using User Preference in Ubiquitous Computing Environment

HwaYoung Jeong¹ and BongHwa Hong²

¹ Humanitas College of Kyunghee University, Hoegi-dong, Seoul, 130-701, Korea

² Dept. of Information and Communication, Kyunghee Cyber University,

Hoegi-dong, Seoul, 130-701, Korea

hyjeong@khu.ac.kr,

bhhong@khcu.ac.kr

Abstract. In this article, a learning system in ubiquitous computing environment is proposed that considers the preference of each student. The system provides information about the learner's preferred learning unit and their score, thus allowing learners to set up an optimized learning environment. A topic preference vector was applied to calculate the learner's preferences, and the learning units were used as each vector value in the system. For experiment of the proposed system, we applied it to English learning system. And we could show the system is useful for learning and convenient for using to learner.

Keywords: U-learning system, Ubiquitous computing, User Preference, Learning process.

1 Introduction

Recently, there has been a change in the nature of ubiquitous computing. This is due to the advantages of ubiquitous; for instance, such environments are always available, the portable and embedded devices they employ are completely connected and intuitive, and they do not require any pre-deployed infrastructure [1]. Therefore, the proliferation of portable electronic devices and wireless networking is creating a change from e-learning to u-learning [2, 3].

In education, a teacher's role is not just that of an information provider; he/she becomes an advisor or a guide. For their part, learners try to construct their own knowledge; in short, they learn to learn. "Personalized learning" suggests a model that reflects the characteristics of each learner. Chen, Lee and Chen [4] proposed a personalized learning system based on item response theory (IRT), called PEL-IRT, to provide web-based personalized learning services. Furthermore, Chen [5] proposed a personalized learning system that can cultivate learning abilities using a self-regulated learning assessment mechanism that provides immediate feedback response to learners and a heteronomy mechanism that comes from the teacher's reminders.

In this study, we developed a learning system in ubiquitous computing environment. To support efficient learning process to learner, we considered learner's preference with learning units(topics) and difficulty levels, that the system is similar a personalized learning system. In order to apply a learner's preference, we was used a

topic preference vector. To provide learners with constant learning interesting, the system helped to find the learning units and difficulty that learner prefers.

2 Relevant Works

Ubiquitous computing has been investigating since 1993. The widely used definition of ubiquitous computing is the method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user [6]. Ubiquitous environments are increasingly becoming heterogeneous and service rich domains where a diversity of communication devices, such as laptops, portable digital assistances, digital home appliances, automotive telematics, and various sensors, are interconnected each other across different platforms and networks [7]. That is, the emergence of ubiquitous computing technologies was able to shift various objects and people from the physical world into an electronic space [8, 9, 10]. Ubiquitous computing refers to the creation of a computing environment that enables people to get information and to connect to networks anywhere at anytime [11]. From this phenomenon, building a ubiquitous learning environment requires a “ubiquitous” learning device accessible by every learner all times. Consequently, a cell phone is the only candidate among various mobile devices, such as a PDA, tablet PC or laptop [12]. Christian, et al [13] suggested effective computer-classroom integration training is still needed for faculty. They also described the importance using the devices such as PDAs, MP3 players and mobile phones in learning environment. Hwang, et al [14] proposed to develop a context-aware u-learning environment to assist novice researchers in learning a complex science experiment. The research described that researchers have to encourage to develop more related u-learning systems, and to carefully examine their impacts on learner’s learning. To support learning materials efficiently, learning system applied the course with learner preference. The system [15, 16, 17] consider learner/user preferences, interests, and browsing behaviors for personalized services [18]. Cristina, Gladys, & Eva [19] developed a system for selecting suitable learning resources for a given topic according to a learner’s characteristics (knowledge level, learning style and references) and the characteristics of the resources (learning activities and multimedia format).

3 Learning System with Learner Preference

3.1 Learner Preference

In this paper, we used topic preference vector in topic map to calculate learners’ preferences and it used to calculate and suggest a learner-preferred the learning unit rank. The number of topics that interest a learner is limited and affects the learner’s preference for a course of study. The learner’s preference measured with a preference vector represents a set of topics and difficulties. The set of topics, $T=[T(1),...,T(m)]$, has m topics. The learning preference is expressed as a set of vector values of topics and difficulties, and each vector value can be expressed as follows.

[Definition] When $T(x)$ is a topic preference vector value and $D_x(y)$ is a difficulty preference vector value of x 'th contents and items, a learning preference considering topic and difficulty can be expressed in the following equations.

$$\text{Topic preference vector } T(x) : \sum_{x=1}^m T(x) = 1$$

$$\text{Difficulty preference vector } D_x(y) : \sum_{x=1}^m \sum_{y=1}^5 D_x(y) = 1$$

$$\text{Learning preference} = \{(T(x), D_x(y)) \mid \text{where } x = 1 \dots m, y = 1 \dots 5\}$$

Learning preference is expressed as a set of the topic preference vector T and the difficulty preference vector D of item x in the related learning section; x is the number of items in the learning section and y is a five-step difficulty level. Table 1 displays the learning sections and difficulty levels of the English learning course implemented in this paper. In order to calculate the learning preferences of a learner, we created a hypothetical learner's historical learning data.

Table 1. Proposed English sample course with learning section

<i>Learning units (Topics)</i>	<i>Difficulty level</i>	<i>Selection count</i>
Chapter 1	Excellent	5
	Advanced	7
	Intermediate	3
	Intermediate Low	1
	Beginner	0
	Total	16
Chapter 2	Excellent	3
	Advanced	4
	Intermediate	8
	Intermediate Low	2
	Beginner	1
	Total	18
:	:	:
:	:	:

This example presents two topics chosen by a learner, “Chapter 1” and “Chapter 2.” If the learner was interested in “Chapter 1” 16 times and in “Chapter 2” 18 times, the topic preference vector for the learner is [0.47,0.53], indicating that the learner had a greater preference for Chapter 2. The difficulty levels chosen by the learner for Chapter 1 were “Excellent” (five times), “Advanced” (seven times), “Intermediate” (three times) and “Intermediate Low” (one time). Thus, the learner’s preference vector for section difficulty is [31,33,19,6,0]. As a result, we can determine that the learner prefers the “Advanced” difficulty level.

3.2 Design of Learning System

Figure 1 shows the sequence diagram of the system by Figure 1. After connect the system and login, LearningConstruction gets the learning course’s information by LearningCourse and request to get learner’s preference of the course to Preference process. Preference process accumulates the preference data and calculated preference using topic preference vector. After the process, LearningConstruction makes a course recommendation through the preference data and shows it to learner. Learner can select the course items to refer a recommendation from the system.

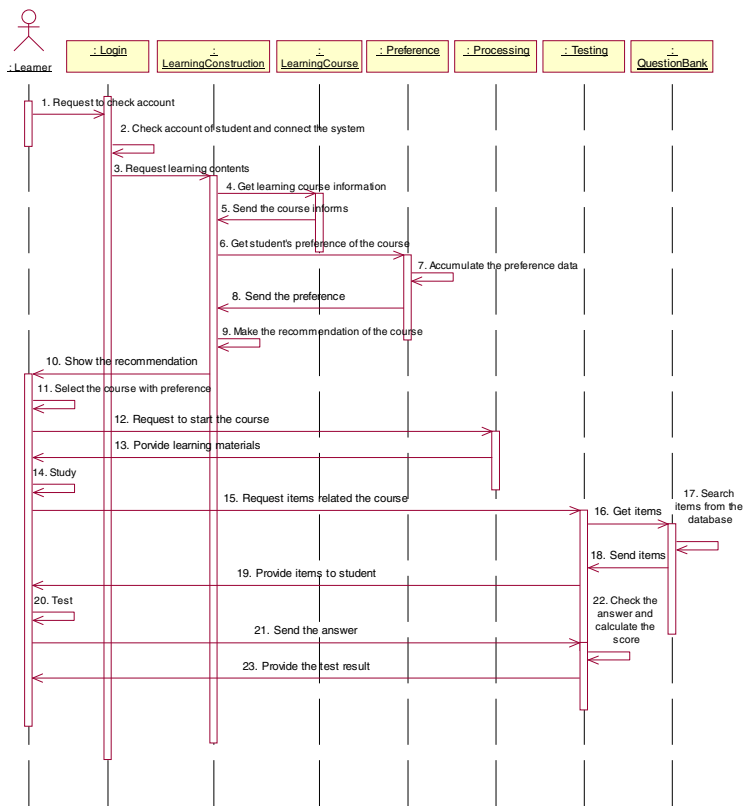


Fig. 1. A sequence diagram of the system

3.3 Implementation of U-Learning System with Learning Preference

We implemented an English learning system according to the course ontology model as shown figure 2. The system can be operated on any PDA, PDA phone, or smart-phone.

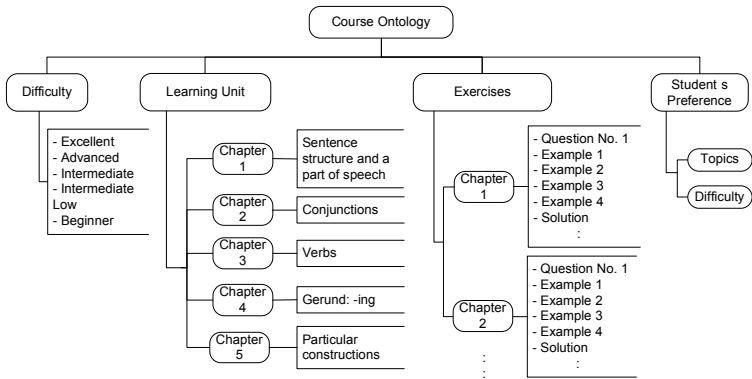


Fig. 2. Course ontology for English learning

After login, the system provide learning course with topic and difficulty status on the screen to learner before learning process, as shown Figure 3 (a). And it also provides a recommendation for the course using calculation of topic preference vector, as shown Figure 3 (b).

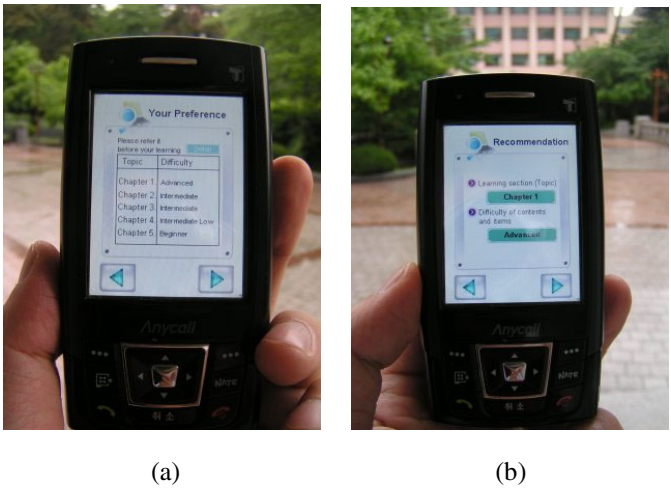


Fig. 3. U-learning system for English course with preference recommendation. (a) Screen of topic and difficulty in the course. (b) Course recommendation related topic preference vector.

4 Experimental Results

Experiments were conducted with university learners enrolled in Kyunghee University. We selected 30 applicants who were interested in studying English and improving their English fluency. Experimental group were divided into two groups

(the control group and the Treatment group) with similar pre-test score distribution. Table 2 shows that the mean score and standard deviation between the control group and the experimental group are almost the same.

Table 2. Pre-test results in the control group and treatment group

Group	N	Mean	Std. deviation	Std. Error Mean
Control	15	17.3333	3.28778	.84890
Treatment	15	17.8667	3.06749	.79202

Table 3 indicates that there is no significant difference under the significant level of 0.05 and $P=0.057>0.05$.

Table 3. Paired sample t-test of the control group between the pre-test and post-test

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Mean		t	dt	Sig. (2- tailed)
				Lower	Upper			
Pre- Post test	-.60000	1.12122	.28950	-1.22091	.02091	-2.073	14	.057

Table 4 shows the paired sample t-test results of the treatment group. The mean score of the post-test increased about 2.5 points. Table 6 indicates that the difference in mean scores is significant under the significant level of 0.05 and $P=0.000<0.05$. Therefore, we can conclude that the proposed system is quite effective.

Table 4. Paired sample statistics of the treatment group between the pre-test and post-test

	Mean	N	Std. Deviation	Std. Error Mean
Pre-test results	17.8667	15	3.06749	.79202
Post-test results	20.3333	15	2.46885	.63746

5 Conclusion

We proposed a u-learning system that accommodates the learning preferences of learners and allows recommendation of the course to them. The lesson content is

constructed according to specified difficulty levels, and the topic preference vector is used to calculate the learner's learning preference. By accessing and managing the learner's historical data, areas requiring additional attention are analyzed, and the learner can refer to this information when selecting the lesson. This function makes it possible to improve the effectiveness and the score of the entire lesson. Using the proposed method, we implemented English learning system. And we also conducted an experiment that implemented this system with 30 university learners divided into two groups, a control group and a treatment group. The results showed that the proposed system improved the effectiveness of learners' learning.

References

1. Boukerche, A., Ren, Y.: A trust-based security system for ubiquitous and pervasive computing environments. *Computer Communications* 31, 4343–4351 (2008)
2. Lee, M.J.W., Chan, A.: Exploring the potential of podcasting to deliver mobile ubiquitous learning in higher education. *Journal of Computing in Higher Education* 18(1), 94–115 (2005)
3. Wurst, C., Smarkola, C., Gaffney, M.A.: Ubiquitous laptop usage in higher education: Effects on student achievement, student satisfaction, and constructivist measures in honors and traditional classrooms. *Computers & Education* 51, 1766–1783 (2008)
4. Chen, C.-M., Lee, H.-M., Chen, Y.-H.: Personalized e-learning system using Item Response Theory. *Computers & Education* 44, 237–255 (2005)
5. Chen, C.-M.: Personalized E-learning system with self-regulated learning assisted mechanisms for promoting learning performance. *Expert Systems with Applications* 36, 8816–8829 (2009)
6. Wanga, H., Zhang, Y., Cao, J.: Access control management for ubiquitous computing. *Future Generation Computer Systems* 24, 870–878 (2008)
7. Jung, J.-Y., Park, J., Han, S.-K., Lee, K.: An ECA-based framework for decentralized coordination of ubiquitous web services. *Information and Software Technology* 49, 1141–1161 (2007)
8. Fleisch, E., Tellkamp, C.: The business value of ubiquitous computing technologies. In: George, R. (ed.) *Ubiquitous and Pervasive Commerce*, pp. 93–114. Springer, London (2006)
9. Jessup, L.M., Robey, D.: The relevance of social issues in ubiquitous computing environments. *Communication of the ACM* 45(12), 88–91 (2002)
10. Sugiyama, M.: Security and privacy in a ubiquitous society. *I-Ways, Digest of Electronic Commerce Policy and Regulation* 27(1), 11–14 (2004)
11. Kim, C., Oh, E., Shinc, N., Chae, M.: An empirical investigation of factors affecting ubiquitous computing use and U-business value. *International Journal of Information Management* 29, 436–448 (2009)
12. Chen, G.D., Chang, C.K., Wang, C.Y.: Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques. *Computers & Education* 50, 77–90 (2008)
13. Wurst, C., Smarkola, C., Gaffney, M.A.: Ubiquitous laptop usage in higher education: Effects on student achievement, student satisfaction, and constructivist measures in honors and traditional classrooms. *Computers & Education* 1, 1766–1783 (2008)

14. Hwang, G.-J., Yang, T.-C., Tsai, C.-C., Yang Stephen, J.H.: A context-aware ubiquitous learning environment for conducting complex science experiments. *Computers & Education* 53, 402–413 (2009)
15. Lee, M.-G.: Profiling students' adaptation styles in web-based learning. *Computers and Education* 36, 121–132 (2001)
16. Papanikolaou, K.A., Grigoriadou, M.: Towards new forms of knowledge communication: The adaptive dimension of a webbased learning environment. *Computers and Education* 39, 333–360 (2002)
17. Tang, C., Lau Rynson, W.H., Li, Q., Yin, H., Li, T., Kilis, D.: Personalized courseware construction based on web data mining. In: *Proceedings of the First IEEE International Conference on Web Information Systems Engineering*, vol. 2, pp. 204–211 (2000)
18. Chen, C.-M.: Intelligent web-based learning system with personalized learning path guidance. *Computers & Education* 51, 787–814 (2008)
19. Carmona, C., Castillo, G., Millan, E.: Discovering Student Preferences in E-Learning. In: *Proceedings of the International Workshop on Applying Data Mining in e-Learning* (2007)

Introduction of the Art of Korean Traditional Culture: Multimedia Based Pansori

Dong-Keon Kim

Humanitas College of Kyunghee University, Hoegi-dong, Seoul, 130-701, Korea
dehi@khu.ac.kr

Abstract. This research shows Pansori that is an art of traditional culture in Korea. Pansori is too able to call the one of convergent art. Because the actor has to play an active part with sound and sing the story in the Pansori. The research is aimed to introduce Pansori exactly.

Keywords: Pansori, Korean traditional culture, Art of culture, Multimedia based culture.

1 Introduction

This search would like to show the Art of Korean Traditional Culture, Pansori. The Pansori is not general concept even though the Pansori[1-4] is Korean Tradition Culture. So there are many works to expand awareness about that with watching movie, going a public performance, or listing the sounds. So it is very important work to announce it to popularize.

2 What Is Pansori?

The Pansori[1-4] is a vocal music of Korea. It constitutes a ‘Changja’ (singer) who hold a ‘buche’ (folding fan) and a ‘gosu’ (drummer) who play the ‘buk’ (Korea traditional drum). As performance way it composes long plot, ‘chang’ (song), ‘anlri’ (narrative lyrics) and ‘neoreumsae’ (mimetic gesture). When pansori performed, changja sing a song which call ‘danga’ (intro song) at first. It is for tune their voice and know the moods of their audience. After danga changja starts to sing pansori (which we call ‘bonsaga’(the main song)). When Pansori played, it has two way of rendering partly or wholly according to their react of audience or intention of the performer. In the past, pansori wanchang’ which means to sing a whole song in one performance (it takes approximately 3 to 7 hours), has been performed only on a special day. Usually it performs partly which the performer choose specific point base on their crowd taste. The gosu who sit from next changja, undertakes accompaniment as a rule. He enlivens between audience and changja as use ‘chuimsae’ (sounds and words of encouragement) even take a counterpart in the story. Figure 1 shows a public performance of Pansori in a theater.



Fig. 1. A public performance of Pansori

In the pansori field, the audience has different characteristics from general western audience. As doing the 'chuismae', they actively participate on the performance as same as 'gosu'. It makes interactive communication which helps to build Pansori and complication of it. For this reason, the person who complete these things perfectly and has a high value of attention generally call 'guimungchang' (a person who is good at hearing). Yet most important quality is the capacity of changja. Shin Jaehyo who makes a synthesis of pansori articles and educates 'gwangdea' (performer) even tried to specialization of 'chang' based on the modern playwriting and direction theory, mentioned about pansori gwangdea's capabilities and its function in pansori.

Being Pansori performer is difficult and more difficult. It needs four skills first looking second influential speech, third attractive vocal sound and forth gesture. His Gesture must look good, express variety of life and feeling of man also the change of nature. That's why the gesture is so difficult. His Attractive vocal sound should distinguished from five sound be and changeable six musical scales. From this work the body must fit sound for everybody. So it is also hard working.

His Influential speech must use treasure words and decorated by meaningful adjectives. It might be same significance a bride who wait groom at the first night. Also it must make people smile at one time and just single word changes people's emotion. Finally looking or shape is destiny, so it's never changeable. Therefore who has good looking should be a singer.

According to the article, gwangdea needs four conditions; 'inmul' (appearance), 'saseol' (ability to perform article), 'dukuem' (to get attractive and suitable voice) and finally 'neoreumsae' (impromptu gesture).

Among these four conditions, the inmul' (appearance) never changeable. Therefore other three things can be practicable. Basically, 'saseol' means script of pansori which include literary construction of story and lyrics. 'dukuem' means that accomplishment of musicality, and 'neoreumsae' means temperament of dancing in the play filed. According to these four conditions, Shin Jaehyo is the person who clearly understands pansori's multi-synthetic art capacity. Namely, he insisted that only performer who has these four qualities could be impressed to their audience. From his mention, defi-

nite pansori is a multi-synthetic art which contains high quality of the voice, excellent dramatic literature capacity and imaginative temperance of the play and dance.

In the mean time, the title “pansori” combine “pan” (place) and “sori” (song). Until 19th century, pansori is referred to by various names such as ‘sori’, ‘gwangdeasori’, ‘tayrung’, ‘gukga’, ‘changak’, etc. In modern term, it fixed ‘pansori’. ‘Pan’ refers a place where to perform or acting and ‘sori’ refers music. So pansori means that in a specific place which, crowd with audience, play a musical performance.

2.1 Formation and Development Period (17~18th)

Generally, we assume that around late 17th century, pansori started to format. It's base on Jeolla province and Chungcheong province's traditional public performance manner. Therefore the singer and the people who enjoy are public. Around 18th century, pansori influenced to ‘yangban’ (the two upper classes of old Korea) so it extended it's filed and participator. In 1754, after sightseeing Honam district Yu Jinhan wrote “Gasachunhyangibaekgu”(). According to this poem pansori was the populace's art manner inherently.

However after it takes notice from ‘yangban’, it accepts to the material of Chinese poem. It means around 18th century pansori fairly achieved literary and musical capacities. Around this era Ha Handam, Choe Heondal, U Chundae were ‘myeongchang’ (famous performer).

2.2 Prosperity Period

This period appears a lot of ‘myeongchang’ and accepts gladly from not only ‘jungin’ (middle class) but also yangban’. As a result, their social reputation also improves at this period. Because of development and maturity of ‘jangdan’ (rhythm), ‘akjo’ (tone) and ‘deoneum’ (ad-lib narration) it is classified some schools and extended repertory. Literarily, this period, lots of pansori subsection novel were published. In the early 19th century, we call the former term eight ‘myeongchang’s era. Gwon Samdeuk, Song Heungrok, Yeom Gyedal, Mo Heunggap, Ko Sugwan, Sin Manyeop, Kim Jecheol, Ju Deokgi, Hwang Haecheon, Song Gwangrok are the former term eight myeongchang. They developed their own ‘deoneum’ and melody so it enhanced artistry of pansori.

The late 19th century was the later term eight myeongchang's era. Park Yujeon, Park Mansun, Lee Nalchi, Kim Sejong, Woo Songryong, Jung Changeop, Jung Chunpung, Kim Changrok, Jang Jabaek, Kim Chaneop, Lee Changyun and others participate actively. They followed the former term eight myeongchang's skills, thereupon they produced musical classics. For this reason, pansori became popular from populace to middle class even the nobility.

In this period, Shin Jaehyo supported the performers actively, and readjusted former 12 ‘madang’ (episode) to 6 madang. From this arrangement pansori got rationality idiomatic phrases. On the other hands, it made weak about democratic vitality. And it became focus on ‘Seopyeonje’ which has lamentation character. but also national competition which call ‘Jeongukdaesaseumnori’ started at this time. Especially, late 19th century, it was boom to be a sponsor of pansori. Because of it, pansori spreaded

out to other class. And pansori schools also classified as three big schools such as 'dongpyeonje', 'seopyeonje' and 'junggoje'.

2.3 Deterioration Period

The late 19th century, *pansori* met the prosperity. Hundreds of *myeongchang* and unique episode created and performed. As other arts also had same waives, *pansori* also became deteriorate in 20th century. In this term, among *pansori* 12 *madang* only 5 *magang* transmission, while other just missing the song or article. The early 20th century, so call former term of 5 *myeongchang*'s era. Park Gihong, Kim Chang hwan, Kim Chaeman, Jun Doseong, Song Mangap, Lee Dongbaek, Kim Changryong, Yoo Seongjun, Jung Jeongryeol and other *myeongchang* were participated. Early 20th century, modern western style theaters which are named '*Hyeomnyulsa*' and '*Wongaksa*' were built.

For accommodate in this circumstance, *pansori* changed the basic concepts like divide the cast. as a result, it became Korean style musical which call '*Changguk*'. As gramophone popularized, it were possible and boomed to publish record. Especially, because of the realistic reason, lamentation rhythm is amplified. During Japanese imperialism period, including Seoul and other major city established '*gwonbeon*' (Korea geisha institution), and this institution started to educate *pansori*. As a result, lots of *myeongchang* appeared one after another.

In 1964, after the Independence, the government enacted the law which called '*juyomhyeongmunhwajae*' (important intangible cultural asset) for authorization, preservation and transmission. In this period Kim Yeonsu, Yim Bangul, Park Chowol, Kim Sohui, Park Bongsul, Kang Dogeun and other lots of *myeongchang* maintained *pansori* and educated new generation.

3 Musical Composition Element of Pansori

Generally, Pansori music composes 'sungeum' (vocal sounds), 'jo' (tempo or time), 'jangdan' (rhythm, beat or time). First, 'sungeum' is an wave of vocal sound. The basic 'sungeum' of pansori sound is gruff and hoarse sound, which called "surisung". If the sound is more clear and soft than former one we call "chungusung" and we regard it as the best sound. In the 'sungeum' filed, supplication regards as the most important thing. Secondly, 'jo' is a shapes of tone and rhythm. Base on the 'jo' it classifies 'Gyemyeonjo', 'Ujo' and 'Pyeongjo'. First, 'Gyemyeonjo' usually use sadness part. It is similar to western musical term, minor. 'Ujo' uses mostly magnificent and courageous part. 'Pyeongjo' uses peaceful part. These last two things similar to western musical major. Finally, 'jangdan' can compare with western musical term 'time' in generally. Suitable 'jangdan' enhances pansori's ability to express. But in Korean music regards 'jangdan' as the most important part. So if someone misses it, it treats as the worst song. Therefore pansori also regards 'jangdan' importantly. In pansori jangdan' it can divide 'jinyang', 'jungmori', 'jungjungmori', 'jajinmori', 'hwimori', 'eonmori'. Jinyang is the slowest clapper which slow 6 clappers refer one unite, so this 4 unites refer one jangdan. Generally, it uses free scene or very lamentation scene and express solemn scene.

Jungmori usually regards medium tempo rhythm which, 12 clappers refer one jangdan. It uses descriptive way or explanation scene. Jungjungmori which is 12 clappers refer one jangdan, is faster than jungmori. It uses a little bit exciting scene or very sad scene mostly. Jajinmori, which faster than jungjungmori, considers three divisions of one clapper four rhythms. It use as things express in a row, or tense scene mostly. Hwimori is faster than jajinmori. This rhythm seems to be hasten or hurry, so we call “Hwimori” (means hurry or hasten in Korean). It composes very fast 4 clappers. Generally, it uses express events or things in a row. Eonmori composes 5 clappers put together which, means 10 clappers. Two divisions of one clapper or three divisions of one clapper also exist. It describes a meaningful character or strange event.

On the other, according to the transmission system, pansori can divide as three schools. Dongpyeonje where eastern bounds of Seomjin river such as Namwon, Sunchang, Gokseong, Gurye focuses on Ujo and it follows Song Heungrok’s skill. This school does not many use tricks and limits their feelings as possible as. Use “tongsung” which means clear and loud vocal sound. And they make a clear and short ending sound each of the ends. Seopyeonje where western bounds of Seomjin river such as Gwangju, Naju, Damyang, Hwasun, Boseong follows Parkyujun’s skill. This school focuses on sorrow and sadness mostly, so they use gyemyeonjo. Generally they are interested in the trick of their voice, so it makes long ending sound also very sophisticate on neoreumsae. Junggoje where Chungcheong and Gyeonggi province bases on Yeom Gyedal, Kim Seongok. In these days it discontinues to maintain. So we don’t know exact musical characteristics.

4 Masterpiece of Pansori

pansori has 12 madang (episode). Because of Song Manjae and Jung Nosik wrote about history of ‘Joseon’ opera which is the oldest record of pansori episode.

According to this book, around early 19th century when pansori already popular to masses 12 madang were existed. But around mid-19th century, from Sin Jaehyo who wrote a book Pansorisaseoljip (the narration and lyrics of pansori collection) arranges it as 6 madang. In early 20th century, Lee Sunyu published a book named Ogajunjip (the five songs collection) which introduces five episodes of pansori. In this paper I mention 12 madang shortly.

4.1 Derivative of Chunhyangga

The real story of the pansori Chunhyangga, and the kodae sosol Chunhyangjeon, which the manga Legend of Chun Hyang is based on, is different from the comic. Chun-hyang in Chunhyangga is commonly known as a daughter between Wol-mae and Seong champan (Seong being a surname and champan being a title of government position in Joseon Dynasty Korea). Wol-mae is a retired kisaeng and, therefore, Chun-hyang is considered as a low class being (this becomes an issue when Mong-ryong, a yangban (the highest class) (and son of Lee Han-rim, satto of Namwon), tells his parents he promised to marry Chun-hyang and his parents get angered by the fact he wishes to marry a woman of lower rank).

After Mong-ryong's departure due to his father's promotion, Byeon Satto (Byeon being a surname and *satto* being a title of government position, equivalent to

"mayor") comes to the town. Byeon becomes unpopular figure largely due to his greediness and corruption. Byeon also orders to bring Chun-hyang out to make him happy. As Chun-hyang still misses Mong-ryong, she refuses his order and gets jailed. Figure 2 shows movie of Chunhyang in 2000.



Fig. 2. Chunhyang (2000 film). Chunhyang is a 2000 KoreaKorea, directed by Im Kwon-taekIm Kwon-taek and starring Lee Hyo-jeong and Jo Seung-wooJo Seung-woo. It was entered into the 2000 Cannes Film Festival.

4.2 Derivative of Simcheongga

Simcheongga is one of the five surviving stories of the Korean pansori storytelling tradition. The other stories are Chunhyangga, Heungbuga, Jeokbyeokga, and Sugungga. From the National Changguk Company of Korea website: "Simcheong-ga is a story about a girl Sim, Cheong, and her father, Hak-Gyu Sim, who is called Sim-Bongsa(a blind person) by everyone. Sim-Bongsa is blind and to be cared for totally by his daughter Sim, Cheong. The story is filled with sadness, though humor enters occasionally to give balance. "Cheong's mother dies at childbirth and her blind father is left with his daughter, who cares for him with the utmost sincerity and devotion. One day, Sim-Bongsa falls into a ditch but is rescued by a Buddhist monk who tells him that Buddha would restore his sight if he donated three hundred bags of rice to the temple. When Cheong learns that some sailors were offering any price for a virgin sacrifice, she offers herself for three hundred bags of rice. The sailors wanted to sacrifice a virgin to the Yongwang (the Dragon King of the Sea) in order to placate him to guarantee the safety of their merchant ships wherever they sailed. "After being tossed into the sea, she finds herself in the palace of the Dragon King of the Sea who, deeply moved by her filial piety, sends her back to earth wrapped in a lotus flower, which is carried to an emperor's palace. The emperor falls in love with Cheong and makes her his empress. The empress later holds a great banquet for all the blind men of the kingdom with the hope that she would be able to find her father again. When Sim-Bongsa finally appears at the banquet, he is so shocked upon hearing his daughter's voice again that his sight is suddenly regained." Figure 3 shows story and play the Simcheongga.



Fig. 3. Simcheongga. (a) this is one of a scene in the story of Simcheongga. (b) a public performance about Simcheongga.

4.3 Derivative of Heungbuga

Heungbuga is one of the five surviving stories of the Korean [[pansori]] storytelling tradition. It is also called Baktaryeong. *Heungbuga* has some other names such as, “*Yeonuigak*”, “*Nolbujeon*”, “*Baktaryeong*” etc. This story is not just focused on moral subject such as brotherly affection, but also to represents social contradiction by describing poor peasantry. Figure 4 shows story and play the Heungbuga.



Fig. 4. Heungbuga. (a) this is one of a scene in the story of Heungbuga. (b) a public performance about Heungbuga.

5 The References Section

The research shows Pasori. There are many genres as like Sugungga, Jeokbyeokga, byongangsoetaryeong, Baebijangtaryeong, Kangneungmaehwataryeong, ongojip-taryeong, Jangkkitaryeong, waljataryeong, and gajjasinseontaryeong except described items in the research. Pasori is a convergent art with sound, performance, the stage and so on. So many people dose not easy to understand the Pansori sounds and stories. However Pasori is one of lifestyle from a forefather to today in Korea. Pasori is too become an art native to Korea in the world.

References

1. Kim, sun hye Smith: What is P'ansori?: A Genre Comparison With English Renaissance Drama. *Comparative Korean Studies* 11(2) (2005)
2. Park, C.: Korean p'ansori Narrative. *Oral Tradition* 18(2), 241–243 (2003)
3. Kang, R.S.-S., Pansori, K.C.-H., Kenzaburo, O.: Focused on Grotesque Realism in “Ttong-Ba-Da” Exemplified by Oe Kenzaburo. In: *Global Korea Proceedings* (2006)
4. Jang, Y.: P'ansori performance style: Audience responses and singers' perspectives. *Ethnomusicology Forum* 10(2), 99–121 (2001)

A Study on Model Transformation Mechanism Using Graph Comparison Algorithms and Software Model Property Information

Jong-won Ko¹, Hwa-young Jeong², and Young-jae Song¹

¹ Dept. Computer Engineering, Kyung Hee University, 446701 Yong-in, Korea

² School of General Education, Kyung Hee University, Seoul, Korea

{jwko,hyjeong,yjsong}@khu.ac.kr

Abstract. In order to easily port mobile applications suitable for each platform, that have been developed under diverse development environment for individual wireless communication service providers, or redevelop them on a specific platform, it is required to reuse them at software model level that is a software development paradigm for MDA (Model Driven Architecture). The existing model verification approaches have focused on using graph comparison between input model and target model or applying graph pattern by simple version tree types. The graph model transformation mechanism proposed on this paper generates prediction model by defining test Oracle as model transformation rules that may conduct verification test of the generated transformation model, in order to support verification on verification of the converted model through MDA based model transformation mechanism. By comparing this prediction model with the target model, it is possible to execute verification test using graph comparison algorithms on the converted model. we supported verification mechanism of transformation model with model property information,dynamic analysis in this paper. Therefore, by increasing reliability of model transformation and further applying test issues on the software development process to the software mode at software design phase. A case study in AGG tool is presented to illustrate the feasibility of the model transformation verification with model property information.

Keywords: graph model transformation, model transformation verification, model comparison algorithms.

1 Introduction

Regarding the software development, MDA (Model Driven Architecture) of OMG can be regarded as the concept of making an independently-designed model according to the development environment and language and reusing it according to the desired development environment and language by expanding the reusable unit into the software model when developing software. The initial research related to the model converting technology for MDA has been executed by mainly focusing on such factors as the function of generating the source code supporting various converting formats or development languages and the expandability and applicability of the

model-converting mechanism. Recently, some researches for the model-converting method for the optimization of the model by applying the specific metrics or design rules based on the given model and the analysis of the uniformity related to the generated model by executing the model-converting process in various points of view for the same model have been executed. Also, the research for the expandability of the model which can be used to generate an expanded model through the model-converting process automatically when the factors required for the expansion of the model from the base model are applied is being currently executed [1]. In regard to the researches which have been recently executed, some test issues have been applied to support the accurateness of the verification method or transformation rule. In such a case, the accurateness of the converting model can be largely considered in terms of the accurateness for structures and the one for meanings. The accurateness for structures represents the one for the generation of a good target model by following the transformation rule with the input of a good source model, while the accurateness for meanings represents the one for the transformation of the information desired by the developer into the target model by interpreting the subject meaning with various points of view in regard to the generated target model [2]. The previous research about the model-converting verification was executed by mainly focusing on the accurateness for structures. However, at the University of Alabama in Birmingham, USA, researches about the transformation of models by using the C-SAW model-converting engine based on the Aspect-Oriented Software Development for the development of the embedded system and the related testing framework have been executed [1]. The testing framework suggests the model-comparing algorithm which can be used to compare the differences between two graphs by expressing the structures of the input model and the target model with the nodes and edges of the graph. However, according to such a comparison, the software design model is simply expressed in the form of the version tree in the management of the software images. Also, there seems to be not much information for the transformation between different software models with various complicated points of view, while the verification between the two models is no more than the simple comparison of the graph structure. Also, when executing the transformation of the model for one input model with various points of view, it could be difficult to execute the test for uniformity between the software models. As another research for the model-converting verification, there is the research method which applies the graph patterns to the verification on the graph-converting process by independently developing the VIATRA model-converting engine [3]. According to such a research, it is possible to define the compare the input model and the converting model and define the same graph pattern with the sub-graph by expressing the input model with the graph and executing the model-converting process based on the graph patterns. Also, it is possible to execute the model verification by investigating whether the graph model has such a pattern or not [4]. Since such a method is also based on the execution of the model verification through the simple graph comparison, it is impossible to verify various kinds of information contained by the test for uniformity in regard to the previously-mentioned model or the design model.

In this study, in order to improve the problems shown by the model verification method through the comparison of different models in regard to the graph model which is mainly used for the research of the model-converting verification and the

verification method based on the existence of specific patterns by considering the converting models in regard to the possession of the patterns provided by the input models, the information for the property of the related model is added to the previous graph model to define the converting information required for the model-converting verification on the graph model. It is possible to verify the converting information containing the information for the property of the model by using the graph model containing the property of aspects, actions and times for the comparison of various property contained by the software design model in regard to the simple structural graph in the form of the version tree shown by the previous graph model consisting of nodes and edges. Also, by defining the dynamic model showing the dynamic message of the software model with the dynamic specific graph, it is possible to describe the information for the converting model in terms of the structural point of view for the model based on the dynamic model analysis and the point of executing the dynamic message. At the same time, it is possible to verify the information acquired through the converting process. By using such a method, it is possible to verify the uniformity of the converting model in the multiple points of view between the input model and the target model after generating the dynamic model for the message call and the executing order in regard to the verification of the model which used to be dealt only in the structural aspect and executing the test for uniformity between the two models.

2 Related Works

In this chapter, the previous researches about the graph transformation method which is mainly used as the model transformation verification method of the model transformation methods for MDA and the model transformation verification are introduced.

2.1 Model Transformation Verification

With the verification research for the model transformation for MDA, the C-SAW transformation engine which is the AOSD-based model transformation engine for the development of the embedded system has been developed. Also, with the research executed by Jeff [1] who has suggested the testing framework and the verification method for the generated model, two aspects for the verification of the converting model have been approached. As the first aspect, the testing tool called M2MUnit has been developed in order to execute and support the text-based test with the application of the testing information as the code test in terms of the source aspect by completing the test case in which the transformation rule will be tested by inserting the file of the model transformation rule. As the second aspect, the comparison of the graph has been executed in the form of the version tree which is used to simply compare nodes and edges between both models by executing the graph-based comparison of the model between the input model and the target model. Regarding the researches which have been executed until now, there are several studies published in regard to the issues required for the sample codes related to the test case, the definition of the algorithm related to the comparison of the model, and the follow-up comparison of the model.

In the research executed by Varro [2], the verification mechanism based on the graph transformation and pattern has been suggested by using the VIATRA transformation engine. In regard to the model transformation based on XMI and XSLT, the verification mechanism is executed through the identification of the same patterns by analyzing the model transformation between the input model and the target model and comparing the two models with the patterns of the graph model or pattern based on the graph transformation. Throughout such a research, the specific pattern found on the host graph which executes the model transformation process by applying the pattern based on the graph model is considered to be the sub-graph. By applying the sub-graph on the target model, the model transformation is executed. Also, the verification method for the converting model is executed according to the existence of the pattern in a specific form after comparing the input model and the target model by using the pattern. The research for the model transformation and verification has been recently executed by expanding the graph pattern for VIATRA and applying the design pattern before adding a specific design pattern from the input model [9]. The problem found in the previously-executed model transformation and verification research is related to the great limit related to the verification of the structural property or various information contained in the model when the simple comparison of the graph model is executed through the mutual comparison of nodes and edges in the form of the version tree. It is necessary to supplement the verification mechanism through the test for uniformity in regard to the model transformation in various points of view..

3 Model Transformation Verification Based on Graph Model with Model Property Information

3.1 Graph Model for Converted Model Comparison

In order to supplement the simple comparison of the graph, various structural property of the software model and the information for the model property in regard to the transformation technology are defined. In the previously-executed study for the model transformation and verification process based on the C-SAW transformation engine, the graph model consisting of nodes and edges is defined. By referring to this graph model, the aspect, behavior and time factors are defined for nodes, while the relationship with nodes and types required for the definition of the meta-factor are added for edges in order to define the information related to the property. Therefore, in regard to the information of the model property for the comparison of the model suggested in the study, the node (N) contains the information about names (n), types (t), attributes (att), aspects (asp), behaviors and times. The property of the aspects show the information for aspects, while the property of the behaviors show the dynamic message called on the note. Also, the node consists of the model node which includes the sub-node and the atom node which is the actual sub-node. The edge (E) contains the information about names (n), source nodes (s), target nodes (d), relations (R) and types (T). By summarizing the contents shown above, it is possible to have the following results., the following Table 1 shows the information for the graph model factors and the model property

- 1. $N(\text{node}) = \{\text{name, type, kind, attribute, aspect, behavior, time}\}$
- 2. $N \in \{\text{Model node, Atom node}\}$
- 3. $E(\text{edge}) = \{\text{name, src, dst, relation, type}\}$
- 4. $G(\text{Graph}) = \{N, E\}$

Table 1. Graph model Elements and model property

Element	Model Property	Description
Node	name	Node's name
	type	Node's stereo type
	attribute	Node's attribute(value)
	aspect	aspect information
	behavior	message property that call node
Edge	time	node's time property
	name	Edge's name
	source	Edge's start node
	destination	Edge's end node
	relation	relationship between nodes
	type	relationship type

The nodes or edges found in the previous graph model and such property as names and types in addition to the source nodes, target nodes and attributes for the edges have been defined in the previous research. Additionally, the time property for the aspects, behaviors and real-time features are defined with the information for the model property added to the nodes. In order to show the real-time features, the time property can be classified into the property of WCET which is the worst-case execution time of the medium which becomes the node with the sub-property, the priority property for the execution priority order of the medium, and the property of the response time of the medium which becomes the node. The aspect property consist of the sub-property for aspects and the function of the common interest. The joint-point property for a specific dynamic position mainly include the method call or the field-value adjustment as the values of the property. Also, the advice property in regard to the decision for the time to apply the function of the common interest to the core logic are defined before or after calling the method or before starting the transaction. The point-cut property represent the joint points to which the actual advice is applied. The model property related to the node and the sub-property for the additional property of types and relations for the edge are defined in Table 2.

As shown above, the comparison between the target models, which are induced through the source model and the model transformation process for the transformation of the model generated based on the graph model to which the information for the model property defined in Clause 3.1 is added, can generate the graph model which has more necessary information for various types of comparative information and the model transformation and verification process than the previous method based on the simple comparison between nodes and edges. Through the generated graph model and the information for the model property, it is possible to improve the model-comparing algorithm for the model verification process.

Table 2. Model Property and sub-property

	Model Property	Sub-Property	Discription
Node	time	WCET	Worst case execution time
		priority	Object's execution priority
		response time	Object's response time
	aspect	name	Aspect name
		joint point	Aspect joint point,method call, Set Field value
		advice	aspect insert logic
		point cut	adapted joint point by advice
Edge	type	directional	directional edge
		non-directional	non-directional edge
	relation	generalization	generalization relationship between nodes
		association	association relationship between nodes
		dependency	depedency relationship between nodes
		composition	Composition relationship between nodes

3.2 Graph-Comparing Algorithm for Converted Mode Comparison

By referring to the graph-comparing algorithm based on the previous model transformation and verification process, it is possible to include the information related to the mapping process with the information for the model property in the node category. Also, the graph-comparing algorithm which includes the comparison of the graphs between the dynamic graph models is suggested as follows.

Input : Target Model (TM),Prediction Model(PM)
Output : MappingSet, DifferenceSet {N,D,C}
Begin

1. For each vertex v1 in TM
 If there is a vertex v2 in PM mapped to v1
 Mark v1 and v2 mapped
 Add pair(v1,v2) to MS
 Let nf={1,2,3,...N} be a NodeAttribute_Flag
 For each af of v1 and v2,
 If the value of v1(af1) does not equals to
 the value of v2(af2)
 Add (v1,v2,af,af1,af2) to C
 Else
 Add v1 to N

2. For each edge e1 in TM
 If there is a edge e2 in PM mapped to e1
 Mark e1, e2 mapped
 Add pair(e1,e2) to MS
 Else Add e1 to N
 Let ef={1,2,3,...N} be a EdgeAttribute_Flag
 For each ef of v1 and v2,
 If the value of v1(ef1) does not equals to
 the value of v2(ef2)
 Add (v1,v2,ef,ef1,ef2) to C
 Else Add v1 to N

3. For those elements in M2 without mapped mark
 Add them to D

End

Fig. 1. Graph- comparing algorithm with model property

According to the algorithm, it is possible to compare the nodes and edges as well as the information for each model characteristic after changing the target model and the predicted model generated with the function of the test oracle into the graph models with the information for the model property. Through the previous comparative algorithm, the nodes and edges between the two models are simply defined, while flag values are provided for the sub-property in the information for the property of the nodes and edges and the related values of the property and recognized in the actual codes. In the improved model-comparing algorithm, NodeAttribute_Flag and EdgeAttribute_Flag values are used for definitions. As a result, each sub-characteristic has an ID value. The values of the property (af1, af2, ef1, ef2) are compared in regard to the information for the objective graph model and the predicted graph model with the ID values. When the two values are same, the MappingSet information is used for definitions. However, when the two values are different, the DifferenceSet information is used for definitions.

4 Model Transformation between Class Diagram and RDBMS Model Using Model Property Information

For the model transformation and verification process based on the graph-comparing algorithm, the model-to-model transformation process in order to convert the UML model into the RDBMS model has been executed by using the AGG tool [14] which is the graph-converting tool and defining the additional graph models and the transformation rule. First of all, Figure 2-(a) shows the meta-model definitions of the class diagram for UML and the RDBMS model for the transformation process through the function of defining the model types of the AGG tool.

In regard to the selection of the type or characteristic of each graph node or edge suggested in this study, it is possible to select the information for the property of each node through the window for the input of the property, while it is possible to select the information for the property of each class, medium or aspect when selecting the model types for the source model and the target model in regard to the transformation process as shown in Figure 2-(b). Therefore, rather than the previous graph model, the

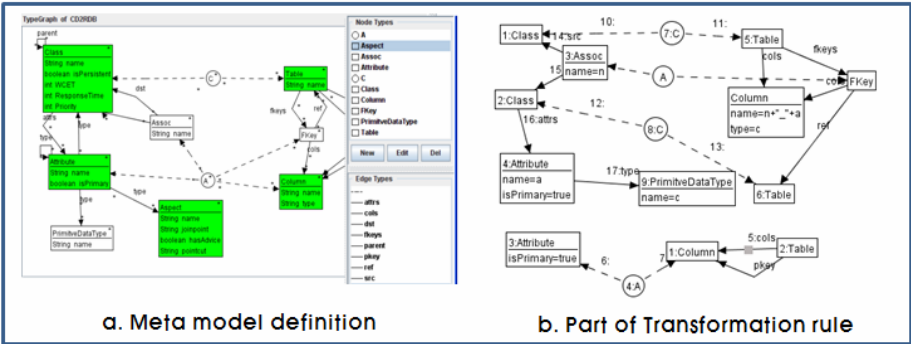


Fig. 2. Definition of Meta-model(a) and Part of Transformation rule(b) between Class diagram to RDBMS model

information for the source model which is subject to the transformation process based on the information for various property of nodes and edges can be expressed more accurately in various ways. By modeling the class diagram which is the source model to be converted on the right bottom window, it is possible to check the automatic transformation of the RDBMS model with the help of the converting menu as shown in Figure 3.

In order to understand the specific transformation rule to convert the class diagram into the RDBMS model, it is necessary to observe the mapping relationship between the graph-transformation rules in regard to various converting factors. Then, as shown in Figure 8, it is possible to show the mapping rule based on the graph type model with a graph. As shown in Figure 3-(b), the class node is expressed as the table node, while the attribute in the class is expressed as the column of the table. Also, the class expressed with the association between different classes is connected to the foreign key on the subject table. Regarding the information for the property defining the attribute, if the value for the isPrimary property is true, the subject column of the converted table is expressed as the primary key. Also, the source for the specific transformation rule is described and provided in the form of the XML file. In Figure 3, the class diagram made with the previous-mentioned source model and the RDBMS model which is automatically converted according to the transformation rule are shown.

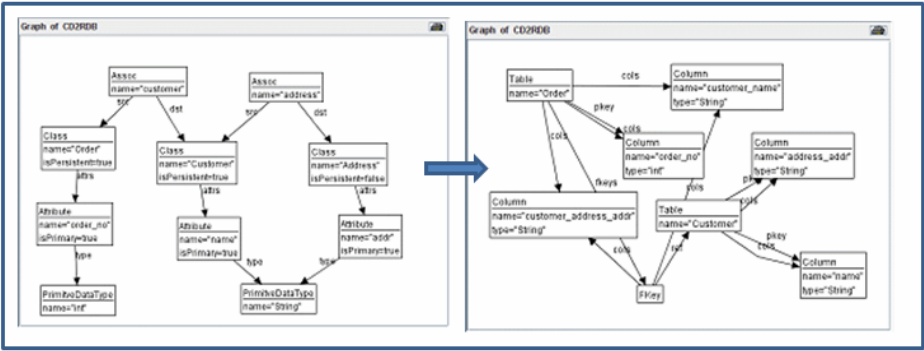


Fig. 3. Model Transformation between Class diagram to RDBMS model

5 Conclusion and Future Works

Those areas with embedded system such as mobile applications as domain are difficult for reuse as it is restricted on a specific domain, however, because of it, reuse is also required. This paper apply MDA based model transformation technology for development of the mobile applications that run on the various platform of the wireless communication providers and mobile equipment. In addition, in order to support verification of the target model generated from the source model, the designer to design mobile application architecture by adding test rules and comparing with prediction model can more easily reuse the architecture model. Moreover, if applying test issues at design stage as like that, there is an advantage to reduce modification cost of error comparing to test at the phase when source codes are almost completed.

However, as perfect mapping of matching APIs for mobile application model transformation is actually difficult, it requires additional code complement works and it has a restriction that model transformation can be applied to only those domains suitable for a specific domain. In addition, it requires studies on scalability to apply test rules diversely and on improvement of comparison analysis between model, implementation of supporting tools to support it is also in progress.

References

1. Lin, Y., Gray, J.: A Model Transformation Approach to Automated Model Transformation, Ph.D Thesis (2007)
2. Varro, D.: Automated Model Transformation for the Analysis of IT System, Ph.D Thesis (2003)
3. Darabos, A., Varro, D.: Towards Testing the Implementation of Graph Transformation. In: GT-VMT 2006 (2006)
4. Csertan, G., Varro, D.: Visual Automated Transformations for Formal Verification and Validation of UML Model. In: SAC 2007 (2007)
5. Czanecki, K., Helsen, S.: Classification of Model Transformation Approaches. In: Workshop on Generative Techniques in the Context of Model-Driven Architecture, OOPSLA 2003 (2003)
6. Agrawal, A., Kalmar, Z., Karsai, G., Shi, F., Vizhanyo, A.: GReAT User Manual (2003)
7. Zhao, G., Kong, J., Zhang, K.: Design Pattern Evolution and Verification Using Graph Transformation. In: Proceedings of the 40th Hawaii International Conference on System Sciences (2007)
8. Varro, G., Schurr, A.: Benchmarking for Graph Transformation. In: Proceedings of the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (2005)
9. Varro, D.: Automatic Transformation of UML Models, Budapest University of Technology and Enconomics (2002)
10. Varro, D.: Towards Formal Verification of Model Transformations, Budapest University of Technology and Enconomics (2004)
11. Matinlassi, M.: Quality-Driven Software Architecture Model Transformation. In: Proceedings of the 5th IEEE/IFIP Conference on Software Architecture (2005)

Video Service Algorithm Using Web-Cached Technique in WLAN

Iksoo Kim, Seoungyeol Lee, and Yoseop Woo

Dept. of Information and Telecommunication Engineering
University of Incheon, 119 Yeonsuku Songdo-dong Incheon, Korea
{iskim, ssheoso, yswooo}@incheon.ac.kr

Abstract. This paper presents the strategy for video service using caching technique in WLAN. The environment of WLAN includes mobile nodes(MN) and an access point(AP) that have memory for store video clips transmitted from AP and video server, respectively. The operation for the proposed system has two modes, the 1st mode is AP mode that all MNs are serviced through AP, and the 2nd mode is an ad-hoc mode that is serviced only through other MNs without AP connecting. The operation of ad-hoc mode is that the requesting MN is serviced directly from the video clip cached on other MN without the connection to AP or multimedia server(MS). Otherwise the proposed system uses AP mode that requesting MN is serviced under the control of AP, thus MN connects to AP's cache or MS.

Keywords: Webcaching, Multimedia, VOD, Wireless and Ad-hoc network.

1 Introduction

In recent, vide service is the one of the most popular one through the Internet. To support multimedia service in wire/wireless network, it should develop a high speed network that has the vast bandwidth, huge capacity of storage devices and intelligent multimedia server because multimedia data are tending to be more. Especially, the problems of deficiency of network bandwidth and electrical power are increasing in wireless networks[1,2,3]. Although those are solving at present, video services including VOD go with excessive loads of multimedia servers and inefficient use of network resources accompanying network delaying and transmitting redundant data[4, 5]. These problems are critical ones in wireless network, thus video service designers in wireless network have to work hard to solve these problems.

In wireless network mobile devices, such as smart phones and wireless laptops, are rapidly increasing the demand for multimedia services. In WLAN environments, the Access Points (APs) are connected to the Internet via a wired network, and mobile devices(nodes) are connected to the AP in order to connect the Internet. Though mobile nodes(MN) are serviced for multimedia through this technique in WLAN, they have to overcome inefficient use of deficient network resources[6, 7].

In generally there are two of the best ways for video service on the Internet in wire/wireless environment, they are multicast delivery technique and web caching

[2, 3, 8, 9, 10]. These strategies reduces excessive load of multimedia server, and uses network resources effectively. This paper using caching technique provides services through cached data on proxy which is located near the clients in WLAN.

The wireless mobile networks in WLAN consists of a number of mobile nodes(MN) and an access point(AP). Williamson[1, 11] suggested caching technique on AP and MN for video service, and a requested MN was serviced video clips from the cache on AP without connection to multimedia server(MS). But there are no requested video clips in the cache on AP, requested MN connects to MS through AP. But this technique always operates under the control of AP, thus the load of AP is increasing in proportion of service request of MNs. And that system does not support any advantage in the viewpoint of use effectively for deficiency of network resources.

Thus, this paper proposes the techniques of using ad-hoc mode among MNs when requested video clips are stored already in the cache of other MNs in WLAN. This mechanism leads to the minimization of service delay and effectiveness of using deficient network resource because of there is no connection of AP. Therefore, this paper presents the method of minimizing the delay of service, the load of AP and multimedia server and using deficient network resources effectively in wireless network using caching technique and ad-hoc mode.

The rest of this paper is as follows: Section 2 describes the structure of wireless mobile network and explains the operation of proposed system, Section 3 addresses 4-algorithms including the strategy of the replacement of cache according to the case-by-case for web-cached multicast system, and Section 4 deals with simulation and analysis of the results. Finally, we discuss our conclusion.

2 The Structure and Operation of the Caching System in WLAN

The structure of caching system in WLAN consists of a number of mobile nodes(MN) and access point(AP). The operation of existing WLAN is that all the MNs can connect to internet under the control of AP. The proposed caching system in WLAN can communicate among MNs without control of AP in ad-hoc mode operating when requesting video clips of a specific MN are stored on the cache of other MN in WLAN.

All the MNs in WLAN have a small memory called the cache that is stored video clips transmitted from AP, and also AP has the cache that is stored video clips transmitted from multimedia server. The cache located on MNs and AP can minimize the load of multimedia server and use effectively insufficient network resources. And AP has a small range of computing power that can grouping for multicast that support the same advantages, as mentioned earlier.

In WLAN, there are 2-modes, AP mode and ad-hoc mode. The ad-hoc mode is peer-to-peer mode without connection to AP and AP mode is traditional communication mode in WLAN if it does not employ caching techniques. And ad-hoc mode supports communication only among MNs that interchange stored data on their cache without connection to AP or video server.

The operation of the proposed system for video services deals with three scenarios as follows,

- i) Requested video clips are not stored on the cache of AP and MNs
- ii) Requested video clips are stored on the cache of AP
- iii) Requested video clips are stored on the cache of a specific MN

The 1st scenario is called initial state that no MNs have been requested before or no video clips stored on the cache of WLAN components. Thus, this state has not advantage for caching technique. In this case, the cache on AP stores them in prior to transmit to MN when the 1st MN requests video clips to MS. Next AP transmits them to requesting MN. The transmitted video clip is stored on the cache of the 1st request MN only. Thus, the 1st scenario has no advantage in comparison with traditional system.

The 2nd scenario is that video clip requested is already stored on the cache of AP. In this case, as soon as AP receives request of a specific MN, AP checks the contents of caching table. AP transmits the requested video clips immediately because they cached already. And the transmitted video clip is stored on the cache of the 1st request MN only just like scenario 1.

The 3rd scenario is called ad-hoc mode that requested video clips are stored on the cache of a specific MN in WLAN. In ad-hoc mode, when a specific MN request for a video clip transmits to AP and other MNs in WLAN, AP checks requested video clip is stored the cache of itself or MNs. And storing requested video clip MN only sends it to the requesting MN, AP does not send it to the WLAN because it knows that one of MNs in WLAN stores it. Thus, ad-hoc mode has the shortest delay without connection to AP, can be called peer-to-peer mode.

In the three scenarios of mentioned above, because all the MNs dynamically move around WLAN, we assume that the video clips stored on the cache of MN are discard when MNs leave the boundary of AP. And the strategy of replacement of cache on AP employs that the video clip stored on the cache of MNs are removed the first, and then employs the combination of LFU(least frequent used) and LRU(least recently used)[3, 12].

3 The Algorithm for Video Service in WLAN

As mentioned section 2, the proposed video service using caching technique in WLAN operates through 3-scenarios and strategy of the replacement of cache on AP. The proposed system have to support caching technique in ad-hoc mode because these scenarios support that effective use of network resources, the minimization of the load of multimedia server(MS) and delay. The operation provision to support multimedia service smoothly in initial state that no MNs have been requested before or no video clips stored on the cache of WLAN components is as follows,

Algorithm for the 1st scenario;

- i) In WLAN, MNs within the boundary of AP try to connect to AP*
- ii) AP assigns local IP address to MNs according to the order of connection through DHCP*
- iii) A specific MN requests service for video clip to multimedia server (MS) through AP*
- iv) AP checks its cache table whether it has requested video clip or not if it has not it yet, then it requests the video clip to the MS*
- v) AP stores it its cache and transmits to the requesting MN, and command to store transmitted clip to MN*
- vi) MN that stored video clip updates its contents for store*
- vii) MN service requested displays video clip v_i and then finish*
- viii) MNs that joined multicast group display video clip*

Step i) and ii) in algorithm for the 1st scenario show the procedures that all the MNs in WLAN try to connect to AP, and next they can connect Internet including multimedia server system. Step iii) indicates a specific MN_i request video clip(v_i) that stored only MS.

The operation provision to support video service smoothly in 2nd scenario state that video clip v_i requested is already stored on the cache of AP when a specific MN_i in WLAN requests them is as follows,

Algorithm for the 2nd scenario;

- i) A specific MN requests service for video clip to multimedia server (MS) through AP*
- ii) AP checks its cache table, and if it has not requested video clip on its cache then it requests video clip to MS*
- iii) MS transmits video clip requested to AP, and store video clip on cache of AP*
- iv) AP transmits to the requesting MN, and commands to the requesting MN to store transmitted clip*
- v) MN that stored video clip updates its contents for storing*

Step i) and ii) in algorithm for the 2nd scenario are the decision state whether video clip v_i requested is stored or not on the cache of AP. The successive steps are the same as that the 1st scenario .

The operation provision to support multimedia service smoothly in 3rd scenario state called ad-hoc mode that video clip v_i requested is already stored on the cache of AP and other MN in WLAN when a specific MN_i in WLAN requests it is as follows,

Algorithm for the 3rd scenario;

- i) A specific MN requests service for video clip to multimedia server(MS) through AP in the manner of broadcast*
- ii) AP and MNs in WLAN check their cache tables, AP and one of the MN identifies that they have requested video clip*

- iii) *AP releases the control to MN that has the video clip*
The MN that occupied the control checks whether the same
video clip is requested or not from other MNs

Step i) and ii) in algorithm for the 3rd scenario are the decision state that use video clip v_i requested that stored a MN or AP. The successive steps are the same as that the 1st and 2nd scenarios.

Now explain the strategy of the replacement of cache on AP following 3-scenario for video services. In the caching technique that decrease the load of video server, one of the most important problem is that the size of cache(proxy) memory. Though the price of buffer memory becomes cheap rapidly, the designers select generally size to small compare with video server. Thus the caching systems always have to deals with the strategy of the replacement of cache. In the proposed system, applying the starategy of replacement for cache on AP is the combination LFU and LRU, and the contents stored on cache of MN(in the case of the same contents stored on that of AP). The operation provision to support video service smoothly in cache flushing state that no memory space for store video clips transmitted from video server is as follows,

Algorithm for the replacement of cache;

- i) *AP stores all the transmitted video clips from multimedia server(MS)*
on its cache and updates cache table
- ii) *AP initiates strategy of the replacement of cache when the cache*
is flushing
- iii) *AP checks MN's cache table located in AP and applies LFU*
using AP's cache table
and if the applicable video clips are already stored in the cache of MNs
then replaces them and finish cache the replacement
else AP initiates LRU for data on cache of itself,
and removes the applicable video clips

In algorithm for the replacement of cache, step i) and ii) indicate cache flushing state that there is no space for store any video stream from video server according to the successive request for them. Step iii) indicates that have priority remove the side of a specific MN in the case of a specific video clip v_i stored on both of the cache on AP and a MN in WLAN. Step iii) shows the condition of applying the strategy of combination of LFU and LRU in the case that two video clip v_i and v_j stored on AP are already stored on two MNs' cache, respectively.

4 Analysis of Simulation Results for the Proposed System

As shown in Fig. 1, the environments of simulation are assumed that the bandwidth between AP in WLAN and video server is 10Mbps, and is 1Mbps between AP and each MNs. And video server transmits 100 video clips and the length is 20-minutes long. This simulation employs the popularity of media contents based on Zipf-like distribution.

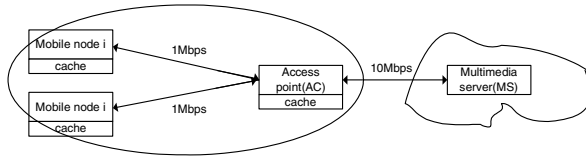


Fig. 1. Simulation topology

Using zipf distribution, if the MNs service request rate to the media server in the network(Internet) through AP that is connected in wireless network is λ , the service request rate of the video item for i_{th} frequency(eq. 1) among N multimedia video items is $\lambda_i = \lambda \rho_i (\rho_i = m/i)$, where $m = 1/(1 + 1/2 + 1/3 + \dots + 1/N)$. Its rate based on popularity is used to determine the weighted parameter for each multimedia item. Thus, this request rate λ for video clips follows Poisson distribution. The service request rate that follows Poisson distribution is as follows,

$$f_x(x = k) = \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!} \quad (2)$$

Fig. 2. Shows the comparison with the cache hit ratio for AP mode and ad-hoc mode for the proposed system according to the service request rate($\lambda = 30$). It indicates that the comparison with the cache hit ratio for AP mode and ad-hoc mode according to the size of cache.

Fig.3 shows the number of channel of multimedia service for unicast, AP mode(AP multicast) and ad-hoc mode. The unicast means that service without web-caching and multicast, AP mode shows that service for web-cached multicast on AP(scenario 2) only, and ad-hoc mode means that service for web-cached multicast on AP and MNs (scenario 3). Thus, ad-hoc mode shows the best performance among the 3-mode, and it confirms the best way to use effectively deficient network resources.

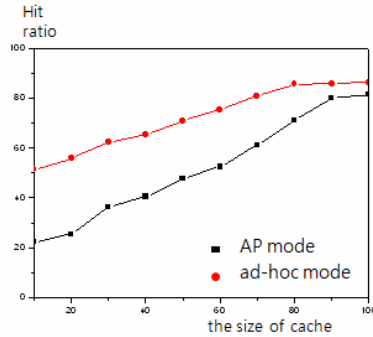


Fig. 2. The comparison with cache hit ratio for AP mode and ad-hoc mode

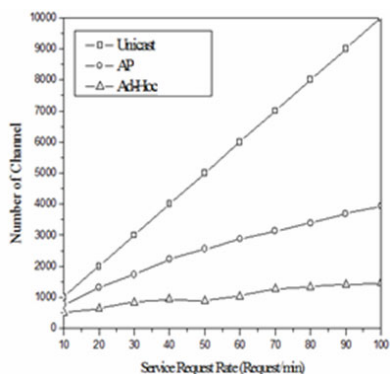


Fig. 3. The comparison with the number of channel needed for service according to the service request rate

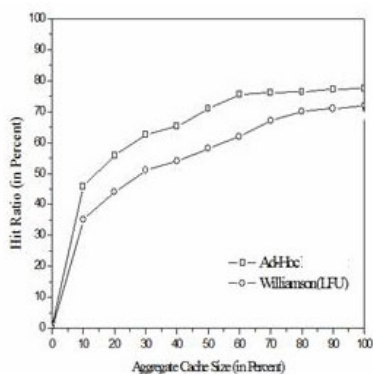


Fig. 4. The comparison with the hit ratio using cache replacement according to the cache size

Fig. 4 shows cache hit ratio according to the cache size, the cache hit ratio indicates more than 50% when the size of cache is above 15% of serviced video clips. As shown Fig. 4, cache hit ratio of proposed system is better than Williamson [1] through the whole range.

5 Conclusion

For supporting minimization of the load of video server and use effectively deficient network resources, this paper presents the caching technique for video service using 2-modes (AP mode and ad-hoc mode) in wireless network. In WLAN, the proposed system confirms superior in performance according to the number of channels needed for service and the cache hit ratio when it compare with traditional method. Also, this paper confirms that it can reduce the number of hops using ad-hoc mode free from connection to AP within WLAN. This model will be expanded easily to the larger wireless system when a number of APs are supported in WLAN.

Acknowledgments. This work was supported by the University of Incheon Research Grant in 2010.

References

1. Goma, H., Messier, G., Davies, B., Williamson, C.: Media Caching Support for Mobile Transit Users. In: Proceedings of IEEE WiMob 2009, Marrakech, Morocco, pp. 79–84 (October 2009)
2. Kim, I., Kim, B.: Content Distribution Strategy using Web-Cached Multicast Technique. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3983, pp. 1146–1155. Springer, Heidelberg (2006)

3. Lee, S., Kim, I.: Multimedia service using Web-Cached Multicast in Wireless Network. Accepted for Journal of Korean Institute of Information Technology 9(4) (2011)
4. Chen, K., Nahrstedt, K.: Effective Location-Guided Overlay Multicast in Mobile Ad Hoc Networks. *International Journal of Wireless and Mobile Computing (IJWMC)*, Special Issue on Group Communications in Ad Hoc Networks 3 (2005)
5. Almeroth, K.C., Ammar, M.H.: On the use of multicast delivery to provide a scalable and interactive VoD service. *Journal on Selected Areas of Communication (JSAC)* 14(6), 1110–1122 (1996)
6. Rajaie, R., Estrin, D.: Multimedia proxy caching mechanism for quality adaptive streaming applications. In: *Proceedings of IEEE INFOCOM*, Tel-Aviv, Israel, pp. 980–989 (March 2000)
7. Cui, Y., Li, B., Nahrstedt, K., Stream, O.: Asynchronous Streaming Multicast in Application-Layer Overlay Networks. *IEEE Journal of Selected Areas in Communication*, Special Issue on Recent Advances in Service Overlays 22(1), 91–106 (2004)
8. Kang, S.H., Kim, I.S., Woo, Y.S.: VOD Service using Web-Caching Technique on the Head-END Network. In: Kumar, V., Gavrilova, M.L., Tan, C.J.K., L'Ecuyer, P. (eds.) *ICCSA 2003*. LNCS, vol. 2668. Springer, Heidelberg (2003)
9. Sen Mazumder, A., Almeroth, K., Sarac, K.: Facilitating Robust Multicast Group Management. In: *Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Skamania, Washington, USA (June 2005)
10. Kim, I., Kim, B.: Content Delivery with spatial Caching Scheme in Mobile Wireless Networks. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) *ICCSA 2006*. LNCS, vol. 3983, pp. 68–77. Springer, Heidelberg (2006)
11. Jardsh, A., Papagiannaki, K., Belding, E.M., Almeroth, K.C., Iannaccone, G., Vinnakota, B.: Green WLANs: On-Demand WLAN Infrastructures. *Mobile Networks and Applications (MONET)* Journal special issue on Recent Advances in IEEE 802.11 WLANs 14(6) (December 2009)
12. Cao, J., Williamson, C.: On Workload Merging and Filtering Effects in Hierarchical Wireless Media Streaming. In: *Proceedings of ACM WMuNeP 2008*, Vancouver, BC, Canada (October 2008)

A Design and Implementation of Mobile Puzzle Game

Seongsoo Cho¹, Bhanu Shrestha¹, Kwang Chul Son¹, and Bonghwa Hong²

¹ Faculty of Electronic Engineering, Kwangwoon University, Seoul, Korea
{css, bnu, kcsn}@kw.ac.kr

² Department of Information Communication, Kyung Hee Cyber University, Seoul, Korea
bhhong@khcu.ac.kr

Abstract. This study has actualized Beadz Puzzle, a beads puzzle game, on the mobile where everyone can enjoy regardless of the location and time. Beadz Puzzle, increases the children's ability in spatial perception, patience, and concentration and allows them to develop cognitive ability in different situations and enhance their intellectual capacity. Moreover, adolescents and adults can develop their ability to process things quickly; the aged can maintain quick thinking and prevent Alzheimer disease. Each question has different ways to solve so that various solutions can be examined. This study suggests a Beadz Puzzle that has been modified and supplemented from the original beads puzzle game by adding layers of levels, a time limit and life system. That is, this game is constructed as interactive games using Beads Puzzle, therefore, it also can induce more participation and interest of elderly persons by using various configurations of games.

Keywords: Mobile puzzle game, WIPI, Virtual game, Interactive game.

1 Introduction

The physical interactions of human experience, that is, the direct physical sensation is more efficient than information recognized by multidimensional. In order to provide effective experience to the users, we have to combine with experience through real environment and in the meantime, computer games have been distinguishing by confirming the virtual and real. However, recently, various interactive games are developing by using various interfaces to make enjoy people as per demand of game market [1-2]. Until late 1990s a mobile phone was simply a portable telephone that allowed people to make phone calls while moving. But now mobile phone is not just a device that allows a phone call. We can now watch movies, play mobile games, and listening to music and much more[3]. Among them, mobile game has the most remarkable features. Mobile games are, unlike PC games or console games, accessible and mobile without being restricted with time and location. And because of the limited display size and simple key pad system, mobile games which are simple to operate and have simple interface are dominating [4]. In terms of the contents of the game, mini games that can be played during short float time are more appropriate than one that requires a long period of time.

Therefore, we have selected puzzle game in order to be able to make a mobile game that can be played with the simple key pads in a short amount of time. Regardless of age and sex, everybody can enjoy puzzle game. In addition, it can help to develop the brain, moreover maintain mental health as well. On the search for a puzzle game that is possible to be developed to a mobile game we crossed beads puzzle game in the market. We have realized the beads puzzle game that had been rather cumbersome to play in real life and programmed Beadz Puzzle to be easy and simple to play yet in an enhanced way.

2 Related Studies and Technologies

2.1 Recent Trends of Mobile Puzzle Games

Most of games for old aged persons were family friendly games where all families can enjoy and developed for brain exercising in order to protect from dementia. The brain exercising games were targeted for elderly persons, for instance, Nintendo's portable game dual screen (DS) 'brain training for adults' is operated from the console [5]. This game is simple, easy operation and interesting for elderly persons using simple interface which is different from the conventional one. This game can give effective for dementia by stimulating the brain but it cannot perform physical activities effectively. To play this game, elderly persons should sit down and perform games by watching small 3 inches display terminal which leads to weaken their eye-sight and face some problems physically when they play for a long time. The main family friendly games are Curball and AgeInvaders etc. These games which are based on virtual reality system, are mainly performed by both children and elderly persons. The Curball is targeted for elderly persons developed as an application of ubiquitous computing [6-7].

With the many puzzle game being the new blue ocean, hundreds of highly developed puzzle games are being developed and put into the market. 'Puzzle heaven' [8] introduced by AK communication in July 3, 2008 was successful by providing three puzzles – Pic-a-Pix, Battleship, and Sudoku. It able the users to play the games that people have been familiarized in magazines sold in Bus terminals and etc at in the one program. In addition 150 Pic-a-Pix puzzles, 210 Battleship puzzles, and 210 Sudokus puzzles are provided by Concept is Puzzles [9], World Puzzle Championship's official sponsor. The general perception is that it is easy enough for everyone to understand and play.

2.2 WIPI

WIPI (Wireless Internet Platform for Interoperability) made by special department for mobile platform in KWISF (Korea Wireless Internet Standardization Forum), as the standard mobile platform, defines the environment for running application programs downloaded via internet after loading them in the mobile device. KWISF's special department for mobile platform carries out standardization to writing WIPI according to the claims postulated by telecommunication firms, mobile platform developing company, mobile phone manufacturing companies, and contents developing companies [10].

WIPI started as national project starting from 2001 under the goal to reduce overhead development through forcing telecommunication firms use same platforms. Wireless Internet Platform is the basic software that plays the role of that of Operating System (OS) in personal computers in mobile phones.

Each telecommunication firm in South Korea had developed and used its own Wireless Internet Platform making it inevitable for the contents developing companies to make same contents into several platforms. Therefore in the developing and service of contents unnecessary factors occurred, and WIPI emerged in order to reduce those kinds of unnecessary factors on the national level.

Conditions for mobile platform considered when establishing WIPI are as following.

- A general-purpose that can accept various mobile devices, application programs and contents.
- A user interface that is stable, fast and convenient
- An independency and economic in terms of realization, transplant, and upgrade
- A consideration of the upgrade costs in response to the traffic increase and platform development
- Guaranteeing contents compatibility between platforms and providing users with abundant contents service
- An establishment of platform where all the telecommunication firms, CP, mobile manufacturing companies and users can gain a win-win situation.
- A building system that can reasonably support development cost

The characteristics of WIPI are as following.

- Accepted claims of the standard platform postulated by telecommunication firms
- Selected as Telecommunications Technology Administration (TTA)'s group standard
- Easily implemented and transplanted on independent hardware to realize regardless of the mobile phone hardware or OS
- Supported by both C and Java
- Provides Memory Compaction and Garbage Collection

2.3 Embedding Executable File on Mobile Phones

The method to mount the Beadz Puzzle executable file created by WIPI onto a real mobile phone is as follows. We used the XCE developer community site "<http://developer.xce.co.kr>" exclusive for SK Telecom. It is relatively open to test the files at SKT in comparison to KTF which allows only the people who have gained authority to KT developer through WIPI developer site only for KT to create binary files and download the enrolled test onto a mobile phone, which is precisely the reason we have selected SKT.

The process from downloading the game to the mobile phone to running them is downloading jar file (wmls file) and Script file (content description file) that compresses the game driver files from the server and the game becomes executable on the phone [11].

Create an account on the site “<http://developer.xce.co.kr>”, and click download test of Tech Support. Upload the completed jar files onto the contents download test page for general members, and transmit the files to the mobile phone for a demonstration by clicking the ‘sms’ button. When the message arrives to your mobile phone, access the wireless internet and download the jar file, having done you can run the game on the phone.

3 Beadz Puzzle Game Scenario

The goal of this game is to place different shaped beads into a two-dimensional square/triangle space. There is a total of 12 pieces of beads and in accordance to the game level the numbers of beads that are prefixed on the square/triangle board are different. The beginner level of the game is level one which has only one bead piece that is out of the picture among the twelve pieces and is to be placed on the right spot. In the second level, ten pieces are fixed in the puzzle and the player is to find the correct place to locate the remaining two pieces. The levels are divided as explained, and if the beads are placed in the board in the given amount of time the game proceeds to the next level. If the bead pieces are not placed in the limited time the game is over. The once used beads are not to be cannot be used again.

After selecting the pieces with the directions key and pressing the OK button will place the bead piece on the place. The placed bead can be moved with the direction keys and once it is located the OK button will fix the bead on that spot. The additional control buttons are as follows.

- Button 5 : rotate beads
- Button 0 : reset the located beads
- Button ← : flip beads
- Button # : go to next level

When the game first begins three lives are given. When the game is over due to the time limitation the player may try three more times. Every time the player fails the heart icon that represents the remaining life will reduce and when all three of them are gone the game is over. The modified Beadz Puzzle game has a bit of speed game style with its implementation of a time limit in addition to the original game that only requires fulfilling the entire board. In a user’s point of view, simply solving repetitive patterned beads can be boring. Therefore this game implemented a system that allows the points to differ based on the time of solving the problem; the quicker the problem is solved the higher points the player may receive. In general mobile games are played alone, so this enables the player to challenge their own highest rank or record that they have set. The following table 1 shows the point distribution by the set time. For example, if player clears the first level within 19 seconds he will receive 480 points, and if cleared within 17 second 460 points will be given.

Table 1. Score system according to time

Level	Time (second)	Gap of Score (seconds)	Score difference by the gap	Lowest score
1	20	2	20	500
2	40	4	40	800
3	180	10	180	1500
4	600	12	250	4000
5	1200	15	400	8000
6	2100	30	800	16000

4 Realization

JAVA2 SDK 1.3.1_20 and WIPI Emulator v1.0.1.1 were used to realize Beadz Puzzle. The most recent JDK is v1.5 but then programmed through v1.5 it is yet impossible to upload onto the phone so v1.3.1 was used. Adobe Photoshop CS was used for game interface designing, Adobe Illustrator CS for character designing, and 3DMAX6 was used to create the beads pieces image.

```

Begin
    If, there is not vacant spot on the board then,
        Answer game cleared, open game cleared message
        Level up to the next level Else, the level has not
        been cleared so continue the level
End
    
```

Fig. 1. Algorithm inspecting the success of the game

-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Fig. 2. The game board of initialization in which every bead is vacant

The algorithm to inspect whether the bead placing is successful or not used in Beadz Puzzle is as Figure 1.

In the square mode the board is 10x5 containing 50 spots to place the beads; hence the mose is initialized as mode -1. All 12 blocks are given its own number and the numbers are set according to the shape of the beads shape. Therefore, whenever the player places a blocks onto the game board, this algorism inspects whether there is a -1 (a vacant spot) and when the blocks have filled the board a 'Cleared' message is sent and proceeds to the next level.

-1	-1	-1	-1	-1	-1	-1	4	2	2
-1	-1	-1	-1	-1	-1	4	4	2	2
-1	-1	-1	-1	-1	4	4	3	3	1
-1	-1	-1	-1	-1	-1	-1	3	3	1
-1	-1	-1	-1	-1	-1	-1	3	1	1

Fig. 3. A board where some block are located

5 Conclusion

Beadz Puzzle is the first mobile game that has transformed the educationally used beads game onto the mobile space. The game is playable in any location regardless of time and space without the actual game tools. Without the users having to place the beads onto the board makes it convenient unlike the original game. Beadz Puzzle is a simple game that only asks you to place the 12 blocks in the vacant places on the board but is also a brain game that requires concentration and problem solving skills. In this sense, it provides the users with unlimited sense of accomplishment for the user must think and put in effort in order to solve the puzzle. Due to the fact that everyone has mobile phones and it is not difficult to operate it is forecasted to be played by people from every age and sex. Such Bead Puzzle game is generally satisfied by most of the users. For further research of the game, first, we need to improve of the result of this game and dissatisfied pace of game etc. This should be basically improved dissatisfactions caused by the mobile equipment operation in the game design and user interface etc. The research will be helpful to apply for senior-only interface design in digital TV market.

References

1. Falk, J.: Tangible Mudding: Interfacing Computer Games with the Real World. MIT Media, Cambridge (2002)
2. Mazalek, A., van den Hoven, E.: Tangible Play II: Exploring the world of hybrid games. In: Computer Human Interaction (2007)
3. Kim, Y.-G., Kwon, G., Yu, T.-Y.: WIPT Mobile Games Programs. Daerim Publishing (2005)
4. Choi, Y.-C., Yim, S.-B.: Mobile Multimedia. Life & Power Press (2007)
5. <http://www.nintendo.com/>

6. Kern, D., Stringer, M., Geraldine, Albrec htSchmidt: Curball - a prototype tangible game for inter-generational play. In: 15th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (2006)
7. Cheok, A.D., Lee, S., Kodagoda, S., Tat, K.E., Thang, L.N.: A Social and physical Inter-generational Computer Game for the Elderly and Children: Age Invaders. In: 9th International Symposium on Wearable Computers (2005)
8. <http://mobilepack.hangame.com/page/view.nhn?SHOW=&CID=26660&c=0&t=&pc=0&WHERE=>
9. Chen, M.X., Lee, S.W., Jang, E.S.: New Texture Coordinate Representation Exploiting Texture Patch Formation. In: Computer Games, Multimedia & Allied Technology 2009, Singapore, May 10-13 (2009)
10. <http://www.wipi.or.kr/>
11. <http://blog.naver.com/blue7water/10013737129>

Implementation of AUV Test-Bed

Soon T. Kwon, Woon Kyung Baek, and Moon G. Joo

Pukyong National University, Busan 608-737, Korea
Happy072@nate.com, {baekwk, gabi}@pknu.ac.kr

Abstract. This study deals with the hardware architecture of developing autonomous underwater vehicle and an application software to test its test bed. Overall hardware consists of a main control processor, a sensor data gathering processor to give an accurate position and posture from IMU, GPS, depth sensor by using extended Kalman filter, and a actuator control processor to control a thruster and four rudder motors. An application program was also designed to test communication with a host computer and to test functions of each processor.

Keywords: AUV, Autonomous underwater vehicle, System design.

1 Introduction

AUV (Autonomous underwater vehicle) runs automatically by using its embedded power in marine situations being affected by unknown variables, which have changed marine conditions or are not considered, performs a duty and returns to the craft carrier. AUV has been developed by advanced countries. Norway's REMUS series [1], [2], India's MAYA, and Singapore's STARFISH are representative.

In Korea, OKPO [3], [4] and ISiMI [5], [6] have been developed. This study deals with the hardware architecture of developing autonomous underwater vehicle and an application software to test its test bed.

2 Main Subject

The appearance of AUV hull in this study is shown in Fig. 1 and the composition of the inside electronic system is shown in Fig. 2.

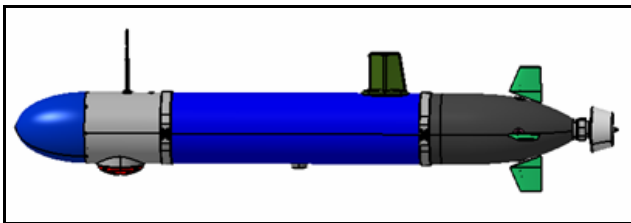


Fig. 1. Appearance of AUV

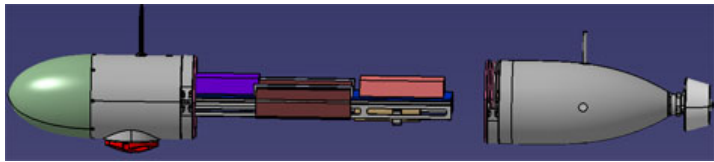


Fig. 2. Electronic System of the AUV

2.1 Driving Gear

AUV driving gear of this study consists of a thruster to propel the hull and four Servo-motors to control rudders. Thruster is BLDC (Brushless DC) motor and has internal pressure waterproof structure, and this runs by 24V-325W to meet power for moving AUV. As a motor to control a rudder, coreless servo-motor that can create 16kg/cm torque at 5V is used. Specification of thruster is as shown Table 1, and those of servo-motor are in the following Table 2.

Table 1. Thruster specification

Trust	8.2kg Forward 3.6kg Reverse
Input power	450-500 Watts 24-330vdc
Weight	1.0kg in air 0.7kg in water
Size	~11.1cm nozzle 22.6cm length

Table 2. Servo-motor specification

Type of Gear	Metal gear
Motor type	Coreless motor
Operating Voltage	4.8 V ~ 7.4 V
Neutral pulse	1.5 ms(Adjustable)
Travel angle	40~45° at ±400μS(Adjustable)
Dead Band	1 ~ 2 μS (Adjustable)
Torque	(kg/cm at 4.8V) 16.00 (kg/cm at 6.0V) 20.00
Weight (g)	50
Dimensions (mm)	40x20x34

2.2 Sensors and Communication System

A depth sensor in AUV measures the depth of water from surface. GPS (global positioning system) is installed to track a precise position on the sea. An underwater hybrid navigation system using DVL (Doppler velocity log) and IMU (inertial measurement unit) is used under the water. CCD camera was added to secure image during the exploration and navigation.

R/F Modem in order for users to control through computer is installed and wire/wireless LAN to acquire data without dismantling AUV is equipped.

2.3 Control System

The control system of AUV consists of an exterior host computer and AUV internal main controller, motor controller, and sensor processor.

AUV operation modes are manual mode and autonomous mode. Regarding manual mode, main controller receives the order sent by the exterior user computer and delivers it to motor controller and sensor processor. Autonomous mode runs automatically by operation algorithm installed in AUV’s main controller and saves image, which is sent from CCD Camera installed in the bows, to SD card.

Specifications of main controller are shown in Table 3. Windows CE 6.0 of the main controller supports hard RTOS so that users can set up system performance efficiently and replace operation algorithm easily.

Motor controller is designed by using Atmel’s ATxmega series. ATxmega series have on board A/D, D/A converter and multiple communication ports, so connection with other modules is efficient and miniaturization is possible due to little additional circuits. In addition, motor controller controls and manages each DC/DC converters, conducts motor control orders received from the main controller, detects residual quantity of battery and leakage.

Table 3. Main controller

CPU	32Bit RISC ARM1179JZF - 667Hz
RAM	128MB
Flash	NAND Flash : 128MB
SD Memory	SD Card 16GB compatibility
RS232	4CH
I2C	1CH
SPI	1CH
Ethernet	10 Mbps
WLAN	USB Type Wireless LAN
Power	DC 9~24V(9V/500mA)

2.4 Power System

Fig. 3. indicates the block diagram of AUV’s power system. The power supply of AUV consists of a 35.9V-17A lithium polymer battery for motor and a 22.2V-5A lithium polymer battery for other electronic devices. Lithium polymer batteries have high energy density so their capacity, discharge rate, and safety are high. The motor part and the electronic part use respectively 5V, 12V 24V that are input by six DC/DC Converters. Each DC/DC converter has individual ON/OFF control, and if an emergency situation is detected, the system forces to cut off the battery and the DC/DC converters to protect each component. There is an external charging terminal for users to charge without disassembling the case.

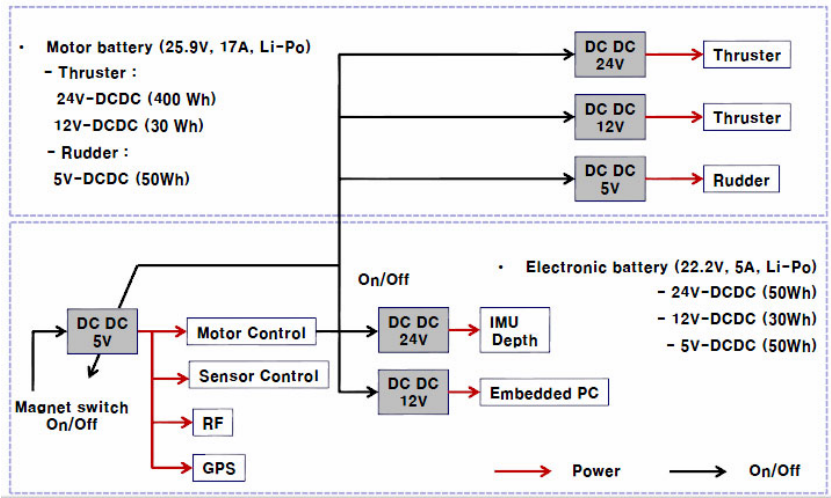


Fig. 3. Block diagram of AUV power system

2.5 Software for Operation Test of Each Part

To test operation of each part, LabVIEW application program shown in Fig.4 is developed.

LabVIEW application program installed in external user computer gives and takes orders and data through wireless communications with AUV by using R/F modem. It can control rudder and thruster, current status of rudder and thruster, residual quantity of battery, and ON/OFF condition of DC/DC Converter. The type of software for test is as in the following Fig.4.

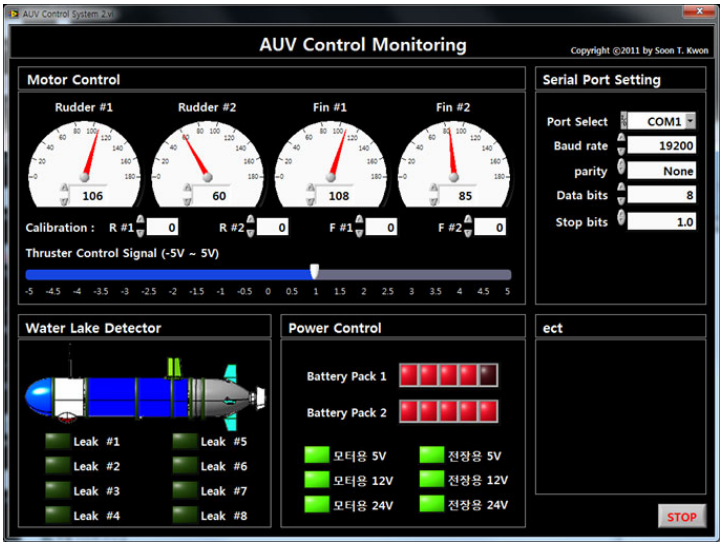


Fig. 4. Software for the test

3 Conclusion

This study deals with the hardware architecture of developing autonomous underwater vehicle and an application software to test its test bed. The developed AUV can measure roll, pitch, yaw, and pitch variation rate and control attitude by using one thruster and four rudders. To test operation of Test Bed, we conducted the test of rudders, propellers, and communications.

References

1. McCarthy, K.: REMUS - A Role Model for AUV Technology Transfer. In: International Ocean Systems (November/December 2003)
2. Griffiths, G., Brito, M., Robbins, I., Moline, M.: Reliability of Two Remus-100 AUVs Based on Fault Log Analysis and Elicited Expert Judgment. In: Proceedings of International Symposium on Unmanned Untethered Submersible Technology, Durham, New Hampshire, August 23-26 (2009)
3. Woo, J.: Development of AUV OKPO-6000 and deep sea trial. In: UE (1999)
4. Lee, P.M., Hong, S.W., et al.: Development of a 200m class AUV. In: KRISO Report UCN038-2064 (December 1997)
5. Lee, F.Y., Jun, B.H., Lee, P.M., Kim, K.: Implementation and test of ISiMI100 AUV for a member of AUVs Fleet. In: Proc. of IEEE/MTS Oceans, Quebec, Canada, pp. 1-6 (2008)
6. Jun, B.H., Park, J.Y., Lee, F.Y., Lee, P.M., Lee, C.M., Kim, K., Lim, Y.K., Oh, J.H.: Development of the AUV 'ISiMI' and a free running test in an Ocean Engineering Basin. *Ocean Engineering* 36(1), 2-14 (2009)

Communication and Computation Overlap through Task Synchronization in Multi-locale Chapel Environment

Bongen Gu¹, Weikuan Yu², and Yoonsik Kwak^{1,*}

¹ Dept. of Computer Engineering, Chungju National University,
50 Daehak-ro Chungju-si Chungbuk 380-702, Republic of Korea
{bggoo, yskwak}@cgnu.ac.kr

² Dept. of Computer Science and Software Engineering, Auburn University,
Auburn, AL 36849-5347
wkyu@auburn.edu

Abstract. Parallel processing systems use data parallelism to achieve high performance data processing. Data parallelism is normally based on data arrays, which are distributed to separate nodes. Therefore, efficient communication between nodes is required to initialize the distribution. In this paper, we propose a computation and communication overlapping technique to reduce the overhead of communication during array distribution. Our overlapping technique uses task parallelism for the initiate task and the worker tasks, and also the synchronization mechanism supported by Chapel. To show our overlapping technique is effective, we design and develop a parallel version of the Mandelbrot set program, and then evaluate the benefit of overlapping against the execution time of Mandelbrot. From our comparison, the overlapping technique proposed in this paper is effective in that it reduces the impact of communication in the initial array distribution.

Keywords: Synchronization, Chapel, Parallel, Array Distribution, Overlap.

1 Introduction

The need for powerful computing leads to the evolution of high performance computing (HPC). Representative HPC applications include combustion simulation, seismic modeling [3], weather forecasting, etc. To achieve high performance computing, parallelism is used in application programs and the runtime system. Parallelism is not only supported by the programming model, but also by computing resources, such as multicore CPUs, clusters, and MPP systems.

There are two types of parallelism supported by parallel systems: task parallelism and data parallelism. Task parallelism consists of parallelization of computer code blocks in the parallel environment, and focuses on distributing processes [6]. Data parallelism consists of parallelization of computing, and focuses on distributing data [6].

Data parallelism is normally based on the data stored in an array. To use data parallelism based on an array, the array element initialization and processing steps are

* Corresponding author.

required in the parallel program. The array element initialization step assigns values to the array elements, and, in general, is executed by the initiate task which lets the worker task start. The values assigned to the array elements are from the disk, the external input, and some value set at the initiate task side. The array element processing steps are executed by the worker tasks to calculate the results from values of the array elements. The worker tasks can concurrently operate in the parallel system or environment to do so.

In the case of large-scale parallel systems, the computing environment consists of many nodes connected by high-speed networks such as Gigabit Ethernet or InfiniBand, and one of the nodes executes the initiate task to begin the parallel processing. The node for the initiate task can also execute the worker tasks according to the system configuration. In this environment, the array elements are distributed to nodes with some data distribution strategy supported by the programming language or the runtime system [1][8]. As the performance of parallel program is influenced by the data distribution strategy, new data distribution strategies are proposed to get high access locality and parallelism [7][8]. When the array initialization is preceded by the initiate task, communication is required because the initiate task determines the values of array elements from the data sources as previously described, and portions of the array are stored in nodes connected to the node running the initiate task.

The communication overhead incurred by initializing to initialize the array elements distributed across the nodes may be mandatory, but it makes the execution time of the parallel processing long, so that the performance of the parallel processing system is negatively affected. If we make the communication time for initializing overlap with the computation time for processing array elements, the performance of the parallel processing may be enhanced.

In this paper, we propose overlapping the communication with the computation to reduce the negative affection by the communication. In our technique, we use task parallelism for the initiate task and the worker task. We also use a synchronization scheme between the initiate and worker tasks because each worker task can begin the processing after initializing the array element stored in the node, which executes the initiate task. Furthermore, we use data parallelism for the initializing and processing of the array elements. To realize our overlapping technique, we use the Chapel language, which supports task and data parallelism, as well as an easy synchronization method.

The rest of the paper is organized as follows. In Section 2, we briefly introduce Chapel, and describe in detail the overlapping technique proposed in this paper. In Section 3, we describe the effect of our overlapping technique using the execution time of Mandelbrot Set program, which is implemented using our overlapping technique. In Section 4, we summarize our technique and the effectiveness of it. Additionally, we suggest future research items related to our overlapping technique.

2 Overlapping Technique

2.1 Chapel Language

Chapel is a relatively new programming language for massively parallel processing systems. There are many parallel programming languages for parallel processing

systems. Design goals vary greatly from language to language in terms of such features as programmability, performance, abstraction levels, etc.

The design goal of Chapel is to provide the programmer with programmability and high-level abstractions for data- and task-parallelism, while the performance of the applications programmed by Chapel is not affected by the programmability and the high-level abstraction. The main features of Chapel are the following [4][5]:

1. Global-view parallel language
2. General parallelism
3. Better separation of algorithm and data structure

The meaning of the term “global-view” language is that the programmer doesn’t use any special index or descriptor to access array elements stored in remote nodes. Each node is called a locale in Chapel. The programmer uses an index defined in the domain, and does not consider where the array element is stored. Furthermore, Chapel supports general parallelism such as data parallelism, task parallelism, and nested parallelism at the level of the language specification [2][5]. Data and task parallelism are previously described. Nested parallelism means that we can use inner parallelisms in outer parallelisms. Chapel clearly separates the algorithm and the implementation. This means that the programmer makes his program from a global viewpoint, not in terms of the low level implementation. The low level implementation is the responsibility of the compiler.

2.2 Overlap between Communication and Computation

As previously described, many parallel applications, such as scientific simulations, seismic modeling, and weather forecasting, process data stored in arrays. For example, seismic modeling is used for oil and gas exploration. In this application, the seismic wave is generated by air guns placed at the surface [3]. The seismic wave generated is propagated downward in the subsurface. The refracted wave at the interfaces of the geological layers is propagated upward, and recorded by sensors. The sensor data are stored in an array for processing and modeling. Then, the array elements are distributed to all nodes or processing elements.

In large-scale parallel systems consisting of nodes, the initialization of the array elements distributed to these nodes requires communication between the initiate node and the worker nodes. The worker tasks executed in the nodes can begin the data processing only after the array elements have been assigned; this is why all worker tasks cannot begin the data processing immediately after creation.

If we let the worker task begin the data processing immediately after the array element stored in its node is assigned the initial value by the initiate task, while the other array elements are still being initialized, the execution time of the parallel application can be reduced. In this case, the computation at the worker task in which the array elements are already initialized and the communication at the worker tasks in which the array elements are being initialized can be overlapped. To do this, synchronization between the initiate task and the worker tasks is required.

Chapel supports the sync data type for synchronization between tasks[2]. The sync type variable in Chapel uses the empty/full protocol. Any task that reads the sync type variable has to wait until another task writes the value to it, and the variable is

changed to the full state. The full state sync variable is read by the task, and changed to the empty state. On the other hand, the task cannot write the full state sync variable, and has to wait until the variable is changed to the empty state.

We use the sync type array variable to synchronize between the initiate task and the worker task. The initiate task assigns the initial value to each array element, and changes the state of the corresponding sync variable to the full state. While the initiate task is assigning the values to the array elements, the worker tasks wait until the sync variable enters the full state. After the sync variable becomes full, the worker task begins to process the data stored in the array element corresponding to the variable. The following template code written in Chapel shows our overlapping technique between communication and computation.

```

1)  D : domain( rank ) dmapped ... = ...
2)  var data : [D] data_type;
3)  var flags : [D] sync bool;
4)  cobegin {
5)      forall idx in D do {
6)          data[idx] = initial_value;
7)          flags[idx] = true;
8)      }
9)      forall (element, flag) in (data, flags) do {
10)         on element {
11)             flag;
12)             element_processing...
13)         }
14)     }
15) }
```

In the template code for the overlapping technique, line 1 defines the domain and the data distribution. Line 2 declares the array for storing the data, and line 3 declares the array for storing the sync data to synchronize between the initiate and the worker task. The template code for the initiate task is shown between lines 5 and 8. In this code segment, the initiate task assigns the initial value to the array element at first, and causes the sync variable to enter the full state by assigning its value. The template code for the worker task is shown between lines 9 and 14. In this code segment, the worker task on the locale waits until the sync variable enters the full state at line 11. When the sync variable is made to enter the full state by the initiate task in line 7, the worker task wakes up and continues processing at line 12. To concurrently execute the initiate task and the worker tasks, we use task parallelism issued by the ‘cobegin’ statement in line 4.

In our template code, line 6 in the initiate task requires communication between the initiate task node and the worker task node(s). Line 12 in the worker task executes the computation task to process the data. These two tasks are concurrently executed by using the task parallelism supported by Chapel. Therefore our technique overlaps the communication with the computation, and contributes to a reduction in the execution time of the parallel program.

3 Performance Evaluation

To show that our overlapping technique is effective in reducing the execution time of the parallel program, we compare the execution times of two versions of the parallel program. In this paper, we implement two parallel versions of the Mandelbrot set calculation with a 1024×1024 -element array of complex numbers. The Mandelbrot set is a set of points in the complex plane and a well-known application of fractal theory. This is a computing-intensive task, which processes an array of complex numbers.

One of two parallel versions of the Mandelbrot set program does not use our overlapping technique. This parallel program executed on multiple locales has communication and computation time pattern shown in figure 1(a). Figure 1(a) shows communication and computation times for tasks executed on locales. The initiate task on Locale0 assigns initial values to array elements evenly distributed into locales. At that time, the communications between Locale0 and other locales are needed because array elements initialized by the initiate task on Locale0 are stored in remote locales. In the figure 1(a), the time for these communications is depicted by ' $comm\_time$ '. After initializing the whole array elements, multiple worker tasks executed on each locale compute begin to process array elements stored in the local locale to compute Mandelbrot set values. The time for these computations is depicted by ' $comp\_time$ ' in figure 1(a). In other words, the initiate task first initializes all array elements, and then the worker tasks begin to process the data.

The other uses our technique; that is, the initiate task and the worker tasks are concurrently executed according to the synchronization scheme previously discussed. Parallel program with our overlapping technique executed on multiple locales has communication and computation time pattern shown in figure 1(b). Figure 1(b) shows

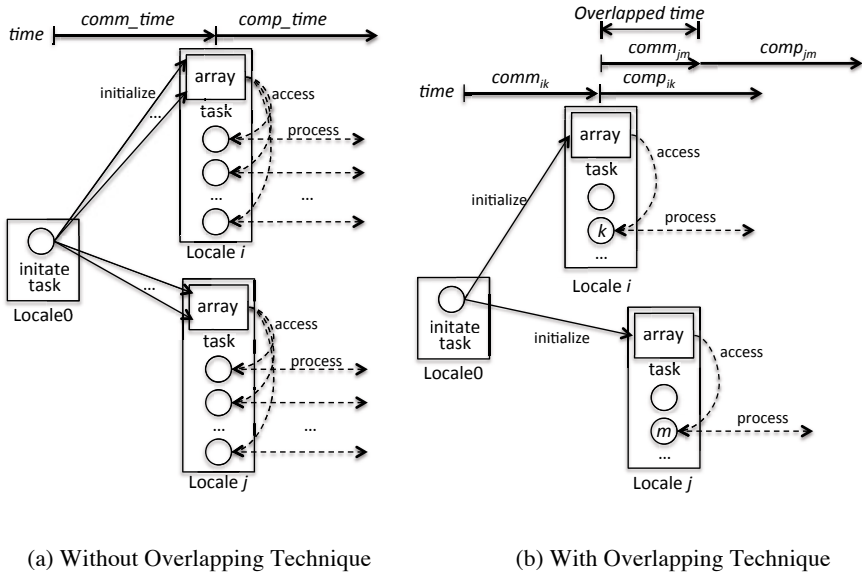


Fig. 1. Comparison of Communication/Computation Pattern between parallel programs without/with Overlapping Technique

communication and computation time for the initiate task, worker task T_{ik} , T_{jm} . The meanings of T_{ik} , T_{jm} are a task k on locale i , a task m on locale j , respectively. After the initiate task on Locale0 initializes array elements stored in locale i , and makes synchronization variables full states, the task T_{ik} for the initialized array elements begins to compute. In this scenario, we call the communication time between initiate task and T_{ik} to initialize array elements as ‘comm_ik’, and the computation time for T_{ik} as ‘comp_ik’. While T_{ik} is computing, the initiate task communicates with T_{jm} to initialize array elements and to make synchronization variables the full states. The time called as ‘comm_jm’ in figure 1(b) can be overlapped with ‘comp_ik’. So, in our technique tasks can immediately begin to process array elements after array elements for them are initialized. As previously described, a task knows whether array elements for it are initialized or not through synchronization variables. As tasks on every locale concurrently are executed, overlapping communication with computation is occurred very frequently.

The parallel programs used for this performance evaluation are executed on a cluster configured for 2-, 3-, or 4-locales. Each node is called a locale in Chapel. Each node in the cluster has the following specifications: two Intel Xeon(E5506) CPU, 8GB main memory, and InfiniBand (Mellanox MT26428).

Table 1 shows execution times of the Mandelbrot set programs. All execution times in Table 1 represent average execution times across four executions for the given configuration. In the table, the ‘Non-Overlapping’ column shows the execution times of the program, which does not use the overlapping technique proposed in this paper. The ‘Overlapping’ column shows the execution times of the program with our overlapping technique. From Table 1, the communication overhead increases in proportion to the number of the nodes. The communication overhead in the execution time is about 24% in the case of 2 locales, but increases to about 41% in the case of 4 locales.

However, the execution time with overlapping is reduced to the time shown in Table 1. In the case of the 2 locales, the execution time of the ‘Overlapping’ is about 1.27 times faster than that of ‘Non-Overlapping’. And in the case of the 4 locales, the ‘Overlapping’ is about 2.1 times faster than the ‘Non-Overlapping’. The reasons of these results are the following. In the case of ‘Non-Overlapping’, all tasks on locales cannot begin to process until the whole array elements are initialized. However parallel program with overlapping technique can absorb communication time for the array initialization into computation time. And as the beginning times of multiple tasks on each locale are different from each other, program with overlapping technique can efficiently use the system resources, such as CPU time and network bandwidth, etc.

Table 1. Execution Time of Mandelbrot Set Program w/wo Overlapping

	Non-Overlapping (sec.)		Overlapping (sec.)	
	Comm. Time	Computation Time	Exec. Time	Exec. Time
2 Locales	22.56	70.68	93.24	73.43
3 Locales	28.22	54.25	72.47	49.12
4 Locales	32.85	35.20	68.05	32.41

As a result, our overlapping technique proposed in this paper contributes to reduce the affect of the communication due to the array initializations, and is effective in reducing the total execution time of the parallel program in spite of the synchronization overhead.

4 Conclusion

The processing of many applications on large-scale parallel processing systems is based on the data stored in arrays, and these data are distributed to the nodes or locales. Chapel supports data distribution mechanism to distribute array elements into locales. In Chapel program with data distribution, the initiate task normally assigns initial values to the whole array elements distributed into locales, and then, worker tasks begin to process array elements. As the initiate task initializes array elements on remote locales, communications between locales are needed. Therefore the overhead of communication issued by array initialization is needed to reduce for enhancing performance of parallel program.

In this paper, we proposed an overlapping technique which communication phases are overlapped with computation phases. In our technique, task parallelism is used to concurrently execute the initiate task and the worker tasks. The initiate task assigns initial values to parts of array elements on a locale, and then, worker tasks immediately begin to process initialized array elements. And we use the sync type variable supported by Chapel to synchronize the initiate task with the worker tasks.

To show that our technique contributes to a reduction in the effect of communication on array initialization, we compared the execution times of two versions of a parallel Mandelbrot Set program. One of them cannot use our technique. Another one uses our overlapping technique. From our evaluation, we showed our overlapping technique is effective in enhancing the performance of the parallel processing system used. The reason of our experimental results is that parallel program with overlapping technique can absorb communication time for the array initialization into computation time, and this absorbability makes execution of parallel program more efficient.

References

1. Diaconescu, R., Zima, H.P.: An approach to data distributions in Chapel. *International Journal of High Performance Computing Applications* 21(3), 313–335 (2007)
2. Cray Inc.: Chapel Specification, 0.795 ed. Seattle, WA (April 2010)
3. Abdelkhalek, R., Calandra, H., Coulaud, O., Roman, J., Latu, G.: Fast Seismic Modeling and Reverse Time Migration on a GPU Cluster. In: *International Conference on High Performance Computing & Simulations*, Leipzig, pp. 36–43 (2009)
4. Steven, J.D., Bradford, L.C., Choi, S.-E., David, I.: Five Powerful Chapel Idioms, Cray User Group 2010 (2010)
5. Chamberlain, B.L., Callahan, D., Zima, H.P.: Parallel programmability and the Chapel language. *International Journal of High Performance Computing Applications* 21(3), 291–312 (2007)

6. <http://en.wikipedia.org>
7. Chamberlain, B.L., Deitz, S.J., Iten, D., Choi, S.-G.: User-Defined Distributions and Layouts in Chapel: Philosophy and Framework. In: USENIX Workshop on Hot Topics in Parallelism (2010)
8. Bikshandi, G., Guo, J., Hoeflinger, D.: Programming for Parallelism and Locality with Hierarchically Titled Arrays. In: PPOPP 2006: Proceedings of the Eleventh ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, pp. 48–57. ACM Press, New York (2006)

Compensation System for RF System-on-Chip

Sung-Woo Kim¹, Jee-Youl Ryu¹, Jae-Hwan Lim¹, Seung-Hun Park¹,
Jung-Hoon Lee¹, and Min Yoon²

¹ Dept. of Information and Communications Engineering, Pukyong National University,

² Dept. of Statistics, Pukyong National University,

Busan, 608-737, South Korea

ryujy@pknu.ac.kr

Abstract. This paper presents a novel programmable compensation system (PCS) for a radio frequency (RF) System-on-Chip (SoC). The PCS is integrated with 0.18- μm BiCMOS SiGe process. It consists of RF Built-In Self-Test (BIST) circuit, Resistor Array Bank (RAB) and digital signal processor (DSP). To verify performance of the PCS we built a 5-GHz low noise amplifier (LNA) with an on-chip RAB using the same technology. Proposed system helps it to provide DC output voltages, hence, making the RF system chain automatic. The PCS also compensates abnormal operation due to the unusual PVT (Process, Voltage and Thermal) variations in RF circuits.

Keywords: System-on-Chip (SoC), programmable compensation system (PCS), low noise amplifier (LNA).

1 Introduction

A recent trend in advanced wireless communication system is to integrate all of the radio frequency (RF) and mixed-signal devices on a single chip. This trend contributes to high density system-on-chip (SoC) with higher performance and reliability. RF and mixed-signal test engineers and designers in SoC field are mainly concerned with unusual PVT (Process, Voltage and Thermal) variations in RF circuits since these variations are hard to distinguish from acceptable variations and lead to a significant performance loss that violates the circuit's specification [1-2]. Therefore, once chips are fabricated and packaged, some sophisticated compensation techniques need to be applied to adjust performance from chips with the PVT variations. In spite of the considerable research underway to solve these problems [2-5], the IC industry is still seeking more suitable technique.

This paper presents a novel low-cost alternative for RF SoC testing to adjust PVT variations. The alternative approach utilizes programmable compensation system (PCS). The main part in proposed system involves RF Design-for-Testability (DFT) circuit and Resistor Array Bank (RAB). The RF DFT circuit helps it to provide DC output voltages, hence, making the compensation system automatic. Proposed system is very useful for concurrent RF ICs in a complete RF system environment.

2 Programmable Compensation System

Fig. 1 shows the PCS set-up. Our proposed PCS provides DC outputs so that all of the measurements can be done by on-chip data converter within the SoC, hence making the measurement automatic. It contains RF circuits with RAB and RF DFT circuit all on a single chip. The test structure can be automatically configured by the external data acquisition hardware. The complete test-bed contains RF source (v_{in}) with source resistance (R_s), RF relays ($S1$, $S2$ and $S3$), load impedance (Z_L), and a data converter board. The measurement set-up contains very low-loss RF relays and input transmission matching to the RF and DFT circuits. The positions of the relays were controlled to measure the output DC voltages and phase, V_{T1} , V_{T2} , and θ_T through the DFT circuit, respectively. These relays were controlled by DeMux chip on external board.

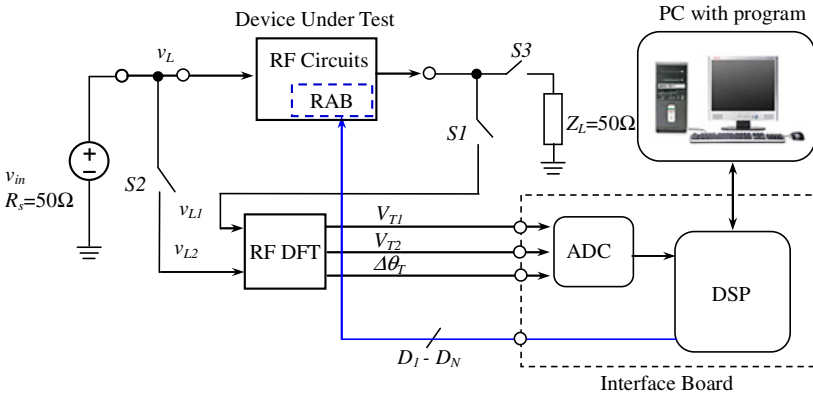


Fig. 1. Measurement set-up of the PCS for an RF circuits

The RF BIST circuit contains a test amplifier (TA), a band-gap reference circuit, two RF peak detectors (PD1 and PD2) and a RF phase detector (PHD) as shown in Fig. 2. This additional circuit occupies a very small area less than 5% on the SoC, and it helps to measure LNA performance without expensive external equipment. Two RF peak detectors and a RF phase detector are used to provide DC output voltages (V_{T1} and V_{T2}) and phase difference ($\Delta\theta_T$), respectively.

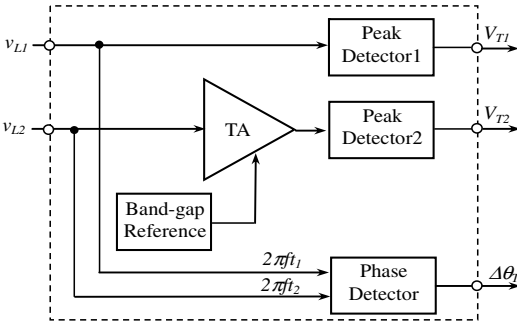


Fig. 2. RF BIST circuit

The proposed RF BIST circuit is shown in Fig. 3, and it is designed using 0.18 μm SiGe technology. It consists of TA, PD1, PD2 and PHD circuits. The PD1 circuit is also a part of the BIST circuit and it has the same topology as the PD2 circuit shown in Fig. 3(a). The test amplifier is designed with the input and output impedances of 50 ohms, respectively. The gain of the test amplifier is designed to be 3 to increase the output voltage level. The RF peak detectors are used to convert RF signal to DC voltage. To detect phase difference (θ_T) a RF phase detector are used. The bias stage utilizes a band-gap reference circuit for a low-supply voltage and a low-power dissipation. The inductor (L_{c01}) is used for matching input and output impedances. The bias resistors (R_{05} and R_{06}) shown in Fig. 4 are used to keep transistor Q_{04} in the active region so that the transistor acts as a rectifier. The diode connections have the advantage of keeping the base-collector junction at zero bias [2]. To reduce the output-ripple voltage, large values were chosen for R_{07} and C_{05} .

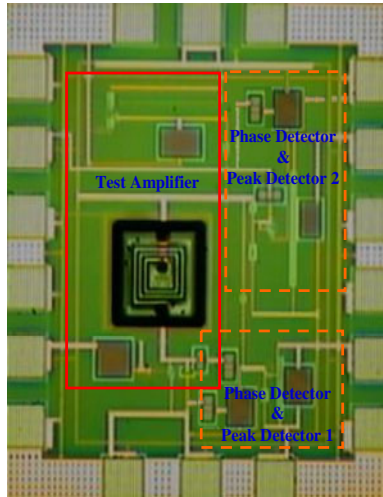
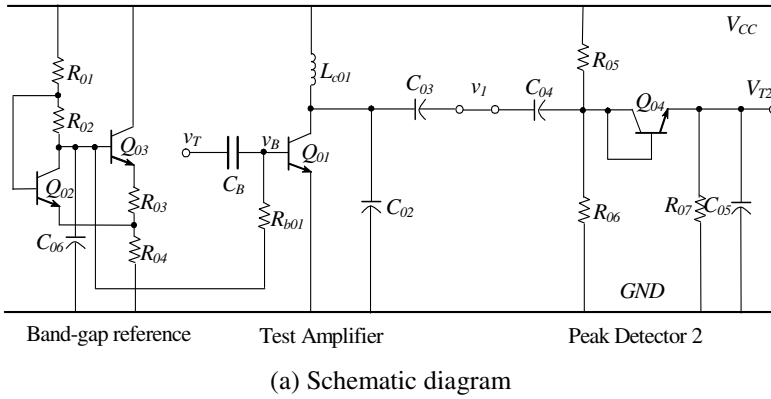


Fig. 3. Schematic diagram of an RF BIST

Fig. 4 shows details of the proposed RAB. It has N-bit resistor banks to accurately compensate an RF amplifier performance. In this approach, we have designed an 8-bit RAB considering a chip area overhead. The resistor bank is controlled by using digital signals ($D_8 \dots D_2 D_1$) from the digital signal processor (DSP). The input data streams of ($D_8 \dots D_2 D_1$) = (0...01) for $8R_b$ and (1...11) for R_b have been used to compensate RF amplifier performance, respectively. The R_b is under defect-free value. It was designed with LNA on a single chip using 0.18 μm SiGe technology to demonstrate this idea. It is powered by 1.8-V supply voltage. It was designed to have separate supplies for RF and digital sections of the chip to isolate the RF circuitry from the switching noise introduced by the digital supplies. The chip divided into RF and digital sections with different substrate grounds is divided to attenuate noise coupling from one area of a chip to another. The distributed gate resistance of the MOS devices contributes to the thermal noise [2]. To minimize this resistance, the transistors $M_1 \sim M_8$ were laid out as a parallel combination of many narrower devices. The transconductances of the transistors were minimized to reduce the input-referred noise voltage related to the thermal noise. These transistors in MOS switches, are designed for operating in the deep triode region so that they exhibit no dc shift between the input and output voltages. The resistors, $R_{D1} \sim R_{S8}$ were used to control a dc bias voltage of MOS switches.

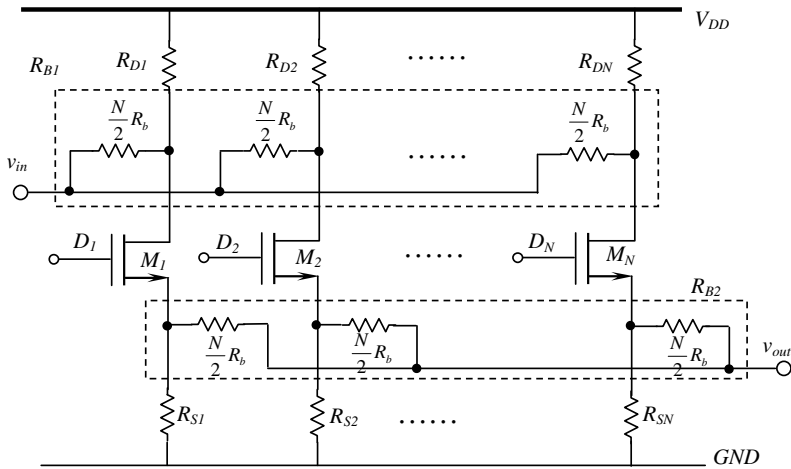


Fig. 4. N-bit RAB circuit

3 Results

A. Voltage Variations and Compensations

The compensation results for a -0.2V ($V_{cc}=1.8\text{V}$) voltage variation in LNA is shown in Fig. 5. The gain compensation shown in this figure was also performed at 5.25GHz. We identified a variation of 0.08dB in the LNA gain from the -0.2V voltage variation. To compensate a 0.08dB gain variation, the input data stream of ($D_8 \dots D_4 D_3 D_2 D_1$) = (1...1111) providing $R_B = R_b$ was applied. The -0.2V voltage

variation showed 0.211dB variation in the noise figure. As can be seen from Fig. 5, the proposed PCS can compensate the gain and noise figure of RF amplifier due to the voltage variation.

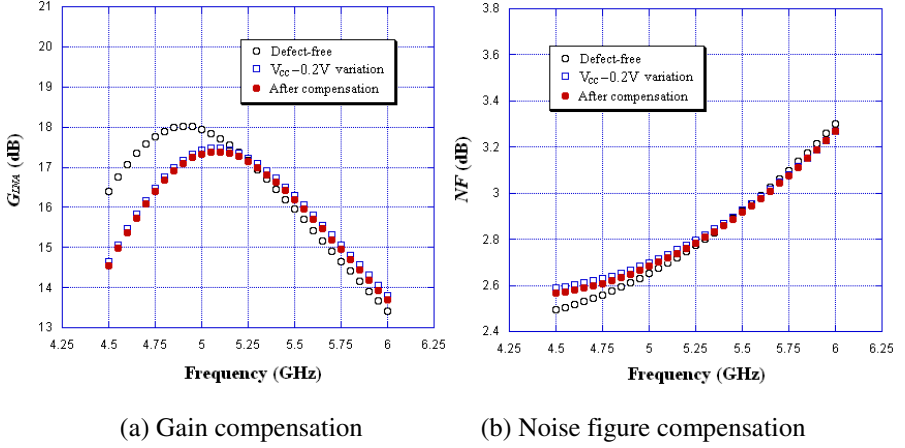


Fig. 5. Compensation for -0.2V variation of supply power, V_{cc}

B. Temperature Variations and Compensations

Fig. 6 shows the gain variation and its compensation results for the $+10^\circ\text{C}$ temperature variation. The gain compensation was done at 5.25GHz. We identified a variation of 0.101dB in the gain from the $+10^\circ\text{C}$ variation. The input data stream of $(D_8...D_2D_1) = (0...01)$ providing $R_B = (8/5)(R_b)$ was applied to compensate a 0.101dB gain variation. The $+10^\circ\text{C}$ temperature variation showed 0.106dB variation in the noise figure. As shown in Fig. 6 the proposed PCS can compensate the gain and noise figure of LNA for the temperature variation.

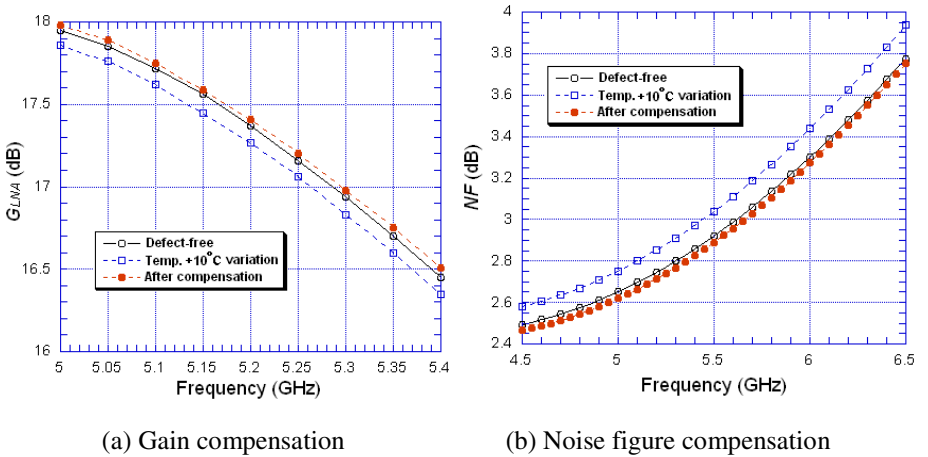


Fig. 6. Compensation for $+10^\circ\text{C}$ variation of temperature

4 Conclusions

In his paper, we proposed a new programmable compensation system (PCS) for an RF System-on-Chip. The PCS was integrated with 0.18- μm BiCMOS SiGe process. The system contained RF Built-In Self-Test circuit, Resistor Array Bank and digital signal processor. To verify performance of the PCS we built a 5-GHz low noise amplifier with an on-chip RAB using the same technology. The proposed system compensated abnormal operation due to the unusual PVT (Process, Voltage and Thermal) variations in RF circuits. The PCS also provided successful measurement results for LNA chips with Resistor Array Bank. We believe that this new capability will provide industry with a low-cost technique to test and compensate RFIC chips.

Acknowledgement

This work is the result of the "Human Resource Development Center for Economic Region Leading Industry" Project, supported by the Ministry of Education, Science & Technology(MEST) and the National Research Foundation of Korea(NRF).

References

1. Ryu, J.Y., Kim, S.W., Lee, D.H., Park, S.H., Lee, J.H., Ha, D.H., Kim, S.U.: Programmable RF System for RF System-on-Chip. *Communications in Computer and Information Science* 120(1), 311–315 (2010)
2. Ryu, J.Y., Noh, S.H.: A New Approach for Built-In Self-Test of 4.5 to 5.5GHz Low Noise Amplifiers. *ETRI Journal* 28(3), 355–363 (2006)
3. Pronath, M., Gloeckel, V., Graeb, H.: A Parametric Test Method for Analog Components in Integrated Mixed-Signal Circuits. In: *IEEE/ACM International Conference on Computer Aided Design*, pp. 557–561 (2000)
4. Liu, H.C.H., Soma, M.: Fault diagnosis for analog integrated circuits based on the circuit layout. In: *Proceedings of Pacific Rim International Symposium on Fault Tolerant Systems*, pp. 134–139 (1991)
5. Segura, J., Keshavarzi, S.A., Hawkins, J.C.: Parametric failures in CMOS ICs – a defect - based analysis. In: *Proceedings of International Test Conference*, pp. 90–99 (2002)

A Study on the SNS (Social Network Service) Based on Location Model Combining Mobile Context-Awareness and Real-Time AR (Augmented Reality) via Smartphone

Jang-Mook Kang¹ and Bong-Hwa Hong^{2,*}

¹ Electronic Commerce Research Institute Dongguk University, 707 Seokjang-dong,
Gyeongju, Gyeongsangbuk-do, 780-714, Korea
mooknc@gmail.com

² Dept. of Information and Communication,
Kyunghee Cyber University, Hoegi-dong, Seoul, 130-701, Korea
bhhonh@khcu.ac.kr

Abstract. The advent of the Internet heralded network age in the 1990s. I-Phone's launch into the market in Korea in 2010 opened up the country to smartphone age in a mobile environment. smartphone provides such functions as GPS sensing module and navigation in a mobile environment. Handheld computers that are linked to web mobile platform on a real-time basis enables personalized user information such as location, mail address, mobile phone numbers of friends and call duration to be provided in mobile context-awareness system. In particular, context of personalized mobile context-awareness combines with AR technology, which fuses with cross media producing, storing, distributing, re-processing, spreading and disseminating various information based on personal context, to create a visual image. This study will suggest a model showing how a social network service (Facebook, Twitter, Gowalla, foursquare etc.) exploding like a volcano splits up into AR (Augmented Reality) technology and the model proposed will be demonstrated to the details via personalized location-based service.

Keywords: smartphone, mobile, context-awareness, AR (Augmented Reality), SNS (Social Network Service).

1 Introduction

Changes set off by technology in the contemporary age of network spread to all corners of society such as politics, culture and education to form a tight link [1] and networks, in particular, are evolving into human relations system [2]. In Korea, more than 6.8 million smartphones manufactured by Samsung Electronics, Apple and Pantech are in service as of Dec. 5th, 2010 and the number crossed the seven million mark by year-end. The number of smartphones enabling social network soared by 24 times vs. 280,000 units in 2008 and by 8.5 times vs. 0.8 million units in 2009 [3].

* Corresponding author.

Outlook of smartphone market changes almost every quarter as its expansion is seemingly beating market expectations. The number of smartphones around the world is forecasted to reach 620 million units in 2015 and it is estimated to post 102.4% in annual average growth in 2012 in Korea, which means that 46.3% of the population or 25 million people will be carrying smartphones.

This new trend started off when Apple's iPhone made inroads into the market last year. The hardware supply is becoming increasingly popular in Korea and software growth is also remarkable. Rapid increase in the number of smartphone users is putting mobile service using sensor under limelight.

This study will analyze growth trend of smartphone market in Korea and will verify how growth will enable mobile-based application service. This study will specifically touch on cases of application service both home and abroad and discuss its future direction based on which social network service applicable in smartphone will be designed. SNS proposed is delivered by reprocessing information obtained from mobile-based sensor and particularly realizes AR, which is hot in the market, into SNS model.

2 Domestic and Global Smartphone Market

Mobile sensor-based smartphone market is exploding in Korea as indicated in the fact that more than 3,000 iPhones were sold in 2010. iPhone by Apple was able to flood into the market by Apple store's applications, which Apple opened for win-win with developers. Such an open policy by Apple motivated venture start-ups and individual programmers to develop applications and jump into the bandwagon with enthusiasm. Such market changes and successful cases also encouraged Korean telecom providers such as SK and LG to open up and share application store.

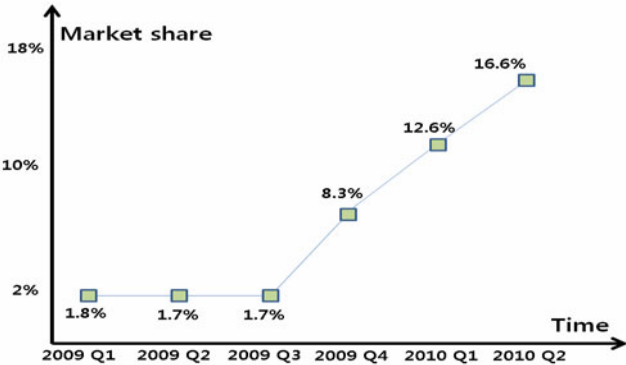


Fig. 1. Share of smartphone in Korea's mobile phone market [4]

Domestic smartphone market trend in <Figure 1> shows a rapid growth of smartphone sales from 3Q, 2009. Wider penetration of smartphone into the public means growth potential of mobile-based social web service since mobile service evolves alongside mobile OS policy that is based on smartphone. <Table 1> below offers overview of mobile OS providers both home and abroad, and their outlook.

Table 1. Overview of mobile OS providers [5]

OS Providers	Operating System	Kernel Type	Official App Store
Nokia	Symbian	Symbian OS	Nokia OVI
RIM	Blackberry	Proprietary	Blackberry App World
Google	Android	Linux	Android Market
MS	Window Mobile	Windows CE	Windows Marketplace
Apple	Max Os X	OS X	App Store
Palm	Web OS	Linux	Palm Web OS App Store

Source: National IT Industry Promotion Agency, Shinhan Investment Corp, Seung Yun Lee (2010), "Smartphone, Now", Mobile Platform Trends.

Korean mobile OS providers put focus on supporting hardware loaded to major OS (android, Apple etc.) and thus inevitably puts them behind their overseas competitors on software development and application.

Their overseas peers develop highly-involved services on mobile OS basis with Facebook (www.facebook.com), Twitter (www.twitter.com) and other SNS providers. The Research Foundation will develop a new system and service capable of formulating a big story by taking domestic contents' mixture level a step higher and will leverage AR technology to display in layer on map or camera subject with a real-time location information by linking with Facebook, Linkedin (www.linkedin.com), Me2Day (www.me2day.com), Yozm (<http://yozm.daum.net>) and other SNSs.

3 Application Technology Based on Smartphone

3.1 Handheld-Based Technology

Handheld technology allows smartphone, PDA, MP3, laptop, PSP (Playstation Portable) and other various devices to be used in a mobile environment. Mobile or application based on smartphone technology has been chosen as the subject for research due to the following reasons.

As shown in <Figure 2>, information moves and space in which information processing takes place changes. IT device is no longer a mere machine but an added value creator by facilitating information communication and processing. It, hence, functions as a hub or portal in which information is collected, stored and re-processed, and the way it processes information has expanded from a fixed desktop to an Internet-based portal. This is demonstrated in more smartphone users sending e-mail and SMS, searching information, making phone calls and doing simple document editions in their smartphone loaded with widget.



Fig. 2. Smartphone based on Hand-held

3.2 AR Technology

Smartphone market expansion is driving development of killer software application. As analyzed, source of competition dominating the market is shifting to software development from hardware functional expansion that started off with mobile phone and now to smartphone.

<Figure 3> below is the scheme of rich expressions made possible by AR's display technology classified into 'web as software interface', 'web as hypertext system', 'task-oriented' and 'information-oriented'.

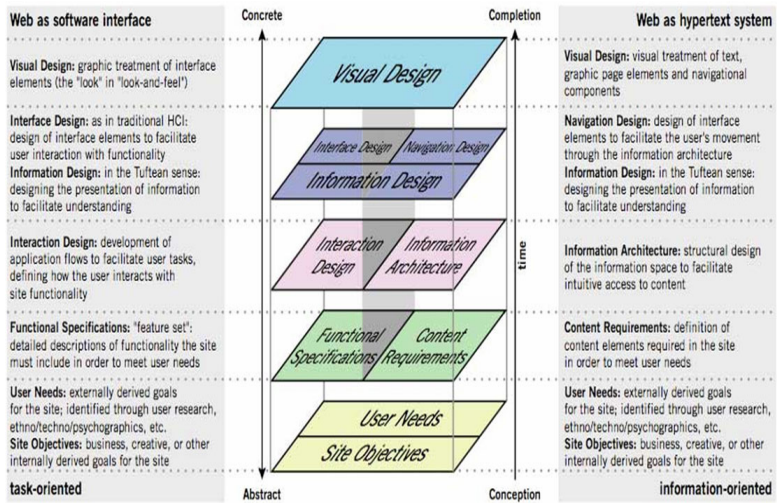


Fig. 3. Display technology delivered with rich interface technology [6]

As illustrated in <Figure 3>, contents in rich emotional context are expressed in AR-based display technology. Above <Figure 3>, which is the scheme of rich technology, shows how a user's demand is expressed into his/her demand for detailed functions and contents. It is then expressed into interaction design and information architecture, in which information is designed as navigation and interface. The final design in visual expression becomes rich in emotions with touch technology. What

kind of technology is required to develop system and service enabling such information processing? The answer lies in AR system and SNS technology based on location information.

3.3 Location-Based SNS

Information gained from smartphone (location, sound, tag, e-mail, address book, etc.) fuses with information on the mobile-web to be processed into a new information that is meaningful. Information obtained from sensing of smartphone is used to beef up or correct limited information that moves around in the real world. This information, in turn, becomes more useful for people's daily lives or offers a new insight to individuals and groups.

Information expressed in AR technology tightens connection with social point of contact among individuals, and such highly-involved participation as well as bond among individuals offers social, political and economic drivers.

Social network analysis technologies have been a popular research subject in many areas eager to analyze relations among users. [7]

First, AR is a technology used to make expressions by combining medium (text, image, sound, video, etc.) linked to the real world. Expression based on medium-orientation puts humans on the periphery and limits itself only to expressing a relation map forged around objects. Human-oriented expression, on the other hand, puts humans at the center to express mediums and objects via context-awareness and contextual reasoning. Services based on AR only analyzes log information (ex. e-mail, call records, etc.) of a communication medium to compose a network based on probabilistic model. These, however, are incapable of dynamic network composition when a user makes behavioral change. [8]

Second, objects and mediums are newly composed in a social aspect. In other words, personal profile, phone number, address book, e-mail and other social information stored in a smartphone are combined with profile information in a social network service picked from the web to create a social meta information, which in turn restructures the map, location information and AR expressed in smartphone into 'human' orientation. In research [9] and [10], which employ social network analysis method in an ubiquitous computing environment, user profile entered earlier or user survey results after a specific event are used as the basis to form social network. While this speeds up relations sampling through rapid prototyping, it challenges creating a network that can respond to user's behavioral change on a real time basis. [8]

Such limitations can be overcome by developing it into a large story, which is the history of a group or an individual inferred by a unique place, theme, time and family members, and is beyond merely mixing individual stories with other mediums.

4 SNS Model Development

4.1 Existing Service Model

<Figure 4-6> below gives overview models served on AR basis and their limitations based on which this research will augment strength and make up for weaknesses of the service model to be proposed.



Fig. 4. Ineedcoffee (www.ineedcoffee.com)

<Figure 4> shows a nearby coffee shop via smartphone camera. The coffee shop's image moves on top of the smartphone's subject based on location information. Information is indicated in a small size on the layer arranged as north, south, east, and west at the bottom of screen layout. Information UI (user interface) and expression are displayed in AR, which enables users to check where the coffee shop is located and its direction. The future, however, will call for a more detailed information processing such as user review, rank and order of most frequent visitors instead of a mere location information.

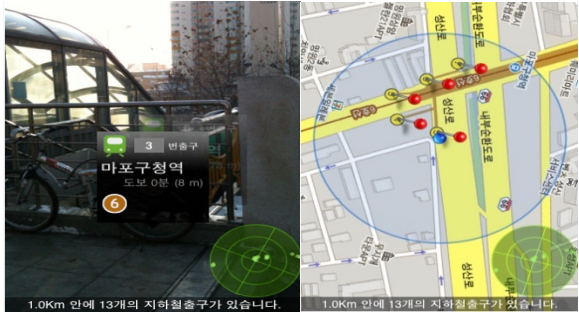


Fig. 5. odiyar (<http://itunes.apple.com/kr/app/id350727509>)

<Figure 5> shows how smartphone camera offers subway information via AR. Its layout also displays direction at bottom right but information is only limited to projected image, location information and subway station itself. Stations of departure, arrival, social network friends such as Twitter nearby the station and list of friends shall be provided in the future for tighter network.

<Figure 6> provides information related to lifestyle in Korea in package type but it fails to form a social network through personalized service. Distance to and location of a friend nearby are provided when a subject in the camera moves across all directions. Twitter or Facebook members can have access to a more personalized and social service by sharing information by a friend or a friend's friend, and sharing that information.

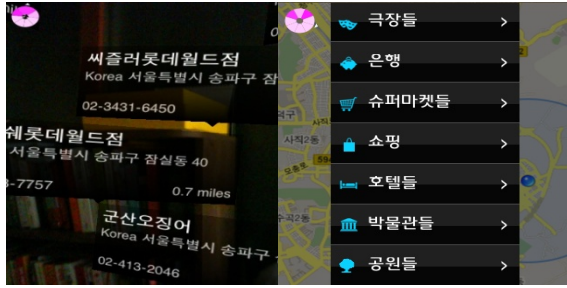


Fig. 6. acrossair (<http://www.acrossair.com/>)

Technology related to AR and application service is limited in that they cannot expand point of contact in a network by creating a large story or tapping into social basis. Hence, this research will propose a new service model that removes such barriers as follows.

4.2 Proposition

(1) Social Layer Method

AR technology enables sound (meta tag), location information (automatic adding tag), mesh-up and display of camera subject as shown in <Figure 7>.

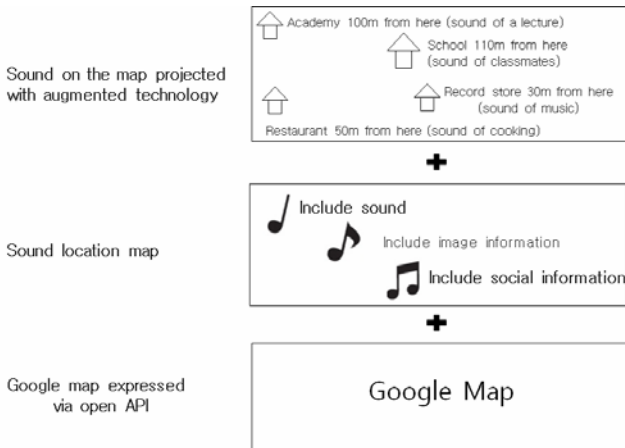


Fig. 7. AR and layered social service

<Figure 7> is a case of how AR is delivered in Google map API. [11] Maps, aerial photos and street view from Google(www.google.com), Naver (www.naver.com), Yahoo(www.yahoo.com) and Daum (www.daum.net) can be retrieved from an external API for display. Location information automatically inserted is mapped based on context when a sound in a map obtained from external API via layer technology is stored and uploaded. User's current location information (GPS), mobile

smartphone's gravity sensor, and navigation information are combined to augment sound's location, distance, direction and information in the user's location for display.

(2) Social Platform Model Proposed as a Case

Information on GPS in which an individual is located needs to be entered as a value to deliver personalized AR. At the same time, major keyword and context information, both of which have a weighted value by individual, are in demand. For example, if an analysis points to banking business as one of the essential lifestyles of an individual, bank's main business and related ads may be indicated via sound from the user's location.

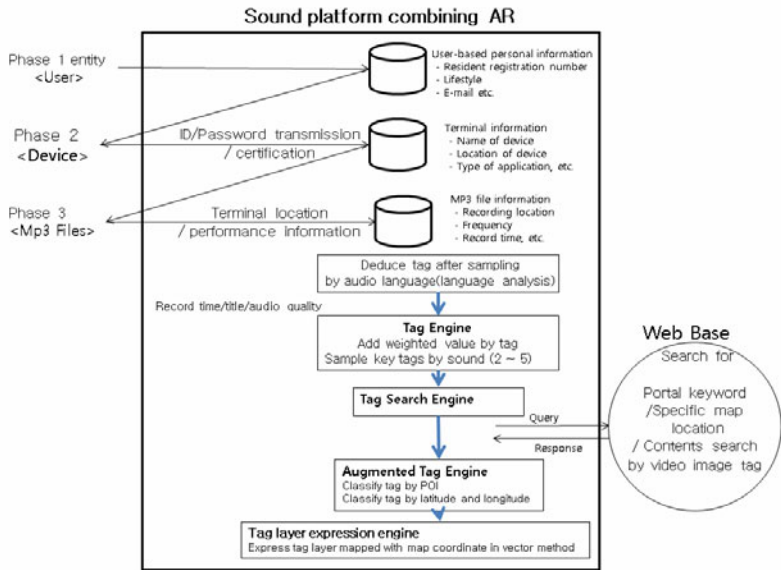


Fig. 8. Sound-based social platform model combining AR

Sound can be marked by notes (mark sound's location, criticality and ranking by applying marking of existing notes and beat), direction (four cardinal points from the user's current location by linking with the navigator) and distance (m, km or else computed with GPS). Display can also take the form of air method in which sound information floats like an air bubble in the visible subject enabled by layer technology or of marking on the radar.

5 Conclusion

This research gave an analysis on limitations of existing service that failed to tell a story via social network or create a large story in terms of context-awareness and information. The study, as an alternative, developed a model with social-based augmented expression and proposed it as a case model, of which its prototype and social platform model can be up for wide application in related business models in the future. Mobile AR, in particular, could serve as a key supporter for studies featuring social mobile web. The

chances of stronger demand will be greater if technology that accurately maps external map data and location information of smartphone becomes more precise. GPS, navigator, gravity sensing retrieved to API (Application Program Interface) and mapping them with the image in the map will display sound-based network information to create new networks and will ultimately establish a social network by linking a certain location, region and contents with sound serving as a meta tag.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0028122).

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-330-B00017).

References

1. Kang, J.-m., Lee, W.-J., Song, Y.-J.: A study for vulnerability analysis and guideline about social personal broadcasting service based on smartphone environment (focus on SNS or U-Health). *The Journal of the Institute of Webcasting, Internet and Telecommunication* 10(6), 162 (2010)
2. Kang, J.-m.: *New Media and Politics of Communication*, p. 15. Hanwool, Seoul (2009)
3. Yonhap News, Seven million smartphones in the market... grew by 8.5 times year-on-year (December 5, 2010),
<http://finance.daum.net/news/finance/world/MD20101205062206411.daum> (search date: December 2010)
4. ATLAS Mobile Index, <http://www.itconference.co.kr/142129> (search date: February 2011)
5. Lee, S.Y.: *Smartphone, Now, Mobile Platform Trends*, 23/48slides (2010),
<http://www.slideshare.net/bluse2/smartphone-platform-trend>;
- Ha, Jo.-d., So, H.-c.: Mobile phone industry, Korea Equity Research Sector report, p. 19 (2009) (mixed date: February 2011)
6. Garrett, J.J.: *The elements of user experience* (2000),
<http://www.jjg.net/elements/pdf/elements.pdf> (search date: February 2011)
7. Zhou, D., Manavoglu, E., Li, J., Giles, C.L., Zha, H.: Probabilistic models for discovering ecommunities. In: *Proceedings of the 15th International Conference on World Wide Web 2006*, pp. 173–182. ACM, New York (2006)
8. Han, J., Woo, W.: Context-based Social Network Configuration Method between Users. In: *HCI 2009*, p. 12 (2009)
9. Axup, J., Viller, S., MacColl, I., Cooper, R.: Lo-Fi Matchmaking: A Study of Social Pairing for Backpackers. In: *Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS*, vol. 4206, pp. 351–368. Springer, Heidelberg (2006)
10. Hope, T., Hamasaki, M., Matsuo, Y., Nakamura, Y., Fujimura, N., Nishimura, T.: Doing Community: Co-construction of Meaning and Use with Interactive Information Kiosks. In: *Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS*, vol. 4206, pp. 387–403. Springer, Heidelberg (2006)
11. <http://code.google.com/intl/ko-KR/apis/maps/index.html> (search date: February 2011)

Design and Implementation MoIP Wall-Pad Platform for Home-Network

Yong-kuk Jung, Min-Su Kweon, Kwang-Seon Heo, Dae-Sung Kim,
and Young-Gyu Choi

Department of Computer Engineering, Chungju National University,
Chungju, 380-702, Korea
{ykjung, kw0410, root, niceguy1314, ygchoi}@cjnu.ac.kr

Abstract. This paper is to implement MoIP platform to send and receive video and audio at the same time by using high-performance Dual Core Processor. Even if Wall-Pad key component of a home network system is released by using embedded processors, it's lacking of performance in terms of multimedia processing and feature of video telephony through which video and voice are exchanged simultaneously. The main reason could be that embedded processors currently being used do not provide enough performance to support both MoIP call features and various home network features simultaneously. In order to solve these problems, Dual processor could be used, but in the other hands it brings another disadvantage of high cost. Therefore, this study is to solve the home automation features and video telephony features by using Dual Core Processor based on ARM 11 Processor and implement the MoIP Wall-Pad which can reduce the board design costs and component costs, and improve performance. The platform designed and implemented in this paper verified performance of MoIP to exchange the video and voice at the same time under the situation of Ethernet network.

Keywords: Home Network, Wall Pad, MoIP, ARM11, S3C6410.

1 Introduction

Home network system is installed Wall-Pad inside the house based wireless devices in the house and integrated management of various sensors and a complex network of systems integration and unmanned patrol, intergenerational communication, and parking management, remote control, and crime prevention inside the house, Disaster Prevention and living closely with and provide a variety of services is an integrated system. Advances in information and communications technology and infrastructure, as the Internet, serial communications and analog transmission technology, using technology to restrict and control of currency used to provide home automation system at a faster rate has been converted to a home network system. But the home network system to perform a key role in Wall-Pad due to performance limitations of embedded systems integration services, except for simple Internet access existing home automation system and the distinctive feature is a reality that does not provide in present. So in this study, with Dual Core Processor ARM11 core is based on

Samsung S3C6410 Processor using video, audio information encoder & decoder MoIP platform to perform at the same time are presented, and the implementation of the proposed platform, the existing as well as in home automation technology, MoIP interactive Internet telephony features and services such as Internet interworking function can be verified that works.

In this study Wall-Pad Main Processor System using the ARM11 Processor and operating system, Microsoft's Windows CE 6.0 R2 were used for development environments with platforms using Visual Studio 2005 is implemented. And fabricated using a sample digital two-way video calls through the IP network can be verified that. However, one of the features in this study Wall-Pad's RS-485, PLC, Zigbee, Bluetooth, and control and metering devices used parts were excluded from the test item.

Chapter 2 of this paper provides an overview of embedded systems as research platforms and related technologies, and particularly describing an overview of the ARM11 core used in this study and describes the main features and the home network system configuration and technical characteristics, and Chapter 3 in this study, designed for ARM11 Processor platform was described in detail. Chapter 4 Implementation of a video call with a sample board to show results, and finally in Section 5 conclusions of this study and future research directions are described.

2 Related Research

2.1 Embedded System

What is embedded systems, typically for performing specific tasks, including hardware and software system that can be application specific. In other words, a microprocessor or a microcontroller with an integrated circuit consisting of only rather than simply to perform a particular function is entered the program has a built-in system. For example, as changes in consumer electronics products, the information processor, operating system, applications, application software is shipped with a mounted TV, audio, refrigerators, washing machines and household appliances and medical devices, navigation devices, multimedia players, home networks, etc. There are many thousands. Many techniques are applied in the field of embedded systems have a close relationship with our lives and the remarkable thing about the future in all areas of our lives that is expected to be used more and more deeply.[1][4]

Embedded processors, 68K, x86, PowerPC, SPARC, and embedded IP (Intellectual Property) is ARM, MIPS, SuperH and so on. Motorola's 68000 (68K) series of embedded processor market is one of the oldest products. Intel's x86 is a PDA based on a WinCE or Linux-based Desktop-class software, such as set-top box that acts as application was used for PC-in-drag. SUN Microsystems 32/64 bit RISC architecture, the SPARC (Scalable Processor ARChitecture) processors used in workstations as well known. ARM low-power consumption requirements of the mobile market is unique in the embedded processor on the market today that boasts a powerful processor. Intel's ARM core family has Xscale micro-architecture derived from StrongARM architecture and PXA26X, PXA250, PXA800F are also based on micro-architecture. In addition, also based on Xscale micro-architecture, for network storage

applications, network processor product, IXP2800, IXP2400, IXP425 has been supplying. In this study, published in ARM11-based Samsung S3C6410 processor is used.

2.2 ARM Processor

ARM Core is an acronym of Advanced RISC Machine it uses general purpose in embedded system, RISC(reduced instruction set computer) Core for mobile it has cheap price and low-cost of electricity. These products are proven solution by several companies and researches its range of usage keep spreading. It produces by OEM(Original Equipment Manufacturing) and associates with several companies. In Korea, Samsung and Hyundai & some companies have partnered. And according to each company's model is provided many peripheral controller from a single CPU(Central Processing Unit) chip, it changing rapidly because technological advances by demands on the Time.

ARM 11 micro-architecture, it adopts ARMv6 architecture have many advantages to a software developer. This includes various media processing order that increases audio/video application's speed, memory system architecture realizes high-performance OS, a new order increases real-time efficiency and speeding interrupt responds. Besides, it is easy to do a data sharing with another processor, an application's porting from non-ARM processor because of more new applications require multi-processor formation that is multiple ARM core, ARM + DSP system and so on.

ARM11 micro architecture is changed to 8-state pipeline instead of previous ARM10's 6-state pipeline. ARM11 micro architecture's 8-state pipeline realizes 40% volume of processing enhanced by comparison to existing ARM core. And 8-state pipeline executes 8 stages of process at the same time.

Change to ARM 11, meanwhile ARMv4T's command structure what has been used to 32/16bit RISC architecture that improved to Powerful ARMv6 command set architecture, especially developed for the needs of Next Generation User's machine, Wireless communication, Networking, Automobile products. Especially ARMv6 architecture focuses on improvement of four abilities. One, twice faster MPEG4 encoder & decoder and audio, DSP(Digital Signal Processor) can be exist because expand of multimedia processing. Two, cache architecture's betterment causes decline of overload from cache and context exchange those have physical address. Three, improvement of exceptional process and interrupt process had ability enhancement in Real-Time task. Four, supporting unsorted & unarranged data and mixed bite system data, and also simple data supply & application porting. These cause memory economization. And time for exceptional process & interrupt process shortened. In this study, add SRS & RFE commands below do more effective status saving & restoration course when mode changes. Then add new CPSID & CPSIE commands can do mode changes except User Mode.

ARM 11 improves SIMD(Single Instruction Multiple Data) media processing that extraordinarily increases 3D graphic acceleration, contains 16KB command & Data L1(Level 1) cache memory per each, and also has MMU(Memory Management Unit), it designed for hardware but designed for decide to use or not through register setting-up. Hereupon add each L1cache produces 64bit interface so that makes faster access.

S3C6400 is optimized to PDA, 2.5G/3G high-performance cell phone, PMP(portable media player). It has stand on the basis of ARM1176EJF-S CPU core by Samsung. S3C6400 works at 400/533MHz and contains internal bus structure, that is consist of 64/32bit AXI, AHB and also have DRAM port, two external memory ports which can connect Flash/ROM/DRAM. Besides it has variety of peripheral hardware which can be easily expanded. S3C6400 supports 4096 X 4096 resolution, zoom, rotation, color space conversion, LCD controller, various functional-direct connect, etc.- camera interface, MPEG4 & H.264/AVC supported LCD controller, video port processor, TV encoder, audio interface contained. And also have external expanded 4-channel UART, I2C, I2S, 2Ch SPI, HIPI HSI, IrDa and so on various interface, MMC/SD Host, USB1.1 Host, USB 2.0 OTG(On-The-Go) and so on supports.[3]

Using S3C6410 ARM11 main processor, MoIP home-network Wall-Pad platform, takes charge phone call function is realized Fig 1 below.

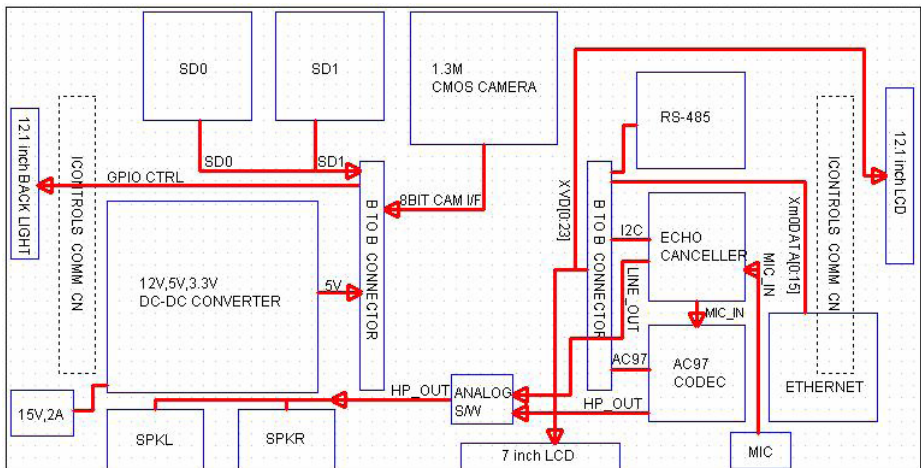


Fig. 1. Base Board block planning

① ARM Core : ARM1176JZF-S core, OS and Application program can be execute.
System Peripheral : System's important part, DMA, PLL, RTC, etc contains.
Multimedia Acceleration : Divided into Multimedia Block / Codec part.
Multimedia Codec is made up of separated ARM core.

Connectivity : An interface to use various peripheral.

Power Management : Manage S3C6410's CPU Mode & each block's power.

- ⑥ TFT LCD Controller : TFT LCD Controller supports both system interface and video interface.

Memory Subsystem : Memory Subsystem supports NOR Flash, SRAM, SDRAM, DRR, NAND Flash and so on various memory.

3.1 Design Audio Device Driver

In board, this paper uses WM9713 by Wolfson as external codec through S3C6410's AC97 interface then print PCM Data of Microsoft Windows CE. And then WM9713 do whether input or output through various path. S3C6410t supports AC97 Ver 2.0. AC97 controller communicates with AC-Link. CPU's AC97 controller transfer stereo's PCM data to external codec. External codec(WM9713) transform PCM data to Analog Waveform through DAC(Digital-to-Analog Converter). Analog waveform as transformed through an amplifier to speaker.

WM9713 has two ADC(Analog-to-Digital Converter) and five DAC. ADC's role is input data from mic or Aux, Line, Mono and so on then transform into PCM digital data. Next, Windows CE gets ready to record through AC97 interface. And five DAC have to do various output Windows CE's PCM digital data through AC97 interface.

Windows CE can be output to a variety of sound formats. From a simple wave MP3, G.711, ADPCM compressed sound, such as interpreted by the software after the AC97 codec but interface is output through the speakers. Windows CE does not support the hardware volume control but a software volume control.

MoIP calls received from the other party G.711(GSM610) digital audio data using a software codec output to main speaker, and micro-PCM data inputted by using the software codec G.711 (GSM610) and the other sent to the encoding to become.

3.2 Embedded System

S3C6410 has a camera interface and board a digital CMOS (Complementary metal-oxide-semiconductor) camera is connected. Digital CMOS camera is a camera module using Micron's MT9M112. MT9M112 integrated with the image sensor chip and ISP. Inside the Sensor Core, Image Flow Processor Camera Control and Image Flow Processor Color pipe there. For Image Flow Processor Camera Control can control the sensor, Auto exposure or Whit balance, Flicker, such as sensors to control incoming through the lens.

Image Flow Processor Image Flow Processor Camera Control in a Bayer RGB data is converted to accept or image resizing BT 601/656 Image Resizing and the same treatment. The sensor is connected to the core of the MT91M112 S3C6410 via I2C MT9M112 can control the internal registers.

3.3 Ethernet Device Driver Design

The board DM9000B used for Ethernet. DM9000B the Cable Auto Detect (Cross, Direct) and automatic detection is 10M/100M. DM9000B LED 2 indicates the state that supports ports and the MAC Address in EEPROM through EEPROM interface

can be read. The registry is stored in the EEPROM read the MAC Address because board has non EEPROM. If you don't have MAC Address Registry E-boot loader stored in the TOC area, read the MAC Address.

DM9000B is connected to the memory controller of the S3C6410 DM9000B S3C6410 interrupt pins should be connected to the external interrupt pins. S3C6410's one kind of restriction on the external interrupt pins should be connected to more than No.8.

4 Operation and Performance Verification

4.1 Test Preferences

Implementation platform, the display of samples of the 7 "800x480 TFT LCD, the camera CCD (Charge Coupled Device) camera was used. By creating two sets with each set through RJ-45 Jack UTP-Cross Cable is connected to. After booting the system then Fig 2 below MoIP Call Test Program will be executed.

MAC Address and System ID No. 1 of the NAND block are stored in the TOC area. In this system, sample number 1, "Room 101-101", sample two times, "Room 101-102" was set. Fig 2 below the test program is executed the screen.

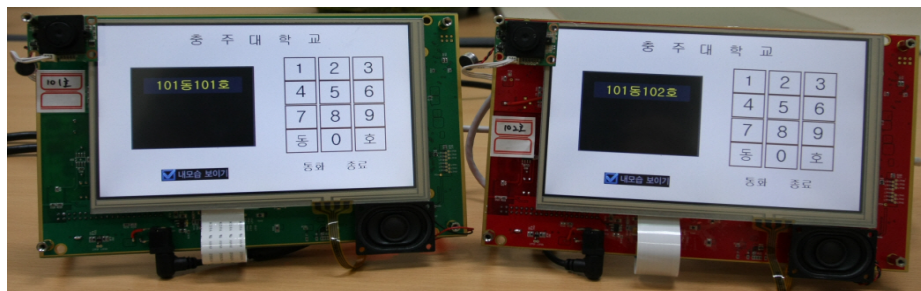


Fig. 2. MoIP Call Test Program

4.2 Results and Performance Evaluation

Fig 3 below MoIP of running through the other side of the screen we can see that the image is displayed. Set the left (101-101) and the right set (101-102) and the other using the camera to show the big screen at the bottom of the right to self-image as a small screen that shows the implementation of real-time video calls to check directly that could be.

However, the following are a few things to note.

First, 100M Speed Hub and that you must use to support the stable 30fps 320x240 size screen as the quality of interactive video can display.

Second, the voice call at least a microphone and speakers, a distance of 15cm or more to maintain spacing between sets should be used in a separate space echo does not occur.

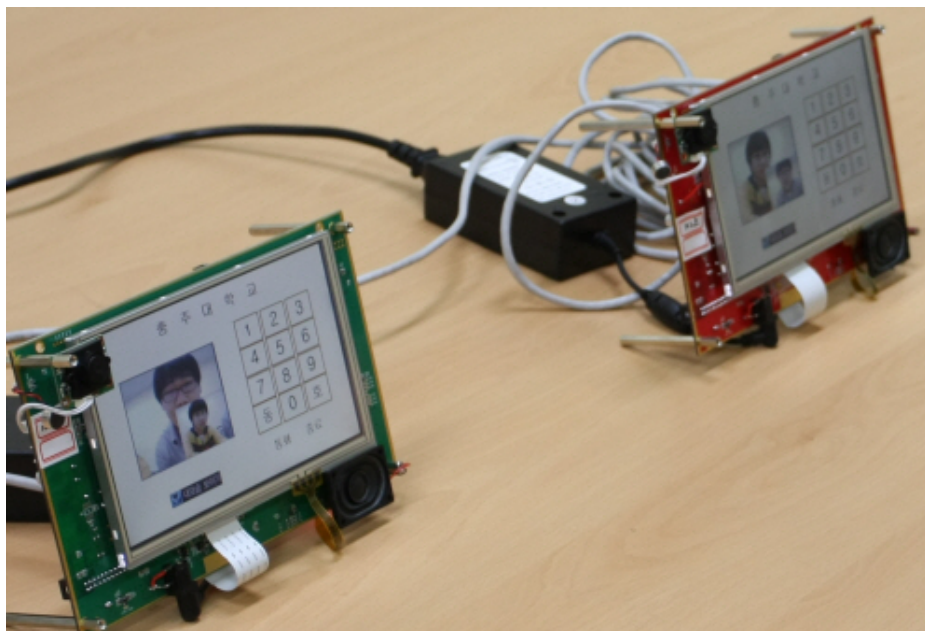


Fig. 3. MoIP execution screen

As a result, implemented a set of images in a separate room when the call seamlessly from video and voice calls and video calls to perform functions based MoIP were found.

5 Conclusions

In this paper, we have implemented Wall-Pad ARM11 Dual Core Processor based platforms, such as the purpose of apartments in multi-family housing is widely disseminated as part of the home network technologies, as well as home automation features and fusion of broadband IP-based video telephony systems embedded processors MoIP Wall-Pad was based on the platform to implement. Main Processor S3C6410 of included used Samsung. By using this platform to the Ethernet network system over conventional analog video calls and video calls for a separate pipeline, which does not require wiring has the advantage. In addition, the processor than when using a unified Dual Processor were able to reduce cost, low power, high-performance Wall-Pad was able to develop.

Recently a growing number of smart phone users, so you can see inside the house Wall-Pad system is equipped with an additional camera Wall-Pad server roles within the family as a smart phone by allowing you to transmit audio and video surveillance, and home monitoring is available control of home appliances available to All-IP based platform will be able to develop.

References

1. Jo, J.h.: Design and implementation MoIP Wall-pad platform based on ARM9 dual processor, Joong-ang University (2009)
2. Kim, D.W.: A Real-time Intelligent Home Network Services Control System design. Gyeongsang National University (2009)
3. Kang, S.b.: A Study on Implementation for Wireless Gas Sensor Data Transmission Platform using ARM11 and Linux, Ajou University (2009)
4. Sin, J.w.: Design and implementation of home automation system using embedded system, Kyungsung University (2005)
5. Written by D. Bowling, Translated by Shin, j.c., Lee, J.Y., Windows Embedded CE 6.0 programming. ACORN (2009)
6. Go, j.g.: Windows Embedded CE programming: Introduction, Jungbomunhwasa (2008)
7. S3C6410X RISC Microprocessor USER'S MANUAL REV 1.10, Samsung Electronics (2008)
8. WM9713 Data Sheet Rev 3.2. Wolfson Microelectronics plc (2008)
9. MT9M12 Data Sheet, MICRON (2009)
10. DM9000B Data Sheet, DAVICOM (2007)
11. <http://www.arm.com>
12. <http://www.msdn.com>
13. <http://www.kldp.org>
14. <http://cafe.naver.com/wincepro>

Outlier Rejection Methods for Robust Kalman Filtering

Du Yong Kim¹, Sang-Goog Lee², and Moongu Jeon¹

¹ School of Information and Communications,
Gwangju Institute of Science and Technology, Korea
{duyong,mgjeon}@gist.ac.kr

² Department of Digital Media,
Catholic University of Korea, Korea
sg.lee@catholic.ac.kr

Abstract. In this paper we discuss efficient methods of the state estimation which are robust against unknown outlier measurements. Unlike existing Kalman filters, we relax the Gaussian assumption of noises to allow sparse outliers. By doing so spikes in channels, sensor failures, or intentional jamming can be effectively avoided in practical applications. Two approaches are suggested: median absolute deviation (MAD) and L_1 -norm regularized least squares (L_1 -LS). Through a numerical example two methods are tested and compared.

Keywords: Kalman filtering, median absolute deviation, L_1 -norm optimization.

1 Introduction

An estimation problem is a fundamental task in communications when the messages or signals are corrupted by noises. It has been investigated in statistical signal processing field intensively by designing filters. Among many signal estimation techniques, if we can model the signal in linear equation with Gaussian random variables, the celebrated Kalman filter is notably utilized [1]. Kalman filter is known to be optimal when the time evolution of the signal is linear and the noise is assumed to be Gaussian. However, it is very strict assumption often violated in practical applications. Linearity and Gaussianity have been relaxed by many researchers to develop more general or dedicated usages [2].

Considering the Gaussianity, there have been attempts to generalize a single Gaussian to more complicated distribution, mixture form of Gaussian approaches are suggested [3]. However, in this case, multi-modality of signal is vague and even if it is meaningful in practice it is hard to get the sophisticated model due to the tuning of latent variables [3]. Particularly in many cases, it is sufficient to model only the outlier based on the single Gaussian model. Then, the problem is how to model outliers and efficiently reject them to improve the performance of filter.

In practical applications outliers occur due to the unmodeled channel uncertainty, spikes, sensor failures, or intentional jamming. Outliers handling has been studied in the literature, for instance, robust statistics [4], median absolute deviation (MAD) [5], and L_1 -norm optimization [6]. However, those methods require batch process and time consuming computations which restrict real-time applications.

In this paper, we propose two novel outlier rejection methods that are incorporated into Kalman filtering to be used in real-time applications. The first method utilizes MAD algorithm to trim the measurement to cope with the outlier in Kalman filtering update. The second method is suggested using L_1 -norm regularized least square (L_1 -LS) which is more robust and consistent than the first method. To verify the efficacy of the proposed methods, a target tracking problem is simulated.

The remainder of the paper is organized as follows. Section 2 formulates the estimation problem. Proposed algorithms are explained in Section 3. A numerical example is provided in Section 4 to demonstrate the advantage of algorithms over the standard Kalman filtering. Finally, conclusion is made in Section 5.

2 Problem Formulation

The evolution of signals over time or object motion can be represented by using dynamic system model. Most of dynamic system models in the estimation problem are given as linear systems or linearized non-linear systems to be easily implemented in the celebrated Kalman filtering. In this paper, we consider a time-invariant linear dynamic system with Gaussian noise as

$$\begin{aligned} x_{t+1} &= Ax_t + w_t, \\ y_t &= C_t x_t + v_t, \end{aligned} \quad (1)$$

where $x_t \in \mathbb{R}^n$ is the state (to be estimated) and $y_t \in \mathbb{R}^m$ is the measurement of the state at time t , respectively. $A \in \mathbb{R}^{n \times n}$ is the system matrix which describes the time evolution of signals, $C_t \in \mathbb{R}^{m \times n}$ is the measurement matrix. $w_t \in \mathbb{R}^n$, $v_t \in \mathbb{R}^m$ are system noise, and measurement noise, respectively. Usually, the process noise w_t is independent identically distributed (iid) $N(0, W)$ and the measurement noise v_t^i is iid $N(0, V)$. Assume that all the noises in different time and different kind of noises are independent each other. Then, the main goal is to estimate the state x_t given measurements up to time t , i.e., $y_{1:t} = \{y_1, \dots, y_t\}$.

In the linear dynamic system estimation, this problem is exactly solved via Kalman filter as known as the optimal solution. Kalman filter gives us the complete posterior probability density function (pdf) of the state $p(x_t | y_{1:t})$ using probabilistic mean $E(x_t | y_{1:t})$ and its covariance $P(x_t | y_{1:t})$ in the recursive form. To make the paper self-contained, Kalman filter equation is composed of prediction stage and update stage, summarized as follows.

Given $E(x_{t-1} | y_{1:t-1}) \triangleq \hat{x}_{t-1|t-1}$, and $P(x_{t-1} | y_{1:t-1}) \triangleq P_{t-1|t-1}$

Prediction : $\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}$,

$$P_{t|t-1} = AP_{t-1|t-1}A^T + W, \quad (2)$$

$$\begin{aligned}
\text{Update : } \hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t \left(y_t - C_t \hat{x}_{t|t-1} \right), \\
P_{t|t} &= (I - K_t C_t) P_{t|t-1}, \\
K_t &= P_{t|t-1} C_t^T \left(C_t P_{t|t-1} C_t^T + V \right)^{-1},
\end{aligned} \tag{3}$$

where $(\bullet)^T$ is the matrix transpose, K_t is the Kalman gain.

To apply Kalman filter equations in the practical estimation problem, we need to model the dynamic system as (1). When there is an inevitable modeling error, we try to reflect the uncertainty in the form of system noise term w_t which is Gaussian. Here, we assume that there is no unmodeled uncertainty, i.e., modeling mismatch is not considered.

The practical challenge of our interest is outlier measurement which means the modeling mismatch in measurement noise model. The modeling error in measurement noise model is related to the sensor failure, spikes, or jamming which is not Gaussian. To simulate these unmodeled measurement uncertainty, we add a sparse non-Gaussian error term z_t . Then the measurement model is rewritten as

$$y_t = C_t x_t + v_t + z_t. \tag{4}$$

Note that the sparse error also can be represented as the Tukey's gross error model which is contaminated Gaussian model [6]. However, we consider the sparse error as the additional term to handle it separately in the proposed algorithm.

3 Proposed Algorithms

3.1 Measurement Outlier Rejection via MAD

The intuition of the first algorithm is that if we take a bunch of samples and calculate its median and test the absolute deviation of each sample, we can easily detect the gross errors. In other words, we reject outliers using statistical threshold test.

Here, the important point is that we use the median of a set of samples instead of using the mean. The statistical mean is popularly used in statistics because in many cased random variable is assumed to be Gaussian. However, outliers cannot be sampled using Gaussian distribution so that we use the median instead of mean. Consider a set of absolute values of residual samples in ascending order as $\{\tilde{y}_{(1)}, \tilde{y}_{(2)}, \dots, \tilde{y}_{(M-1)}, \tilde{y}_{(M)}\}$ where $\tilde{y}_t = y_t - C_t \hat{x}_{t|t-1}$.

Then, the median of sample is represented as

$$\text{med } y = \begin{cases} \tilde{y}_{(l+1)}, & M = 2l + 1, \\ \frac{\tilde{y}_{(l)} + \tilde{y}_{(l+1)}}{2}, & M = 2l, \end{cases} \tag{5}$$

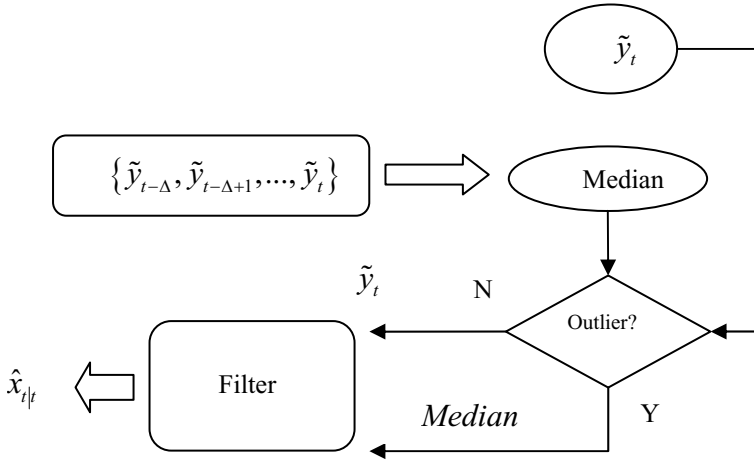


Fig. 1. Algorithm flow of MAD based Kalman filtering

where "med" represents the median of sample set, and M is the number of samples. However, this process is inherently a batch process that means we require a set of sample set to calculate the median. Unlike the mean calculation, a recursive form of median calculation does not exist.

To apply the median test in the sequential estimation, i.e., Kalman filter, we suggest to use a sliding window approach that can be thought of as a semi-batch method. In the sliding window approach, we consider a set of recent Δ measurements set in time interval $[t-\Delta, t]$ which is receding in the next time step, i.e., $[t-\Delta+1, t+1]$. Hence, the oldest measurement is discarded and a new measurement is included in the window. The median is calculated based on the samples in the sliding window and the median test is done with the measurement of time t . According to the outlier test using the median and the current measurement, Kalman filter update of (3) considers the median as its measurement if the measurement is detected as the outlier or original current measurement, otherwise. The overall procedure of the algorithm is described in Figure 1.

3.2 Measurement Outlier Rejection via L_1 -LS

In the second method, we utilize the L_1 -norm least square approach to avoid outlier measurements. As mentioned in the first method, mean is not useful when the outlier is associated. That is because a large magnitude of outlier can easily distract the mean value. In Kalman filter, deviation of the measurement error is considered in Kalman gain calculation K_t as inverse of measurement error covariance V^{-1} . However, outlier measurements cannot be handled here so the first method switches the outlier measurement as the median when the outlier is detected.

Unlike the first method, we set the estimation problem as a L_1 -norm regularized least square (L_1 -LS) and solve it with convex optimization algorithm to obtain the estimate of the sparse error z_t .

If we consider Kalman filtering as a least square problem, the cost function is described as

$$(v_t)^T V_i^{-1} v_t + (x_t - \hat{x}_{t|t-1})^T P_{t|t}^{-1} (x_t - \hat{x}_{t|t-1}), \quad (6)$$

subject to $y_t = C_t x_t + v_t$.

The defined cost function (6) is optimally minimized using Kalman filtering equation given in (2)-(3).

However, if the sparse error z_t is considered and using L_1 -LS form, we redefine the cost function as

$$(v_t)^T V_i^{-1} v_t + (x_t - \hat{x}_{t|t-1})^T P_{t|t}^{-1} (x_t - \hat{x}_{t|t-1}) + \lambda \|z_t\|_1, \quad (7)$$

subject to $y_t = C_t x_t + v_t + z_t$, where $\|\cdot\|_1$ denotes L_1 -norm and λ is the regularization parameter. In the optimization literature, the minimization of the cost function (7) is solved via convex programming that is known as computationally expensive.

Recently, advances in computer hardware and sophisticated algorithm enable us to solve L_1 -norm optimization problem in real-time [6]. We adopt the robust Kalman filtering approach suggested in [6] but additionally consider the thresholding procedure to enhance the speed and the accuracy.

We used the converted cost function of (7) as

$$(e_t - z_t)^T Q (e_t - z_t) + \lambda \|z_t\|_1, \quad (8)$$

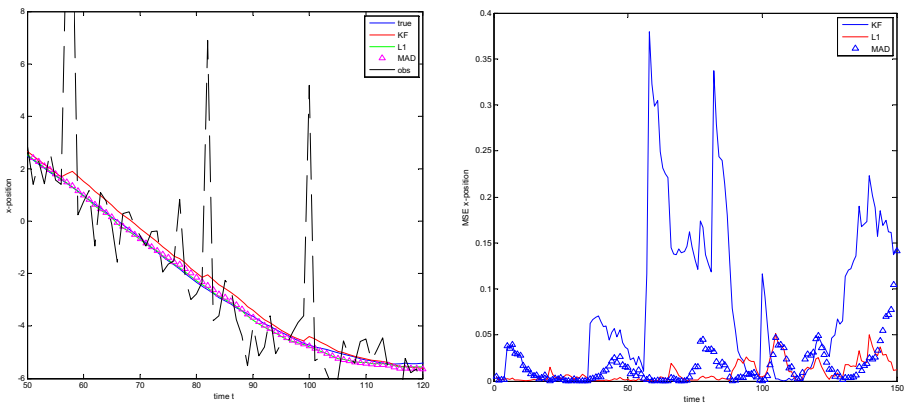


Fig. 2. Estimated trajectory comparison and its MSE

subject to $e_t = y_t - C_t \hat{x}_{t|t-1}$, where $Q = (I - C_t K_t)^T V^{-1} (I - C_t K_t) + K_t^T P_{t|t}^{-1} K_t$ and I is the identity matrix of appropriate dimension. Then, the update state is given by $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (e_t - z_t)$. Note here that we solve (8) only when the threshold testing in the first method detects the outlier. In both algorithms, outlier detection with threshold testing is given by (9)

$$|y_t - \text{med} \{y_{t-\Delta t}\}| > \gamma \text{MAD}\{y_{t-\Delta t}\} \quad (9)$$

where $\text{med} \{y_{t-\Delta t}\}$ is the calculated median of a measurement set $y_{t-\Delta t} = \{y_{t-\Delta}, \dots, y_t\}$, and γ is the tuning parameter which is usually defined as 3. $\text{MAD}\{y_{t-\Delta t}\}$ represents the median absolute deviation given in (10).

$$\text{MAD}\{y_{t-\Delta t}\} = \text{med} |y_{t-\Delta t} - \text{med} \{y_{t-\Delta t}\}| \quad (10)$$

4 Experimental Results

To verify two outlier rejection algorithms, a circular movement of an object is given with model parameters,

$A_0 = 2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $B_0 = 5^2 I_2$, $A = I_2 + \varepsilon A_0 + \frac{\varepsilon^2}{2} A_0^2 + \frac{\varepsilon^2}{6} A_0^3$, and $B = \varepsilon B_0$. In addition, I_2 is

a 2×2 identity matrix which is a discretized model with a step-size $\varepsilon = 0.015$, and the initial position and uncertainty are $x_0 = (15, -10)^T$, and $P_0 = 10I_2$, respectively.

The measurement matrix is $C_t = [1 \ 0]$ which measures x-axis position. Outlier measurements are simulated with the probability i.e., $P(u(0,1) < 0.05)$ where $u(0,1)$ is the uniform probability density function with interval $(0,1)$.

Simulation results are displayed in Figure 2 including estimated trajectories and Mean Square Error (MSE) of three algorithms (KF: standard Kalman filter, MAD: MAD-based Kalman filter, L1: L1-LS based Kalman filter). The figures demonstrate that the two proposed algorithms robustly avoid outlier measurements compared to the standard Kalman filter. Note that the window length of the MAD-based Kalman filter is set as 5 and the threshold for detection is given as the three sigma of the measurement noise.

Based on the experiment, MAD-based Kalman filter is computationally efficient than L1-LS based Kalman filter, however, not consistent in the performance. We think that it is because updating with calculated median is not always reliable.

5 Conclusion

Simple and robust Kalman filtering algorithms are proposed using MAD and L1-LS, respectively. MAD-based Kalman filter proposes the semi-sequential calculation of

the median with a finite length of sliding window to implement in Kalman filtering. On the other hand, outlier is estimated using L_1 -LS algorithm to exclude undesirable effect in Kalman update step. As illustrated in the experimental result, two algorithms are much more robust than the standard Kalman filter in filtering accuracy. Based on this result our future research is directed to the extension toward multi-sensor environment.

Acknowledgments. This work is partly supported by National Research Foundation of Korea Grant funded by Korean Government (No. 2010-001-2932) and by GIST Systems Biology Infrastructure Establishment Grant.

References

1. Kalman, R.E.: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 35–45 (1960)
2. Kailath, T.: *Linear Systems*. Prentice-Hall, Inc., Englewood Cliffs (1980)
3. Merwe, R., van der Wan, E.: Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing* (2003)
4. Huber, P.J.: *Robust Statistics*, 2nd edn. John Wiley & Sons Inc., Hoboken (2009)
5. Nguyen, N.-V., Shevlyakov, G., Shin, V.: MAD robust fusion with non-Gaussian channel noise. *IEICE Transactions on Fundamentals*, 1293–1300 (2009)
6. Tukey, J.W.: Contributions to Probab. and Statist. In: Olkin, I. (ed.), pp. 448–485. Stanford University Press, Stanford (1960)
7. Mattingley, M., Boyd, S.: Real-time convex optimization in signal processing. *IEEE Signal Processing Magazine* 27, 50–61 (2010)

A Study on Receiving Performance Improvement of LTE Communication Network Using Multi-hop Relay Techniques

Chan-Hong Park, Yong-Seok Choi, Young-Hwan Han, and Hyeon-Kyeong Seong

School of Computer, Information, and Communication Engineering, Sangji University,
Woosan 660, Wonju, Kangwon 220-702, Korea
{pchnaya1, yhhan, hkseong}@sangji.ac.kr,
macbeth-choi@hanmail.net,

Abstract. This paper made research to boost reception performance in link-down transmission method of LTE system, which is the next-generation mobile communication technology standard underway in 3GPP. To this end, it proposed installing relay (RS) in between station(BS) and terminal(MS), set the distance between BS and RS at 500m and 1,000m, each, and chose OFDMA and SC-FDMA as transmission method of RS. The paper found SC-FDMA to be better choice in RS when it is closer to BS and OFDMA to be a better choice in RS when the distance between BS and RS is farther. The system's reception performance improved when the most appropriate transmission method fitting the circumstances was used in the middle between BS and MS.

Keywords: 4G, LTE, OFDMA, SC-FDMA, Multi-hop relay.

1 Introduction

In 3GPP LTE-advanced, hybrid OFDMA/SC-FDMA is recommended for its technological capability to make up for the performance gap between OFDMA and SC-FDMA and to make PAPR more efficient[1]. OFDMA is used in LTE downlink in order to make frequency more efficient and raise cell capacity but OFDM struggles with a high PAPR owing to its use of multi carrier wave. LTE uplink, on the other hand, employs SC-FDMA, which is similar to OFDMA, but makes up for OFDM's big PAPR in mobile stations where electricity efficiency is critical[2]. This paper proposed relay to make up for the performance gap between OFDMA and SC-FDMA and performed research to improve reception performance by using OFDMA and SC-FDMA as transfer modes for stations and relays depending on the distance between station and relay. The result of the paper is expected to improve reception performance in the overall system and provide smooth communications service by reducing shadow domain and expanding BS' cell coverage.

2 Hybrid OFDMA/SC-FDMA

2.1 OFDMA

Semi-carrier waves are grouped into sub-channels, which are a larger unit, and these sub-channels are grouped into bursts to be allotted to wireless users. Each burst allotment can be changed in every frame within the order of modulation and this enables dynamic adjustment of bandwidth use according to what the current system in a station requires. Furthermore, power consumption by each user can also be adjusted according to the current system's requirement since each user consumes only part of the whole bandwidth[3].

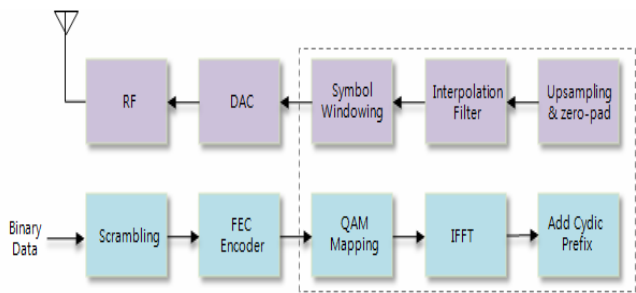


Fig. 1. Transmitters/receivers block diagram of OFDMA downlink system

2.2 SC-FDMA

SC-FDMA, which stands for Single Carrier Frequency Division Multiple Access, can stave off frequency selection attenuation and phase distortion as both FFT and IFFT are applied to both transmitter and receiver [4].

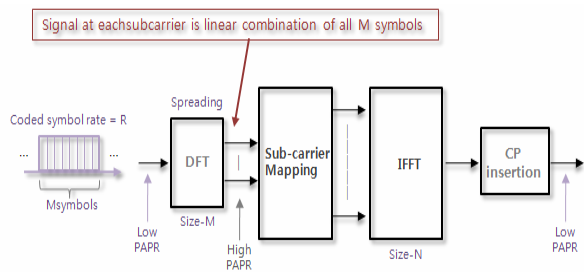


Fig. 2. Structures of SC-FDMA transmitter unit

3 Multi-hop Relay

Relay methods can be split into fixed relay and selective relay based on data relay method. This paper applied DF method in fixed relay, which decodes received signal

up to byte and re-transfers it after coding and modulation[5]. RS, which employs DF method, decodes received signal and transfers re-coded and modulated signal to MS. Signal received by D is as follows.

$$y = h \hat{x} + n \tag{1}$$

$\hat{x}$ is transfer signal in RS that has been re-coded and modulated via DF. Channel capacity with smaller SNR out of BS→RS and RS→MS become that for the system applied with DF method, of which the formula shall be as follows.

$$C = \min \left\{ \frac{1}{2} \log (1 + P), \frac{1}{2} \log (1 + P) \right\} \tag{2}$$

Assuming that valid SNR of both channels is equal, channel capacity of DF method has more gain than channel capacity under AF method as comparison shows.

4 LTE Reception Performance Improvement Test via Multi-hop Relay Technology

This paper proposes making a selection between SC-FDMA or OFDMA depending on each terminal environment in RS in order to bring about optimal performance under LTE environment. Relay installation to optimize performance in shadow domains and the subsequent protocol needed is also proposed in the paper. In cases where transmitter unit and receiver unit is in short distance, OFDMA is used to ensure frequency efficiency and boost cell capacity. In cases of long-distance, SC-FDMA employed in LTE uplink is used to make up for OFDMA's weakness and boost system's performance.

Table 1. The proposed relay protocols

	Protocol	Existing LTE (HSDPA)	Proposal	
			Protocol 1	Protocol 2
Time slot 1 (Phase 1)	BS → Relay	OFDM	OFDMA	OFDMA
Time slot 2 (Phase 2)	Relay → MS	OFDM	OFDMA	SC-FDMA

The method proposed in this paper enabled smoother communication environment by putting RS in between BS and MS to obtain consistently optimal performance under LTE environment. It also proposed ways to properly select between SC-FDMA or OFDMA depending on where BS, RS and MS are located and made simulations to check how these methods turned out. In the simulation environment, which is in Table 2.

Table 2. Simulation parameter

Parameter	OFDM	SC-FDMA
FFT	256	256
Guard Period	FFT/4	FFT*Q/4
Modulation	QPSK, 16QAM	QPSK, 16QAM
Pathloss Model (NLOS)	$27.7+40.2\log_{10}(d)$	$27.7+40.2\log_{10}(d)$
Tx power(dBm)	27 (back off = 6,8dB)	27
Noise power(dBm)	-114	-114
Coding	Convolution (1/2, 1/4)	Convolution (1/2, 1/4)
Channel Compensation	ZF	ZF

5 Simulation

5.1 OFDMA and SC-FDMA Employed as Transfer Mode of RS and MS Assuming 500m Distance between BS and BS

Fig. 3 shows reception performance of transfer modes selected based on RS when MS is 500m from BS. Performance of OF-SC is better than that of OF-OF in 100~200m in which BS and RS are in close distance while RS and MS are in a longer distance assuming that center is 250m. This demonstrates better RS performance when using SC-FDMA than using OFDMA as the distance to RS and MS is relatively long. This also shows that OFDMA mode offers better performance than SC-FDMA when RS and MS are closer to each other.

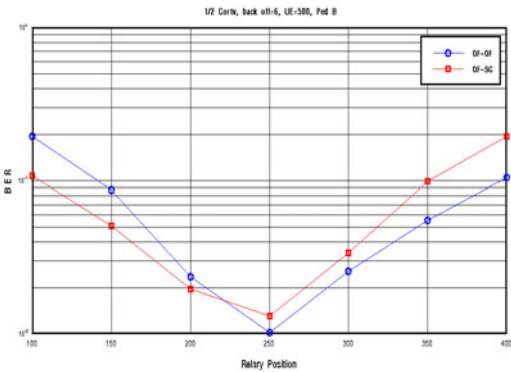


Fig. 3. BER in MS location of 500m and OFDMA transfer mode between BS and MS

5.2 OFDMA and SC-FDMA Employed as Transfer Mode of RS and MS Assuming 1,000m Distance between BS and BS

Fig. 4 is the reception performance of transfer modes selected based on RS's location assuming distance between BS and MS at 1,000m. The simulation fixed OFDMA as the transfer mode from BS and RS and tested both OF-OF and OF-SC by switching transfer mode from RS to MS to OFDMA and SC-FDMA. OF-SC performed better than OF-OF when RS was located in 100m~300m but both modes hardly differed when BER of both modes were compared from 350m to 500m in the center.

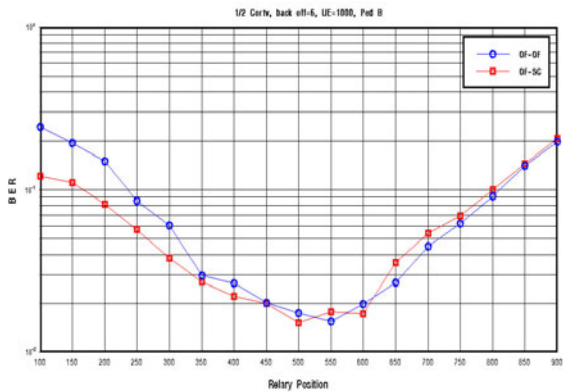


Fig. 4. BER in MS location of 1000m and transfer mode between RS and MS

6 Conclusion

This paper proposed a combination of two transfer modes to fill up the performance gap between OFDMA and SC-FDMA and to improve reception performance of LTE system's downlink transfer mode. The paper also proposed setting up RS in between BS and MS to enhance system's performance and coverage, and ran a simulation of the proposed idea. A simulation to appropriately select one out of OFDMA and SC-FDMA based on RS set with BS was carried out by setting the location between BS and MS at 500m and 1000m, respectively, and setting RS in between BS and MS. Simulation revealed that OFDMA is a better option in BS and SC-FDMA in RS when RS was more closely located to BS. This was just the opposite as in the longer distance between BS and RS where SC-FDMA performed better in BS and OFDMA in RS. In the center between BS and MS, an improvement in the system's reception performance was foreseeable by selecting transfer mode befitting particular situation.

References

1. Dahlman, E., Parkvall, S., Skold, J., Beming, P.: 3G Evolution: HSPA and LTE for Mobile Broadband, 2nd edn. Academic Press, London (2008)
2. 3GPP TSG RAN WG1, 3GPP TR 25.892 v6.0.0; Feasibility Study for Orthogonal Frequency Division Multiplexing(OFDM) for UTRAN Enhancement(Rel-6) (June 2004)
3. Zhang, J., Huang, C., Liu, G., Zhang, P.: Comparison of the Link Level Performance between OFDMA and SC-FDMA. IEEE CNF (October 25-25, 2006)
4. Holma, H.: LTE FOR UMTS - OFDMA And SC-FDMA Based Radio Access. John Wiley & Sons, Ltd, Chichester (2009)
5. IEEE 802.16MMR-06/005, 802.16 Mobile Multihop Relay Tutorial (March 2006)

Implementation of Linux Server System Monitoring and Control Solution for Administrator

Sung-Won Jang, Byeong-Ho Park, Sang-Ju Park, and Hyeon-Kyeong Seong

School of Computer, Information, and Communication Engineering, Sangji University,
Woosan 660, Wonju, Kangwon 220-702, Korea
{bestpolarbear, eden200, yhhan, hkseong}@sangji.ac.kr

Abstract. Linux server offers various kinds of service including web, FTP, and SSH. The users of these kinds of service are trying to hack by making use of it. That's why some countermeasures are required for the security of the server. In this thesis, each type of service log of multiple Linux server was analyzed, and a solution was developed to monitor and control the multiple Linux server system not based on Linux but based on Windows.

Keywords: Linux server, Multi-Linux, Linux based on Windows.

1 Introduction

Users can have access to web, FTP, SSH and other various kinds of services by taking advantage of Linux server but these services may be prone to hacking risks since there may be mass users who would rather abuse their adverse effects over the right functions. Such calls for server security one of which is analyzing log of each service to determine appropriate counteractions [1-3]. If, however, this requires an administrator to manage multiple Linux servers, not just one, and to analyze log for each server, it would exhaust a long time since it means analysis of service log in text-based random order and appropriate counteractions [4]. Such an inconvenience naturally raised the need for a system control enabling analysis of logs of multiple Linux servers with ease and visual demonstration of subsequent output to the administrator. Administration via Windows server, which is a frequent scene these days, allows administration in Windows-based server without the need for integrated control with multiple Linux servers [5].

2 Multiple Linux System Structure

This solution takes a client/server structure since multiple Linux server needs to be controlled under Windows and client module as shown in Table 1 below needs to be established under Linux development environment and server module under Windows development environment. Visual Studio 2008 was used as the tool to develop server module for the purpose of providing GUI to users via MFC. Oracle10g is used for database management program and ODBC is connected in between server program and database to enable access. Pro*C/C++ is used as the tool to develop client

module, which ensures access to database without using unixODBC and delivery of client module with C language. In addition, TCP data communication is employed for data communication with client and server, which is composed with GNU Socket library and Windows socket library.

Table 1. Statistics for CM and taken out moves

	server	client
Operating System	Windows XP SP3	CentOS 5.2
Development Tool	Microsoft Visual Studio 2008 v9.0.30729.1 SP	Pro*C/C++ Release 10.2.0.1.0
DBMS	Oracle 10g	Oracle 10g

2.1 Total Structure of Client/Server Module

The entire structure is as shown in Fig. 1. Server-wise, Linux server mounted on server administration program is controlled via 1:1 or 1:N shell command, which is then printed in the form of graph or table to users via log data stored and sampled from database on mounted server list. Client-wise, each Linux server is mounted on server module via client module, which is then put under control by shell command.

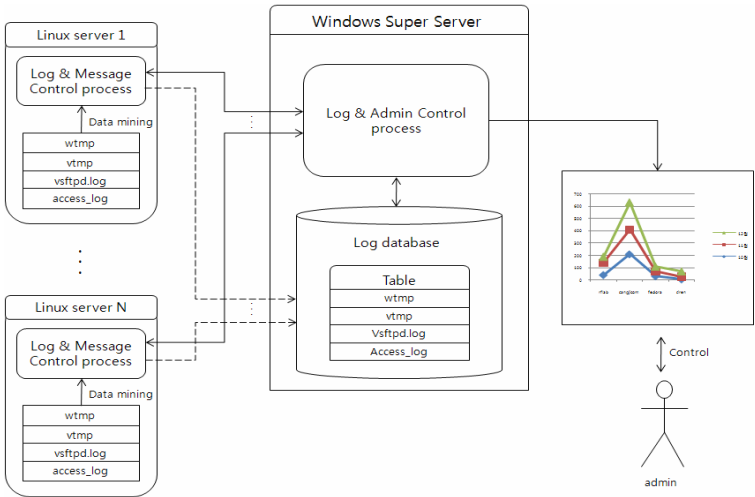


Fig. 1. Client / Server module of the entire configuration

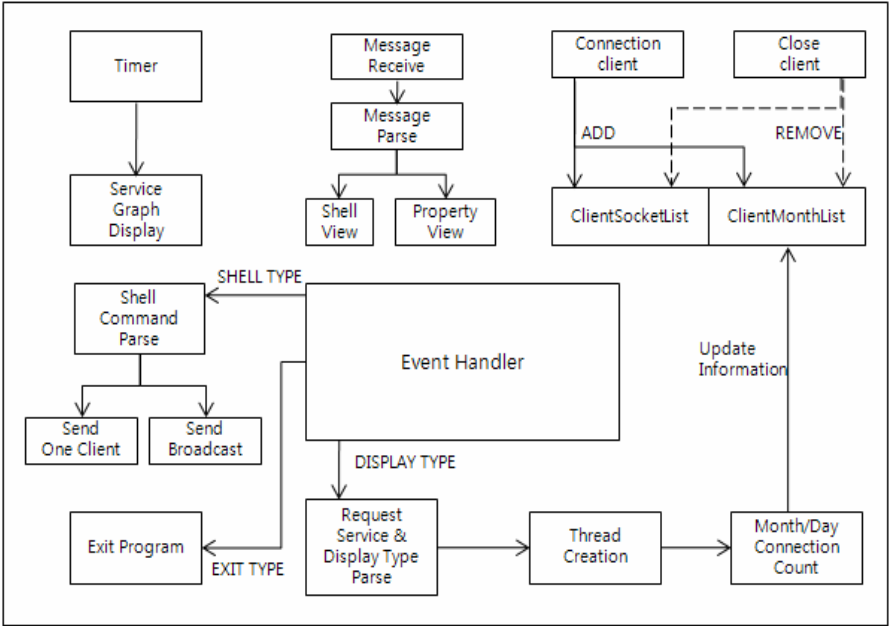


Fig. 2. Server module configuration

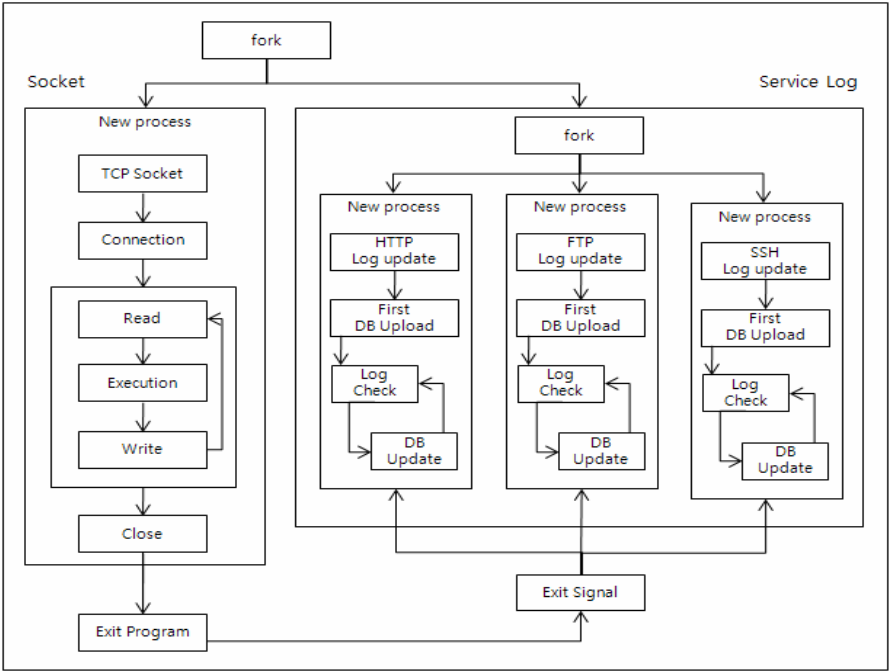


Fig. 3. The client module configuration

2.2 Server Module Structure

Server module shows access results by service via graph or table on a real-time basis via timer since it needs to be controlled 24 by 7. Server module is always on a standby for mounting client module with TCP socket. Once mounted, the client's information is added to the monthly list computing socket list and access data, with which the client can be controlled. In addition, it embarks on multi-thread upon display request from user to fend off delay caused by a single operation. ‘

2.3 Client Module Structure

Client module, which is compatible in Linux environment, should be able to make transmission to database by analyzing log of HTTP, FTP and SSH service, and to mount on the server for system control. This is why a total of five processes are created for operation. Log analysis database update function for each Linux service facilitates log analysis of the current service when running first module, transmits analysis data to server database, detects length of log per second and analyzes log with a newly revised length and eventually updates it to the server database.

3 Monitoring and Control Solution Implementation

Fig. 4 is the implemented server application. RFLab Linux server is currently mounted and shows total traffic for HTTP service in November. It also demonstrates that transmission of shell command to RFLab server via shell commander enables control.

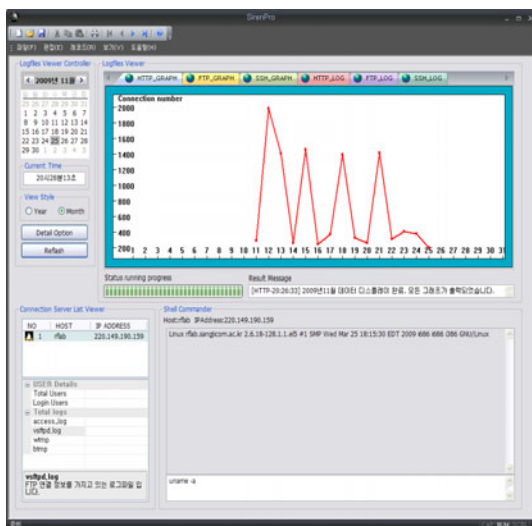


Fig. 4. November RFLab server HTTP connection information screen

Furthermore, multiple Linux server access is viable under this application as indicated in the following Fig. 5, which is the graph on FTP access to Sangjicom server and RFLab server in 2009. The server system's information is printed out in Property window through 1:1 shell access via Sangjicom server.

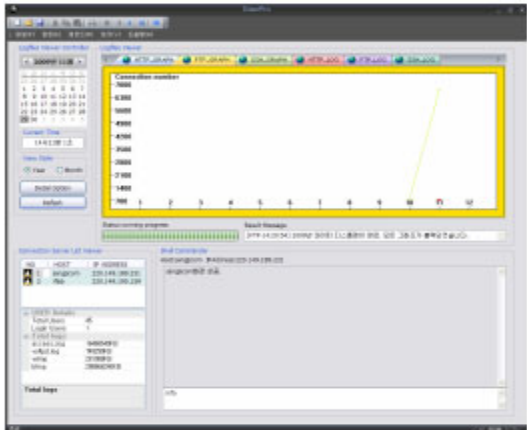


Fig. 5. RFLab Sangjicom servers and server FTP connection information screen 2009

Client application is in text-based composition and mounted on server on initial implementation. It gets updated after automatically logging into the database. Log is checked and transmitted every second for update and the current status is printed in text and displayed in the screen.

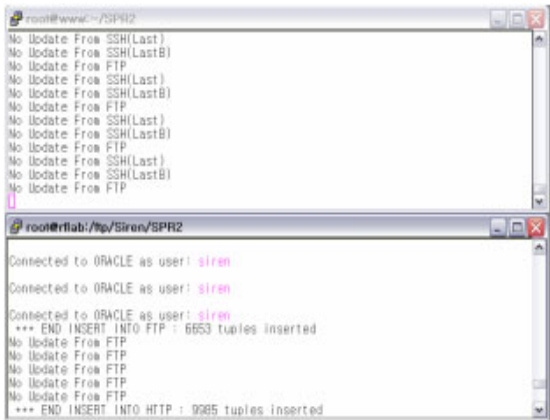


Fig. 6. Sangjicom RFLab server and client applications running on the server screen

4 Conclusion

The administrator displays traffic accessed to HTTP, FTP and SSH service provided in multiple Linux server via implemented application in graph and table. The administrator can take advantage of shell command to control security policy and system in each Linux server. Such being, it guarantees higher efficiency than others that analyze text-based log file and control system. Moreover, broadcast message transmission enables multi-processing. Such functions, however, cannot be offered in Linux server since they are delivered as Windows-based application. They will be independently executed in the platform if an application is delivered to enable use of web programming language like JSP to be displayed on the web.

References

1. Kim, T.-Y.: CentOS Linux Construction & Administration, Super User Korea
2. Spanosa, S., Melionesb, A., Stassinopoulousa, G.: The internals of advanced interrupt handling techniques: Performance optimization of an embedded Linux network interface. *Computer Communications* 31(14), 3460–3468 (2008)
3. Salah, K., Kahtani, A.: Performance evaluation comparison of Snort NIDS under Linux and Windows Server. *Journal of Network and Computer Applications* 33(1), 6–15 (2010)
4. Athanas, M., Ogg, M.: An evaluation of PCs for high energy physics under Windows NT and Linux. *Computer Physics Communications* 110(1-3), 225–229 (1998)
5. pro*C/C++ Precompiler Programmer's Guide, <http://www.otn.oracle.com>

A Smart Personal Activity Monitoring System Based on Wireless Device Management Methods

Jugeon Pak and Keehyun Park

Department of Computer Engineering, Keimyung University,
Daegu, South Korea
{corea,khp}@kmu.ac.kr

Abstract. Lack of physical activity can cause obesity, diabetes, hypertension and other cardiovascular diseases. To prevent these diseases, experts recommend regular physical exercise. In this paper, a Smart Personal Activity Monitoring System (SPAMS) is proposed in order to enhance the effectiveness of physical activities. The proposed system consists of a Physical Activity Measurement Device (PAMD), a Physical Activity Computation Device (PACD) and a Device Management Server (DMS). The proposed system offers personalized exercise programs based on individual body characteristics, and a remote device management service operated by system administrators, using a wireless device management method, specifically Open Mobile Alliance (OMA) Device Management (DM), as it is the international de facto standard for mobile device management. In this paper, we present a practical method for using OMA DM to personalize and manage the SPAMS and also present the implementation results.

Keywords: personal health device, physical activity monitor, OMA DM.

1 Introduction

Obesity, diabetes, hypertension and other cardiovascular disease rates have been increasing over the years and the main reason seems to be a lack of exercise. To prevent these diseases, experts recommend regular physical exercise[1, 2]. Reflecting this trend, various activity measurement devices are being released all the time[3].

Recently, a significant volume of research on activity monitoring systems has been carried out. Some studies presented methods for measuring physical activities[4, 5], while others proposed methods for increasing the accuracy of calculations [6-8]. However, common activity monitoring systems only offer a uniform exercise program delivered without regard to the physical characteristics of the user. Consequently, the benefits of exercise are not being optimized. Moreover, the systems provide no internal management functions. To overcome these issues an efficient device management method for activity monitoring systems is essential.

In this paper, we propose a Smart Personal Activity Monitoring System (SPAMS) which uses a wireless device management method (specifically OMA DM). We propose a practical method of using OMA DM to personalize and manage the SPAMS. The differences between SPAMS and other systems are that our system uses

biometric information to provide a personalized exercise program for users and the system is managed remotely by administrators without users' intervention, over wired or wireless networks.

The rest of the paper is organized as follows. Section 2 presents related works on device management methods and describes an overview of the OMA DM protocol. Section 3 illustrates the architecture of the SPAMS and section 4 describes our proposed method. In section 5, we present the implementation results of the SPAMS. Finally, we conclude in section 6.

2 Related Works

To date, several device management methods have been proposed. For network and desktop system management, the Internet Engineering Task Force (IETF) has released the Simple Network Management Protocol (SNMP) [9]. The Distributed Management Task Force (DMTF) defines the Web Based Enterprise Management standard (WBEM) [10]. WBEM defines a Common Information Model (CIM) as a data model. OMA has developed a DM protocol for mobile devices by extending SyncML Data Synchronization (DS). Currently, the OMA DM protocol is the international de facto standard for mobile device management [11, 12]. In recent years various studies on OMA DM have been performed. Jugeon et al. [13,14] presented a device management system for WIPI-based mobile devices and Jieun et al. [15] proposed a device management system for WiBro mobile devices. Some software management methods were proposed for mobile devices using OMA DM. Hongtaek et al. [16] presented a software release management system based on OMA DM and Redhat Package Manager (RPM) technology. Joonmyung et al. [17, 18] proposed remote software fault management and debugging systems. In these studies, methods for debugging and correcting software faults using the collected information received from DM Clients were introduced. In addition to the above studies, research has also been carried out in various fields and aspects of OMA DM such as network management [19] and vehicle management[20].

However, only a limited number of studies on managing personal health devices (PHD) including activity monitors have been carried out. Since an activity monitor has limited computing resources and is closely correlated to a user's health, an efficient device management method for activity monitoring systems is essential.

3 System Architecture

The SPAMS proposed in this paper is comprised of three parts: the physical activity measurement device (PAMD); the physical activity computation device (PACD) and the DM server (DMS). Figure 1 shows the detailed architecture of the system.

The PAMD is a device for measuring a user's physical activities in order to calculate the exercise intensity and to estimate caloric expenditure. Since the PAMD is usually used during exercise, its portability is one of the most important requirements. Accordingly, it should be small and light weight, and therefore it is limited by constraints such as low power, limited memory and low bandwidth. To address these

constraints, we propose a method that measures a user's raw motion data on the PAMD with calculations and analysis performed on PACD. The PAMD measures raw motion data using a 3-axis accelerator. It also measures steps taken and exercise time. It usually accumulates measured data until communication with the PACD is available. When communication between the two devices is possible, the PAMD transmits the accumulated raw motion data to the PACD over Bluetooth.

The PACD, which is located in mobile devices such as PDAs or smart phones, plays two important roles in the SPAMS: it functions as an activity calculator and as a device manager. The PACD calculates basal metabolic rate (BMR) and exercise intensity, and estimates caloric expenditure using pre-determined equations. Furthermore, it also manages itself through the DMS's management commands. To achieve this, the PACD exchanges several OMA DM messages with the DMS.

The DMS plays a leading role in managing the PAMD and the PACD. Medical or fitness staff analyzes the user information and calculated activity data received from the PACD and determine a personalized exercise program including details such as recommended exercise intensity and caloric expenditure. System administrators analyze the system configurations or MOs(Management Object) received from the PACD and determine management operations. If a personalized exercise program is updated or the system configuration should be managed, the DMS sends an OMA DM message containing OMA DM commands and the value of the specific MO.

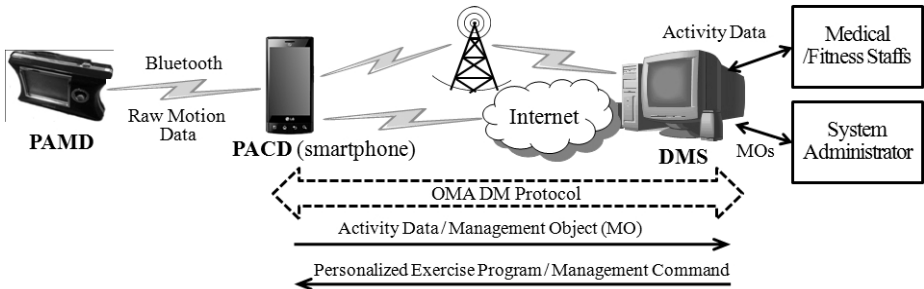


Fig. 1. The architecture of the SPAMS

4 Proposed Method

According to the OMA DM protocol, all data to be managed should be defined as MOs, and the MOs should be structured in a hierarchy tree form called a DM Tree. The MOs and DM Tree used have been discussed and defined in a previous study[21].

- *User_Info*: This node has five child nodes and these nodes are entered by the user manually.
- *Measured_Data*: This node has two child nodes and these nodes are updated whenever data is received from the PAMD. *Measured_Data* and its child nodes are not transmitted to the DMS but used in calculations.

- *Calculated_Data*: This node has four child nodes, and they represent the data which is calculated from the measured data. The calculated data is transmitted to the DMS at regular intervals (specified in the node *./SPAMS/System_Conf/Report_Interval*) or when the DMS requests it.
- *Aimed_Data*: This node has two child nodes and the values of the nodes are not transmitted to the DMS but determined by medical or fitness staff.
- *System_Conf*: This node has three child nodes and the values of the nodes are pre-set.

We have designed the following management operations: activity data transmission; system configuration management; software management; and error reporting. For these operations the PACD and the DMS exchange several OMA DM messages during a management session. A management session starts at regular intervals (the value of the node *./SPAMS/System_Conf/Report_Interval*), whenever an error occurs (whenever the node *./SPAMS/System_Conf/ERR* is generated) or the DMS alerts the PACD to initiate a DM Session with Pkg. #0. When a management session starts, the child nodes belonging to the node *./SPAMS/Calculated_Data* are transmitted to the DMS automatically, while the child nodes belonging to the nodes *./SPAMS/User_Info*, *./SPAMS/Aimed_Data* and *./SPAMS/System_Conf* are transmitted only when the DMS requests them. On the other hand, if an error occurs, then the child nodes belonging to the node *./SPAMS/System_Conf/ERR* are transmitted immediately. Figure 2 shows the management operations for the SPAMS.

Figure 2(a) shows the activity data transmission phase. When the DMS alerts the PACD or a system clock reaches the reporting time, the PACD initiates a management session sending the calculated activity data. The DMS analyzes the data and determines the recommended exercise intensity and caloric expenditure for the user. Figure 2(b) shows the system configuration management phase. When the DMS wants to manage specific configurations, it alerts the PACD. Before managing the configurations, the DMS checks the assigned values of the target configurations first, and then sends management commands to add, delete or replace the target configurations. Figure 2(c) shows the software management phase. When a new, improved version of software is released, the DMS updates the software. The DMS requests the current version of the software which is installed on the PACD, and then informs it of the download URL for the new version of software. The PACD downloads the software from the URL. Once the PACD finishes downloading, the DMS sends an execution command to the PACD to install the downloaded software. The PACD installs the downloaded software, and the management session is finished. Figure 2(d) shows the error report phase. When an error occurs in the PACD or PAMD, the *./SPAMS/System_Conf/Err/Err_Code* and */Err_Contents* nodes are generated dynamically. The PACD is programmed to initiate a management session sending the nodes after they are generated. The DMS analyzes the error code and contents and then considers a proper recovery method such as resetting the system configurations or reinstalling software.

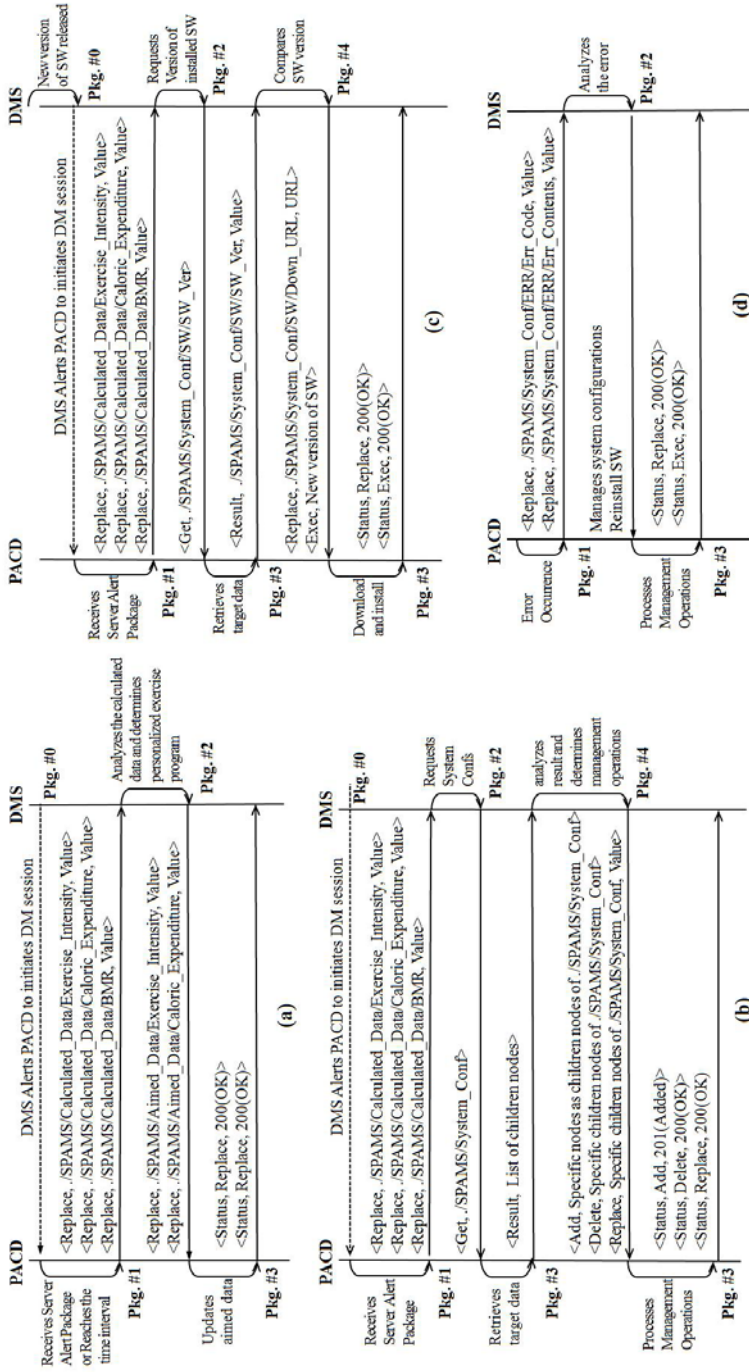


Fig. 2. Management operations: (a) activity data transmission, (b) system configuration management, (c) software management, (d) error report

5 Implementation Results

We developed the PACD to measure a user's raw motion data and send it to the PACD over Bluetooth, not to perform complicated calculations as this is the PACD's role. Figure 3 shows the prototype of the PAMD.

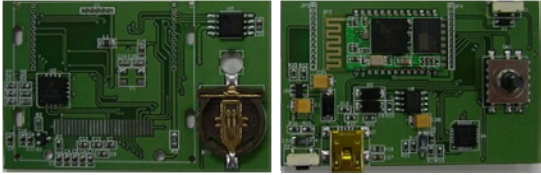


Fig. 3. The prototype of the PAMD

The PACD calculates exercise intensity and caloric expenditure, and exchanges several OMA DM messages with the DMS. Figure 4 shows the architecture and the screenshot of the implemented PACD. The PACD was implemented on a smart phone, LG-LU3000 called Optimus Mach. It operates on 1 GHz TI OMAP 3630 processor and runs on Android 2.2. In addition, an embedded SQL database, SQLite[22] was used for the database. The size of the DS server program is 104 KB.

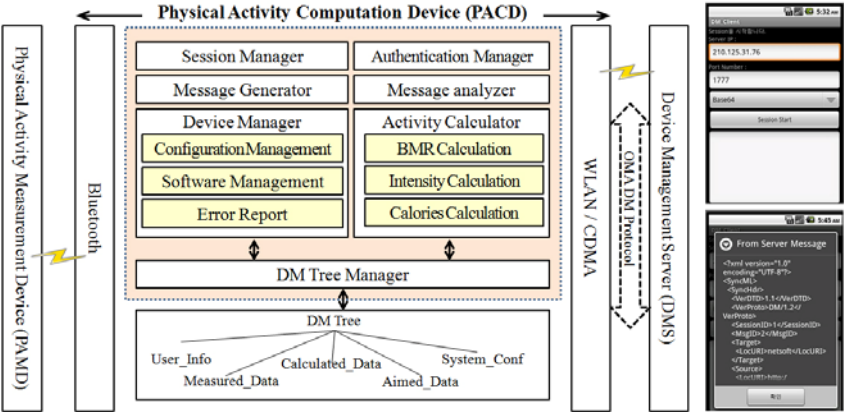


Fig. 4. The Architecture and the screen shot of the PACD

- *DM Tree Manager*: extracts/stores the activity data and MOs
- *Device Manager*: performs management operations following the DMS's management commands
- *Activity Calculator*: calculates BMR and exercise intensity, and estimates caloric expenditure
- *Message Generator/Analyzer*: generates/Analyzes OMA DM messages
- *Session Manager*: generates the header part of the OMA DM message and appends it to the front of the body part.
- *Authentication Manager*: authenticates the credential information.

We developed the DMS to manage the PACD. The architecture of the DMS is very similar to the PACD. Figure 5 shows the screenshot of the implemented DMS. As shown in Figure 5, the DMS shows the status of the management session and the exchanged message with the PACD. The DMS was implemented in C# on a desktop computer with an Intel Core 2 Duo processor (at 2.66 GHz) and 1024 MB of RAM. In addition, MSSQL was used for the database. The size of the DMS is 302 KB.



Fig. 5. The screen shot of the DMS

6 Conclusion

In this paper, we proposed a SPAMS based on a wireless device management method. We proposed a practical method of using OMA DM to personalize and manage the SPAMS. The SPAMS consists of the PAMD, PACD, and DMS. The main feature of the proposed system is to self-manage by following the DMS's management commands. To achieve this, we defined the MOs and designed the following management operations: activity data transmission, configuration management, software management and error reporting. We discussed how to design the management operations using OMA DM commands.

For future work, we plan to consider security issues which may arise during the exchange of OMA DM messages.

Acknowledgments. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (No. 2010-0016454).

References

1. Dena, M.B., Crystal, S.S., Vandana, S., Allison, L.G., Nancy, L., Robyn, L., Christopher, D.S., Ingram, O., John, R.S.: Using Pedometers to Increase Physical Activity and Improve Health. *JAMA* 298(19), 2296–2304 (2007)
2. Butler, R.N., Davis, R., Lewis, C.B., Nelson, M.E., Strauss, E.: Physical fitness: benefits of exercise for the older patient. *Geriatrics* 53(10), 46–62 (1998)

3. Chao, C., Steve, A., Abdelsalam, H.: A Brief Survey of Physical Activity Monitoring Devices. Technical report, University of Florida (2008)
4. Mihee, L., Jungchae, K., Kwnagsoo, K., Inho, L., Sunha, J., Sunkook, Y.: Physical Activity Recognition Using a Single Tri-Axis Accelerometer. In: World Congress on Engineering and Computer Science 2009, vol. 1 (2009)
5. Bouten, C.V., Westerterp, K.R., Verduin, M., Janssen, J.D.: Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine & Science in Sports & Exercise* 26(12), 1516–1523 (1994)
6. Zhi, L.: Exercises Intensity Estimation based on the Physical Activities Healthcare System. In: International Conference on Communications and Mobile Computing 2009, vol. 3. pp. 132–136 (2009)
7. Heather, H., Denise, F., Richard, B., Rita, F., Charles, S.: Accuracy of a custom-designed activity monitor: Implications for diabetic foot ulcer healing. *Journal of Rehabilitation Research and Development* 39(3), 395–400 (2002)
8. Jungeun, L., Ohoon, C., Hongseok, N., Dookwon, B.: A Context-Aware Fitness Guide System for Exercise Optimization in U-Health. *IEEE Transactions on Information Technology in Biomedicine* 13(3), 370–379 (2009)
9. Case, J., Fedor, M., Schoffstall, M., Davin, J.: A Simple Network Management Protocol (SNMP). RFC 1157, IETF Network Working Group (1990)
10. Distributed Management Task Force, Web-Based Enterprise Management, WBEM (2008), <http://www.dmtf.org/standard/wbem>
11. Open Mobile Alliance (OMA), <http://www.openmobilealliance.org>
12. Uwe, H., Riku, M., Apratim, P., Peter, T.: SyncML Synchronizing and Managing Your Mobile Data. Prentice Hall PTR, New Jersey (2003)
13. Jugeon, P., Keehyun, P., Daejin, J., Myungsook, J., Jongjung, W.: Design of DM Agent based on the WIPI. *Journal of Society of Mobile Technology* 4(1), 61–67 (2007)
14. Jugeon, P., Keehyun, P., Daejin, J., Myungsook, J.: Design and Implementation of Wireless Device Management Agent based on OMA DM. *Journal of Korea Institute of Information Scientists and Engineers* 14(4), 363–368 (2008)
15. Jieun, L., Sunghak, S., Byungduck, J.: WiBro Device Management System based on OMA DM Protocol. *KNOM Review* 10(2), 1–11 (2007)
16. Hongtaek, J., Keehyun, P., Daeuk, B.: Software Release Management System: ThinkSync DM-SoftMan for Wireless Device based on OMA DM. *Journal of Korea Information Processing Society* 13(5), 641–650 (2006)
17. Joonmyung, K., Hongtaek, J., Mijung, C., James, W.H., Jungu, K.: OMA DM-based Remote Software Fault Management for Mobile Devices. *International Journal of Network Management* 19(16), 491–511 (2009)
18. Joonmyung, K., Hongtaek, J., Mijung, C., James, W.H.: OMA DM-based Remote Software Debugging of Mobile Devices. In: Ata, S., Hong, C.S. (eds.) *APNOMS 2007*. LNCS, vol. 4773, pp. 51–61. Springer, Heidelberg (2007)
19. Mijung, C., James, W.H., Hongtaek, J.: XML-Based Network Management for IP Networks. *ETRI Journal* 25(6), 445–463 (2003)
20. Hyunki, R., Sungrae, C., Shiquan, P., Sungho, K.: The Design of Remote Vehicle Management System Based on OMA DM Protocol and AUTOSAR S/W Architecture. In: *IAdvanced Language Processing and Web Information Technology 2008*, pp. 393–397 (2008)
21. Jugeon, P., Keehyun, P.: Design of an OMA DM-Based Remote Management System for Personal Healthcare Data Devices. *Internet Technologies & Society* (2010)
22. SQLite, <http://www.sqlite.org>

A Study on Demodulation System Design of the VOR Receiver

Jeong IL Park and Hwa Se Park

Electronic-Communications Dept, Daelim University College. 526-7, Bisan-Dong,
DongAn-Gu, Anyang-Si, Gyeonggi-Do, Korea
{jipark,hwasepark}@daelim.ac.kr

Abstract. In this paper, we studied VOR receiver system designed for digital communication system. A VOR provides the user with a bearing to the station. Digital hardware was used to determine the phase relationship between the two 30Hz signal. The design in this paper acquires the phase of the 30Hz variable and reference signals from the composite audio output of a VOR receiver using DSP and FPGA methods implemented in software. VOR signal generator using a frequency 108MHz, -70dBm to receive input power to the system, the developed system was verified by operation.

Keywords: VOR, Bearing, Demodulation, DSP, FPGA.

1 Introduction

VOR(Very high frequency Omni directional Range)is system which provides azimuth information that aircraft to reach to the destination safely. VOR uses Omni Antenna from central system to transmit modulated signal. VOR transmitting station utilize microwave to send two types of radio wave and uses its phase difference and calculate bearing. VOR consists of CVOR (Conventional VOR) and DVOR(Doppler VOR). Difference between two methods is reference phase of CVOR is FM modulation, variable phase of it consists of AM modulation and reference phase of DVOR is AM modulation, variable phase of it consists of FM modulation. Nowadays uses DVOR instead of DVOR to reduce bearing error. Operation of DVOR is calculated by reference phase and variable phase those which are called carrier wave modulated by two 30Hz signal's phase difference and calculate bearing. Reference phase signal is given by 30Hz sinusoidal signal which is amplitude modulated carrier. This amplitude modulated signal is radiated omnidirectionally to central carrier wave antenna to horizontal plane. Radiation pattern is circle, make 30Hz independent phase signal of bearing to receiver of aircraft. Variable Phase signal is given by 9,960Hz frequency modulated sub carrier to carrier by amplitude modulating, amplitude modulation of this carrier is given by radiated upper-sideband and lower-sideband signal's space diversity from omnidirectional radiating carrier and sidebands antenna's ring. Upper-sideband and lower-sideband signal is be off by average 9,960Hz when correct phase is added to carrier and make 9,960Hz amplitude modulated composite signal. Sub carrier is frequency modulated by 30Hz rate. 30Hz signal from receiver of aircraft is extracted from 9,960Hz FM carrier wave.

VOR receiver which is installed in plane senses phase difference of two different signals and mark location of plane on measuring instrument. To judge reliability of received signal, those signal received plane put morse code to transmission frequency and check whether pilot has received information of bearing from VOR transmitting station[1].

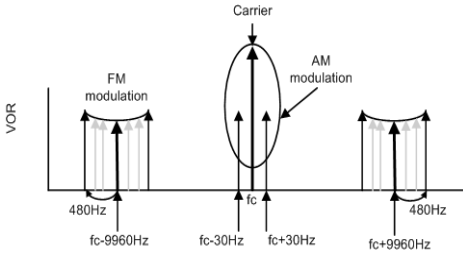


Fig. 1. Spectrum of VOR Modulation Signal

In this paper, designed receiving system that uses receiving VOR signal which is IF signal processing system that tells bearing of aircraft to DSP and FPGA.

To achieve this, we designed digital receiver system consist of DSP TMS320F2812(TI) and FPGA EP3C40F324 (ALTERA), and to do verification efficiency of suggested digital method, using VOR/ILS signal generator to do simulation consist of VOR frequency 108MHz, and -70dBm if reception power.

2 VOR Receiver

A block diagram of a standard VOR receiver is shown in Fig 2.

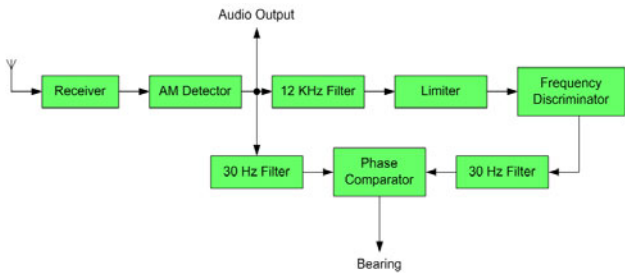


Fig. 2. Spectrum of VOR Modulation Signal

A standard analog VOR receiver includes a RF front end and AM detector. The output of AM detector is referred to as the composite audio signal. The signal is then split into two parts using 30Hz(variable signal) and 12KHz (reference signal) filter. Then detect AM signal and through the 12KHz filter, extract FM frequency band which is located 9,960Hz far from carrier. Finally, for demodulation by passing

amplitude limiter extract change of carrier frequency to output voltage. The reference portion is then frequency detected and bandpass filtered. The resulting signal is then fed to a phase comparator, along with the variable the variable signal. The output of phase comparator is the bearing from the VOR transmitter.

Sending signal from VOR transmitting station can be mathematically shown as below.

$$s(t) = m(t) \cos(2\pi f_c t) \quad (1)$$

In this equation, $s(t)$ is transmission signal, $m(t)$ is modulation signal and f_c is carrier frequency. Modulation signal $m(t)$ can be shown as below equation.

$$m(t) = 1 + 0.3 \cos(2\pi f_m t + \theta_1) + 0.3 \cos[2\pi f_{sc} t + 16 \cos(2\pi f_m t + \theta_2)] \quad (2)$$

In this equation f_m is 30Hz, f_c is sub-carrier of 9960Hz and $\phi = \theta_1 - \theta_2$ is bearing.

3 Digital Demodulation System Design of VOR

Structure of demodulation system of VOR receiver we are going to develop can be shown in Fig 3 [2],[3].

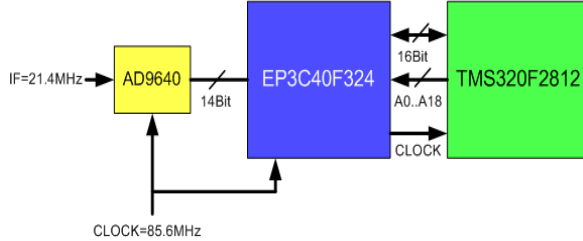


Fig. 3. Digital demodulation system architecture of VOR receiver

Analog system input to digital signal processing part is converted to digital signal by AD converter. This signal demodulated by FPGA. DSP find phase difference by reading result of AM and FM saved in FPGA memory. Implementation of hardware used in receiver system is consist of AD9640 (Analog Device : 14Bit, 150Msps, AD Converter), TMS320F2812 (TI,DSP) and EP3C40F324 (ALTERA,FPGA).

The method of operating system shown in Fig 3 received external signal go through RF part and do down conversion 21.4MHz IF signal. Sampling frequency of AD converter uses 85.6 MHz to under sample and save output of AD converter to DPRAM in FPGA. By using saved signal, FPGA going to consist PLL loop for frequency synchronization for coherent demodulation made of 21.4MHz signal. Then create Baseband signal from carrier signal which consist of demodulated carrier signal is demodulated as receive signal. Put this baseband signal to filter and decimator then

created AM and FM signal of 30 Hz save in DPRAM of FPGA. To get phase difference between tow signal, read AM and FM 30Hz signal saved in DPRAM of FPGA and calculate phase of two signals by using DSP. Fig 4 and Fig 5 show hardware block of internal system of FPGA of VOR system that we are going to develop.

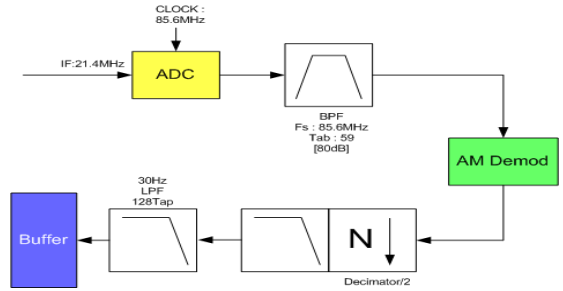


Fig. 4. Block diagram of the AM demodulation in the FPGA

Hardware block in Fig 4 shows inside block of FPGA for AM demodulation. Down convert RF signal of 108 MHz as IF signal of 21.4 MHz then save receive signal which came through by AD converter to DPRAM of FPGA. Saved signal will going to pass through BPF(Band Pass Filter), get rid of DC component and plus 4096 sample power value by received signal then do AM demodulation. The signal which is demodulated let go through Decimator, Half band Filter and 30Hz LPF and then save result of AM signal in DPRAM of FPGA.

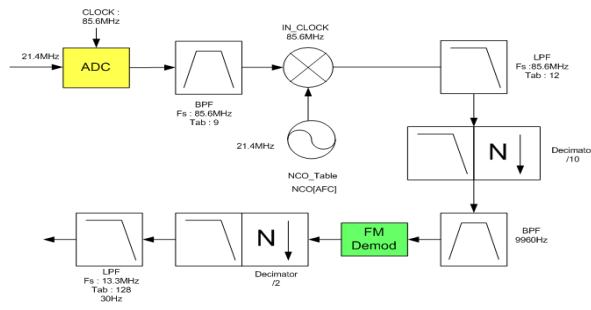


Fig. 5. Block diagram of the FM demodulation in the FPGA

Fig 5 shows FPGA internal block for FM demodulation. To get rid of DC component of the input signal through the AD converter, it goes through BPF. Let down conversed signal to go through NCO and LPF and current input signal down to baseband and get rid of sampling noise. Use decimator and HPF to down sampling input signal. Then to demodulate FM signal located at 9960 Hz let signal converted to baseband go through 8.5 KHz BPF then use DPLL technique to demodulate to FM signal. Demodulated 30Hz FM signal and demodulated 30Hz AM signal is saved through LPF and DPRAM of FPGA. Read saved results in DSP and get phase different of two signals[4],[5].

4 Experiment and Consideration

For experiment, by using VOR/ILS signal generator[6] to make better experimental environment so put frequency of designed module to 108 MHz, receiving power to -70dBm. Result that below figure shows demodulated signal of FPFA to uses Code Composer from TI to makes and shows results by blocks which happened by demodulated AM, FM.

Fig 6 shows modulation signal that is saved in DPRAM which modulation signal input in VOR signal generator go through AD converter.

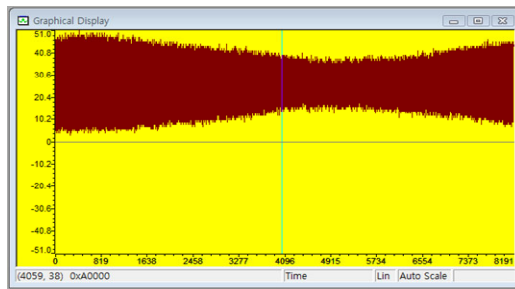


Fig. 6. VOR input signal after AD Converter

Fig 7 shows 30Hz AM input signal which has passed BPF in input signal status and go through down sampling and frequency spectrum of signal that has input.

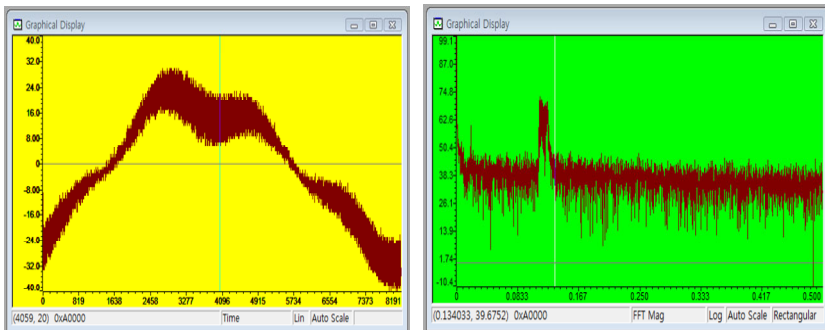


Fig. 7. 30Hz AM Spectrum and modulation signal

Fig 8 like in Fig 4 per 4096 sample of input signal add power value and demodulate AM. After that result signal goes through HBF then decimate and pass 30Hz LPF and shows result signal. Fig 8 shows first period of result of 30Hz. To get first period it has points of 720 samples.

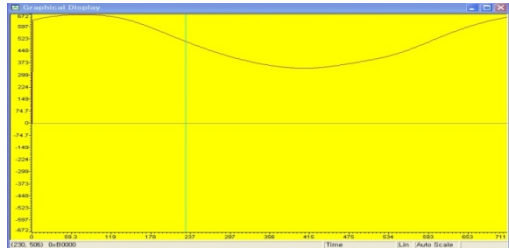


Fig. 8. 30Hz AM demodulation receive signal

Below Fig 9 shows the results to get 30Hz FM signal down sample 21.4MHz IF signal then to extract 9960Hz signal let the down sampled signal to go through BPF and we have demodulated signal's 30Hz FM signal and spectrum.

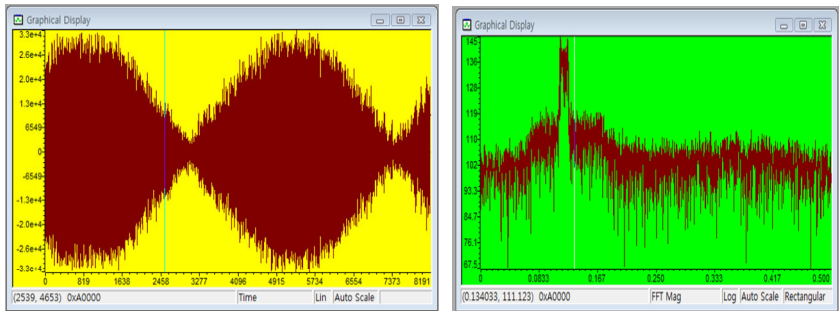


Fig. 9. 30Hz FM Spectrum and modulation signal

Fig 10 shows result of like in Fig 5 through designed FM demodulator signal has been demodulated. As appears by Fig 10, we designed system to have points of 720 samples to reduce error of phase difference as minimum.

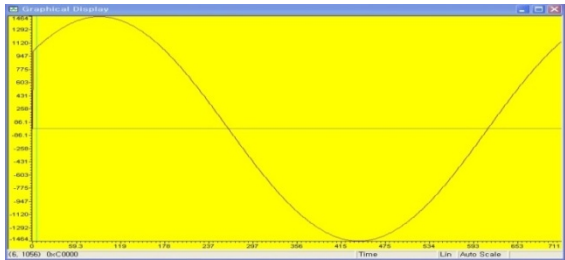


Fig. 10. 30Hz FM demodulation receive signal

Also in this paper by using Zero crossing algorithm to calculate phase difference between AM and FM demodulation to reduce error to minimum between phase difference between two signals by using DSP. In results, phase signal set in VOR/ILS Signal Generator has error about ± 0.5 degrees.

5 Result

Nowadays most of commercializing VOR receiver system uses analog system structure. This VOR receiver is now converting to digital system □ digital system. On this thesis we have described not analog method about VOR system but lay out methods about digital methods and then made hardware system and verified about designed system. Input RF signal in receiver use Mixer to down converse to IF signal, instead of current analog design method which is by using down conversed IF signal to demodulate AM and FM signal suggest that by using ADC, FPGA and DSP to demodulate VOR receiving signal which is digital VOR system. Also, manufacture hardware of designed system and utilize VOR/ILS only signal generator to investigate efficiency of receiving signal. To get phase difference of two signals we made phase error to minimum by using zero crossing method. Results of current experiment has input 108 MHz of frequency and -70dBm and measured it. However, from now on we are going to develop signal sensitivity of demodulated system to receive signal even though higher than -110dBm.

References

1. Yoon, J.: Aircraft Information and Communication Engineering, Kyohaksa
2. Lillington, J.: Wideband Spectrum Analysis using Advanced DSP Technique
3. Rabiner, L.R., Gold, B.: Theory and Application of Digital Signal Processing. Prentice Hall, Englewood Cliffs (1975)
4. Abdulla, M., Svoboda, J.V., Rodrigues, L.: AVIONICS MADE SIMPLE, M. Abdulla (2005)
5. prisaznuk P.J.: Integrated module avionics. In: Proceedings of IEEE National Aerospace and Electronics Conference, vol. 1, pp. 39–45 (May 1992)
6. <http://www.rohde-schwarz.us>

China's Electronic Information Policy

Won-bong Lee

Dept. of Chinese Studies, Kyung Hee Cyber University
1 Hoegi-Dong, Dongdaemun-Gu, Seoul, 130-701, South Korea
wblee@khcu.ac.kr

Abstract. Through its reform and opening policy since early 1980s, China has achieved rapid economic growths. Ever since the reform and opening policy, the electronic information technology has been advanced and national IT infrastructure was constructed. Since 1995, the electronic information industry in China has grown at the fastest pace and secured its position as a major industry. Since 2006, China became the largest IT devices producer in the world. In particular, the electronic information technologies were applied vigorously in the fields of finance, automobile, medical treatment and military. In China, the development of electronic information technology and industry will play a key role in the enhancement of its national power.

Keywords: China. Electronic Information, Information Society, Information policy, high-tech industry, IT.

1 Introduction

Contemporary society has been dubbed as Information Society or Era of Information Revolution. Cutting-edge info-telecommunication networks and computing technologies have emerged as new elements of the national power, and they also affected the advance of the scientific technology and means of war. In the early 1990s, the Central Intelligence Agency (CIA) of the U.S. added the market share of the semiconductor production into the index of the national power. Furthermore, PC penetration ratio, the number of high-speed Internet users, and the number of the server computers connected to Internet also emerge as new factors indicating national power (8).

Its economic development has heavily depended on the investment from foreign firms and overseas trade. Trade liberalization and the development of the IT technologies helped the area of business activities go beyond the borders (10). With its economy performance, Chinese government embarked on its strategic development of the high-tech industry. The high-tech industry includes high-level scientific technology industry and electronic information industry. In this context, this study first provides an analysis on Chinese government's policy direction on the electronic information industry. It also illuminates the features and problems of the Chinese electronic information industry by looking at the current situation. Then, based on comprehensive analysis of Chinese government's policy on the electronic information industry, the paper will explore features and problems, future outlook on the industry.

2 China's Economic Policy and Industrial Policy

2.1 China's Economic Policy and Electronic Information Policy

From the year 1978, China set up new policies for the economic development. Deng Xiao Ping took reform policies at home and opening policies abroad. From 1980s onward, Chinese economy has enjoyed a remarkable growth. It eagerly introduced the investments from advanced countries through the opening-door policy and also accepted the scientific technologies of the advanced countries (2).

Ever since the reform and opening policy, the electronic information technology has been advanced and national IT infrastructure was constructed. Multiple applications of the electronic information technologies have been tried in various fields. In particular, the electronic information technologies were applied vigorously in the fields of finance, automobile, medical treatment and military. From the outset of the 21st century, Chinese government has set up and made public its economic plans. They include the 10th five-year plan (2001~2005), the 11th five-year plan (2006~2010) and the 12th five-year plan (2011~2015). Through these five-year plans, efforts have been made to develop cutting-edge technology industries and electronic information technologies in a steady manner.

The basic course of the 12th five-year plan that starts from 2011 is to expand the domestic consumptions. It represents a diversion from its erstwhile export-oriented economy. In addition, it aims to nurture new strategic industries such as biotechnology, alternative energy, development of new materials and next-generation IT technologies (12).

The 12th five-year plan proposed six directions of the industrial restructuring. One of them is to expand the application of the IT technologies. China also adopted six fields of the new key industries, and the IT industry is one of them. The 12th five-year plan nominated the electronic information manufacturing industry as its key industrial development strategy (5).

2.2 Changes in the Industrial Policy and IT Policy

Chinese industrial structure has transformed into that of the high-tech industry. Since 1995, the high-tech industry has posted rapid average annual growth of 25.2%. The high-tech industry took the lead in the sales of new products and oversea exports. This can be ascribed to the improvement of the research and development capabilities (5).

Since the 11th five-year plan was launched in 2006, the basic direction of the Chinese industrial policy has become the advancement of the industry structure by strengthening technological capabilities. Manufacture of PC and electronic communication facilities is one of the high-tech industries.

In February 2006, Chinese government announced "National Mid-and Long Term Scientific Technology Development Plan." The plan aims to secure world-class scientific technologies and IT core technologies. In 2009, the government came up with "2050 scientific technology Roadmap. This roadmap introduced a plan to achieve China's full-scale entry into the e-society until 2020. "10 Industrial Structure Promotion Plan" was announced in 2009. The electronic information industry was

selected as one of the 10 industries In February 2010, "Strategic new industries" was announced (2). Next-generation information industry was selected as one of the 7 strategic new industries.

Table 1. China's major industry policy (5)

major policy	year
Information industry scientific technique development 11.5 Plan(2005)	Enhancing of Industry through reinforcement of technical skills
National Long-term scientific and technical development 11.5 Plan: 2006-2020(2006)	Secure world-class scientific technique Secure world-class Information industry scientific technique
2050's Chinese scientific and technical development road map (2009)	Suggest road map of scientific and technical development 2020 year enter into e-society
Core High major project	Core electronic components, Emphasis on promotion of CPU
10 kinds industry structure promotion plan(2009)	electronic information, car, steel, steel, petro chemistry, light industry, textile manufacturing, construction machinery, nonferrous metals, distribution industry
Strategic new industry(2010)	①energy conservation & protection of the environment next generation information industry Bio-related Industries new energy new energy car ⑥high equipment industry new material
China's next generation of information industry	NGN(Next Generation networking), triple network fusion(telegraph network, television network, Internet), High-performance semi-conductors & Advanced software, New flat panel displays, The Internet of Things

3 Electronic Information Industry of China

Since 1995, the electronic information industry in China has grown at the fastest pace and secured its position as a major industry. In particular, IT sectors including PC, cellular phone and semiconductor achieved a remarkable advance, and PC and

cellular phone emerged as major export items (3). The development of the Chinese electronic information industry was attained through attraction of investments from foreign firms (1).

Table 2. Import and export progress of China's electronic information Industry (5)

Indicator unit	2003	2004	2005	2006	2007	2008
Import (hundred million U.S. dollars)	1,321	1,809	2,205	2,877	3,452	3,637
Export (hundred million U.S. dollars)	1,421	2,075	2,682	3,640	4,595	5,218

From 2000 and onward, the technological capability of the Chinese electronic information industry(especially, telecommunication, high-performance PC and Digital TV) has improved a lot. Cellular phone, PC, Color TV, Display attained the world number one market shares (5). Since 2006, China became the largest IT devices producer in the world and, in 2008, its IT production accounted for 21% of the world market(11). Foreign-invested firms played a leading role in the electronic information industry of China. China became a major production site of the multinational companies due to following reasons. First of all, it is easier to secure low-cost labor force. Chinese government's active efforts to attract foreign investments are also one of the reasons (10).

The amount of production and sales of the electronic information industry in China has steadily increased. At present, the electronic information industry is occupying the core position in the Chinese economy. China has become world's number one in the field of export of high-tech and Information & Communication Technology (ICT) products. The level of the information(cellular phone, TV, PC owners) is also advancing at an astonishing pace (7).

The high-tech industry in China has been enjoying a rapid growth thanks to following factors: First, active manpower fostering at the national level; second, government's policies to actively promote industry, one of which is technology industry; and third, the massive domestic market in China (1).

Since 2010, the electronic information industry in China has transformed its growth pattern from rapid one to stable one. It has kept growing to be one of the strategic new industries and widened its inroad into the domestic market. New display industries (LCD, PDP, OLED) reached their peaks (5).

The electronic information industry has a big technological ripple effect on other industries (11). The electronic information industry in China has recently been applied to other sectors in the economy. Information in the industries and firms were getting faster. Import of applications by such industries as telecommunication, finance, power industry and traffic exceeded 25 percent in 2010 (7). Expansion of the application of the electronic information technologies is likely to contribute to the development of the Chinese economic growth to a great degree.

4 Electronic Information Policy

4.1 Electronic Information Policy of China

Chinese government has set up the 'Plan for Fostering Semiconductor industry'. And five-year plan also included policies for fostering the semiconductor industry. In 2000, the 'Incentives for the development IC Industry' which aims to support the semiconductor industry only was announced (11).

In 2005, the Ministry of Information Industry of China announced direction of the IT Industry policy. Its main concerns included strengthening of self-reform capability and scale-up of IT enterprises. The IT industry policy put its emphasis on the qualitative development and it also seeks to enhance the competitive edge (4).

The Ministry of Information and Industry announced list of core technologies and major product of the IT industry. At the same time, it came up with various support policies for the industry. IT firms in China are seeking ways to strengthen their competitive edge. Research and development spending of the IT firms are on an upward path (4). On the 10th five-year plan announced in 2001, digitization, networking and intelligence in the electronic information industry were suggested as major targets (3). The "11th Five-year plan for IT Industry" and "Plan for the Promotion of IT Industry Restructuring" have been implemented as the major IT industry development policies since 2006.

4.2 The 11th Five-Year Basic Direction for Electronic Information Industry

In 2005, Chinese government made public the 'Five-year Basic Direction for Electronic Information Industry'. This plan was spearheaded by the Ministry of Information Industry. The 11th five-year direction was implemented starting 2006 and it took aspect of more positive policy supports for the IT industry. As a means for carrying out the direction, 10-item plan for information industry were prepared.

The 11th Five-year Basic Direction pointed out two core sectors in the electronic information industry: telegraphy and electronics. Based on this, the policy goal was set up to construct 'the powerful nation in the field of telegraphy and electronics'. Main focus of the 11th Five-year Basic Direction is to intensively nurture third-generation telecommunication service, Digital TV, semiconductor, and software. At the same time, it also aims to construct information infrastructure that can integrate wide-area telecommunication network, Digital TV network and Internet network. Overseas expansion of Chinese firm was emphasized. Fostering China's own brands and multi-national corporations equipped with first-class quality and first-class service were suggested as an important task (3).

At the early stage of the reform and door-opening era, Chinese government set up 'Development Fund for the Electronic Information Industry' in order to boost the electronic information industry, and it is still maintained. The 'Development Fund for the Electronic Information Industry' selected software and semiconductor (Integrated Circuits) as its major investment targets during the period of the 11th Five-year Basic Direction. The Fund also expanded its support for the development of such application fields as electronic government, electronic commerce, autocttive

electronics, medical electronics, and electronic finance (3) This is a policy that seeks to enhance the improvement of productivity by integrating the electronic information industry with other industries.

4.3 Plan for the Promotion of IT Industry Restructuring

In February 2009, Chinese government announced its "Plan for the Promotion of IT Industry Restructuring". Basic principles of the plan are as follows: ① Promotion of stable growth; Coordinated combination of market operation and government directives; Combination of 'Self-reliance on Technological Innovation' and international cooperation. The 'Self-reliance on Technological Innovation' refers to China's endeavor to achieve technological self-reliance through its efforts to develop industrial technologies (6).

The Plan for the Promotion of IT Industry Restructuring provided 6 key projects related with the electronic information industry (5). Major contents of the plan consist of four categories. ① Ensuring stable growth of core industries; The core industries' major technology innovation; Fostering new growth points; Policy measures (6). Major contents are as shown in Table 3.

Table 3. Promotion plan of electronic information industry restructuring (6)

Planes	Main content
6 major projects	Integrated circuit upgrade Changes in new display & color TV industry Passing of the new mobile communications industry Penetration of digital TV & Movies Computer upgrade & next-generation Internet applications ⑥Development of software & information services
Ensure the stable growth of key industries	Strengthen the competitiveness of the computer industry, Accelerate upgrade of electronic components & products, Audio & Video industry's digital transition promotion
The core industry's major technology innovation	Integrated Circuit Systems Maintenance Industry, Overcoming the challenges of the new display industry, self-reliance development capability of the software industry
Fostering new growth points	Dedicated to the development of telecommunications infrastructure manufacturers, Develop new management methods and new models of information services, Strengthening Information Technology Convergence Applications

5 Conclusion

On the back of the rapid economic growth since its reform and opening policy, China has emerged as the world's biggest production site of the electronic information products. Its export of the electronic information products has also been increasing steadily. The export volume is almost up to a third of China's whole export.(4) It is forecast to be continued for the times to come.

However, the electronic information industry of China is also saddled with some chronic problems. The competitiveness of the industry still remains weak. Capability of technological self-reliance is also far from being sufficient. The growth of the industry has been led by foreign-invested firms. Chinese firms are not likely to secure market dominance within short period of time. Though R&D investments in China have been increasing in a steady manner, great portion of the investments were made by the foreign-invested firms. China has expanded domestic firms' overseas M&A in an attempt to narrow the technological gap, and it is expected to continue in the future.(4)

In its effort to address problems of the electronic information industry and promote harmonious growth, Chinese government started to implement the policy of expanding domestic consumption. Growth rate of domestic consumption has been hovering around 10% since 2009. Key production sites started to expand into middle-west region of China from coastal region.(7) Major cities in the middle-west region include Sichuan, Shaanxi, Henan, Hunan, and Anhui.

Amid the ever-increasing economic growth of China, the consumption of the electronic information products has also been increased steadily. It is forecast that the consumption of the electronic information products will keep expanding thanks to information strategies including Chinese government's electronic government effort, enterprise information policy, and metropolitan information policy.

Possibility of industrial growth through convergence of information with industrialization seems to be very high. Through the convergence efforts, the productivity of traditional industries could be enhanced and the consumption of energy and raw materials could be diminished. Information systems for service industries including finance, insurance and traffic is likely to undergo a generation shifts and investments in these fields are also expected to increase. In China, the development of electronic information technology and industry will play a key role in the enhancement of its national power.

References

1. Kim, J.-w.: Float of Chinese High Tech Enterprise and Its Implications, Samsung Economics Reserch Institute, SERI management notebook, no. 38 (2010)
2. Cho, J.-h.: China's Economic development & 21 century Development Strategy. Busan University press (2003)
3. Lee, M.-h.: Industrial Policy Direction of China's 11.5 Plan & Its Implications, KIET (2005)
4. Kotra North-East Team, Change of Chinese Industrial Policy, 4 Kinds Responce Point, Planning Investigate 06-053 (2006)

5. Cho, C.: Prospect of China's Structure Change & Korean Industrial Long-term Responce Strategy, KIET (2010)
6. Lee, G.-g.: The Trend of China's Electronic Information & Its Implications. Trend 21(8) (2009)
7. Korea International Trade Association(ShangHai), The Current Situation of China's Electronic Information Industry & Prospect, Report 10-04
8. Yoo, H.: The Understanding of International Situation. HanWool Academy, Seoul (2007)
9. Gong, Y.-i.: The Overseas Expansion Trend of Chinese IT Industry. Electronic Information Policy (2005)
10. Lim, J.: A Study on Float of Chinese IT Industry & Inhancing Strategy of Korean IT Industry. East-West Study 19(1) (2007)
11. Jeong, D.-y.: A Float of China's Semiconductor LCD Industry & Action, amsung Economics Reserch Institute, CEO Information, no. 733 (2009)
12. Chosun.biz, <http://biz.chosun.com>

3D Content Industry in Korea : Present Conditions and Future Development Strategies

Yunkyung Lee¹ and Youngsik Kwak²

¹ Senior Researcher, Korea Culture & Tourism Institute,
#827, Bangwha-3dong, Gangseo-gu, Seoul, Korea
Tel.: +82-2-2669-6938
yunky@kcti.re.kr

² Associate Professor, Gyeongnam National University of Science and Technology,
#150, Chilam-dong, Jinju, Korea
Tel.: +82-55-751-3454
yskwak@gntech.ac.kr

Abstract. The purpose of this study is to formulate a direction of development for acquisition of global competitiveness by understanding the status of Korea's 3D content industry and to diagnose problems. For this, the study has collected and analyzed literature related to 3D content industry and grafted them onto Porter's Diamond model. As a result, the study has suggested key success factors such as the fostering of professional 3D personnel from a long term perspective, networking of industry and academia, consolidating skilled technology able to lead 3D project, acquiring government support for reducing high investment risk, etc. Also, the study has revealed a necessity for the government, businesses, and schools to formulate an integrated developmental strategy by cooperating with each other to solve such issues.

Keyword: 3D content industry, Korea content industry, Porter's Diamond model.

1 Introduction

The movie, *Avata* was immensely successful not only in the global market but also in the Korean market. The total number of audiences in Korea who watched *Avata* exceeded 13 million, and it achieved a record result of an accumulated retail amount of 120 billion won. As one reason for such a success, one cannot leave out the evolution of method of expression due to the development of 3D technology. Because of impressive technology utilizing computer graphics, the audience was able to view a variety of special effects which they had not seen before.

Yet, in the case of contents which are not armed with a story, it is difficult to succeed however impressive the technology that has been grafted is. 3D content which is not grounded in creative imagination will not be able to give inspiration to the consumer. Attending the 'Seoul Digital Forum' convened in 2010, director James Cameron emphasized that 3D content is a key factor for invigorating the 3D industry.

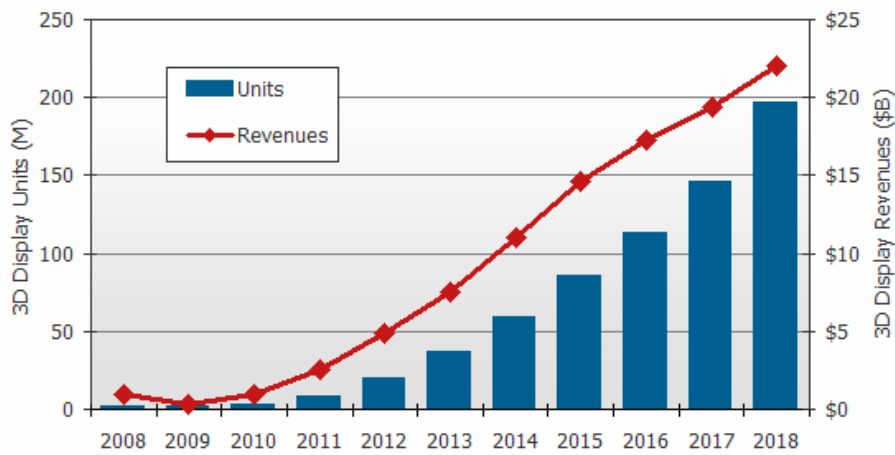
As such, solid content supported by a well weaved story, lively and moving characters, etc. are the key to 3D image revolution.

Yet, there is a side to the direction of improvement of 3D in Korea which focuses only on the technical aspect. Human resources are mainly concentrated in the area of basic converting technology that converts 2D content to 3D and which neglects the sphere of creating and distributing 3D content.

Thus, this study seeks to diagnose the status of Korea’s 3D content industry that is trying to move onto the global market as a successful competitor of the 3D image revolution and to suggest a direction of development for a balance between technology and storytelling of the content.

2 Status of Korea’s 3D Industry in Global Market

The competition for 3D industry in the global market is fierce. In the case of 3D display market among 3D contents worldwide, one forecasts a rapid growth to an annual average of 40% from 900 million dollars in 2010 to 15 billion dollars in 2015.¹ Countries around the world are competing to secure technological development to corner the 3D market and secure the initiative for standardization, and for this, cooperation among companies are actively being implemented.



Source : DisplaySearch(2010), 3D Display Technology and Market Forecast Report

Fig. 1. 3D Display Forecast

Amidst such a global competition, one forecasts that Korea’s 3D content industry will form a market of 2.5 trillion won by 2015, after growing at an annual average of 90% from a market size of 30 billion won in 2009.² It is expected that the Immersive Media such as 3D image and others will create an effect on production inducement of

¹ DisplaySearch (2010), 3D Display Technology and Market Forecast Report .
² Ministry of Culture, Sports and Tourism (2010), Plan for Fostering 3D Content Industry .

approximately 88 trillion won by 2027 via service and equipments in the Korean market. Moreover, the prospect is that an added value of approximately 2 trillion will be induced and employment for 490,000 man/year will be created.³

Table 1. Expectation of economic effect for Korea's 3D media service

Category	`08~`12 year	`13~`17 year	`18~`22 year	`23~`27 year
Effect on production inducement (Unit: Billion won)	25,129	141,075	303,371	417,081
Effect on value added inducement (Unit: Billion won)	10,021	51,045	98,861	132,408
Employment inducement (Unit: Person)	11,909	75,518	169,798	237,671

Source : ETRI(2009), Analysis of Consumer's Adoption Intentions and Prospects of Industry for Immersive Media

Still, in the case of Korea's 3D industry today, because entertainment contents including special video for theme parks, PR video, etc. make up more than 90%, and the market for movie broadcast and so on are still in its initial stages.⁴ Yet, popularization of 3D is gaining momentum as global media companies begin operations in Korea starting with the 3D broadcast of skylife. In 2010, a trial broadcast of 3D TV was implemented by a terrestrial broadcasting station for the first time in the world.

For movies as well, 3D theaters are increasing rapidly in number. An entertainment company in Korea, CJ CGV converted 30% of its screens to 3D in 2010. Megabox likewise plans to increase the number of 3D screens by five times in 2010.

3 Direction of Improvement for Korea's 3D Content

In the global market, Korea is considered as an IT powerhouse. According to the 'Information and Communications White Paper 2010' issued by Japan, the national IT competitiveness of Korea is evaluated as number 1 in the world.⁵ Nevertheless, in terms of 3D content industry, the gap with the developed countries is significant. This is because existing IT competitiveness has not been linked with new media technology for development. To analyze the reasons, this study seeks to use Porter's Diamond model which recognizes national competitiveness as a comprehensive concept of a variety of factors.

Porter's Diamond model is composed of 4 factors.⁶ First, the factor condition refers to output factors such as skilled workers for infrastructure. Second, the demand-

³ ETRI(2009), Analysis of Consumer's Adoption Intentions and Prospects of Industry for Immersive Media .

⁴ Ministry of Culture, Sports and Tourism (2010), book above.

⁵ Japan Ministry of Internal Affairs and Communications (2010), Information and Communications White Paper 2010 .

⁶ Porter,M.E.(1990), Competitiveness Advantage of nation , The Free Press, New York.

ing condition denotes characteristics of demand for goods and services. Third, the related and supporting sector refers to the conditions of related and supporting sector. Fourth, the strategy, structure & rivalry are a diagnosis of the working environment of a company.

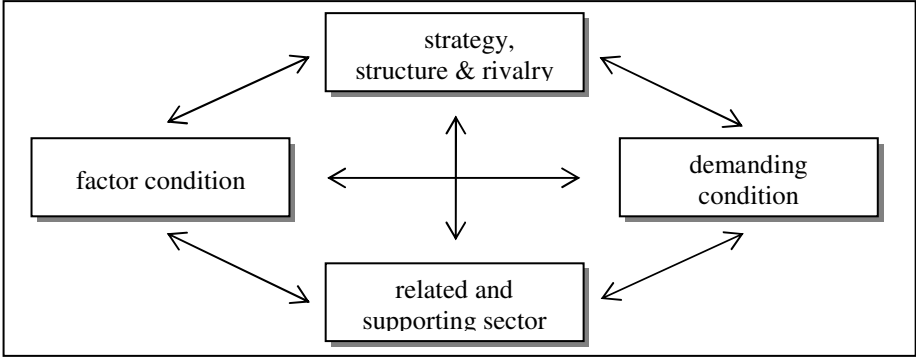


Fig. 2. Porter's Diamond Model

First, the competitiveness of Korea's 3D industry which is related to the factor condition can be evaluated to be very weak. In the case of human resources for 3D content industry, despite the forecast that it will expand from 6,000 persons in 2011 to 15,000 persons in 2015, dearth of human resources creates a serious situation because among the personnel in broadcast, movies, games, animation, and CG, the personnel for 3D production stands at the level of 1.3%.⁷ To solve this problem, an analysis of the structure of supply and demand of human resources and a job analysis must precede. Further, there is a need for a long term plan for development of human resources based on such a survey.

Also, there exists a serious GAP between the capacity of graduates from schools and the capacity as required by companies because of a lack of actual industry and academia cooperation in terms of human resources development. According to one survey, 86% of content businesses of Korea responded that the capacity of graduates does not meet the required level of companies.⁸ According to another survey, in respect to content industry education, it was found that there was a lack of onsite supervising professors as well as a curriculum for learning the tasks required on industrial sites.⁹ Therefore, cooperation among the government, schools, and companies are necessary so that field oriented education can be offered by the active participation of company personnel in education.

⁷ Ministry of Culture, Sports and Tourism, Korea Creative Content Agency (2009), A Survey of Domestic and Overseas CG Status and a Study of a Plan to Start Operation Overseas .

⁸ Korea Creative Content Agency (2010), A Forecast of Supply and Demand of Human Resources for Content Industry and a Survey of Benchmarking of Case Study of Overseas Developed Countries .

⁹ Korea Creative Content Agency (2010), An Analysis of Content Educational Environment and a Study of a Plan for Improving Education .

Second, in terms of demanding condition, because there are many consumers familiar with industries related to IT, it exerts positive influence upon 3D industry. For example, approximately 33% of Korean audiences who watched the movie, *Avata* visited a 3D theater by paying for a higher expenditure. Among these, 13% watched again, thus showing a higher commitment than for general content. As such, there is a need to develop a plan for utilizing such a positive demanding condition.

Table 2. Number of audiences for the movie, *Avata* (Unit: Thousand persons)

Total number of audiences	Number of 3D audiences	Number of repeat audiences
Approximately 800	Approximately 260	Approximately 34

Source : Maxmovie aggregate data (as of January 8, 2010)

Third, there is inadequate technical support for development of 3D industry from the perspective of related and support sector. While the strength is that technology exists for converting 2D content to 3D content, a gap is observable in generic technology such as filming, calibrating, editing, workflow, etc. Hence, there is a high tendency to rely on overseas countries for supply and demand of stereoscopic image content for developing 3D content. We need to train again field specialists to furnish them with the basic capacity to create 3D content rather than foundational technicians in charge of converting. Based on this, we need to be able to generate high added value by developing human resources at the level of Technical Director, who can lead a project.

Fourth, one sees negative elements from the perspective of strategy, structure, and rivalry. When compared with the high investment list of 3D industry, there is no active investment due to an absence of a business model. As such, the industry is not invigorated based on fair competition. The current situation is that 69% of the business world of domestic video digital content show less than 1 billion in retail, because the investment environment as compared to a required expenditure that is 2 to 3 times greater than for 2D is inadequate. To improve this, we must make an effort to reduce the risk caused by market uncertainty when a content business first seeks a share of the market. The government must support the decision for investment association to secure source of revenue and find ways to improve the system for invigorating overseas investment.

4 Conclusion

The 3D content industry of Korea is now in its beginning stages. It is expected that in the future, the market size of Korea's 3D content industry will continue to expand and will create high economic effect. Nonetheless, such an expectation will be realized only if several key conditions have been improved.

To diagnose key conditions, this study has specified areas of improvement by utilizing Porter's Diamond model. To become a globally leading 3D country, we require development of professional human resources in 3D from a long term perspective. Also, development of professional human resources must have as its basis, the networking of industry and academia in order to reflect the demands of worksite.

Second, from the technical perspective, one finds the need to reinforce skilled technique that is able to lead a 3D project. Third, government support is needed in order to reduce the high investment risk associated with 3D industry. The present key task for development of Korea's 3D industry is to foster a capacity to plan for and create 3D content that brings about high added value. For this, what is important above all is that companies, schools, and the government cooperate with each other to formulate an integrated implementation plan.

References

- DisplaySearch, 3D Display Technology and Market Forecast Report (2010)
- ETRI, Analysis of Consumer's Adoption Intentions and Prospects of Industry for Immersive Media (2009)
- Japan Ministry of Internal Affairs and Communications, Information and Communications White Paper 2010 (2010)
- Korea Creative Content Agency, An Analysis of Content Educational Environment and a Study of a Plan for Improving Education (2010)
- Korea Creative Content Agency, A Forecast of Supply and Demand of Human Resources for Content Industry and a Survey of Benchmarking of Case Study of Overseas Developed Countries (2010)
- Ministry of Culture, Sports and Tourism (2010), book above
- Ministry of Culture, Sports and Tourism, Plan for Fostering 3D Content Industry (2010)
- Ministry of Culture, Sports and Tourism, Korea Creative Content Agency, A Survey of Domestic and Overseas CG Status and a Study of a Plan to Start Operation Overseas (2009)
- Porter, M.E.: Competitiveness Advantage of nation. The Free Press, New York (1990)

A Study of Mobile Application Usage in after Service Management for Consumer Electronics Industry

Wonkyu Ji and Xu Wang

Digital Marketing Group, Samsung OpenTide China
kay.ji@opentide.com.cn

Abstract. This motive of this study was to introduce how an influence Mobile STS can be; with a vision of advanced technology brought to simple everyday business use. Interviews with major companies' ASC managers were set up, to verify the problems at hand and to accurately help change the current situation. A prior variable in all this was the test trial we ran; study showed the new adjustment in the ASCs had a high ratio of success. Each situation solved in a significant short amount of time, thus proving the Mobile STS will be an important element in businesses.

Keywords: Consumer electronics, After service, MID, Mobile, Application, Android, LBS, AR, Navigation, GPS, 3G network.

1 Introduction

21st century, is a time where theories come to life, visions are taken into action, and concepts beyond belief are reality [1]. It is a time, where advancing all possibilities to the next level are enacted. This proposal presents a whole new system, in which ASC's (After Service Center) will enhance and improve Management and Service quality exceedingly. The Mobile STS (Mobile Service Tracking System) is a system that is programmed to help Companies reduce expenses in all different heights, ameliorate work efficiency and improve GPPM services. Confrontations of before and after use of Mobile STS are included in this study; aiming to bring an understanding of how a phenomenal role it can constitute in this generation and generations to come.

2 Problems

This study was started base on the following problems.

2.1 Management

With a global sized electronics company, it comes with a massive amount of personnel; an amount where operating, is impossible to prefect. Erudition Complications, being one of the imperfections, has been a headache hard to comprehend. Shortage of resources and time to cultivate mechanics on contrivance skills have resulted to

constant re-visitation and customer's dissatisfaction. Correspondence between ASC and mechanic has been a complex expedient. Every assignment and situation was contacted by telephone, culminated with confused and unhappy employees; indicating the next "imperfection", Navigation situations. With China, Beijing in particular, growing in rapid speed, traffic geography has become a maze. An amiss circumstance mechanics are sent off without any accurate map to destination, causing tardiness. Delay not only being a problem of direction, but also of complication of schedules. Time is valuable and a gift we cannot control, thus being delayed is unavoidable. There is not an efficacious way to record the exact amount the mechanics use and receive, or even to track their whereabouts.

2.2 Expenses [2]

Mechanic travel expense: Average of 10,000/month in Feb 2010 to Jan 2011

After service center 1 expense based on estimated data: 3,400/month

After service center 2 expense based on estimated data: 8,300/month

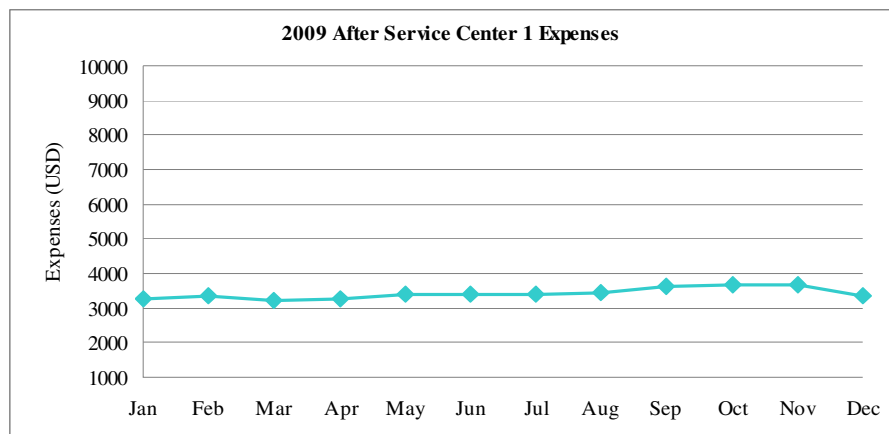


Fig. 1. After Service Center 1 Expenses in 2009

2.3 Customer Discontent

There are various cases, in which a given assignments could not be or is not accomplished, lack of needed items, misunderstanding of where the problem resided, traffic conditions or even because of limited training of the mechanics. Every year 8.7% of VOC records show repairing delay and customer's dissatisfaction; 11% of re-visits are caused by technical reason and repairing delay. Fraud also plays a huge role in customer's complaints. Without knowing the exact time and actual reason of visitation, it is insurmountable to prevent.

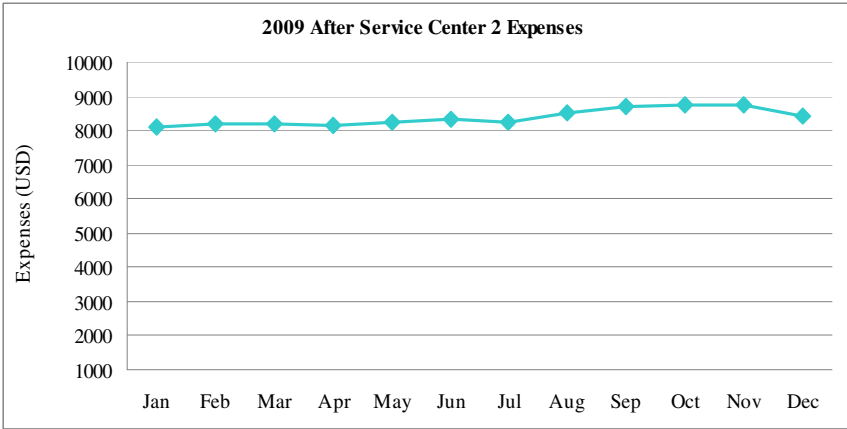


Fig. 2. After Service Center 2 Expenses in 2009

2.4 Example Aka Point of View

The estimated amount of personnel nation-wide is 5,600, a massive number to manipulate. Not only so, there are over 50,000,000 products sold every year from multi-electronic companies [3]. Without an effectual system to organize, the profits are not satisfying. The complex handling every assignment by hand is chaos; receiving task from phone calls, to getting the needed information to the mechanics, and so forth steps of the current system. There is not middle connecting point. In which, it is easily noticeable that it is in immediate requirement.

Each day on the job, mechanics are given assignment without a precise direction or clear understanding of what the exact problem is that the customer has stated. Though given two days training, it is not enough for the mechanics to undergo all situations, without a full understanding of the skills needed, it makes the process cycle much slow. Also with new products brought into the market in a exceedingly high accelerating rate, it is close to impossible to update manually to every mechanic each time a new product is introduced.

Case_1. In the morning, a mechanic arrives at ASC to receive his first assignment of the day. Without a precise direction of the destination, he sets off to the given address. Unfortunately, he hits traffic on his way, preventing him to make the designated time.

Case_2. Upon arriving, he inspects the problem, realizes the problem was not located where the customer had said it was, yet to have resided in a different area, in which he had no idea how to approach. Misfortune upon misfortune, he apologizes and returns without completing his assignment to ASC.

3 Analyze

3.1 Management Issues

Quoting from Mr.Yun, a Senior Manager at ASC [4], “A mechanic can’t really be equipped 100% of all problems that might occur on all the merchandises. The rate of

new merchandise brought out into the market is like lightening speed.” Thus resulting to and revealing clearly to surface, that mechanics ‘lack of skill’ to be one of the problems in hand. If Mobile STS provides more resources, mechanics can look up needed guides on merchandise mantling, shortening the time to accumulate them on learning the manuals or placing a phone call each time there is a situation to ASC headquarters.

Communication/ Schedule management, a huge worriment for both mechanics and ASC managers; for each assignment mechanics are to return all paperwork (customer’s signed receipt, item ordering forms, etc) back to the office when completed. Sudden changes of schedule and announcements are also a problem in need of serious change. Many a time, unable to update mechanics on time causes unnecessary misunderstandings to occur. By letting mechanics know their daily task on the STS with instant messaging system (similar to modern day SMS), excludes the waste of time of returning each time to main headquarters. Documents can be delivered back to the system in an alternative way: email, online management system, and so forth.

Navigation/ GPS service, the process of monitoring and controlling the movement of a craft or vehicle from one place to another. It is also the term of art used for the specialized knowledge used by navigators to perform navigation tasks. All navigational techniques involve locating the navigator’s position compared to known locations or patterns a convenient tool also provided in STS, that way they can avoid traffic and visitation delays.

Embezzlement, one to the things companies worry about immensely. Because there is no existing surveillance system to record the activities of mechanics, it is hard to prevent any unacceptable behaviors. Yet, with a surveillance system inside of the STS, we can ensure there are no scandals or breaking of company policies. With Internet connected digital mobile device, it is easily reached and recorded correctly.

3.2 Customer’s Discontent Issue

Incomplete task, to say it is a problem is harsh, like mentioned [5] before schedule changes and merchandise problems are never predictable. From a consumer’s view point, it is not an excuse understood. Mechanics find it difficult when trying to repair a digital product, sometimes hard to clarify whether the needed part is in stock or not. Causing frustration and resulting to un-happy customers. Considering all the factors, STS proffer a platform to help mechanics repair product, step by step, even providing an instant item checking system.

Punctuality, by human nature is respect we all want to perceive. Though not always intentionally, incidence occur and not always can we be on time. Vis versa Mechanics could not find location of destination swiftly of with overlapping tasks, decreased the efficiency of customer’s content. The navigation system arranges the nearest route, locating all the mechanics for the ASC, to manage task allocation and avoid overlapping repair areas.

3.3 Expense [6]

The percentage of budget reduction improved on both travel expenses and Customer satisfaction.

After applying the Mobile STS in daily use, the average monthly expenses of After Service Center 1 was 20% reduced in second half of 2010 compare to the first half of 2010, from \$3,400/month to \$2,720/month.

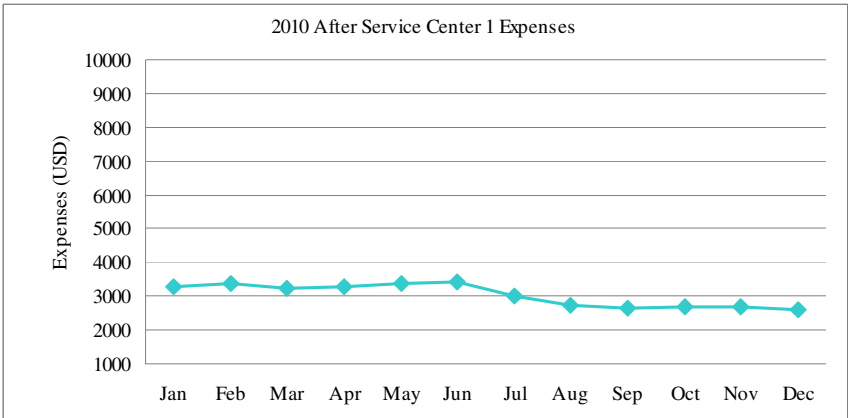


Fig. 3. After Service Center 1 Expenses in 2010

In addition, it lowered the cost of After Service Center 2 as well, the average monthly expenses was 20% reduced in second half of 2010 compare to the first half of 2010, from \$8,300/month to \$6,640/month.

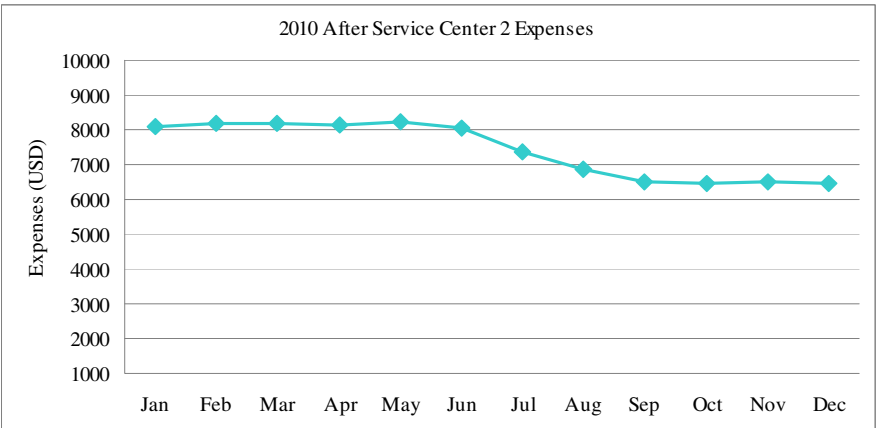


Fig. 4. After Service Center 2 Expenses in 2009

4 Functions Induction

A 7inch Touch pad, with CPU ARM A9 800MHz, memory of 256, capacity of 16G, and a USB opening; it is an Android 2.2 operating system, using a specifically designed, able to handle dynamic content by executing code written in JAVA [7].

Task management Program is concluded by three parts: *schedule & repairing data, job completion form & product part stock status, and Push notification*. The program digitalized all factors, overtaking accustomed flaws, as in schedule forms not to be acquired by sheet but with STS. Repair forms which customers sign will also be improved, thus SC will be able to get accurate information of the process. On Mechanics are able to transfer data by uploading pictures when inputting info in STS's batch mode.

Navigation extends its needs in *job allocation management, mechanics traveling system, and customer's confirmation system*. After every job, each customer's information is inputted into S-pad, and into internal STS system (WEB). The navigation system is searched by using Google Map API, which improves the accuracy of time and location; it can calculate automatically the travel expenses consumed, resulting to reduce false traveling expenses.

Barcode (+QR) Scanner is installed in the S-Pad so by using it mechanics can simply scan barcode on distribution sheet or box. Using S/N barcode of product can reduce error and use of ERP interface to confirm quality product and service information by checking the barcode.

E-Learning Program is a learning program for mechanics, based on text, images, videos, and even quizzes. Instructions of all products are installed and are updated constantly. E-manuals can be downloaded and read offline.

Cameras on the S-pad are used to take picture of products before and after repair information. It is sent back in batch mode to ASC instantly by 3G or WIFI.

Video Phone gives a direct contact with ASC, resulting to minimum complications. With a camera on both sides of the STS it makes it a convenient and swift process.

AR (Augmented Reality), is an advanced program that helps the mechanic get a clear image of the problem at hand; being able to get repairing instructions by clicking on the problem area on image taken by STS-camera.

5 Conclusions

The use of mobile and digital technology is not limited in marketing; it can be used as a business tool to help improve corporate service performance. Mobile STS is a network system that has easy access and has exceptional results. The outcome of this study encapsulates into eight goals we aim to bring to mobile STS users.

1. Accurate route distance from ASC to customer home (reduce travel expenses)
2. Get Accurate repair TAT (turn-around-time) by inputting completion time instantaneously in customer home.
3. Save time by repairing schedule management
4. Reduce repeat visit to customer home. Real-time customer satisfaction survey by mechanic.

5. Improve service quality by collecting actual data.
6. Engineer education in spare time. Improve engineer working ability.
7. New engineer learn working knowledge by eLearning.
8. Understand service process by eLearning.

References

- [1] RalSchauerhammer: From Spring 2002 21st Century issue. Why There Really Are No Limits to Growth (2002),
<http://www.21stcenturysciencetech.com/articles/Spring02/NoLimits.html>
- [2] Test run_before data
- [3] Interview with Mr. Yang manager at SC
- [4] Interview with Mr. Yang
- [5] Interview Mr. Yun, Senior Manager
- [6] Test run after data
- [7] Shin, S.-W., Kim, H.-K.: Lightweight Framework For Supporting Mobile Web Development (August 2009)

A Study on the Mechanical Properties Analysis of Recycled Coarse Aggregate Using Polymer Impregnated

Hyun-gi Ryu^{1,*} and Jung-sik Kim^{2,*}

¹ Professor, Chungju National University, Architectural Engineering, #123,
Kumdan-ri, Chungju, Korea, +82-43-841-5207
hgryu@cjnu.ac.kr

² Professor, Chungju National University, Industrial management Engineering, #123,
Kumdan-ri, Chungju, Korea, +82-43-841-5308
jskim@cjnu.ac.kr

Abstract. This study aims at comapring and analyzing the mechanical properties of concrete when the recycled coarse aggregate impregnated. Water-soluble polymer is used to solve the problems of recycled coarse aggregate, because when recycled coarse aggregate is produced. The amount of air in the unhardened concrete was biggest when it was inpregnated for one hour in all replacement rate of 40% and 60% of W/C, and when the smallest amount of air was impregnated for 24 hours. Its per unit capacity mass was bigger than that of plain with 40% of W/C in case of 1 hour and 6 hours, and was generally smaller than that of plain at 60% of W/C. This is because when polymer is impregnated, it plays the role to fill in the gap for a short time, but after a certain amount time, it makes small but more gaps with solid powder, which lowers the adherence strength.

Keywords: Recycled coarse aggregate, Polymer, Impregnation time.

1 Introduction

In case of old building, neutralization is accelerated by a falling-off in quality and environmental contamination. As the result, the durability of buildings declines and reconstruction and redevelopment is magnified [4]. The volume of construction waste by reconstruction was 5.419 million ton in 2004 and it will be twice in 2010 [2]. Through construction waste keeps increasing, the recycle is not very active. It is just recycled as embankment, the earth bermed and subbasal. Recycle ratio for cement and concrete is only 1.8%. The need of using recycled aggregate is getting more and more important. The government legislated “the law of encouragement of recycled construction waste” in 2003. “Standard quality of Recycled aggregate” was released by Ministry of construction and transportation in 2005. Thus, the quality standard is established to manage of recycled aggregate according to the purpose. However, cement paste on the aggregate surface is not perfectly removed during the production process of recycled aggregate, so it becomes main reason of strength performance of concrete

* Corresponding author.

[5]. The concrete using recycled aggregate increases unit amount, and falls durability and strength compared to natural aggregate. As the reason, the quality standard of recycled aggregate suggested to use recycled aggregate less than 30% of total aggregate [1]. Therefore, in this study, we aim at comparing and analyzing the mechanical properties of concrete when the recycled coarse aggregate is impregnated by waste soluble polymer because cement pastes of recycled coarse aggregate gives the strong effect for strength performance of aggregate.

2 Design of Experiment and Experiment Method

2.1 Design of Experiment

Table 1 shows the design of experiment (DOE). Experiment factors are consisted of mixture and experiment item. Mixture items are varied based on table 1. Experiment item contains fresh and hardening concrete. To figure out the properties of fresh concrete and hardening concrete, experiment is performed with table 1 based on KS standard.

Table 1. Design of Experiment

Experiment factor		Types	
Mixture item	W/C(%)	2	40 , 60
	Target slump(㎜)	1	180±25
	Target air content(%)	1	4.5±1.5
	Recycled aggregate coarse replacement ratio(%)	4	0(Plain), 20, 30, 40
	Polymer impregnation time(hr)	4	0, 1, 6, 24
Experiment item	Fresh concrete	4	-Slump and slump loss (0, 1, 6, 24 hrs) -Air content -Unit volume weight
	Hardening concrete	4	-Compressive strength (3,7,28 days) -Tensile strength (3,7,28 days) -Changing length of Drying shrinkage (3,7,14,28,56 days) -SEM images

2.2 Materials

In this study, portland cement produced by “S” company in South Korea was used and fine aggregate is obtained from river sand in Chungju, South Korea. In addition, coarse aggregate has maximum size 25mm from Mt. Chungju and recycle coarse aggregate with maximum size 25mm is obtained from “H” company in Gyunggi, South Korea. Waste ceramic was obtained from Gyunggi, South Korea. Ethylene propylene rubber produced by “S” company was used as polymer.

3 Results and Analysis

3.1 Fresh Concrete

3.1.1 Slump and Slump Loss

As shown in Fig 1, slump loss of plain at entire pass time and recycled coarse replacement ratio represents the largest value. As recycled aggregate coarse replacement ratio increases, slump loss dramatically decreases. Slump loss at 20% recycled aggregate coarse replacement ratio shows the largest change at 1 hour and the smallest change at 0 hour. On the other hand, slump loss at 30% shows the smallest change at 6 hours, and slump loss at 40% shows smallest changes at 1 hour. The slump loss at W/C 60% has the largest value at 24 hours impregnation time compared to that of plain and at 0 and 1 hour polymer impregnation time, it has the smallest value for entire recycled aggregate coarse replacement ratio. It was analyzed that the loss width of slump loss according time pass showed the largest value at 40% recycled aggregate coarse replacement ratio.

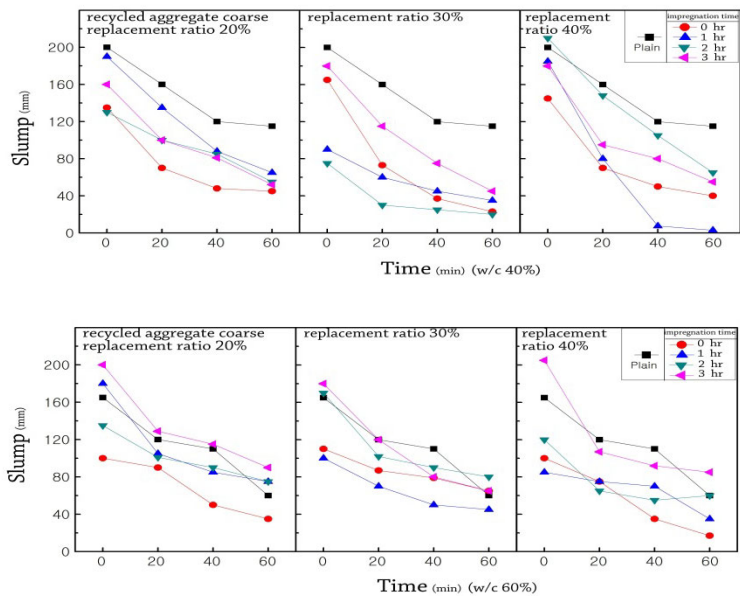


Fig. 1. Slump loss test result

3.1.2 Air Content and Unit Volume Weight

As impregnation time of recycled aggregate coarse replacement ratio passes, air content tends to increase. In addition, At 40% W/C, air content is 5.4% at 1 hour and it is 3 times larger than that of 24 hours. Air content at 60% W/C is twice than air content at 40% W/C. Unit volume weight is the highest at 1 hour impregnation time and the lowest at 24 hours impregnation time.

3.2 Hardening Concrete

3.2.1 Compressive Strength and Tensile Strength

As shown in Figure 2, compressive strength at 40% W/C had the largest strength performance - 21MPa, 27MPa and 30MPa for respectively 3 days and 7 days early age, and 28days standard age with 1 hour polymer impregnation time. And each case had 20MPa, 24MPa, 29MPa with 6 hours impregnation time. Compressive strength at 60% W/C had relatively good strength performance (1.69~1.89MPa higher) at 3 days and 7 days early age as recycled aggregate coarse replacement ratio increases le strength performance – 18.35~21.78MPa at 1 hour and 6 hours impregnation time. Compressive strength (F_c) and tensile strength (F_t) shows a similar trend. Strength performance rate (F_c/F_t) at 40% W/C was 1/9~1/11 and it was 1/8 at 60% W/C.

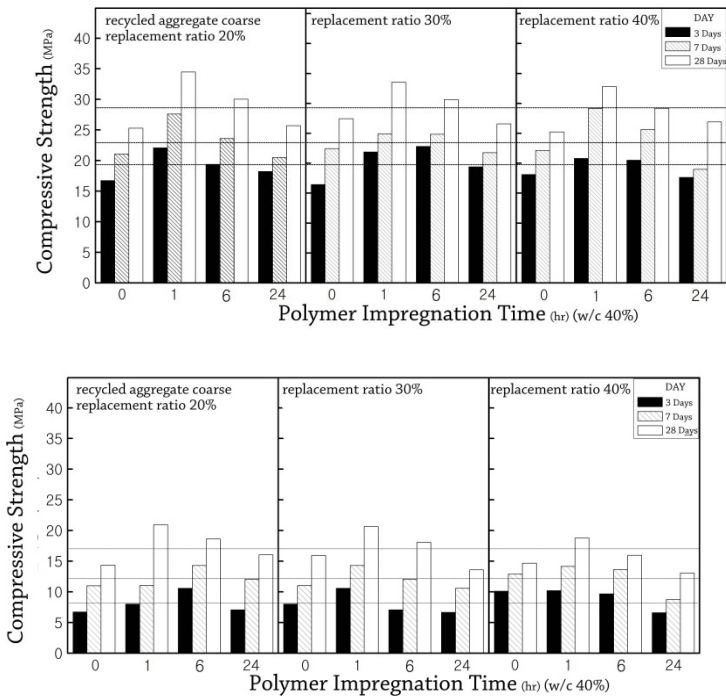


Fig. 2. Compressive Strength of polymer impregnation time

3.2.2 Changing Length of Drying Shrinkage

The largest changing length was 28mm at plain of 40% W/C. At the 56 days of 30% recycled aggregate coarse replacement ratio, the difference of changing length was -10~25mm according to impregnation time. It was -5mm according to impregnation time at 60% W/C.

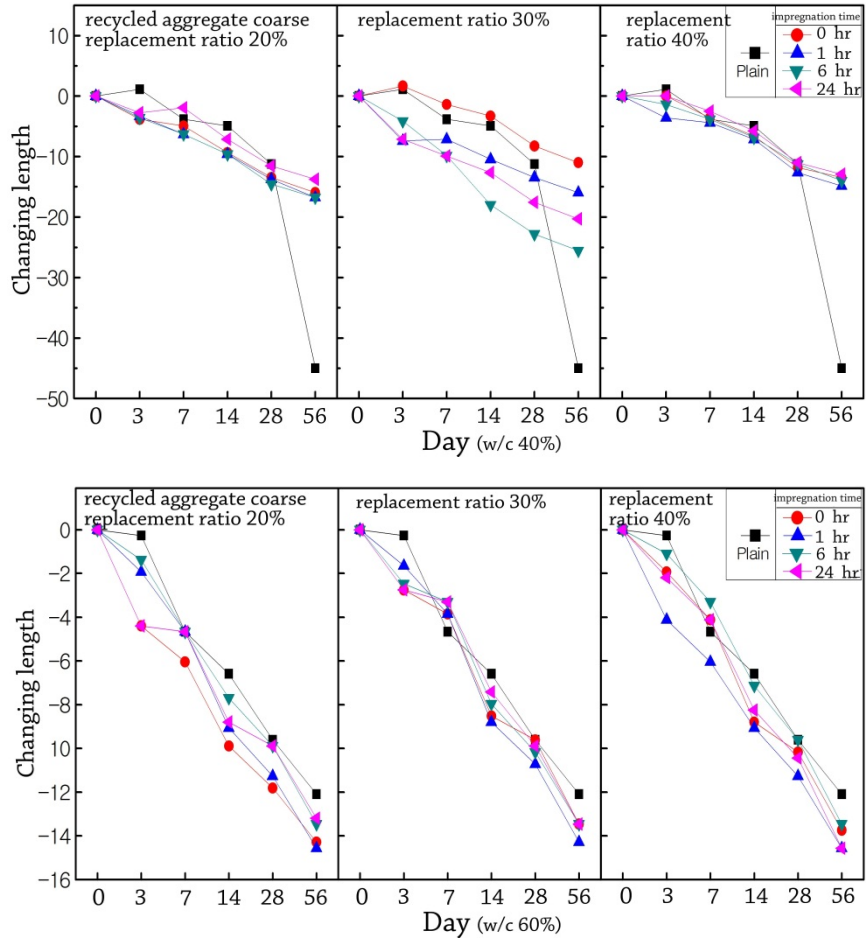


Fig. 3. Changing length of polymer impregnation time

3.2.3 SEM Images

Figure 4 and 5 show SEM images of recycled aggregate. The recycled aggregate with 1 hour impregnation time has many large pores. However, the aggregate with 24 hours impregnation time has many small pores because large pores were filled with solid contents of polymers and the contents are attached on the surface of aggregate.

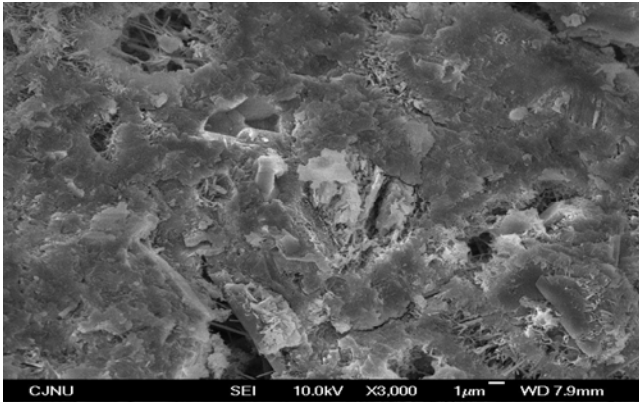


Fig. 4. W/C 40% and 1 hour impregnation time (X 3000 SEM)

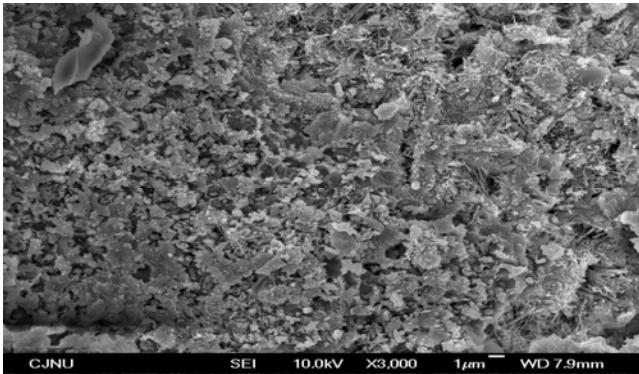


Fig. 5. W/C 40% and 24 hour impregnation time (X 3000 SEM)

4 Conclusion

Slump loss at 40% W/C dramatically declines as recycled aggregate coarse replacement ratio increases. The slump loss at W/C 60% has the largest value at 24 hours impregnation time. Air content increases as impregnation time increases. Unit volume weight tends to decrease as impregnation time increases.

Compressive strength at 40% W/C had the largest strength with 1 hour polymer impregnation time. Compressive strength at 60% W/C had relatively good strength performance (1.69~1.89MPa higher) at 3 days and 7 days early age. Compressive strength (F_c) and tensile strength (F_t) shows a similar trend. Strength performance rate (F_c/F_t) at 40% W/C was 1/9~1/11 and it was 1/8 at 60% W/C.

Compressive strength (F_c) and tensile strength (F_t) shows a similar trend. Strength performance rate (F_c/F_t) at 40% W/C was 1/9~1/11 and it was 1/8 at 60% W/C.

Changing Length of drying shrinkage was larger at 40% W/C and it was small (-5mm) at 60% W/C. SEM images showed that the aggregate with 24 hours impregnation time had good quality and it had only small pores.

References

1. Korean Standard Associations, KS Standard
2. Ministry of Environment, Basic recycling plan of construction waste, pp. 11–28 (December 2006)
3. Han, C.: Properties of concrete and design of mixture, Gimoondang (1998)
4. Yoo, D.: The study of Engineering properties of latex recycled aggregate, pp. 21–39 (2007)
5. Moon, D.-J., Moon, H.-Y.: Evaluation on Qualities of Recycled Aggregate and Strength Properties on Recycled Aggregate Concrete. The KSCE Journal of Civil Engineering 22(1-A), 141–150 (2002)
6. Choi, M.-K., Park, H.-g., Paik, M.-S., Kim, W.-J., Lee, Y.-D., Jung, S.-J.: An Experimental Study on the Utilization of Recycled Aggregate Concrete. Architectural Institute of Korea 25(1), 269–272 (2005)

A Study on the Strength Improvement of Recycled Aggregates Using Industrial Waste

Jung-sik Kim^{1,*} and Hyun-gi Ryu^{2,*}

¹ Professor, Chungju National University, Industrial management Engineering, #123,
Kumdan-ri, Chungju, Korea, +82-43-841-5308
jskim@cjnu.ac.kr

² Professor, Chungju National University, Architectural Engineering, #123,
Kumdan-ri, Chungju, Korea, +82-43-841-5207
hgryu@cjnu.ac.kr

Abstract. Objective of this study is to identify properties on strength increase of hardened concrete and fluidization of non-hardened concrete using waste ceramics generated by construction waste. Determine length change ration caused by drying shrinkage during substitution of recycle aggregate and waste ceramics, and whether they can be used as concrete compounds. Slump of non-hardened concrete exhibited the best fluidization and formability at recycled aggregate's replacement ration of 60% driven by higher substitution ration of recycled aggregate and waste ceramics while air content met the KS requirement when substitution ration of waste ceramics was 4,000cm²/g. Compressive strength of hardened concrete exceeded the requirements at early age and standard age and temperature dropped by roughly 6~10°C less than the standard at maximum temperature in adiabatic temperature increase. .

Keywords: Aggregate, Ceramics, Fineness, Heat of hydration, Temperature.

1 Introduction

Recently, Concrete waste has been dramatically increasing. It was reported that concrete waste was 1.5million ton in 2000 and it will be reached more than 10million produced from rebuilding of time-worn buildings in 2020[1]. Many studies have been conducted to recycle concrete waste, however, the produced recycle aggregate contains a lot of cement pastes and forging substance so that strength, resistance to freezing and thawing, drying shrinkage and chemical resistance decline. It is the main reason not to be used the recycling aggregate from concrete waste in industry. Korean government established KS F 2573 (The standard about recycle aggregate for concrete) to suggest the handling method of concrete waste through producing recycling aggregate with high quality. Additionally, it is encouraged to use recycle aggregate (30% of total) for the new building constructed by government [4]. In this paper, we study the effect of degree of particle size of recycle aggregate for strength of concrete. Construction waste and waste ceramics are used for the experiment.

* Corresponding author.

2 Design of Experiment and Experiment Methods

2.1 Design of Experiment

Table 1 shows the design of experiment (DOE). Experiment factors are consisted of mixture and experiment item. Mixture items are varied based on table 1. Experiment item contains fresh and hardening concrete. For the fresh concrete, change of slump loss every 15 minutes, air content, unit volume weight, time of setting of concrete mixtures by penetration resistance are measure to monitor slump and slump loss. For hardening concrete, compressive strength, tensile strength, changing length of drying shrinkage, water absorption ratio and adiabatic temperature rise of concrete according to scheduled age [5]. The experiment is performed based on KS standard [3].

Table 1. Design of Experiment

Experiment factor		Types		
Mixture item	W/B(%)	1	55	
	Target slump(mm)	1	150±25	
	Target air content(%)	1	4.5±1.5	
	Blast Furnace Slag ratio (%)	1	30	
	Recycled Aggregates ratio(%)	3	40, 50, 60	
	Waste pottery blain fineness ratio	6	4,000 (mm /g)	5, 10, 15
6,000 (mm /g)				
Experiment item	Fresh concrete	5	Slump and slump loss (every 15 minutes) Air content Unit volume weight Time of setting of concrete mixtures by penetration resistance Bleeding	
	Hardening concrete	5	Compressive strength (age 3, 7, 28, 91 days) Tensile strength (age 3, 7, 28, 91 days) Changing length of Drying shrinkage (age 1, 3, 7, 14, 28, 56, 91 days) Water absorption ratio Adiabatic temperature rise of concrete	

2.2 Materials

In this study, Portland cement produced by “S” company in South Korea was used and fine aggregate is obtained from river sand in Chungju, South Korea. In addition, coarse aggregate has maximum size 25mm from Mt. Chungju and recycle aggregate with maximum size 25mm is obtained from “H” company in Gyunggi, South Korea. Waste pottery was obtained from Gyunggi, South Korea and it was grinded as 4,000cm²/g and 6,000cm²/g using a 60kg grinder. Figure 1 shows the particle of waste pottery.

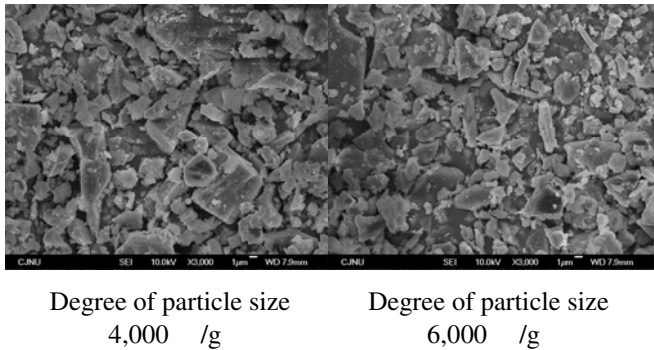


Fig. 1. The Particles of waste pottery (SEM X3000)

3 Results and Analysis

3.1 Fresh Concrete

3.1.1 Slump and Slump Loss

Figure 2 shows the change of slump loss according to time.

The change of slump loss according to pass time gradually decreases, as recycled aggregates ratio and waste pottery blain fineness ratio increase. At 4,000cm²/g degree of particle size according to pass time gradually decreases compared to at the degree of particle size 6,000cm²/g. Generally, lower slump trend shows every substitution ration compared to the slump loss of plain.

3.1.2 Air Content and Unit Volume Weight

Air content decreases mostly compared to that of plain. At 5% waste pottery blain fineness ratio air, high air content was observed and, as waste pottery blain fineness ration increases air content tends to decrease. It is probably due to stuffing of microsite of large volume of micro particles. Unit volume weight decreases mostly compared to that of plain.

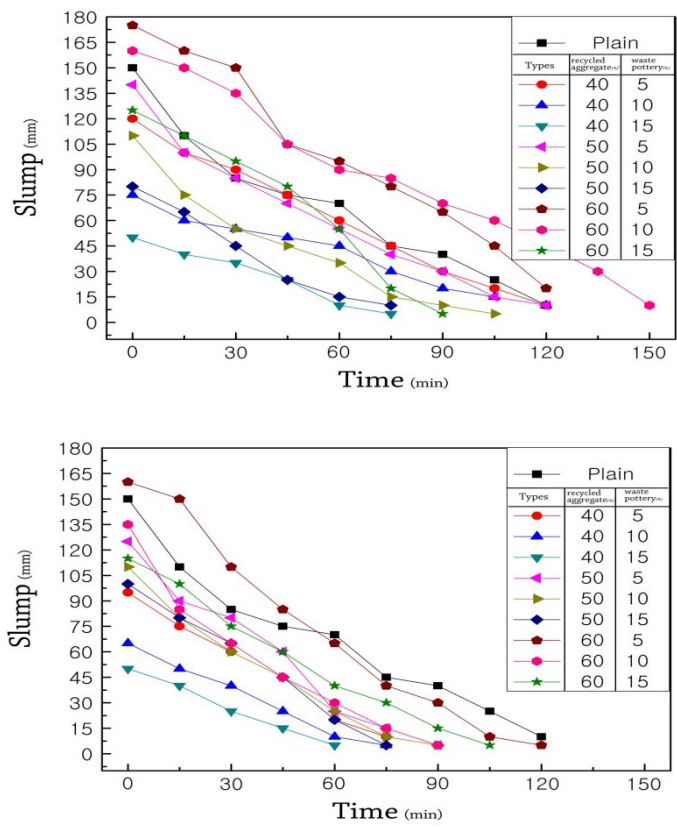


Fig. 2. Slump and slump loss of waste pottery ratio

3.1.3 Time of Setting of Concrete Mixtures by Penetration Resistance

Figure 3 shows the graph of time of setting of concrete mixtures by penetration resistance according to waste pottery blain fineness ratio.

At 40 and 60% recycled aggregates ratio with 4,000cm²/g degree of particle size, initial set happened when pass time was between 6 and 7 hours. It was 1 or 2 hours earlier compared to that of plain. At 50% recycled aggregates ratio, initial set appeared after 10 hours passed, which was 2 hours later than that of plain. In case of 6,000cm²/g waste pottery blain fineness ratio, at 60% recycled aggregates ratio, initial set appeared after 10 hours passed, which was 2 hours later than that of plain. Initial set happened 2 hours earlier compared to that of plain. Final set started 5 hours passed for plain. It shows similar trends in 4,000cm²/g and 6,000cm²/g degree of particle size. Final set started after 5.5 hours passed for both cases regardless of recycled aggregates ratio.

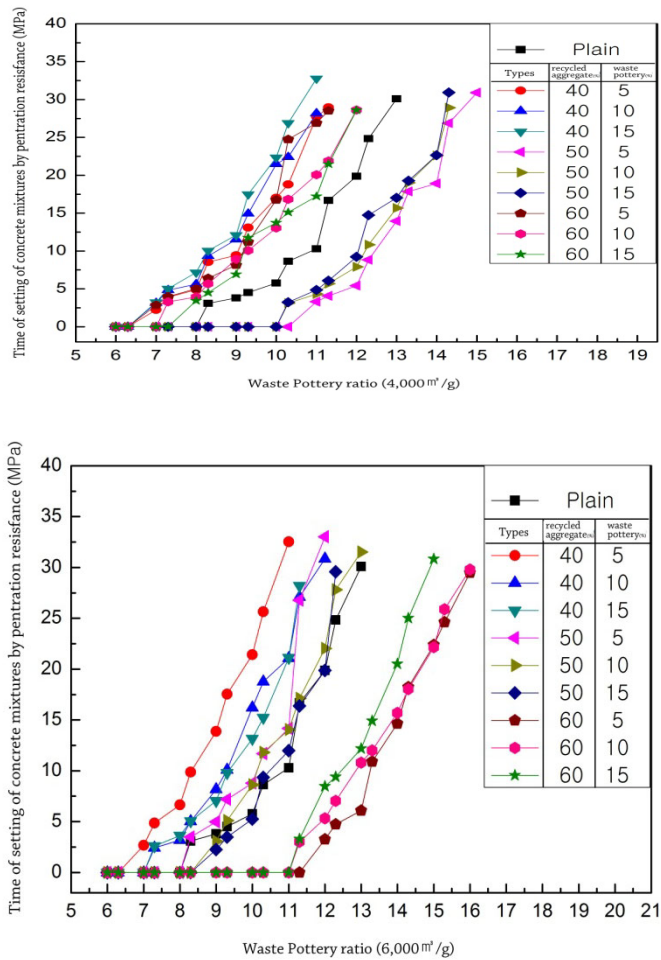


Fig. 3. Time of setting of concrete mixtures by penetration resistance of waste pottery ratio

3.1.4 Bleeding

Bleeding shows lower value every waste pottery blain fineness ratio, however, higher bleeding was observed at 60% and 5% recycled aggregates ratio compared to that of plain.

3.2 Hardening Concrete

3.2.1 Compressive Strength and Tensile Strength

Compressive strength was measured according to waste pottery blain fineness ratio as shown in Figure 4.

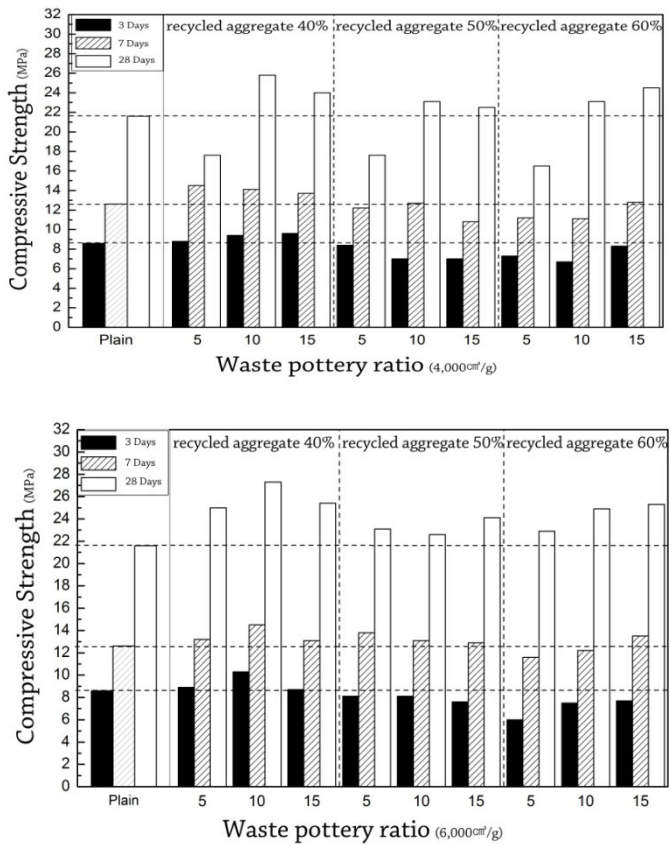


Fig. 4. Compressive Strength of waste pottery ratio

In case of 4,000cm²/g waste pottery blain fineness ratio with 3 and 7 days early age, a little higher strength performance appeared at 40% recycled aggregates ratio with 5, 10, and 15% waste pottery blain fineness ratio. But 50% and 60% recycled aggregates ratio, strength performance was smaller than that of plain. At 28 days standard age, every recycled aggregates ratio except 5% waste pottery blain fineness ratio shows higher strength performance compared to 21.6MPa for plain. 28.5MPa highest strength performance was observed at 40% recycled aggregates ratio and 10% waste pottery blain fineness ratio. In case of 6000cm²/g degree of particle size, higher strength performance was measured compared to that of plain at 40% recycled aggregates ratio for 3 early age and at 40% and 50% recycled aggregates ration for 7 days early age. At 28days standard age, higher strength performance was observed mostly compared to that of plain. 27.3MPa highest strength performance was observed at 40% recycled aggregates ratio and 10% waste pottery blain fineness ratio. Compressive and tensile strength shows a similar trend.

3.2.2 Changing Length of Drying Shrinkage

Figure 5 shows the changing length of drying shrinkage according to waste pottery blain fineness ratio.

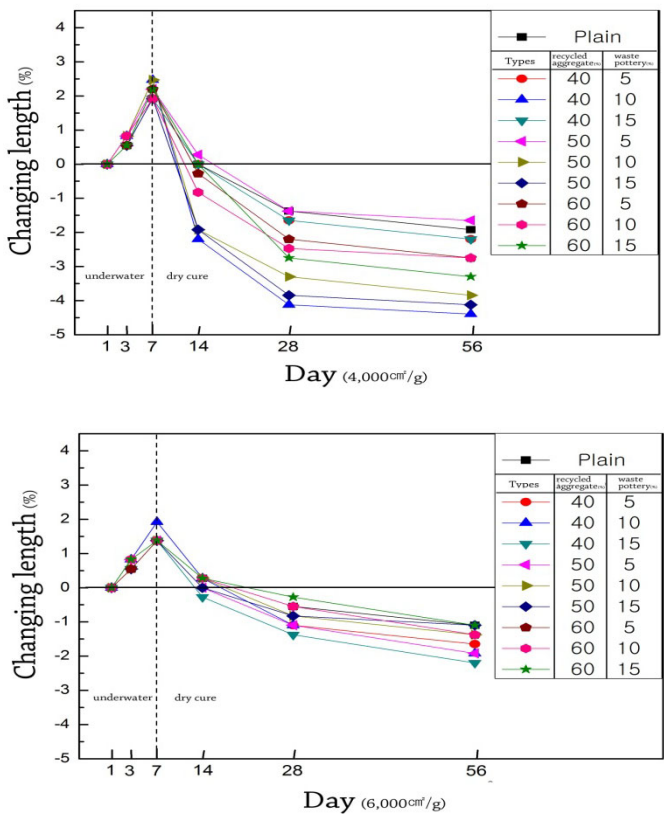


Fig. 5. Changing length of waste pottery ratio

First of all, it mostly expanded due to underwater curing until 7 days age for 4000cm²/g degree of particle size. In addition, highest change length was observed at 40% recycled aggregates ratio with 10% waste pottery blain fineness ratio on 7 days age. Between 7days and 14days age, the length was dramatically shrunk at 40% recycled aggregates ratio with 10% waste pottery blain fineness ratio. However, it was shrunk gradually because of evaporation of moisture after 14 days age. It shows similar trend with 4000cm²/g degree of particle size until 7 days age for 4000cm²/g degree of particle size. Between 7days and 14days age, the highest length was shrunk at 40% recycled aggregates ratio with 15% waste pottery blain fineness ratio. However, it was shrunk gradually after 14 days age. In addition, the change length was -0.66mm at 4000cm²/g degree of particle size and -0.22mm at 6,000cm²/g degree of particle size.

3.2.3 Absorbed Amount

The absorbed amount at 4,000cm²/g degree of particle size with 60% recycled aggregates ratio was higher than that of plain. However, the absorbed amount was similar or lower than that of plain at 40% and 60%. The absorbed amount at 6000cm²/g was higher than that of plain for every recycled aggregates ratio.

3.2.4 Adiabatic Temperature Rise of Concrete

Adiabatic temperature rise of concrete was observed according to waste pottery blain fineness ratio as shown in Figure 6.

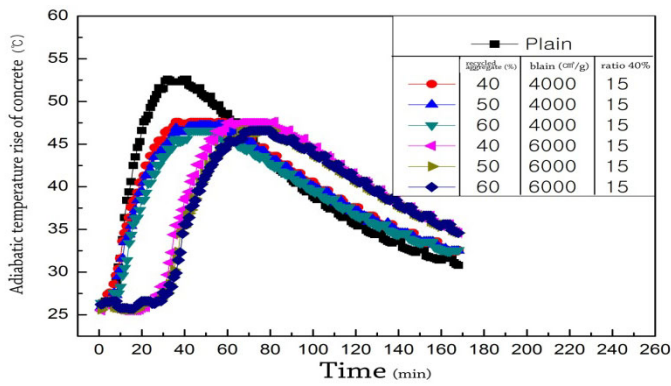


Fig. 6. Adiabatic temperature rise of concrete of waste pottery ratio

Highest temperature change with 4000cm²/g degree of particle size was 26.5°C on plain right after concrete was driven. Second highest temperature change was 26.2°C at 40% recycled aggregates ratio and 15% waste pottery blain fineness ratio. The temperature change at 50% and 60% recycled aggregate ratio was about 25°C, which is 1~1.5°C lower than that of plain. Highest temperature rise was 53.6°C at plain, 47.5°C at 40%, 46.7°C at 50% and 45.9°C at 60% recycled aggregates ratio. The highest temperature rise at recycled aggregates ration was 6~8°C less than that of plain. Temperature change with 6000cm²/g degree of particle size was less than that of plain regardless of recycled aggregates ratio. At the highest temperature rise, 8~10°C lower adiabatic temperature was generally observed compared to that of plain.

4 Conclusion

Flowability of fresh concrete tends to increase as recycled aggregates ration increases and it tends to decrease as Waste pottery blain fineness ration increases. It is revealed that air content is satisfied with the limited range of KS standard. Unit volume weight decreases as recycled aggregates ration increases and it increases as degree of particle size increases. Showed a slightly faster setting time of all initial set and final set, Bleeding tends to decrease.

Compressive strength of hardening concrete increases as Recycled aggregates ratio and waste pottery blain fineness ration increase. Tensile strength tends to act similar to compressive strength and highest strength development at 40% recycled aggregates ratio with standard age.

Length of drying shrinkage is gradually changed as waste pottery power increase. In addition, Adiabatic temperature rise of concrete show 6~10°C less value from highest temperature as recycled aggregates ratio and waste pottery power increase. Thus it gives the effect of crack reduction due to heat of hydration.

Absorbed amount tends to increase as recycled aggregates ratio and waste pottery powder increase.

References

1. 2004 Present condition of production of waste and handling, Ministry of Environment (2005)
2. Korea concrete Institute, The newest concrete engineering, Gimoondang (1997)
3. Korean Standard Associations, KS Standard
4. Ministry of Environment, Basic recycling plan of construction waste, pp. 11–28 (December 2006)
5. Kim, Y.-R., Jung, Y.-H., Lee, D.-B., Khil, B.-S., Yoon, K.-H., Han, S.-G.: An Experimental Study on the Development of Low Heat Concrete Using Hydration Heat Reducing Agent based Latent Heat. Architectural Institute of Korea, 345–348 (2007)
6. Kim, J., Jeon, C.-K., Shin, D.-A., Yoon, G.-W., Oh, S.-K., Han, C.-G.: Temperature History of Mock-up Mass Concrete Considering Different Heat Generation Due to Mixture Adjustment. The Korea Institute of Building Construction 5(1) (2005)

Design and Analysis of Optimizing Single-Element Microstrip Patch Antenna for Dual-Band Operations

Dong-Hee Park^{1,*} and Yoonsik Kwak^{2,*}

¹ Professor, Chungju National University, Information & communication Engineering,
#123, Kumdan-ri, Chungju, Korea, +82-43-841-5364

dhpark@cjnu.ac.kr

² Professor, Chungju National University, Computer Engineering, #123,
Kumdan-ri, Chungju, Korea, +82-43-841-5345

yskwak@cjnu.ac.kr

Abstract. A novel frequency selectable microstrip patch antennas for dual-band WLAN is presented. The proposed design is based on the shorting pin antennas, which have ten shorting pins at a distance from the radiating edges on patch. The designed antennas can perform in IEEE 802.11.a/b and be adopted for a near region RF wireless network systems. As a result, this paper presents possibility for the design of optimizing to select the required frequency as controlling the number of shorting pins.

Keywords: Microstrip antenna, dual band antenna, multi-band antenna, frequency selective antenna, shorting pin antenna.

1 Introduction

In the last few years, the rapid progress of Ubiquitous Sensor Networks(USN), such as Wireless Local Area Networks(WLAN), has greatly increased the development of wireless communication systems. A small size and multi-band antennas are required as one of the important factor in mobile communication terminals(MCT). The microstrip patch antennas(MPA) are widely being used for its low volume and thin profile characteristics.

Therefore, these MCTs are being researched and developed to use more than two communication service band in one terminal. Especially, the literatures [1], [2] and [3] can not be optimized in terms of the size of slot antennas that are designed to be applied throughout the 2.4/5.2/5.8 GHz band. In literature [4], the slot structure and the size of ground plane 40 mm ×40 mm was not able to optimize.

In the previous studies pertaining to the IEEE 802.11.a/b standard operating frequency 2.4 GHz and 5 GHz, various types of antennas have been developed in the personal wireless communication systems. Then the frequency ranges are used 2.4 GHz ~ 2.4835 GHz frequency band at the IEEE 802.11.b and 5.725 GHz ~ 5.825 GHz frequency band at the IEEE 802.11a.

* Corresponding author.

Recently, various dual-band antenna structures have been proposed due to dual T-junction stub using strip line feeder on FR4 substrate and wide-band circular polarization microstrip antenna [5], [6]. And a dual-band Solar - Slot antenna using the Si substrate is presented for the 2.4/5.2 GHz WLAN applications [7].

In this study, we propose shorting pin patch antennas supporting in dual-band (2.4 GHz and 5.8 GHz) WLAN operations using Taconic TLC substrates. We use ten shorting pins at arbitrary positions from radiating edges in patch in order to select the frequencies of 2.4 GHz and/or 5.8 GHz.

This paper is organized as following. Section II presents a simple description of antenna configuration and the antenna design. Then, in Section III, the simulation results are described. Finally, the entire work is summarized in Section IV.

2 Antenna Design

In this paper, we have design and analysis a single microstrip patch antenna with dual-band operations applicable in the 2.4 GHz and 5.8 GHz to IEEE 802.11a/b systems. The type of substrate we have investigated dual-band antenna design is Taconic TLC.

A schematic diagram of the microstrip patch antenna is indicated in Figure 1. The substrates has thickness $h_1=0.79$ mm and $h_1=1.14$ mm. Also dielectric permittivity are $\epsilon_r = 3.0$ and $\epsilon_r = 3.2$, and loss tangent $\tan\delta = 0.003$. Also, the metallization layers were realized using $t=0.018$ mm and $t=0.035$ mm at each substrate thicknesses.

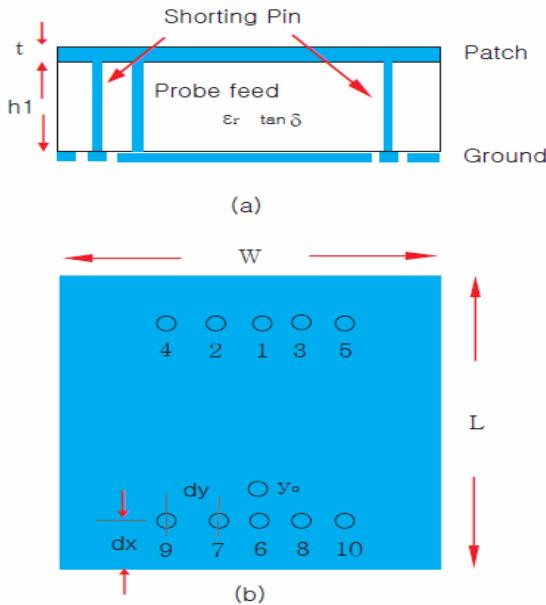


Fig. 1. Microstrip patch antenna with ten shorting pins from the radiating edges (a) Side view with shorting pins, (b) Top view of coaxial feed patch with shorting pins

Fig. 1. Shows one of optimized patch antennas for multiband operation. The structure consists of a patch antenna with ten shorting pins and one coaxial feed. In order to select a specific frequency of dual-band, we are used shorting pins located symmetrically with respect to the width of the patch from radiating edges.

A radius of shorting pins is r_s , shorting pins are located with a distance dx from the radiating edge, and the distance between of pins are dy . The feed point with a $50\ [\Omega]$ coaxial line is located in the central line of the patch with a distance of y_o from the radiating edge. For analysis suggested dual-band antenna, we are used the Designer tool.

3 Simulation Results

We are designed the dual-band patch antennas in order to depend on the final design process. The first step is to calculate the dimensions of the dual-band patch antenna at the operating frequencies 2.4 GHz and 5.8 GHz using relative dielectric constant $\epsilon_r = 3.0$ with a thickness of $h_1 = 0.79$ mm.

The dimensions of patch and ground are $35.5\text{ mm} \times 55\text{ mm}$ and $71\text{ mm} \times 110\text{ mm}$, respectively. The position of the feeding point y_o is 11.5 mm. The computed input return loss of this antenna is shown in data 1 of Fig. 2. The second step is designed to have ten shorting pins within the patch for a variable frequency band. The dimensions of the patch and ground are $36\text{ mm} \times 32.5\text{ mm}$ and $72\text{ mm} \times 65\text{ mm}$ respectively. The feeding point y_o is 11 mm. The all shorting pins are loaded at the $dx = 7\text{ mm}$ from the radiating edge and the distance between of pins are $dy = 6\text{ mm}$. Then the all shorting

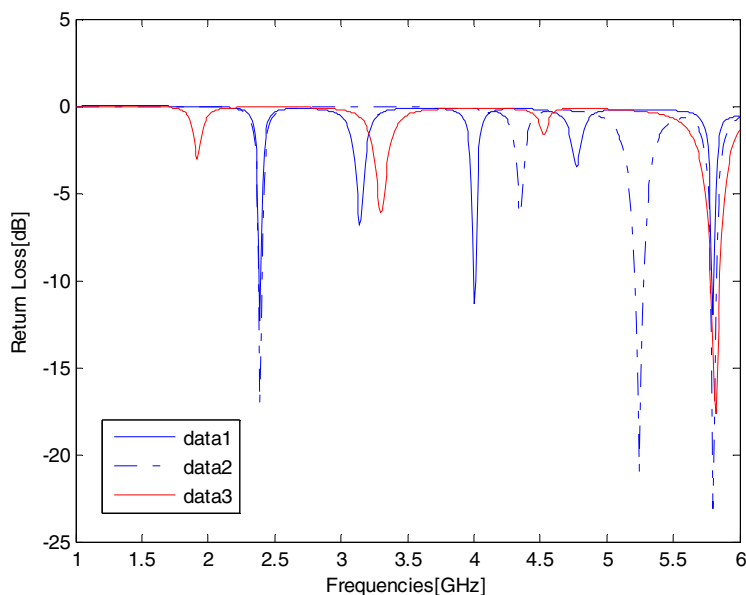


Fig. 2. The return loss of single patch antennas with central frequencies 2.4 GHz and 5.8 GHz in dielectric thickness $h=0.79$ mm

pins are opened on the ground plane to choose both bands 2.4 GHz and 5.8 GHz. The computed input return loss of the shorting pins patch antenna is shown in data 2 of Fig. 2. The third step is grounded between the conducting patch and the ground plane as the numbers of shorting pin 1, 9, and 10 in order to select 5.8 GHz only. Then the return loss of the antenna is shown in data 3 of Fig. 2.

Next, we consider the second modification, viz., the introduction of the shorting pins at arbitrary positions from radiating edges. Generally, the shorting pins have been employed for the purpose of reducing the size of the radiating structure. However we find that they also help to select the single frequency band between two frequency bands. Therefore we are shown how the dx value can be chosen. Fig. 3 shows the variation of return loss with frequency for different dx when the dy is 6 mm. Here the dy is the distance between of pins. From these results we are known that the central frequencies of the first frequency band are moving to 2.4 GHz as the dx gradually increases from 3 mm to 7 mm.

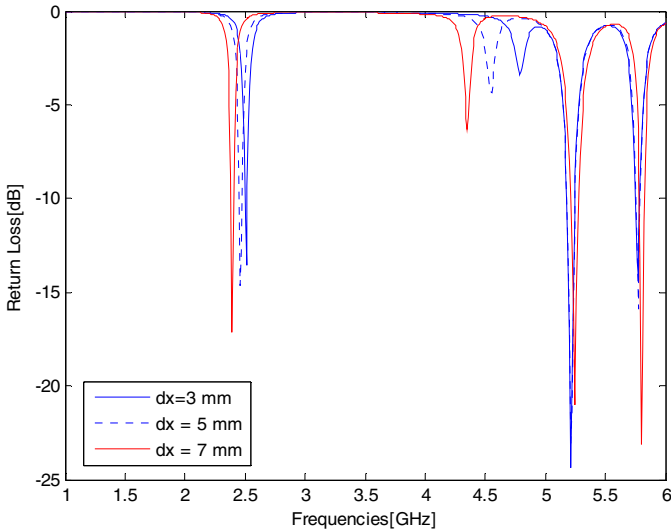


Fig. 3. Variation of return loss with frequency for different dx when the dy is 6 mm

Also, we are shown how the dy value can be chosen. Fig. 4 shows the variation of return loss with frequency for different dy when the dx is 7 mm. Here the dx is the distance at arbitrary positions from radiating edges in patch. From these results we are known that the central frequencies of the second frequency band are moving to 5.8 GHz.

In this study, another dielectric substrate is used having a relative dielectric constant of $\epsilon_r = 3.2$ with a thickness of $h_1 = 1.14$ mm. The dimensions of patch and ground are 35.5 mm 57 mm and 71 mm 114 mm, respectively. The position of the feeding point y_0 is 9.5 mm. The computed input return loss of this antenna is shown in data 1 of Fig. 5. Next we are designed to have ten shorting pins within the

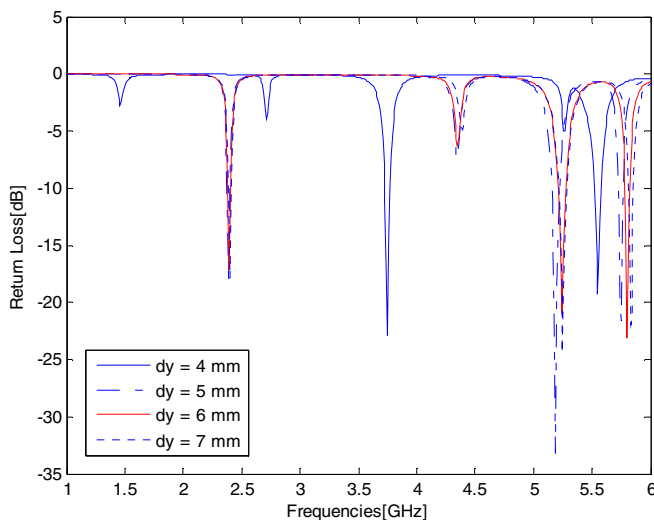


Fig. 4. Variation of return loss with frequency for different dy when the dx is 7 mm

patch for a variable frequency band. The dimensions of patch and ground are 35.8 mm 36.5 mm and 71.6 mm 73 mm, respectively. The position of the feeding point y_0 is 10.5 mm. The position of shorting pins are loaded at the $dx = 5.2$ mm from the radiating edges and the distance between of pins are $dy = 3.5$ mm.

And we are grounded between the conducting patch and the ground plane as the numbers of the shorting pin 1, 2, 4, 6, 8, and 10 to choose both 2.4 GHz and 5.8 GHz. The computed input return loss of the antenna is shown in data 2 of Fig. 5. In order to

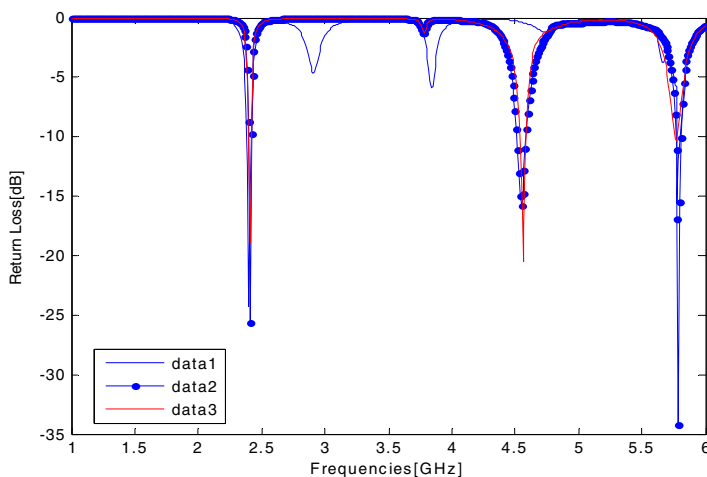


Fig. 5. The return loss of single patch antennas with central frequencies 2.4 GHz and 5.8 GHz in dielectric thickness $h=1.14$ mm

select 2.4 GHz, we are shorted to the ground plane the shorting pins 6, 7, 8, 9, and 10. The return loss of the antenna is shown in data 3 of Fig. 5.

From above analysis results, we have designed patch antennas with dual-band characteristics. Also, we can be verified that a central frequencies are changed add ten shorting pins in the patch. As a result, this paper proposes enable design of dual-band patch antenna using optimized single patch for near RF network.

4 Conclusions

In this paper, we are designed and analyzed the dual-band patch antenna pertaining to the IEEE 802.11.a/b standard operating frequency 2.4 GHz and 5.8 GHz. The future of researches of this paper is continuously required on the design of multi-band frequency selectable antennas for near RF Network using RF cognitive techniques.

As a result, this paper proposes possibility for the design of optimizing to select the required frequency as controlling the number of shorting pins in the single patch for near RF network.

References

1. Su, C.M., Chen, H.T., Chang, F.S., Wong, K.L.: Dual-band slot antenna for 2.4/5.2 GHz WLAN operations. *Microwave and Optical Technology Letters* 35, 306–308 (2002)
2. Wu, J.W.: 2.4/5-GHz dual-band triangular slot antenna with compact operation. *Microwave and Optical Technology Letters* 45, 81–84 (2005)
3. Wu, Y.J., Sun, B.H., Li, J.F., Liu, Q.Z.: Triple-band omni-directional antenna for WLAN application. *Progress In Electromagnetics Research, PIER* 74, 21–38 (2007)
4. Ren, W.: Compact Dual-band Slot Antenna for 2.4/5 GHz WLAN Applications. *Progress in Electromagnetics Research B* 8, 319–327 (2008)
5. Lin, Y.-C., Hung, K.-J.: Design of dual-band slot antenna with double T-match stubs. *Electronics Letters* 42(8) (April 13, 2006)
6. Boissbouvier, N., Le. Bolzer, F., Louzir, A.: A compact radiation pattern diversity antenna for WLAN applications. In: *IEEE AP-S Int. Symp. Dig.*, vol. 4, pp. 64–67 (2002)
7. Shynu, S.V., Roo Ons, M.J., Ammann, M.J., Norton, B.: Dual band a-Si: H solar-slot antenna for 2.4/5.2 GHz WLAN applications. *Radioengineering* 18(4), 354–358 (2009)

An Analysis of U-Healthcare Business Models and Business Strategies: Focused on Life Insurance Industry

Dong-Hwan Cho¹ and Jin-Hyuk Hong²

¹ Assistant Professor, Department of Venture and Business,
College of Business and Economics,
Gyeongnam National University of Science and Technology,
150, Chilarm Dong, Jinju City, Gyeongsangnamdo, Republic of Korea

² Part-Time Instructor, Department of Venture and Business,
College of Business and Economics,
Gyeongnam National University of Science and Technology,
150, Chilarm Dong, Jinju City, Gyeongsangnamdo, Republic of Korea

Abstract. Recently major developed countries such as U.S. and EU selected U-Health as future growth engine industry, and concentrated their capabilities. Major companies all around the world also entered this U-Health business. Most of the researchers in this field confined their research focus mainly on the elementary technologies for infrastructure buildings. This research systematically investigates U-Health business models with various objectives in diverse sub-areas. For this purpose, the value chain of U-Health industry is analyzed, and then various business models were explored all around the world. Finally, the future changes possible to occur in the life insurance industry were analyzed.

Keywords: U-Health, U-Healthcare, Business Models, Value Chain, Life Insurance.

1 Introduction

The world medical market has recently seen enormous growths in the side of technology and market by focus on both doctor and medical institution, and has been rapidly changing the 21 Century. The U-Health is in the immediate middle of the changes. The U-Health is defined that as the provider that link the sector of information community and Health Care, which enables among the prevention, diagnosis, treatment, and aftercare at anytime, anywhere. The disease symptoms of previous patient and layman promote health by the mitigation and treatment, while the trend is to change is to change and extend to disease prevention [1][2].

The Global U-Health industry is regarded as a new business area due to maximization in the U-Health, welfare and accommodation, and to provide both health care and disease prevention. For example, the sphere of IT(Information Technology) infrastructure and Bio-Sensor are the Internet, IPTV(Internet Protocol Television), and WiBro(Wireless Broadband), BT(Bio Technology) and NT(Nano Technology) are also new technologies in the field of high-tech convergence. Under the circumstance into the IT based convergence strategy will expect to choose in the

national alternative, that will solves to discuss in the social issue by the energy problem, the aging society, the traffic problem, the medical service system. Eventually, the U-Health industry will extend to grow in the IT medical convergence, and will quickly turn into the high value industry. It is necessary that U-Health industry requires fostering in the leading-driven policy [3].

Recently major developed countries such as U.S., EU and Japan around of the world push for the advancement of medical industry through IT convergence, and the Korea government also strongly promotes it.

The U-Health research gets accomplished by the Korea academia(e.g. the internal or the external of academic world) mainly has been proceeding about the foundation of U-Health environment for factor technology analysis[7][8][9], the definition and evaluation of conceptual services[10][11][12], the U-Health applied system and application development[13][14][15]. Such as this research, the whole U-Health industry does not provide the effective explanation about how to improve the business performance. Accordingly, this research systematically investigates U-Health business model with various objectives in a diversity of the sub-fields. Through this research is expected to promote in developing the field of U-Health.

This paper has concrete purpose for research as follows. First, the various U-Health Business Model cases are explored all around the world. To achieve the goals, the value chain of U-Health industry is analyzed, and then the Business Model Analysis is done. Second, the future changes predicted to occur in the specific industries (e.g. the life insurance industries) are analyzed.

2 Analysis of U-Health Industry Structure

2.1 Understanding of the U-Health Industry

The paradigm shift of U-Health Care Market is changing through improving the level of consciousness and developing the information technology. Focused on the hospital of independent healthcare, passed in the shared healthcare by the hospital and the home, the medical services are available to go into the ubiquitous healthcare age [1][2]. The U-Health services and various aspect of the existing healthcare are different. Above all, the service main front is focused that the U-Health is a patient-centered as compare with a medical practice by the existing medical services, the form front is emphasized that the U-Health is a precaution compared with a post-treatment by the existing medical services.

The scale of U-Health Market is expected that it will be increased from a mere 1.68 trillion won in 2010 to 3.03 trillion won in 2014 by the South Korea [16]. Especially, the growth ratio of U-Health industry will likely continue to rise up above 14% since 2014, and the percentage of employment will be expected to create above 39 million people. The world market is also expected that it will be increased from a mere 10 billion dollar in 2004 to 340 billion dollar [17].

2.2 Analysis of the Industry Structure

The value chain was defined to analyze the specific activities through which firms can create competitive advantages, and it is useful to model the firm as a chain of

value-creation activities. This model was first described by Michael M. Porter. The value-chain concept has been extended beyond individual firms. It can apply to whole cost and differentiation[18].

The value chain of U-Health industry is greatly composed of the health care provider, the health care promoting firm, the health solution provider, the service provider, the network provider, the measuring machine, the terminal manufacturer, the customer, as shown in figure 1[1][2][3].

To minutely analyze the value chain of role and activity, firstly is considered as the medical care provider and the health care promoting firm. The health care provider is in the hospital, the pharmacy, the public health center, and is built in the networking with among the health care provider, health Support provider, and the user. It is also provided to achieve the prevention, the promotion of health, the treatment, the aftercare for the health care recipient [6]. In case of the health care promoting company, it is means to provide in the gymnasium and Health Club as like to promote both the health and fitness, and it offers the health information for the health care provider and the user through both the high-technology of exercise equipment with attaching in sensor and networking. So, the health care provider and health care promoting firm is presently finished for the U-Health service development, and is going on the pilot project.

The medical solution providers carry out both the health care provider and health care promoting firm, that it is developed the medical information solution and the equipment including the EDI(Electronic data interchange), the OCS(Ordering Communication System), the PACS(Picture Archiving and Communication System) by EMR(Electronic Medical Record), ERP(Enterprise Resource planning), DW(Data Warehouse), CRM(Customer Relationship Management) and so on. The solution development and the equipment manufacturer spontaneously take part in conduct of medical business field. The medical solution provider continuously were developed the solution for offering in the direct service and promoting in the user convenience.

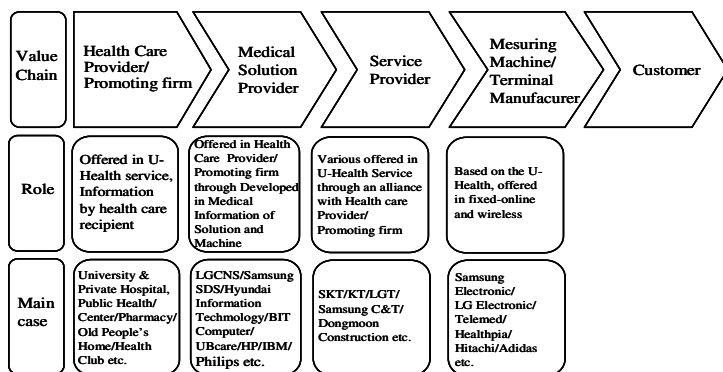


Fig. 1. The value chain of U-Health industry

The service providers have cooperated with the health care provider and promoting firms in order to offer a variety of U-Health services[3]. The communication provider serves in the fixed-online and wireless network among the health care provider, the healthcare promoting company, and the users. Based on this healthcare institution cooperate with the communication provider, and it takes a role to develop efficient and user-oriented business model. With home-network industry is create in the field of complex industry to service as the several information technology by the network and data processing with the ultrahigh speed of infrastructure, includes not the communication provider as KT(Korea Telecom) or SKT but the construction provider as Samsung C&T or Dongmoon Construction. As the home-network provider is the killer application of home-network service, where observe as the U-Health service and build in the linking of home-network by the U-Health, the field of health care is created in the new business domain that it is developed the new service model.

3 Analysis of U-Health Business Models in Korea

3.1 Concepts of Business Models and Research Methods

The business model is generally explained to compose in the main factor of a firm's business, or is used to describe in the specific business [19]. The various field of e-business research is defined that it mainly consist of the system and architecture by the business model and it is fundamentally considered for the information system and the business process [20].

Recently, U-business is defined as the ubiquitous environment in this research[20][22], this U-business is the intellectualization by the customer of business environment, and U-business is the business system by the network. However, the various parts of U-business model are achieved not the user-base approach but technique-base approach, and the requirement of user as do not have reflected on that the services suggested that it gets down the acceptable user and it impedes the diffusion of U-business.

Meanwhile, the U-business is considered as the expansion of e-business rather than the new concept of e-business in the service model [20][22]. Therefore, this research aims to analyze U-health business model based on the methodology of e-business plan suggested by Rayport and Jaworski in University of Harvard [23]. According to their methodology, the systematic e-business model analysis has four steps. As follows—1) the business opportunities analysis through the existing of value system analysis → 2) the business model investigation → 3) the service investigation of business killer → 4) the investigation step of resource system through the business killer service. The researches follow to analyze about the business opportunities analysis through the existing of value system analysis by chapter 2(The U-Health Industry Structure Analysis). The next step of business model investigation and the service investigation of business killer by chapter 3(The Korea Business Model Business) is analyzed, in the last step for the investigation of resource system through the business killer service mainly argue about the Life Insurance Industry by Chapter 4 (The Change of Life Insurance Industry).

Table 1. Type of U-Health Projects

Classification	Exposition	Project Type
Medical Information system	Information and IT infrastructure	• Assets Management System(RFID) • Patient Information System • Medical Telematic • E-Prescription • Remote-EDI Service
Remote-Medical Service	Remote-medical behavior	• Virtual Consultation • Reservation Agent System • Medical Smart Card Service • Mobile Care Service • Infrared Rays Emergence Rescue Service
Health Care Service	Medicine/disease /health care information	• Health Information Insurance System • On-Line Portal Service • On-Line fitness Service • Mobile Health Care Service • Older/Youth Care Service

This research shows the framework of research analysis in the explored method through the expert's interview with the literature review by the field of U-Health, and this study investigates or identifies the detail item, and is supposed to relax by the expert's subjective decision[24]. According to the diffusion of U-Health is composed by the data of interview survey in relation to the change of life insurance industry, for this purpose of the preference data is framed to base on the literature review that it is requested the large-sized Korea Insurance Company(case by the KL firm) of the Business Plan Manager/IT Plan Manager(2-Person) and the Korea IT company(case by the H firm) of the U-Health Manager(1-Person), and the preference data preferentially verified the content validity.

Next, this study has tired the anti-structured in-depth interview. The interviewees have interviews with the large-sized Korean insurance companies(cases by the S, K, KL firm) of the Business Plan Manager/IT Plan Manager(9-Person), the Korea IT company(case by the S, L, H firm) of the U-Health Manager(3-Person), and the Health Care company(case by the B, I, U firm) of the Consultant(3-Person) from the November 2008 to the January 2009.

Finally, this research suggests improving the reliability and appropriation of research consequence, and the expert's group is enforced to exchange the opinion collections. The opinion collectors take part in the 15 persons in all, and the period of in-depth interview was distributed in the priority interview statement in the e-mail by from the 2th to the 20th in the February 2009, was enforced in the additional interview.

3.2 Classifications of U-Health Business Models

Based on methodology and preceding research([4][6]) of Rayport and Jaworski[23], the U-Health Business Model is systematically divide into the recent research shown in table 2. It is classified the character and technology user object in U-Health service.

The U-Health Business Model is classified the medical information service, the remote-medical service, the health care service.

Firstly, the Medical Information System used in the remote service offer and continuous disease control by related in the health care institution information and infrastructure system.

Secondly, the Remote-Medical Service is means that it is disciplined to remote the medical behavior. The remote medical treatment will increase to use through the networking system between the university hospital and the private hospital, and also will offer the moving-type hospital service in the alienate location.

Thirdly, the Health Care Service Model is to service the personal health care and health promotion in the house and moving location. This service offers the medicine information and the disease information, and the private health information.

4 Future Changes of Life Insurance Industry Caused by U-Health

Fundamentally the Life Insurance related in the health care also estimated the fundamental changes, and this part of research deeply is not enough in the actual circumstances. The Life Insurance Industry lately takes an effect on the influence of u-health industry and pervasive effect compare with other industry, and to consider the processing of u-health industry deeply analyzes the future changes of life insurance industry.

The future changes of life insurance industry are the changes and expansion of commodity concept, the changes of insurance operation management as like the investigation and the assessment of insurance ratio, and the changes of insurance distribution channel etc[1][2].

4.1 Change of Life Insurance Concepts

In this source from the propriety of U-Health Industry, the existing life insurance industry is changed the concept of guarantee insurance, and the existing insurance is expected to vanish the uppermost limit. On the other hand, the concept of insurance by hereafter U-Health is expended to include in the disease prevention and health promotion, the aftercare. The total care offers through the whole course of health medical a matter possible, be embossed the positive aspect of insurance.

This extension of insurance concept is changed the wellness aspect, worst of all it will be move into the healthy life through the total care. Hence, the existing of negative aspect is alleviated or vanished; the new concept insurance is embossed. For example, the insurance member will possibly design more benefit the insurance product.

4.2 Change of Internal Operation Management

These days, to increase the aging and chronic disease have required the necessity for utilizing health information. The aging and chronic disease leads to expand markets on care insurance. For this reason, private information is gradually necessary for evaluating insurance. The life insurance and annuity insurance as like the private insurance, the damage insurance as like the auto insurance and general insurance, and

the third part of insurance as like the accident, disease, care insurance that it is raise to use in the health information. The revitalization of U-Health used to the private of realistic health information by the insurance operation management.

Presently, in middle of the distinction of sex and the age by the way of insurance assessment is changed the personal health condition. It is possible to offer in the medical insurance of maintenance gear type. This point is possible to provide the risk fractionation type insurance. The insurance firm position is a positive aspect by the promoting health, the other side the unhealthy person is a negative aspect by the left out of the group.

4.3 Change of Distribution Channel

The changes of insurance distribution channel reduce the power of channel by insurance agents, but the power of online channel will be increased. The most insurance firm of main channel is the insurance planner channel. The ratio of insurance planner channel (The large-sized of firm by the Samsung insurance company, Deahan insurance company, Kyobo insurance company) is 71% in 2006. Recently insurance firms have increased the new distribution channels—the Bankasurance, the telemarketing, and the GA(General Agency). If the insurance join customer will have provided in the U-Health Service, the existing insurance planner quickly will have extended the influence of other distribution channel.

5 Conclusions

This research systematically investigates u-health business model with various objectives in various sub-field. Also, this study analyzed the value chain of u-health industry by the various related industry and the stakeholder. The value chain of U-Health industry is greatly composed of the health care provider, the health care promoting firm, the health solution provider, the service provider, the network provider, the measuring machine, the terminal manufacturer, the customer, and analyzed the activity and the role. The various business model cases systematically are explored over the world in Korea. The business model is classified the medical information service, the remote medical, the health care service. Finally, the future changes possible to occur in the life insurance industries are analyzed. The insurance industry will be change or extend the business concept by the influence of u-health, will improve the centered-customer with the insurance operation management, and also will change the distribution channel by the insurance company.

References

1. Cho, D.H.: The U-Health Business Model and Business Strategy. The Korea Society of Management Information System Conference Call for Papers (Spring 2007)
2. Cho, D.H.: The U-Health Business Model Case Study. Korean Academic Society of Business Administration Conference Call for Papers (2009)
3. Telecommunications Technology Association: The Health Forum Final Research Report 2009, pp. 1–64 (December 2009)

4. Samsung Economic Research Institute: The Advent of U-Health Era, (May 2007)
5. Kim, O.N.: The U-Health is Coming On Us, LG Business Insight, LG Economic Research Institute, pp. 23–41 (August 2009)
6. Kim, S.H.: The Biomedical Signal Monitor Technology for U-Health. *Information and Communications Magazine* 26(8), 3–7 (2009)
7. Kim, J.Y., Kim, Y.H., Ahn, K.S.: An Adaptive Middleware for U-Healthcare. *The Korean Institute of Information Scientist and Engineers Conference Call for Papers* 34(2-B), 291–295 (2007)
8. Choi, E.J., Hwang, H.J.: Multiple User and Service Management Architecture for Medical Gateway. In: *Korean Society for Internet Information Conference Call for Papers*, pp. 315–319 (Spring 2009)
9. Ahn, S.Y., Lee, T.Y., Kim, D.W., Seong, Y.R., Oh, H.R., Park, J.S.: An Implementation of a U-Health Service Space Based on Sensor Network. *The Journal of Korean Information And Communication Society* 35(2), 225–231 (2010)
10. Yu, J.K., Han, J.H., Kim, P.G., Nam, J.H., Jung, J.Y., Yee, Y.H., Seo, D.Y.: Design and Implementation of U-Health Care Service for Infertile Women. *Korea Computer Congress Call for Papers* 36(1-C), 268–273 (2009)
11. Park, M.J., Jung, M.H.: The Observation on Health Indexes of Visiting Health Management Before and After Access to U-Health Care. *Journal of The Korean Society of Living Environment System* 15(1), 42–50 (2008)
12. Kim, J.H., Park, J.S., Jung, E.Y., Park, D.K., Lee, Y.H.: A Diet Prescription System for U-Healthcare Personalized Services. *Journal of Korea Contents Association* 10(2), 111–119 (2010)
13. Song, S.Y., Hwang, H.J.: U-Healthcare Application Framework for Medical Gateway. *Korean Society for Internet Information Conference Call for Papers*, pp. 349–353 (Spring 2009)
14. Kim, J.Y., Oh, B.K.: A Study of Content Design for Healthcare on Mobile Phones for the Old Generation. *InfoDesign Issue* 18, 19–30 (2009)
15. Newsis: Ministry of Knowledge Economy, The Emphasis Upbringing of U-Health Industry... of the Medical Power Speeded Up (May 11, 2010)
16. Kim, S.H.: The U-Health actualization should not be any more delays, *KorMedi* (November 10, 2009)
17. Kim, K.S.: A Study on the Strategic Application of Hotel Information System. *Journal of Tourism Management Research* 9, 24–41 (2000)
18. Hedman, J., Kalling, T.: The Business Model Concept: Theoretical Underpinnings and Empirical Illustrations. *European Journal of Information Systems* 12(1), 49–59 (2003)
19. Hwang, K.T., Shin, B.S., Kim, K.J.: Ubiquitous Computing-Driven Business Models: An Analytical Structure and Empirical Validations. *Journal of Information Technology Application and Management* 12(4), 105–121 (2005)
20. Kim, K.K., Chang, H.B., Kim, H.G., Kwon, K.J.: A Study on the Business Service Design in Ubiquitous Computing: The Case Study in Bookstore. *The Journal of Korean Institute of CALS/EC* 3(2), 165–179 (2008)
21. Rayport, J., Jaworski, B.: *E-Commerce*. McGraw-Hill, New York (2001)
22. Shin, K.S., Suh, A.Y.: *Case Study Research*, Hankyungas, Seoul (2008)

Outage Probability Evaluation of DSF-Relaying Equipped CDD Schemes for MC-CDMA Systems^{*}

Kyunbyoung Ko¹ and Choongchae Woo²

¹ Dept. of Instrumentation and Control Engineering, Chungju National Univ.,
72 Daehangno, Chungju-city, Chungbuk, 380-702 Korea
Tel.: +82-43-841-5463; Fax : +82-43-841-5320
kbko@cjnu.ac.kr

² Dept. of Electronics and Computer Engineering, Hanseo Univ.,
360 Daegok-ri, Haemi-myun, Seosan-si, Chungnam, 356-706
Tel.: +82-41-660-1360
woo9@hanseo.ac.kr

Abstract. In this paper, CDD(Cyclic Delay Diversity)-DSF(decode-and-selectively forward) relay schemes are proposed for MC-CDMA(multicarrier-code division multiple access) systems over multipath Rayleigh fading channels. On the contrary to general DSF schemes, multiple relays equipped CDD scheme can transmit on single channel. The cooperative diversity of relay networks can be obtained as the frequency diversity for MC-CDMA systems with proposed CDD schemes. Furthermore, a semi-analytical approach for the outage probability is proposed for general DSF-Relay schemes and proposed CDD-DSF-Relay schemes. By simulation and numerical results, it is confirmed that the proposed one can be a possible solution to achieve cooperative diversity gain without a reduction of spectral efficiency.

Keywords: CDD, DSF-Relay scheme, MC-CDMA, multipath Rayleigh fading channels, outage probability, frequency diversity.

1 Introduction

Recently, cooperative diversity schemes have been widely discussed in wireless networks. Two main relaying protocols are usually used for cooperative diversity schemes: amplify-and-forward (AF) and decode-and-forward (DF) [1]. A third option is to have relays forward only those correctly decoded messages, which can be considered as decode-and-selectively forward (DSF) schemes [2][3]. Use of DSF schemes presumes incorporation of, e.g., cyclic redundancy check (CRC) codes from a higher layer in order to detect errors. The advantages of general cooperative diversity schemes come at the expense of the spectral efficiency since the source and all the relays must transmit on orthogonal channels (in this paper, different time slots)

^{*} This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2009-0072762, 2010-0002650).

in order to avoid interfering with each other as well [1]. Recent researches are carried out for best-relay selection scheme in which only two channels are need (one for the direct link and the other one for best relay link) [2-4]. However, this scheme needs to additional process or feedback information for the channel state [5]. In this paper, another scheme utilizing only two channels is proposed for multicarrier-code division multiple access (MC-CDMA) systems. It is the cyclic delay diversity (CDD)-DSF-Relay scheme. Note that the proposed one can be regarded as a solution not only to obtain cooperative diversity gain but also to maintain spectral efficiency even if the number of relays is increased.

At first, we focus on the DSF-Relay networks without CDD scheme in order to generally derive a semi-analytical method based on error-events at relay nodes. Then, it is modified to cover CDD-DSF-Relay schemes. Even if the proposed analytical method does not give a closed-form solution, it is confirmed that the proposed one can be used as a tool to verify effects of an erroneous transmission at a relay node on both the received SNR and the outage probability.

2 MC-CDMA DSF-Relay Systems

MC-CDMA DSF-Relay systems have with a source(S), a destination(D), and relay(R)s. The number of relays is R . Each relay node detects the received signal, and only if there is correct detection, then transmits a regenerated MC-CDMA signal on orthogonal channels [1]. At a destination node, a cooperative diversity gain can be achieved by combining signals form R-D links and S-D link.

2.1 MC-CDMA Signal Model

The user data bit is mapped into binary phase-shift keying (BPSK) symbol of $b_m = \pm 1$. The transmitter spreads b_m in the frequency domain by using the m th spreading code $\{c_{m,n}\}_{n=0}^{N-1}$ with $|c_{m,n}| = 1$. The same process is carried out for M symbols. After multiplexing, the signal is converted to the time domain using an IFFT device. Without loss of generality, we can consider a transmission of the 0th MC-CDMA symbol in time domain. Therefore, the discrete time domain signal can be written as $s(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} b_m c_{m,n} \exp(j2\pi nt/N)$ with $t \in \{0, 1, \dots, N-1\}$. Then, a guard period (GP) is added in the form of a cyclic prefix. After a digital to analog conversion, it is transmitted through transmit antenna.

When the number of multipath for the r th link is L^r , the l th path's channel gain is h_l^r with $l \in \{0, 1, \dots, L^r-1\}$. It is assumed for $\{h_l^r\}$ that the magnitudes are Rayleigh-distributed and the phases are uniformly distributed over $[0, 2\pi)$ [6]. The exponential decay factor of multipath intensity profile is $1/L^r$. Then, we can obtain the channel response of the n th subcarrier for the r th link as

$$H_n^r = \sum_{l=0}^{L^r-1} h_l^r \exp(-j2\pi nl/N) \text{ where the normalized channel tap interval is } 1/N. \text{ Note}$$

that $\{H_n^0\}$, $\{H_n^{R+r}\}$, and $\{H_n^r\}$ means frequency domain channel gains of S-D link, r th S-R link, and r th R-D link, respectively.

Conventional MC-CDMA receivers operate using a process that is the reverse of the process for the transmitter. The discrete frequency domain signal for the i th subcarrier can be expressed as $y_i^r = b_0 H_i^r c_{0,i} + \sum_{m=1}^{M-1} b_m H_i^r c_{m,i} + n_i^r$ where $y_i^r|_{r=0} = y_i^0$ means the received signal of S-D link and y_i^{R+r} with $r \in (1, 2, \dots, R)$ is that of the r th S-R link. Also, n_i^r is a complex AWGN term with $E[n_i^r] = 0$ and $E[|n_i^r|^2] = \sigma^2$. Considering performance and complexity, we can use MMSE-C (Minimum Mean squared Error-per subCarrier) as the combining method [7]. Without loss of generality, the desired code is assumed to be the 0th code. Then, the combining weight can be written as $w_{0,i}^r = c_{0,i}^* H_i^{r*} \left[N \left(|H_i^r|^2 + \sigma^2/M \right) \right]^{-1}$.

2.2 Averaged BER for S-R Links

For the r th S-R link, the decision variable can be obtained as

$$v_0^{R+r} = \sum_{i=0}^{N-1} y_i^{R+r} w_{0,i}^{R+r} = \mu_0^{R+r} b_0 + \eta_0^{R+r} \quad (1)$$

$$\text{with } \mu_0^{R+r} = \sum_{i=0}^{N-1} \frac{|H_i^{R+r}|^2 / N}{|H_i^{R+r}|^2 + \sigma^2/M} \text{ and } \eta_0^{R+r} = \sum_{i=0}^{N-1} \left(\sum_{m=1}^{M-1} b_m c_{m,i} H_i^{R+r} + n_i^{R+r} \right) w_{0,i}^{R+r}.$$

Under Gaussian approximation for η_0^{R+r} , we can find that $E[\eta_0^{R+r}] = 0$ and

$$\text{Var}[\eta_0^{R+r}] = \sum_{m=1}^{M-1} \frac{1}{N^2} \left(\sum_{i=0}^{N-1} \frac{c_{m,i} c_{0,i} |H_i^{R+r}|^2}{|H_i^{R+r}|^2 + \sigma^2/M} \right)^2 + \frac{\sigma^2}{N^2} \sum_{i=0}^{N-1} \frac{|H_i^{R+r}|^2}{(|H_i^{R+r}|^2 + \sigma^2/M)^2}. \text{ Note that}$$

$N = M = 64$ (i.e., full-loaded system) leads to $\text{Var}[\eta_0^{R+r}] = \mu_0^{R+r} - (\mu_0^{R+r})^2$.

Therefore, the received instantaneous signal-to-noise ratio (SNR) can be written as $\gamma^{R+r} = (\mu_0^{R+r})^2 / \text{Var}[\eta_0^{R+r}]$. Then, the conditional BER for the r th S-R link can be

expressed as $P_{R+r}(\gamma^{R+r}) = Q(\sqrt{2\gamma^{R+r}})$ with $Q(\sqrt{2x}) = 1/\sqrt{2\pi} \int_{\sqrt{2x}}^{\infty} \exp(-t^2/2) dt$ [6].

By integrating the conditional BER of $P_{R+r}(\gamma^{R+r})$ with respect to random variable γ^{R+r} , the averaged BER can be presented as

$$P_{R+r} = \int_0^{\infty} f(\gamma^{R+r}) P_{R+r}(\gamma^{R+r}) d\gamma^{R+r} \quad (2)$$

where $f(\gamma^{R+r})$ is the probability density function(PDF) of γ^{R+r} .

2.3 Outage Probability of DSF-Relay Schemes

In DSF schemes, the r th relay is only to transmit the regenerated symbol of $\hat{b}_m^r$ when messages are correctly decoded. Note that $\hat{b}_m^r$ can be two values, which are $\hat{b}_m^r = 0$ with the probability of P_{R+r} or $\hat{b}_m^r = b_m$ with the probability of $1 - P_{R+r}$. In order to generally derive the analytical method based on error-events at relays, let us define the p th error event vector as $E_p = [e_p^1 \cdots e_p^r \cdots e_p^R]$ with $p \in \{1, 2, \dots, 2^R\}$ [13] and the total number of error-events is 2^R . For DSF schemes, only when the r th relay correctly decodes, the r th relay transmits the reconstructed MC-CDMA signal on the r th R-D link channel. Therefore, we can define $\overline{e_p^r} = (e_p^r + 1) \bmod 2$ as a retransmitting indicator at the r th relay for the p th error-event. The received signal from the r th R-D link channel can be shown as

$$y_{i^{r,p}} = \hat{b}_0^r H_i^r c_{0,i} + \sum_{m=1}^{M-1} \hat{b}_m^r H_i^r c_{m,i} + n_i^r = \overline{e_p^r} b_0 H_i^r c_{0,i} + \sum_{m=1}^{M-1} \overline{e_p^r} b_m H_i^r c_{m,i} + n_i^r \quad (3)$$

Consider combined signals from S-D link as well as R-D links. The totally combined SNR can be written as $\gamma_p^{TC} = \gamma^0 + \gamma_p^{RD}$ where γ^0 is SNR for the S-D link and $\gamma_p^{RD} = \sum_{r=1}^R \gamma_p^r = \sum_{r=1}^R \overline{e_p^r} (\mu_0^r)^2 / \text{Var}[\eta_0^r]$ is SNR for the R-D link. For the p th error-event, the outage probability can be expressed as

$$P_{out,p}^{TC} = \Pr[\gamma_p^{TC} < \gamma_{th}] = \int_0^{\gamma_{th}} f(\gamma_p^{TC}) d\gamma_p^{TC} \quad (4)$$

where $f(\gamma_p^{TC})$ is the PDF of γ_p^{TC} . Taking into account for all the possible error-events, the outage probability of combining both S-D and R-D links can be expressed as

$$P_{out}^{TC} = \sum_{p=1}^{2^R} P_{out,p}^{TC} \prod_{r=1}^R (1 - P_{R+r})^{\overline{e_p^r}} (P_{R+r})^{e_p^r}. \quad (5)$$

Not that P_{R+r} is the averaged BER of the r th S-R link.

3 CDD-DSF-Relay Scheme for MC-CDMA Systems

In this paper, to overcome disadvantage of general DSF schemes, we propose the CDD-DSF-Relay scheme for MC-CDMA systems. Fig. 1 shows the concept of CDD-DSF-Relay scheme. The 1st relay is same with that of general DSF schemes. But, the 2nd relay constructs the cyclic delayed version in time domain by using the regenerated symbols and then, transmits regenerated MC-CDMA symbol over the same time-slot which is assigned to the 1st relay node's transmission. Other relays

have same process. Only difference is that CDD amount can be different for each relay. Therefore, CDD-DSF-Relay scheme needs two orthogonal time-slots.

3.1 R-D Link Signal Model for CDD-DSF-Relay Scheme

Let D^r be the cyclic delay time of the r th relay. It is assumed that GP is sufficiently larger than total delay spread of R-D link. Then, the channel gain of the n th subcarrier caused by the r th R-D link channel can be expressed as

$$H_{n,CDD}^r = \sum_{l=D^r}^{L^r-1+D^r} h_l^r \exp(-j2\pi nl/N). \quad \text{For the } p\text{th event-vector, the received signal of}$$

R-D links can be the summation of R-D links components as follows

$$y_{i,CDD}^p = \sum_{m=0}^{M-1} b_m \sum_{r=1}^R \overline{e_p^r} H_{i,CDD}^r c_{m,i} + n_i^1. \quad (6)$$

3.2 Outage Probability for Proposed Schemes

For the p th error-vector at the proposed scheme, the combining weight can be obtained by substituting H_i^r with $H_{i,CDD}^p \left(= \sum_{r=1}^R \overline{e_p^r} H_{i,CDD}^r \right)$ as

$w_{0,i}^{p,CDD} = w_{0,i}^r \big|_{H_i^r = H_{i,CDD}^p}$. The decision variable of R-D links can be expressed as

$$v_{0,CDD}^{p,RD} = \sum_{i=0}^{N-1} y_{i,CDD}^p w_{0,i}^{p,CDD} = \mu_{p,CDD}^{RD} b_0 + \eta_{p,CDD}^{RD} \quad \text{with} \quad \mu_{p,CDD}^{RD} = \mu_0^{R+r} \big|_{H_i^{R+r} = H_{i,CDD}^p},$$

$E[\eta_{p,CDD}^{RD}] = 0$, and $Var[\eta_{p,CDD}^{RD}] = Var[\eta_0^{R+r}] \big|_{H_i^{R+r} = H_{i,CDD}^p}$. Therefore, the

instantaneous SNR can be written as $\gamma_{p,CDD}^{RD} = (\mu_{p,CDD}^{RD})^2 / Var[\eta_{p,CDD}^{RD}]$. Let us consider the case of combining signals from S-D and R-D links. For the p th error-event at the relay nodes, the totally combined SNR can be written as $\gamma_{p,CDD}^{TC} = \gamma^0 + \gamma_{p,CDD}^{RD}$. The outage probability can be expressed as

$$P_{out,p}^{TC,CDD} = Pr[\gamma_{p,CDD}^{TC} < \gamma_{th}] = \int_0^{\gamma_{th}} f(\gamma_{p,CDD}^{TC}) d\gamma_{p,CDD}^{TC} \quad (7)$$

where $f(\gamma_{p,CDD}^{TC})$ is the PDF of $\gamma_{p,CDD}^{TC}$. Considering all the possible error-events, the outage probability for R-D links can be expressed as

$$P_{out}^{TC,CDD} = \sum_{p=1}^2 P_{out,p}^{TC,CDD} \prod_{r=1}^R (1 - P_{R+r})^{\overline{e_p^r}} (P_{R+r})^{e_p^r}. \quad (8)$$

Note that evaluations of (2), (7), and (8) can be done by Monte Carlo integration [12][13].

4 Numerical and Simulation Results

In this section, we have shown numerical results of outage probability and verify their accuracy by comparing simulation results. Furthermore, we compare outage probabilities for general DSF-Relay and proposed CDD-DSF-Relay schemes so that we have verified advantages of proposed ones.

We assume that $E\left[\sum_{l=0}^{L^r-1}|h_l^r|^2\right]=1$ for $r=0, R+1, R+2, \dots, 2R$. Also, the total transmit power of R-D links is assumed to be 1 for $r=1, 2, \dots, R$. We consider $N=M=64$ and $N_G=8$ (length of a GP). We use Walsh-Hadamard codes of length N as spreading codes. Average SNR is defined as $1/(\sigma^2/N)$. $L^0=4$ and $L^{R+r}=L^r=2$ are assumed for different r . In order to evaluate MCI over multipath fading channels, $\{h_l^r\}_{r=0}^{2R+1}$ are generated 10^5 times.

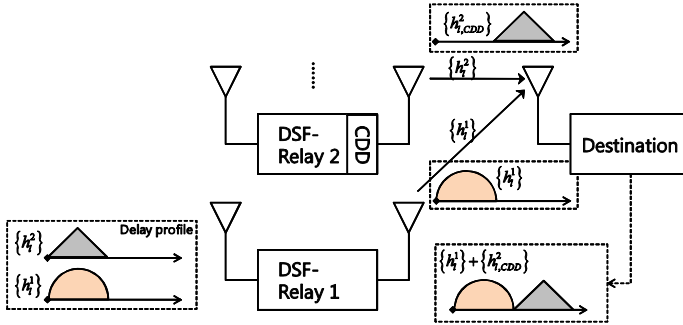


Fig. 2. Block diagram of CDD-DSF-Relay schemes

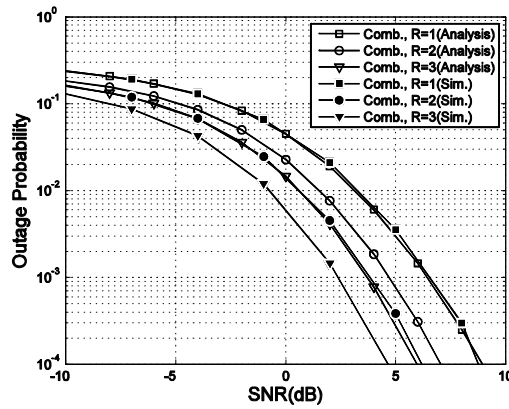


Fig. 2. Outage Probability versus SNR for general DSF-Relay schemes

Fig. 2 and Fig. 3 show the outage probability versus SNR with respect to different number of relay nodes for general DSF-Relay schemes and proposed CDD-DSF-Relay schemes, respectively. These figures indicate that the increment of the number of relays gives the diversity gain and generate outage performance improvement. Note that proposed CDD-DSF-Relay schemes need only two time-slots regardless of R . It means that the proposed scheme can give the diversity gain without lose of spectrum efficiency even if $R > 1$. Even if there is a mismatch, we can say that the proposed semi-analytical results are similar with simulated ones. Consequently, it is confirmed that the proposed semi-analytical method can be used as a technical tool to verify the outage probability for DSF-Relay MC-CDMA systems.

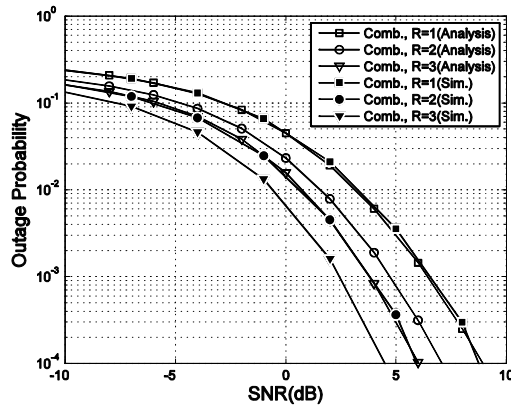


Fig. 3. Outage Probability SNR for proposed CDD-DSF-Relay schemes. ($D^r = (r-1)2$).

5 Conclusions

We have proposed CDD-DSF-Relay schemes for MC-CDMA systems. General DSF-Relay schemes have to use orthogonal channels in proportion to the number of relay links. But, the proposed schemes require the only one channel for R-D links regardless of the number of relay links. Therefore, when there are R relays, total number of orthogonal channels is $R+1$ for general DSF-Relay schemes and 2 for proposed ones.

Furthermore, we propose the semi-analytical method based on error-events at relay nodes for DSF-relay schemes and then, it is modified to cover the proposed CDD-DSF-Relay schemes. Our semi-analytical expressions for the outage probability have been verified by comparing with simulations to be bounds. Consequently, we can say that our analytical approach is easily tractable form to explain the effects of frequency diversity caused by CDD on the combined SNR and outage probability, and can be used as technical tools to verify outage performance for both DSF-Relay and CDD-DSF-Relay MC-CDMA systems. By simulations and numerical results, it is confirmed that the proposed one can be applicable to achieve cooperative diversity gain without a reduction of spectral efficiency.

References

1. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. on Info. Theory*, 3062–3080 (2004)
2. Bletsas, A., Khisti, A., Reed, D.P., Lippman, A.: A Simple Cooperative Diversity Method Based on Network Path Selection. *IEEE Journal on Selected Areas in Commun.*, 659–672 (2006)
3. Kim, J.-B., Kim, D.: Exact and Closed-Form Outage Probability of Opportunistic Single Relay Selection in Decode-and-Forward Relaying. *IEICE Trans. on Commun.*, 4085–4088 (2008)
4. Ikki, S., Ahmed, M.H.: Exact Error Probability and Channel Capacity of the Best-Relay Cooperative-Diversity Networks. *IEEE Trans. on Wireless Commun.*, 1051–1054 (2009)
5. Yang, C., Wang, W., Chen, S., Peng, M.: Outage Performance of Opportunistic Decode-and-Forward Cooperation with Imperfect Channel State Information. *IEICE Trans. on Commun.*, 3083–3092 (2010)
6. Proakis, J.G.: *Digital Communication*, 3rd edn. McGraw Hill, New York (1995)
7. Helard, J.F., Baudais, J.Y., Giterne, J.: Linear MMSE detection techniques for MC-CDMA. *Electronics Letters*, 665–666 (2000)
8. Lee, Y., Tsai, M., Sou, S.: Performance of decode-and-forward cooperative communications with multiple dual-hop relays over nakagami-m fading channels. *IEEE Trans. on Wireless Commun.*, 2853–2859 (2009)
9. Lodhi, A., Said, F., Dohler, M., Aghvami, H.: Performance comparison of space-time block coded and cyclic delay diversity MC-CDMA systems. *IEEE Wireless Commun. Mag.*, 38–45 (2005)
10. Lodhi, A., Said, F., Dohler, M., Aghvami, H.: Closed-Form Symbol Error Probabilities of STBC and CDD MC-CDMA With Frequency-Correlated Subcarriers Over Nakagami-m Fading Channels. *IEEE Trans. on Vehicular Tech.*, 962–973 (2008)
11. Ko, K., Park, M., Hong, D.: Performance Analysis of Asynchronous MC-CDMA systems with a Guard Period in the form of a Cyclic Prefix. *IEEE Trans. on Commun.*, 216–220 (2006)
12. Tranter, W.H., Sam Shanmugan, K., Rappaport, T.S., Kosbar, K.L.: *Communication Systems Simulation with Wireless Applications*. Prentice Hall, Englewood Cliffs (2004)
13. Kim, D., Park, M., Park, J., Ko, K.: Investigation of the AGA Effect on Performance analysis of an MPIC. *IEICE Trans. on Commun.*, 658–661 (2009)

On BER Performance of CDD-DF-Relay Scheme for MC-CDMA Systems^{*}

Jeanyeung Jang¹, Choongchae Woo², and Kyunbyoung Ko¹

¹ Dept. of Instrumentation and Control Engineering, Chungju National Univ.,
72 Daehangno, Chungju-city, Chungbuk, 380-702 Korea,
Tel.: +82-43-841-5463; Fax : +82-43-841-5320

kbko@cjnu.ac.kr

² Dept. of Electronics and Computer Engineering, Hanseo Univ.,
360 Daegok-ri, Haemi-myun, Seosan-si, Chungnam, 356-706,
Tel.: +82-41-660-1360

woo9@hanseo.ac.kr

Abstract. In this paper, CDD (Cyclic Delay Diversity)-DF (Decode-and-Forward)-Relay scheme is proposed for MC-CDMA (Multicarrier-Code Division Multiple Access) systems over multipath Rayleigh fading channels. The advantages of general DF schemes come at the expense of the spectral efficiency since the source and all the relays must transmit on orthogonal channels. In order to mitigate this disadvantage of general DF schemes, we have applied CDD techniques to each relays so that all the relays can transmit on single channel. It means that all R-D link channels can be considered as a single channel which is widely delay spread. By simulations, we have compared proposed one with general DF scheme. Therefore, it is confirmed that the proposed one can be a possible solution to achieve cooperative diversity gain without a reduction of spectral efficiency.

Keywords: CDD, DF-Relay scheme, MC-CDMA, multipath rayleigh fading channels, cooperative diversity.

1 Introduction

Two main relaying protocols are usually used for cooperative diversity schemes: amplify-and-forward (AF) and decode-and-forward (DF) [1]. At the destination, the receiver can employ a variety of diversity combining techniques to benefit from the multiple signal replicas available from the relays and the source. A third option is to have relays forward only those correctly decoded messages, in which case we say that they considered selective relay (SR) schemes [2]. Use of SR schemes presumes incorporation of, e.g., cyclic redundancy check (CRC) codes from a higher layer in order to detect errors. The advantages of general cooperative diversity schemes come

^{*} This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2009-0072762, 2010-0002650).

at the expense of the spectral efficiency since the source and all the relays must transmit on orthogonal channels(i.e., different time slots or frequency bands) in order to avoid interfering with each other as well [1].

Recent researches are widely carried out for SR schemes assuming that the relay node can correctly decode symbol with CRC [3][4]. For the practically attractive DF relay strategy, the authors in [2] derived a high-performance low-complexity coherent demodulator at the destination in the form of a weighted combiner. However, general relay schemes still have trade-off between diversity gain and spectral efficiency. It is a practical issue how to mitigate the reduction of spectral efficiency. Therefore, we propose CDD (Cyclic Delay Diversity)-DF-Relay scheme for MC-CDMA (Multicarrier-Code Division Multiple Access) systems as a solution not only to obtain cooperative diversity gain but also to maintain spectral efficiency even if the number of relays is increased.

2 MC-CDMA DF-Relay Systems

Fig. 1. shows the block diagram of MC-CDMA DF-relay systems with a source(S), a destination(D), and relay(R)s. The number of relays is R . Each relay node detects the received signal and then transmits a regenerated MC-CDMA signal on orthogonal channels[1]. At a destination, a cooperative diversity gain can be achieved by combining signals form S-R-D links and S-D link.

2.1 Multipath Fading Channel Model

Let us consider transmission over a multipath Rayleigh fading channel having a slow fading rate. When the number of multipath for the r th link is L^r , the l th path's channel gain is h_l^r with $l \in \{0, 1, \dots, L^r - 1\}$. It is assumed for $\{h_l^r\}$ that the magnitudes are Rayleigh-distributed and the phases are uniformly distributed over $[0, 2\pi)$ [5]. Furthermore, they can be mutually independent for different l and r .

The exponential decay factor of multipath intensity profile is $1/L^r$. Then, we can obtain the channel response of the n th subcarrier for the r th link as

$$H_n^r = \sum_{l=0}^{L^r-1} h_l^r \exp(-j2\pi nl/N) \text{ where the normalized tap interval is } 1/N. \text{ Note that}$$

$\{H_n^0\}$ means a frequency domain channel gain of S-D link. $\{H_n^{R+r}\}$ and $\{H_n^r\}$ mean frequency domain channel gains of the r th S-R link and the r th R-D link.

2.2 MC-CDMA Signal Model for S-D and S-R Links

The user data bit is mapped into binary phase-shift keying (BPSK) symbol $b_m = \pm 1$, which is the m th code. The transmitter spreads b_m in the frequency domain by using the m th spreading code $\{c_{m,n}\}_{n=0}^{N-1}$ with $|c_{m,n}| = 1$. The signal is converted to the time domain using an IFFT device and a GP is added in the form of a cyclic prefix. Therefore, the transmitted signal for the m th code can be written as

$$s_m(t) = \sum_{n=0}^{N-1} b_m c_{m,n} p(t) e^{j2\pi f_n g(t)}, \quad (1)$$

where $f_n = n/T$, N is the length of spreading code (i.e., size of IFFT) and $p(t)$ is the pulse waveform which is uniform in the interval $[-T_G, T)$, and zero otherwise. With respect to the guard interval in the form of a cyclic prefix, $g(t)$ is defined

$$g(t) = \begin{cases} (t+T) & \text{for } -T_G \leq t < 0 \\ t & \text{for } 0 \leq t < T \end{cases} \quad \text{where } T_G, T, \text{ and } T_S (= T_G + T) \text{ are the guard period, the bit duration, and the MC-CDMA symbol duration, respectively [6].}$$

Conventional MC-CDMA receivers operate using a process that is the reverse of the process for the transmitter. The received signal is sampled at a rate of $1/T_c (= N/T)$, and the guard interval samples are removed. After a serial to parallel conversion, a FFT of size N is performed. When a symbol timing offset is perfectly recovered, the received signal of the i th subcarrier can be expressed as

$$y_i^r = b_0 H_i^r c_{0,i} + \sum_{m=1}^{M-1} b_m H_i^r c_{m,i} + n_i^r \quad (2)$$

where $r = 0$, namely y_i^0 , means a signal of S-D link and y_i^{R+r} with $r \in (1, 2, \dots, R)$ is that of the r th S-R link. Also, n_i^r is a complex AWGN term with $E[n_i^r] = 0$ and $E[|n_i^r|^2] = \sigma^2$.

2.3 Combining Schemes for MC-CDMA Systems

For MC-CDMA systems, there are combining schemes per subcarrier which are MRC (Maximal Ratio Combining), EGC (Equal Gain Combining), ZF (Zero Forcing), and MMSE-C (Minimum Mean squared Error-per subCarrier) [7][8]. Without loss of generality, the desired code is assumed to be the 0th code and the combining weight can be written as

$$w_{0,i}^{r,X} = \begin{cases} c_{0,i}^* H_i^{r*} & X = \text{MRC} \\ c_{0,i}^* H_i^{r*} / (N |H_i^r|) & X = \text{EGC} \\ c_{0,i}^* / (N H_i^r) & X = \text{ZF} \\ c_{0,i}^* H_i^{r*} / \left[N \left(|H_i^r|^2 + \sigma^2 / M \right) \right] & X = \text{MMSE-C} \end{cases} \quad (3)$$

where X is an index indicating each combining scheme. In addition, there are two options (i.e., JD-ZF (Joint Detection-ZF) and JD-MMSE) which can be applicable per code-block. The vector version of (3) can be presented as

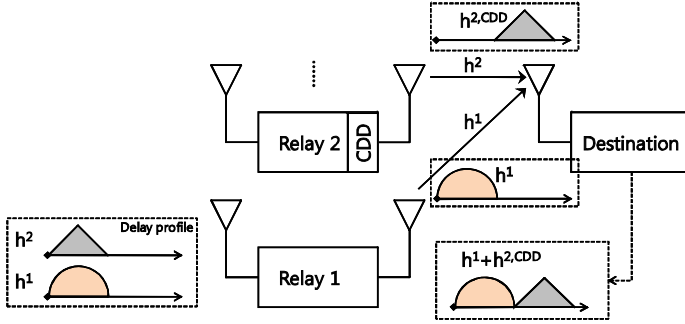


Fig. 1. Block diagram of CDD-DF-Relay system

$$y^r = A^r b^r + n^r \quad (4)$$

and weight vectors of JD-ZF and JD-MMSE are expressed as [7]-[9]

$$W^{r,X} = \begin{cases} [A^{rH} A^r]^{-1} A^{rH} & X = \text{JD-ZF} \\ [A^{rH} A^r + \sigma^2 I]^{-1} A^{rH} & X = \text{JD-MMSE}. \end{cases} \quad (5)$$

Then, the 0th row vectors are the combining weight vectors for the 0th code (i.e., $w_0^{r,\text{JD-ZF}}$ and $w_0^{r,\text{JD-MMSE}}$).

2.4 Combining Process at a Destination Node

For the r th S-R link, the regenerated symbol can be written as $\hat{b}_0^r = \text{sig}\left(\sum_{i=0}^{N-1} y_i^{R+r} w_{0,i}^{R+r,X}\right)$ where $\text{sig}(\cdot)$ is the signum function. Then, the r th relay transmits the reconstructed MC-CDMA signal on the r th R-D link channel and the received signal can be shown as

$$y_i^r = \hat{b}_0^r H_i^r c_{0,i} + \sum_{m=1}^{M-1} \hat{b}_m^r H_i^r c_{m,i} + n_i^r. \quad (6)$$

Therefore, the decision variables for combining R-D links and combining both S-R-D and S-D links are given as $v_0^{SRD,X} = \sum_{r=1}^R \sum_{i=0}^{N-1} y_i^r w_{0,i}^{r,X}$ and $v_0^{Comb,X} = \sum_{r=0}^R \sum_{i=0}^{N-1} y_i^r w_{0,i}^{r,X}$.

3 CDD-DF-Relay Scheme for MC-CDMA Systems

Main disadvantage of general DF schemes is to reduce the spectral efficiency since the source and all relay nodes must transmit on orthogonal channels in order to avoid interfering with each other [1]. In this paper, to overcome this disadvantage of general DF scheme comes, we propose the CDD-DF-Relay scheme for MC-CDMA systems.

Fig. 1 shows the concept of CDD-DF-Relay scheme. The 1st relay is same with that of general DF schemes. But, the 2nd relay constructs the cyclic delayed version in time domain by using the regenerated symbols of (7) and then, transmits regenerated MC-CDMA symbol over the same time-slot which is assigned to the 1st relay node's transmission. Other relays have same process. Only difference is that CDD amount can be different for each relay. Therefore, CDD-DF-Relay scheme needs two orthogonal time-slots. As shown Fig. 1, even if all relays use same time-slot, CDD schemes make an effect of increasing the number of multipath for R-D link [10][11]. It means that we can expect the frequency diversity gain caused by increasing the number of multipath [6].

3.1 R-D Link Signal Model for CDD-DF-Relay Scheme

In the proposed CDD-DF-Relay scheme, let D^r be the cyclic delay time of the r th relay. It is assumed that GP is sufficiently larger than total delay spread of R-D link. Then, the channel gain of the n th subcarrier caused by the r th link channel can be

expressed as $H_n^{r,CDD} = \sum_{l=D^r}^{L^r-1+D^r} h_l^r \exp(-j2\pi nl/N)$. In addition, the received signal of R-D links is shown as

$$y_i^{CDD} = \sum_{r=1}^R \hat{b}_0^r H_i^{r,CDD} c_{0,i} + \sum_{m=1}^{M-1} \sum_{r=1}^R \hat{b}_m^r H_i^{r,CDD} c_{m,i} + n_i^{CDD}. \quad (7)$$

We can obtain the combining weight by replacing H_i^r with $H_n^{CDD} \left(= \sum_{r=1}^R H_n^{r,CDD} \right)$ as $w_{0,i}^{CDD,X} = w_{0,i}^{r,X} \big|_{H_i^r = H_n^{CDD}}$.

3.2 Combining Process at a Destination Node

The proposed scheme can generate the decision variables for S-R-D link and the combining both S-R-D and S-D links as $v_0^{CDD-SRD,X} = \sum_{i=0}^{N-1} y_i^{CDD} w_{0,i}^{CDD,X}$ and

$$v_0^{CDD-Comb.,X} = v_0^{CDD-SRD,X} + \sum_{i=0}^{N-1} y_i^0 w_{0,i}^{0,X}.$$

4 Simulation Results

In this section, we show simulation results of averaged BERs for general DF-relay and proposed CDD-DF-Relay schemes and verify advantages of proposed one by comparing two schemes. We assume for each link that $E \left[\sum_{l=0}^{L^r-1} |h_l^r|^2 \right] = 1$, $L^0 = 4$,

$L^{R+r} = L^r = 2$ for different r , and $R \in \{1, 2, 3, 4\}$. For MC-CDMA parameters, we consider as $N = M = 64$ (i.e., full-loaded system) and $T_G/T = 1/8$. We use Walsh-

Hadamard codes of length N as spreading codes. SNR is defined as

$$E\left[\sum_{l=0}^{L^r-1} |h_l^r|^2\right] / (\sigma^2/N).$$

Fig. 2 shows the averaged BERs versus SNR with respect to different combining schemes of an MC-CDMA system with $R=1$. We can find for full-loaded system that ZF and JD-ZF show same performance and MMSE-C and JD-MMSE also do[5][12]. From here, we consider MMSE-C as the combining method for MC-CDMA systems.

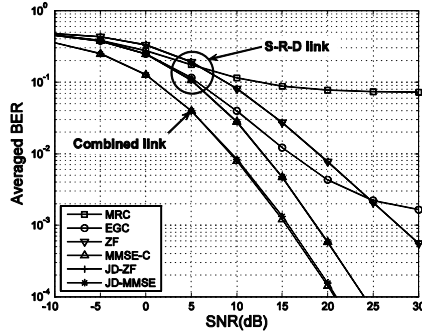


Fig. 2. Averaged BER versus SNR for DF relay systems with respect to different combining schemes of an MC-CDMA system. ($R=1$).

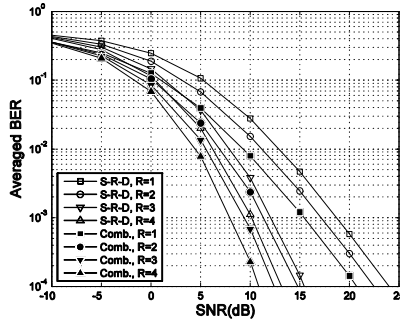


Fig. 3. Averaged BER versus SNR for general DF-Relay schemes with respect to different number of relay nodes. ($X=MMSE-C$, $R \in \{1, 2, 3, 4\}$, $L^0 = 4$, $L^{R+r} = L^r = 2$).

Fig. 3 and Fig. 4 show the averaged BER versus SNR with respect to different number of relay nodes for general DF-Relay and proposed CDD-DF-Relay schemes, respectively. Note that proposed CDD-DF-Relay schemes need only two time-slots regardless of R . It is confirmed from Fig. 3 that general DF-Relay schemes give BER performance improvement by decreasing spectrum efficiency. On the contrary, the proposed scheme can give the diversity gain without lose of spectrum efficiency in proportion to the increase of R . Fig. 5 shows the averaged BER versus SNR with

respect to different number of relay nodes for proposed CDD-DF-Relay schemes. Note that we have assumed $L^0 = L^{R+r} = L^r = 4$ under same other parameters. This condition means the increase of the number of multipath for S-R and S-D links so that we can find larger frequency diversity gain. Also, it covers the case of lager delay spread. Therefore, CDD effect can be larger than GP so that we assign CDD value by operation of modular as $D^r = (r-1)3 \bmod 8$. It means that the multipath components can be overlapped and we cannot fully obtain the diversity.

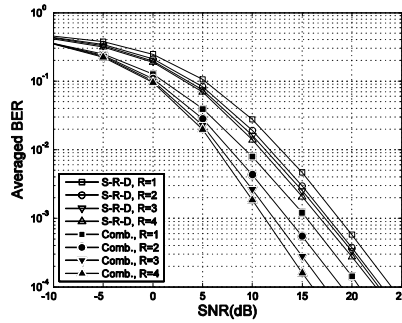


Fig. 4. Averaged BER versus SNR for proposed CDD-DF-Relay schemes with respect to different number of relay nodes. ($X=MMSE-C$, $R \in \{1, 2, 3, 4\}$, $L^0 = 4$, $L^{R+r} = L^r = 2$, $D^r = (r-1)2$).

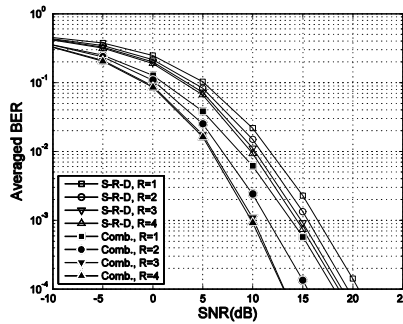


Fig. 5. Averaged BER versus SNR for proposed CDD-DF-Relay schemes with respect to different number of relay nodes. ($X=MMSE-C$, $R \in \{1, 2, 3, 4\}$, $L^0 = L^{R+r} = L^r = 4$, $D^r = (r-1)3 \bmod 8$).

5 Conclusions

We have proposed the CDD-DF-Relay scheme for MC-CDMA systems. General DF-Relay schemes have to use orthogonal channels in proportion to the number of relay links. But, the proposed schemes require the only one channel for R-D links

regardless of the number of relay links. Therefore, when there are R relays, total number of orthogonal channels is $R+1$ for general DF-Relay schemes and 2 for proposed ones. By simulations, we have compared BER performance of proposed one with that of general DF-schemes. It is verified that the proposed one can be applicable to achieve cooperative diversity gain without a reduction of spectral efficiency.

References

1. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 3062–3080 (2004)
2. Wang, T., Cano, A., Giannakis, G.B., Laneman, J.N.: High-Performance Cooperative Demodulation With Decode-and-Forward Relays. *IEEE Transactions on Communications*, 1427–1438 (2007)
3. Bletsas, A., Khisti, A., Reed, D.P., Lippman, A.: A Simple Cooperative Diversity Method Based on Network Path Selection. *IEEE Journal on selected areas in Communications*, 659–672 (2006)
4. Ikki, S., Ahmed, M.H.: Exact Error Probability and Channel Capacity of the Best-Relay Cooperative-Diversity Networks. *IEEE Transactions on Wireless Communications*, 1051–1054 (2009)
5. Proakis, J.G.: *Digital Communication*, 3rd edn. McGraw Hill, New York (1995)
6. Ko, K., Park, M., Hong, D.: Performance Analysis of Asynchronous MC-CDMA systems with a Guard Period in the form of a Cyclic Prefix. *IEEE Trans. on Comm.* 216–220 (2006)
7. Helard, J.F., Baudais, J.Y., Giterne, J.: Linear MMSE detection techniques for MC-CDMA. *Electronics Letters*, 665–666 (2000)
8. Klein, A., Kaleh, G.K., Baier, P.W.: Zero forcing and minimum-mean-square-error equalization for multiuser detection in code division multiple access channels. *IEEE Trans. on Vehicular Tech.* 276–287 (1996)
9. Zhang, K., Guan, Y.L., Shi, Q.: Complexity Reduction for MC-CDMA With MMSEC. *IEEE Trans. on Vehicular Tech.* 1989–1993 (2008)
10. Lodhi, A., Said, F., Dohler, M., Aghvami, H.: Performance comparison of space-time block coded and cyclic delay diversity MC-CDMA systems. *IEEE Wireless Communications Mag.* 38–45 (2005)
11. Lodhi, A., Said, F., Dohler, M., Aghvami, H.: Closed-Form Symbol Error Probabilities of STBC and CDD MC-CDMA With Frequency-Correlated Subcarriers Over Nakagami-m Fading Channels. *IEEE Trans. on Vehicular Tech.* 962–973 (2008)

The Effect of Corporate Strategy and IT Role on the Intent for IT Outsourcing Decision

Dong-Hwan Cho¹ and Jin-Sung Kim²

¹ Venture & Business, Gyeongnam National University of Science and Technology,
150 Chilam-dong, Jinju, Gyeongsangnam-do, 660-758, Republic of Korea

² Management Information System, Gyeongsang National University,
900, Gazwa-dong, Jinju, Gyeongsangnam-do, 660-701, Republic of Korea

Abstract. The Intent for IT outsourcing is commonly categorized as IS Improvement, Business Impact, and Commercial Exploitation. The decision factors for these IT outsourcing include corporate strategy types, IT role, and interaction between these two. Firm size and IT maturity are selected as moderating factors to add the credibility to the research of the effect of IT outsourcing decision making. Analyzing collected data revealed that corporate strategy rather than IT role is more significant factor affecting the intent for IT outsourcing decision. Defenders tend to use IT outsourcing more as a means to improve information systems, compared to analyzers and reactors. Prospectors tend to use IT outsourcing more as a commercial exploitation, compared to analyzers and reactors. Unlike previous researches, the hypothesis that IT role determines outsourcing intent was rejected. It was also rejected that the interaction between corporate strategy types and IT role determines outsourcing intent.

Keywords: IT Outsourcing, Organization Strategy Types, IT Role, IT Maturity, Intent for IT Outsourcing.

1 Introduction

Recently, the Outsourcing is important to manage in the corporation for IS with market scale of the outsourcing of IT keep going growth: [1],[15]. The view point of IT outsourcing goals is categorized as transaction cost [20], taking competency based on strategic part [2],[10] and the theory of social-exchange:[7],[13]. These three view point is that the system of information improvement, influence of business, and used the commercial to representation of outsourcing of intention of strategic [6].

It is very important that the strategic intention of outsourcing is different as each outsourcing of strategic intention is coincidental to reach to purpose. However the most outsourcing had point out that estimate the reduction of costs, business value and improvement could not calculate for effect because customer and vender have problems as partnership, importance of contract:[4],[9],[17]. In addition, individual corporations plays strategic role and IT, maturity remain ignorant of strategic intention of outsourcing decision is made to a cause to point out[5],[16]. The decision makers couldn't determine to intent for purpose appropriate outsourcing that if general

outsourcing made a decision, between the role of corporate strategy and strategic intentions of outsourcing are to be linked to serious inconsistency.

This study will focus various and strategic intentions of outsourcing, corporate strategy and the roles of IT.

2 Theoretical Background

2.1 Strategic Intent for IT Outsourcing

The purpose of the previous outsourcing research of strategic intention approach is mainly to reduce costs as a primarily economic concern[14]. However, these perspectives of reduction costs have critical point that could not effectively describe for recently occurred outsourcing of each type[19]. The strategic intent for IT outsourcing is the improvement of IS, the influence on business and the commercial purpose.

2.2 Strategic Orientation of Corporate

The studies on the strategy consider that the strategic orientation has influence on the performance. Venkatraman[20] classified the corporate strategy types as aggressiveness, analysis, defensiveness, futurity, proactiveness and riskiness. Miles [17] classified the corporations as defender, analyzer, prospector and reactor.

2.3 IT Role in a Firm

The Johnston and Carrico[12] argued that corporate strategy reference the three different roles – traditional role, developmental role, essential role- of IT. The empirical results of previous studies, the strategic roles of diversity of IT appeared that similar to associate the adaptation decision of outsourcing[19].

3 Research Model and Hypotheses

This study aims to investigate whether the intent for IT outsourcing is determined by the corporate strategy types and the IT role. The research model is shown in Fig.1.

3.1 Relationship between Corporate Strategy Types and Intent for IT Outsourcing

The case of defender, prospector and analyzer except reactor of Corporate Strategy Types, Miles [17] shows response to changes around business environment consistently and stably. These types also present same reactions about decision of imposition and how to deal outsourcing. So we check relationship between researched study about outsourcing [19] and corporate strategy types.

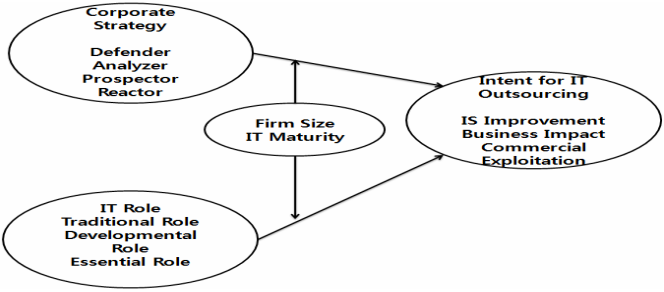


Fig. 1. Research Model

Hypothesis 1: At least one intention for IT outsourcing in corporate strategy types is different.

3.2 Relationship between IT Role and Intent for IT Outsourcing

We can see that imposition of outsourcing depends on informational technology in corporate [18].Corporate that uses IT traditionally decides more frequently to impose outsourcing. On the other hand company uses IT more essentially and expansively determines less to impose outsourcing. Furthermore the role of IT may influence on decision about strategic purpose of IT outsourcing.

Hypothesis 2: At least one intention of IT role in corporate strategy types is different.

3.3 Relationship between Intent for IT Outsourcing and the Interaction between Corporate Strategy and IT Role

The research strategic alignment emphasized that expectation outcome of the company will increase if IT role is connected well with corporate strategy [11]. On the contrary to this, in the information department of the corporate, which does not have good alignment efficaciousness will go down, therefore the result will go lower and lower [3]. The difference between expectation and performance in organization to overcome that considering the outsourcing the intention for IT outsourcing will be different that the difference between expectation and realism.

Hypothesis 3: At least one intention for IT outsourcing by interaction between corporate strategy types and IT role is different.

According to the previous studies [9] and [8], we included control variables that organizational scale and information technology maturity have influenced on strategic intention's decision of outsourcing.

4 Research Method

4.1 Survey Measures

The contents of the variable measurements are shown in Table 1.

Table 1. Defining Variables

Variables		Information	Num of Questions	Measure
Corporation Factors		- Industry - Number of Employees - Budget related to information systems - Private information systems department	5	Nominal Open-ended Open-ended Open-ended
Outsourcing Experience		- Whether outsourcing experience	1	Nominal
Outsourcing intent	Information System Improvement	- Information system costs - Introduction of new IT - Improve the quality of IT - Switch to new forms of IT-based	4	7-point Likert scale
	The impact of business	- Alignment between IT and business - Competency of projects development by IT-based - Changes in business processes - Receiving active support from IT to perform business processes	4	7-point Likert scale
	Commercial use	- IT assets, outside sales - Development of IT products and services - Development of market process and the path - IT-based business founded	4	7-point Likert scale
Type of Corporation strategy		- Defender - prospector - analyzer - Reactor	1	Nominal
Role of Corporation IT		- Traditional roles - Developmental role - Essential role	1	Nominal
IT Maturity		- Level of IT Maturity	12	7-point Likert scale

4.2 Data Collection

This study used the questionnaire to collect data. Questionnaires were sent to 95 companies by mail, 72 questionnaires were returned by mail and fax and 11 questionnaires of companies that have never experienced outsourcing were excluded from analysis. Frequency analysis about demographic variables is shown in Table 2 and Table 3.

Table 2. Information of the corporate respondents

Industry Range	Frequency	Percentage
Distribution	4	6.6%
Banking	8	13.1%
Insurance	3	4.9%
Service	3	4.9%
Manufacturing	37	60.7%
ICT	1	1.6%
Public enterprise	3	4.9%
Other	2	3.3%
Total	61	100.0%
Sales	Frequency	Percentage
Under 100million	2	3.3%
Over 100million ~ under 10 billion	5	8.2%
Over 10 billion ~ under 10 0billion	13	21.3%
Over 10 0billion ~ under 50 0billion	16	26.2%
Over 50 0billion ~ under 50 00billion	14	23.0%
Over 50 00billion	7	11.5%
Non-response	4	6.6%
Total	61	100.0%
Number of employees	Frequency	Percentage
Under 200 person	12	19.7%
Over 200 person ~ under 500 person	15	24.6%
Over 500 person ~ under 1000 person	10	16.4%
Over 1000 person ~ under 2000 person	9	14.8%
Over 2000 person ~ under 10000 person	12	19.7%
Over10000 person	2	3.3%
Non-response	1	1.6%
Total	61	100.0%

Table 3. Distribution of respondents

Position of respondents	Frequency	Percentage
Executives	4	6.6%
General Manager	10	16.3%
Deputy General Manager	12	19.7%
Manager	35	57.4%
Total	61	100.0%

5 Analysis and Results

In this study, data analysis was performed using SPSS for Windows 12.0 statistical package and the hypothesis was proved through MANCOVA analysis.

5.1 Reliability and Validity of the Constructs

In this study, feasibility analysis was conducted through factor analysis. Also the intent for IT outsourcing as a dependent variable and the IT maturity as a moderating variable were each conducted through factor analysis.

Table 4. The result of factor analysis of intent for IT outsourcing

construct	Measurement Items	Factor			Communalities
		IS Improvement	Business Impact	Commercial Exploitation	
IS Improve ment	IT level	0.856	-0.109	-0.004	0.745
	Improvement				
	IT instruction	0.793	0.305	-0.009	0.730
	IT reconversion	0.749	0.323	0.000	0.666
	IT service	0.665	0.007	0.142	0.466
Business Impact	Development				
	Set up IT-based company	-0.002	0.823	-0.102	0.688
	IT-based business capability	0.243	0.737	0.009	0.609
	BPR	0.008	0.702	0.356	0.626
	IT-oriented	0.392	0.69	0.146	0.651
Commer cial Exploitat ion	Business Process Development of Market process	-0.316	-0.297	0.867	0.935
	External sales of IT assets	-0.294	0.364	0.852	0.941
Eigenvalues		2.764	2.620	1.656	
Cumulative Var (%)		23.332	45.060	59.063	

5.2 Hypothesis Testing

The result of hypothesis testing is shown in Table 6.In the case of *Hypothesis 2*, F value is 0.501 and P value is 0.807, is not statistically significant. *Hypothesis 3*, F value is 0.798 and P value is 0.678, is not statistically significant. Because the result of F test hypothesis1 is significant, we analyzed Pairwise comparisons to know the difference between the groups, the result as seen in Table 7. As seen in Table 6. F value of Pillai’s Trace is 1.897, P value is 0.057, *Hypothesis 1* turned out to be statistically significant. As result, null hypothesis is rejected.

Table 5. The result of factor analysis of IT maturity

Measurement Items	Factor Loading	Communalities
IS understanding in quality	.869	.754
Explore to opportunity	.856	.732
Ability of cooperation	.847	.718
IS performance	.829	.687
IT level of competitor	.812	.659
Business understanding of the IT personnel	.799	.639
IS structure	.799	.638
Participation of Users	.757	.573
IT strategy of competitor	.745	.556
Attention of CEO	.734	.539
Business goal & Support Strategy	.720	.518
Get to the IS Goal	.690	.476
Eigenvalues	7.490	
Cumulative Var (%)	62.414	

Table 6. The result of hypothesis testing

Effectiveness	dependent variable	Pillai's Trace	F	P	Univariate Tests F	D.F	P	hypothesis
IT Maturity	IS Improvement	.366	9.029	.000*	21.006	1	.000**	
	Business Impact				11.266	1	.002**	
	Commercial				3.735		.059*	
	Exploitation				2.841	3	.047**	
Corporation strategy	IS Improvement	.312	1.897	.057*	1.390	3	.257	hypothesis 1 adopt
	Business Impact				1.949	3	.134	
	Commercial							
IT role	IS Improvement	.061	.501	.807	.772	2	.467	hypothesis 2 dismissal
	Business Impact				.774	2	.467	
	Commercial				.564	2	.572	
Corporation strategy* IT role	IS Improvement	.266	.798	.678	1.941	5	.104	hypothesis 3 dismissal
	Business Impact				.534	5	.749	
	Commercial				.658	5	.657	

***: p<.001, **: p<.05, *: p<.1

Table 7. Result of the Pairwise comparative analysis inter corporate strategy group

dependent variable	(I) Corporation strategy	(J) Corporation strategy	mean difference(I-J)	Standard error	p-value
IS Improvement	Prospector	Analyzer	.395	.370	.291
		Defender	-.351	.477	.465
		Reactor	.604	.468	.203
	Analyzer	Prospector	-.395	.370	.291
		Defender	-.746**	.326	.026
		Reactor	.208	.311	.506
	Defender	Prospector	.351	.477	.465
		Analyzer	.746**	.326	.026
		Reactor	.955**	.387	.017
Business Impact	Prospector	Analyzer	.358	.556	.522
		Defender	.938	.716	.196
		Reactor	1.142	.703	.111
	Analyzer	Prospector	-.358	.556	.522
		Defender	.580	.489	.242
		Reactor	.784	.467	.100
	Defender	Prospector	-.938	.716	.196
		Analyzer	-.580	.489	.242
		Reactor	.204	.581	.727
Commercial Exploitation	Prospector	Analyzer	1.205*	.624	.059
		Defender	1.255	.804	.125
		Reactor	1.892**	.789	.020
	Analyzer	Prospector	-1.205*	.624	.059
		Defender	.049	.549	.928
		Reactor	.687	.524	.196
	Defender	prospector	-1.255	.804	.125
		analyzer	.049	.549	.928
		Reactor	.637	.653	.196

6 Conclusions

In this study, we figured out how the corporate strategy types and IT roles affect the intent for IT outsourcing independently. And we also examined that the corporate strategy types and IT roles to affect each other bring about different influence of intent for IT outsourcing. The result of our study is as follows. First, the corporate types turned out to be a decisive factor which has an influence on the intent for IT outsourcing. Second, the IT outsourcing decision did not depend on IT roles.

These results imply the intent for IT outsourcing within the corporation, regardless of the IT roles can be more diversity. And when we make the outsourcing decision, we need to consider that corporate strategy rather than the IT roles in the corporation.

This study is cross-sectional study using data that respondents recall their past experiences, we used small sample size with contained the same distribution throughout the entire industry, and we could not adequately control outsourcing scale.

References

1. Ang, S., Straub, D.W.: Production and Transaction Economics and IS Outsourcing: A Study of the U.S. Banking Industry. *MIS Quarterly* 22(4), 535–552 (1998)
2. Barney, J.: Firm Resources and Sustained Competitive Advantage. *Journal of Management* 17(1), 99–120 (1991)
3. Chan, Y.E., Huff, S.L., Barclay, D.W., Copeland, D.G.: Business Strategic Orientation, Information Systems Strategic Orientation, and Strategic Alignment. *Information Systems Research* 8(2) (June 1997)
4. Cross, J.: IT Outsourcing: British Petroleum's Competitive Approach. *Harvard Business Review*, 94–102 (1995)
5. Cullen, S., Willcocks, L.: *Intelligent IT Outsourcing: Eight Building Blocks to Success*. Butterworth-Heinemann, Butterworths (2007)
6. Di Romualdo, A., Gurbaxani, V.: Strategic Intent for IT Outsourcing. *Sloan Management Review* 39, 67–80 (1998)
7. Grant, R.M.: Toward a Knowledge-based Theory of the Firm. *Strategic Management Journal* 17, 109–122 (1996)
8. Grover, V., Cheon, M.J., Teng, J.T.C.: A Descriptive Study on the Outsourcing of Information Systems Functions. *Information Management* 27, 33–44 (1994)
9. Grover, V., Cheon, M.J., Teng, J.T.C.: The Effect of Service Quality and Partnership on the Outsourcing of Information Systems Functions. *Journal of Management Information Systems* 12(4), 89–116 (1996)
10. Hamel, G.: Competition for Competence and Inter-partner Learning within International Strategic Alliances. *Strategic Management Journal* 12, 83–103 (1991)
11. Henderson, J.C., Venkatraman, N.: Strategic Alignment: Leveraging Information Technology for Transforming Organizations. *IBM Systems Journal* 32(1) (1993)
12. Johnston, H.R., Carrico, S.R.: Developing Capabilities to Use Information Strategically. *MIS Quarterly* 12(1), 37–50 (1988)
13. Kern, T., Lacity, M., Willcocks, L.: *Netsourcing: Renting Applications and Services Over a Network*. FT/Prentice Hall, New York (2002)
14. Kogut, B., Zander, U.: Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science* 3(3), 383–397 (1992)

15. Kogut, B.: Joint Ventures: Theoretical and Empirical Perspectives. *Strategic Management Journal* 9, 319–322 (1998)
16. Lacity, M.C., Willcocks, L.P.: An Empirical Investigation of Information Technology Sourcing Practices: Lessons From Experience. *MIS Quarterly* 22(3), 363–408 (1998)
17. Miles, R.E., Snow, C., Meyer, A., Coleman, H.: Organizational Strategy, Structure, and Process. *Academy of Management Review* 3(3), 546–562 (1978)
18. Nunnally, J.C.: *Psychometric Theory*, 2nd edn. McGraw-Hill, New York (1978)
19. Teng, J.T.C., Cheon, M.J., Grover, V.: Decisions to Outsource Information Systems Functions: Testing a Strategy-Theoretic Discrepancy Model. *Decision Sciences* 26(1), 75–103 (1995)
20. Venkatraman, N.: Strategic Orientation of Business Enterprises: The Construct, Dimensionality and Measurement. *Management Science* 35(8), 942–962 (1989)

A Study on IT Organization Redesign with IT Governance: Focusing on K Public Corporation in Korea

Dong Hwan Cho¹ and Hyoung Seop Shim²

¹ Gyeongnam Nat'l University of Science & Technology,
Department of Venture & Business,
Jinju City, South Korea
dhcho@gntech.ac.kr

² Dongguk University,
Disaster & Risk Management Institute,
Seoul, South Korea
hsshim@dgu.edu

Abstract. This study investigated the redesign methods and procedures of IT organization with specific focus on K public corporation. Job analysis was carried out so as to redesign IT organization in K public corporation. Task redesign was also performed and IT organization as well as personnel was assessed. (1) In task analysis, core activities were identified and 96 standard tasks were drawn. (2) IT support, IT delivery, IT operation teams were the target for organization redesign. Bukyoung and Jeju teams were also separated. (3) Previous job organization was restructured.

Keywords: IT Governance, IT Organization, Organization Redesign, Task Analysis, FTE.

1 Introduction

Rapid changes and adaptations in business environment, business efficiency and cost reduction Non-competitive businesses have improved their performance through outsourcing, especially in the field of information systems. They send expertise to the outsourcing companies trying to gain competitiveness. From the mid of 2000s, certain corporate functions or business units which have introduced the concept of BPO (Business Process Outsourcing) have been increasing rapidly [1].

However, business activity plays an important role in the information systems and because many of the outsourcing workforce, if you configure the IT organization, dependant on outsourcing management system, [2]. In addition, accidents in the area of mission-critical IT failures occur; damage will occur many tangible and intangible [3]. Introduce the concept of organizational management in some of the IT organization is working to rebuild [4].

This study investigates K construction's IT organization and issues derived through analysis and enhancements, IT organizations and the job analysis improve IT

organization and IT organization to ensure the professionalism of the organization to redesign a case study.

2 K Public Corporation

2.1 K Corporation's IT Organization

K Corporation's IT organization were non-integrated that classified business division and location as business management division, Seoul ticketing headquarter, Pusan headquarter, Jeju headquarter.

IT organization of K Corporation consistence such as regular workers is 43%, contract workers are 8%, and outsourcing is 49% composed of a total of 101 in table 1. In particular, technical support personnel compared to the regular rate of 31%: 9%: 60%, and most are outside the workforce outside the workforce to be significantly higher proportion were investigated.

Table 1. K Corporation's IT Organization Manpower

Team	Regular Worker	Contract Worker	Outsourcing	Total
Chief of IT	1	-	-	1
Information Management Team	9	1	8	18
Ticketing Technology Team	6	-	6	15
Ticketing Computing Team	8	1	4	13
Technology Support Team	11	3	21	35
BuKyoung Ticketing Team	4	1	5	10
Jeju Ticketing Team	5	2	2	9
Total	43	8	49	101

2.2 Problems of K Corporation's IT System Operation

Problems of K Corporation's IT system operating examined IT management processes, IT management system outsourcing and IT User Satisfaction.

1) Operation focused on the IT business

K Corporation's IT organization was strategy and planning related manpower is small, and most job offers IT services focused on the business of operating. IT organization's manpower levels are displayed in Figure 1.

Strategy and Planning areas were only handful personnel to perform. Looking at each level, level 2 or more people centered primarily responsible for personnel management, and 3-5 level personnel management was focused on the application.

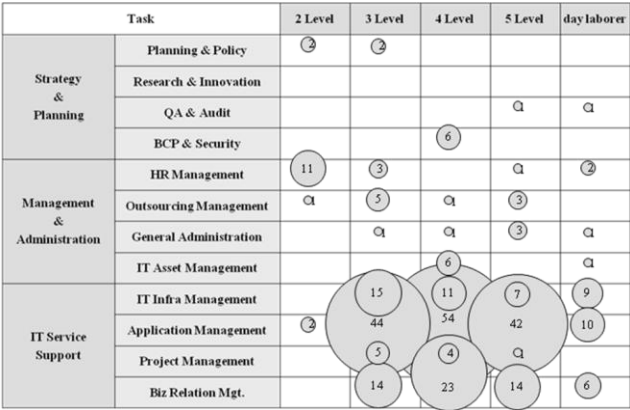


Fig. 1. K Corporation’s IT organization manpower level

2) Problems of IT management processes

K Corporation's IT Management diagnosed in terms of IT planning, IT infrastructure management, development and maintenance, IT operations in Figure 2.3. IT management represented lower level excepted infrastructure management.

IT processes diagnosis has a total of 21 indicators were analyzed, comparing the average target of 2.72. This case was investigated by approximately 62% versus a financial firm (4.47) and B government agency (4.33).

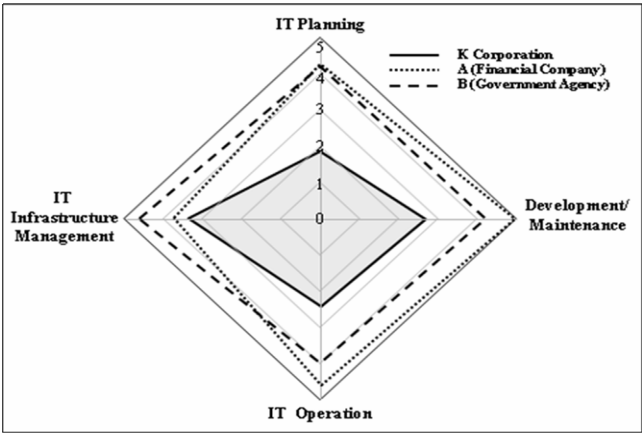


Fig. 2. K Corporation's IT Management diagnosis result

3) Problems of IT outsourcing management system

K Corporation was outsourcing in IT sector for development and operational perspective. But management system managed outsourcing companies and service level management system was inadequate. K Corporation's issue of IT outsourcing projects is as follows. First, K Corporation's primary system was due to high dependence on the maintenance companies, internal and external environment is difficult adapt to changes quickly.

Second, the similar system for the similar job outsourcing companied and K Corporation's business scope was not clear, as business interference and system failure, the responsibility is not clear judge of responsibility.

Third, K Corporation to lack of self-management based on the proposed outsourcing of management standards and procedures are being used, outsourcing is not good agreement with the business.

4) IT users' low satisfaction

K Corporation's IT user satisfaction and service satisfaction in the overall system have been investigated in Figure 3. Analyzed only work support system was high, key system like ticketing systems and an ERP system was lower.

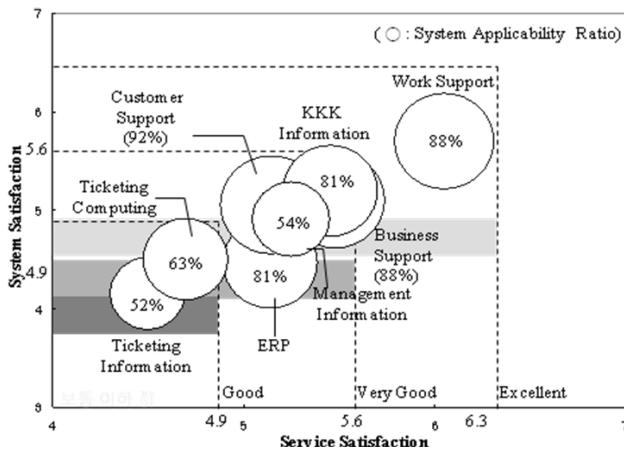


Fig. 3. K Corporation's IT Management Satisfaction

3 IT Organization Redesign

K construction's IT organization redesigned that procedures (1) job analysis, (2) To-Be organization design, (3) job redesign in Figure 4.

3.1 Job Analysis of IT Organization

K corporation's job analysis of IT organization were the JDS(Job Diagnostic Survey) to be developed, a member of the survey was created through. Existing JDS which has

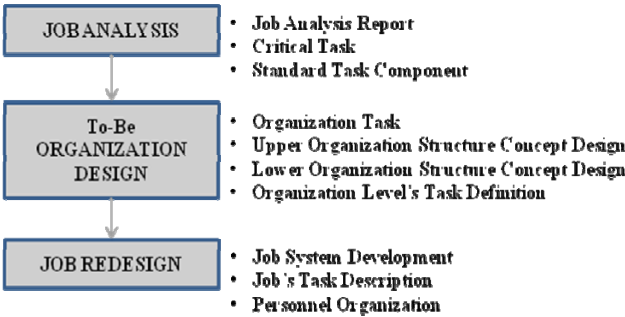


Fig. 4. IT Organization Redesign Process

five kinds of characteristics that key job has to some extent as a means to evaluate the Hackman and Oldham was developed in this study based on this JDS were used in the development of new forms [5].

Duties derived from the existing organization to perform analysis on the extent and distribution of personnel and the need to derive the missing service personnel to analyze the extent becomes.

1) Job analysis Scope of K Corporation's IT organization

K Corporation's job analysis of IT organization to target members of the JDS was conducted. JDS respondents participated 51 people in 49 respondents (96%), were 2 non-respondents in Table 2.

Table 2. K Corporation's JDS Respondents

Team	Respondent		Non-Respondent	Total
	Regular Worker	Contract Worker		
Information Management Team	9	1	-	10
Ticketing Technology Team	6	-	-	6
Ticketing Computing Team	8	1	-	9
Technology Support Team	11	2	1	14
BuKyoung Ticketing Team	3	1	1	5
Jeju Ticketing Team	5	2	-	7
Total	42	7	2	51

According to the JDS results K Corporation's IT organization was (1) strategy and planning, (2) Management and Administration, (3) IT service Support are classified and four kinds of business processes consists of a total 12 were classified as business processes.

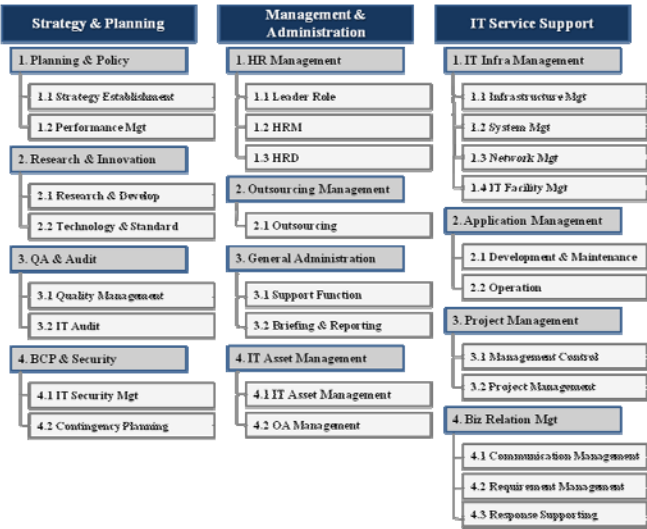


Fig. 5. K Corporation’s IT Organization Standard Task

Table 3. Team’s Key Works

Work			Work Perform					
Mega Process	Main Process	Sub Process	Information Management Team	Ticketing Computing Team	Technology Support Team	Ticketing Technology Team	BuKyoung Ticketing Team	Jeju Ticketing Team
1. Strategy & Planning	1.1 Planning & Policy	1.1.1 Strategy Establishment	●	*	*	*	*	*
		1.1.2 Performance Management	-	-	-	-	-	-
	1.2 Research & Innovation	1.2.1 Research & Development	-	-	-	-	-	-
		1.2.2 Technology & Standardization	-	-	-	-	-	-
	1.3 QA & Audit	1.3.1 Quality Management	-	-	-	-	-	-
		1.3.2 IT Audit	-	-	-	-	-	-
	1.4 BCP & Security	1.4.1 IT Security Management	-	-	*	-	-	-
2. Management & Administration	2.1 HR Management	1.4.2 Contingency Planning	-	-	-	-	-	-
		2.1.1 Leader Role	*	*	*	*	*	*
		2.1.2 HRM	*	-	-	-	-	-
	2.2 Outsourcing Management	2.1.3 HRD	*	-	-	-	*	*
		2.2.1 Outsourcing Adaptation	*	-	●	*	*	-
	2.3 General Administration	2.3.1 Support Function	*	*	*	*	*	*
		2.3.2 Briefing & Reporting	●	*	*	*	*	*
3. IT Service Support	3.1 IT Infra Management	2.4 IT Asset Management	-	-	●	-	*	*
		2.4.2 OA Management	-	-	-	-	*	*
		3.1.1 Infrastructure Management	-	-	●	-	-	*
		3.1.2 System Management	-	-	●●	-	-	*
	3.2 Application Management	3.1.3 Network Management	-	-	●●	-	●	●
		3.1.4 IT Facility Management	-	-	●●	-	●	●●
		3.2.1 Development & Maintenance	●●	●	-	●●	-	-
3.3 Project Management	3.2.2 Operation	●	●●	-	-	-	●	
	3.3.1 Management Control	-	-	-	-	-	-	
	3.3.2 Project Management	*	-	-	-	-	-	
	3.4 Biz Relation Management	3.4.1 Communication Management	-	-	-	-	-	-
3.4.2 Requirement Management		*	*	*	*	*	*	
3.4.3 Response Supporting		*	-	-	-	-	-	
notice: ●: full perform, *: part perform, - non perform, ●●: over perform								

12 business process-specific details task by constructing a total of 96 kinds of sub task were analyzed. Strategy and planning business processes sub task is the result in Figure 6.

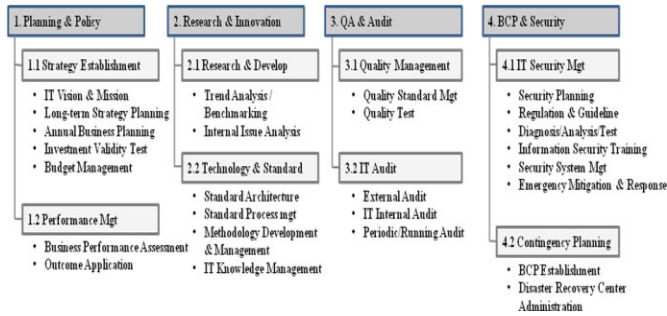


Fig. 6. Sub Tasks of Standard Tasks

2) K Corporation's job analysis of IT organization

Derived based on the task of IT organizations to perform at current level K corporation have been investigated. K corporation does not currently complete task on the sub tasks to be performed weakly 96 Tasks of the 39(40%), respectively. The main task of planning area were quality management, research and development that it does not perform such tasks were analyzed. Table 3 was derived from task unit to investigate the status of job performance.

The following is drawn according to the IT organization's work done in the current construction manpower for K Corporation has been investigated. As part of strategy and planning area and the senior general manager and was primarily responsible for the planning area, management and administration, including a senior manager and team members to handle two tasks for sharing. This area of IT service support personnel performs their duties due to high concentration if divided geographically distributed and was executed.

3) Missing tasks of K Corporation's IT organization

K Corporation's IT organization does not perform missing tasks in Figure 7. IT organizations need to reinforce it during IT organization redesignn.

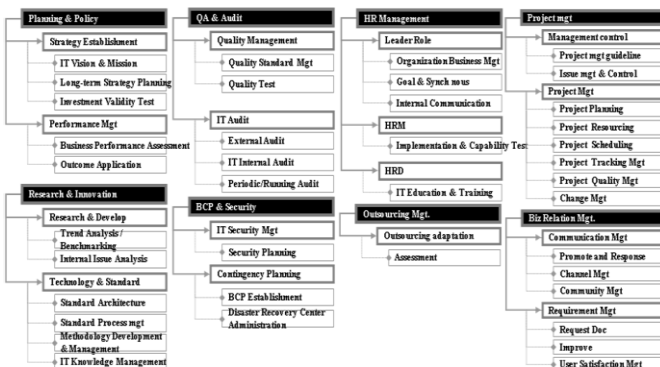


Fig. 7. Missing tasks

3.2 To-Be IT Organization Design

To-Be IT organization design were organizational tasks to derive the upper organization and lower organizations to developed and defined by organization's level.

First, standard IT service organization reflected on the design key task according to the standard IT task.

Second, the higher the organizational structure the conceptual design is configuration and sub-organizational structure. Designed by the front of the sub organizational structure based on the parent organization of expert meetings and working to reestablish the business group, and each team was assigned.

Third, the purpose of assignment organizational objective, the roles and responsibilities, reporting lines, the sequential tasks related (Relationship) consists of the tasks were defined.

1) IT Organization Redesign Concept

K Corporation's IT Organization constructed 5 teams by information management team, ticketing Computing team, technical Support team, IT development team, IT operating team, Bukyung ticketing team, Jeju ticketing team.

K Corporation's IT organization to organization designed based on a standard business (1) Planning, (2) Administration, (3) Project Management, (4) Development, (5) Maintenance, (6) RM(Relationship Management), (7) Application Service Management, (8) Infrastructure Operations were grouped into different functions. Thus K Corporation's IT organization is Figure 9.

2) Task Definition

Definition of Task Subscripted (1) Objective, (2) Role & Responsibility, (3) Report Line, (4) related granites (Relationship) was written consisted. Figure 8 is IT planning and a Business Information medical book of business written is an example.

IT Planning (Strategy & Planning)		
Objective	IT trends and technologies based on research and development for the long-term IT strategy, IT business and project promotion, deployment of resources for the strategic direction Presenting	
Role & Responsibility	- Industry, IT trends, new technology research, analysis, and provide relevant data - Chief Executive Officer, Chief and relevant departments to request specific ITEM research, analysis & reports - Development of IT vision, long-term IT strategy and short-term business plan through strategic Alliance and coordination - IT investment analysis and post-performance management (IT Project Feasibility Analysis and Evaluation of Investment Priorities) - Performance Indicators Pool Management - Information Strategy Committee Schedule / Agenda management and operational subjective - Chief of Information Technology support business activities - Teams, and individual KPI definitions and performance / competency assessment system development / Edit / Update	
Report Line	- Chief of Information Technology - Information Strategy Committee - Chief of IT Support Team	
Relationship	Pre	IT Support (Research & Standard), RM
	Post	IT Organization

Fig. 8. Description of IT Task Definition

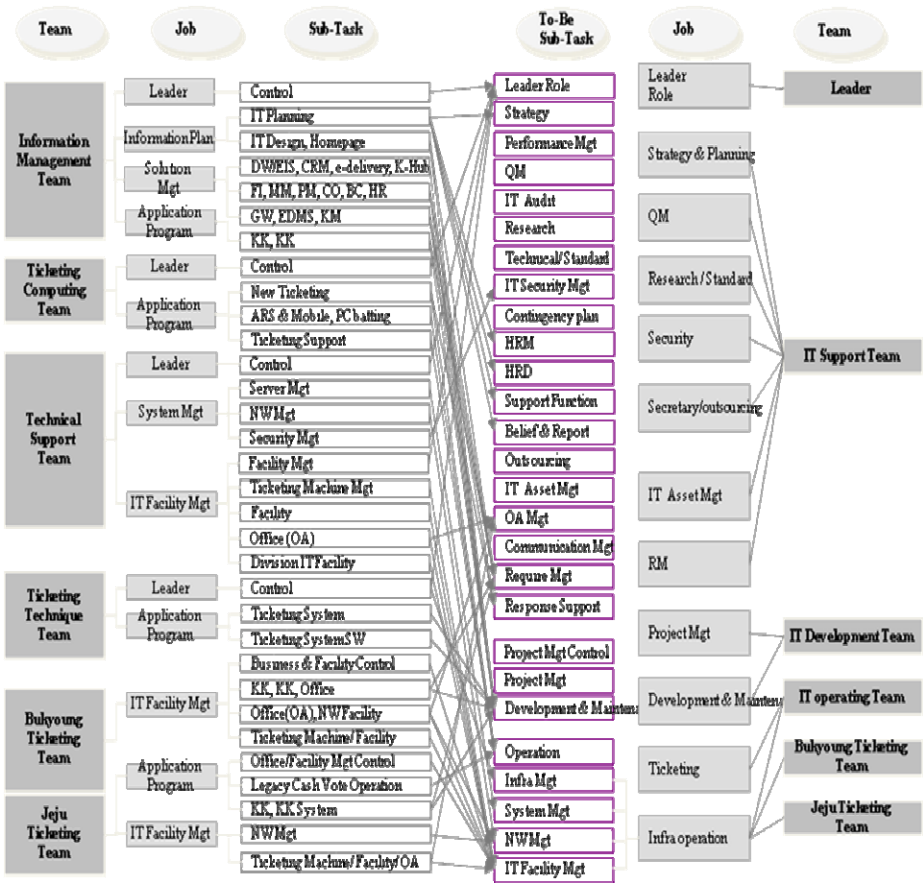


Fig. 9. IT Organization Conceptual Design

3) IT Organization Task Redesign

K Construction's IT organization redesign were IT organizations must re-established in accordance with the principles, To-Be organization designed to target job system is derived from the task definition and task function according to the detailed technical work plan proposes strengthening.

Through, IT organizations are required to estimate the number of people in this study to analyze the organization through change FTE and total costs were estimated by organization personnel.

FTE (Full Time Equivalent) analysis in the proportion of the IT organization's performance for the analysis is presented by the Gartner Group [6]. FTE analysis is IT organization tasks FTE (8 hours / day) was investigated and presented by the Gartner Group IT organization's office in the proportion of Global IT organizations to derive the number of people, Global Gap between the IT organization's personnel are analyzed.

FTE job performance through the analysis portion of the IT organization performance is part performance, separated by an attempted.

Through, IT organizations are required to estimate the number of people in this study to analyze the organization through change FTE and total costs were estimated by organization personnel.

Table 4. K Corporation's IT Organization FTE

Team	Task	FTE
IT Support Team	Strategy/Performance/Contingency Planning	0.7
	Research/Standard	0
	Quality/Audit/Security	1.3
	IT Office	1.05
	Asset/OA/Outsourcing	2.5
	RM	0.1
IT Operation Team	Application Operation	7.89
	System Operation	4.81
	IT Facility Mgt	4.3
IT Development Team	Cash/Account Voting	1.6
	Middleware/ Ticketing	7.6
	KKKK	3.05
	Non-ERP	3.45
Bukyoung Ticketing Team	IT Facility Mgt	2.19
	System Opration	0.66
JeJu Ticketing Team	IT Facility Mgt	2.5
	System Opration	0.3
Total		44

4 Discussions and Conclusion

This study redesigns in order to adapt the needs of IT organizations have been filed, the existing construction of the IT organization's internal staff and K Corporation's outsourcing with the appropriate number of persons depending on the requirements for configuring the system was studied.

Redesigning the IT organization to research the application of tasks were obtained through job analysis, organizational, and business was redefined. In addition, IT personnel in FTE are calculated to estimate task and review.

First an analysis on IT organization of K Corporation was conducted. Internal personnel to conduct job analysis and targeted (1) strategy and planning, (2) Management and Administration, (3) IT services and were divided into 3 kinds, each consisting of 4 business processes, business processes, the total was composed of 12 kinds. Based on this, the details of 96 tasks were derived.

Second, IT organization of K Corporation was designed. Detailed job profiles and perform a number of performance targets Survey was restructured through the IT organization. Existing Information Management, Information Systems Team released, technical support, release engineering team, Bukyoung ticketing team, Jeju ticketing team, including IT support, IT support, IT development team, Bukyeong ticketing team, Jeju ticketing team and five teams were reconstructed. In addition, by developing an organizational structure based on it, each of 10 tasks was assigned to teams.

Third, K Corporation redesigned the job of the IT organization. Job duties and organizational redesign mapping method of calculating the number of FTE are used for the estimated number of IT people. Through the estimated number of people to target three rounds of review, internal personnel with adjustment personnel are 44 people.

This study required a redesign of the IT organization and how to process K Corporation's IT organization redesign business practices to target the IT organization has been studied. Implications are as follows.

First, K Corporation's as a public institution and the organizational and corporate IT personnel, and business and personnel for Outsourcing is the difference between configuration and so forth. Students have many issues raised by outsourcing your IT organization promoting the redesign, but as a result a lot more outsourcing workforce was composed of internal personnel. However, the IT organization's core business is responsible for internal personnel, IT management outsourcing for parts by personnel management in IT operations and management was to ensure continuity.

Second, IT organizations of persons estimated using the FTE calculations, but calculations indicated that there is a limit to the number of people applying. Redesign principles and strengthen the destination FTE, FTE adjustment target, FTE remain subject to review and adjustment and due to the determination of the appropriate personnel were forced to. Due to this limitation in calculating the number of IT organizations to more quantitative methods based on estimating the development of standardized personnel will be needed.

References

- [1] Mani, D., Barua, A., Whinston, A.: An Empirical Analysis of the Impact of Information Capabilities Design on Business Process. *Outsourcing Performance. MIS Quarterly* 34(1), 39–62 (2010)
- [2] Ministry of Public Administration & Security, *IT Outsourcing Operation Management Manual*, (2009)
- [3] National IT Industry Promotion Agency, *Diffusion and Technical Evolution of IT Risk Management*, SW Industry Review, pp.1-7 (2007)
- [4] Yoon, S.B., Lee, S.C.: Restructuring the MIS Department. *Information Systems Review* 3(1), 115–129 (2001)
- [5] Hackman, J.R., Oldman, G.R.: Development of the Job Diagnostic Survey. *Journal of Applied Psychology* 60(2), 159–170 (1975)
- [6] Scardino, L., Young, A., Maurer, W.: Common Pricing Models and best-use Case for IT Services and Outsourcing Contracts, Gartner (September 2005)
- [7] Goo, J.H., Kishore, R., Rao, H.R.: The Role of Service Level Agreements in Relational Management of Information Technology Outsourcing: An Empirical Study. *MIS Quarterly* 33(1), 119–145 (2009)

MF-SNOOP for Handover Performance Enhancement

Chi-Hyun Ahn¹, Hyoungchul Kim², and Jongjung Woo³

¹ Professor, Seowon University, Computer Engineering,
Korea, +82-10-5415-6393
jfsodmes@chol.com

² Professor, Open Cyber University, Information & communication Engineering,
Korea, +82-2-2197-4175
hckim@ocu.ac.kr

³ Professor, Sungshin Women's University Computer Engineering,
Korea, +82-2-920-7193
jwoo@sungshin.ac.kr

Abstract. Wireless network has high BERs because of its path losses, fading, noises and interferences. Particularly, in the TCP and MIP environments, the network is often disconnected because of the handover. To solve such problem, the Freeze-TCP mechanism is proposed, but during the handover in the Network layer, the mobile node can't receive packets. In addition, it cannot handle traffic with high BERs. SNOOP hides packet losses for Fixed Host(FH) and retransmit lost packets in wireless network. However, SNOOP has a weakness for burst errors in wireless network. This paper proposes the MF-SNOOP that loads Enhanced SNOOP modules on the MAP and maintains TCP connectivity during network layer handovers. The Enhanced SNOOP module performs multiple local retransmission for burst errors. By buffering in MAP, MF-SNOOP uses Zero Window Advertisement(ZWA) Messages of Freeze-TCP. Thereafter, MN finishes the handover immediately, and receives its packets.

Keywords: Wireless-TCP, Freeze-TCP, SNOOP, MF-SNOOP.

1 Introduction

In the future, internet users want to use high quality internet services regardless of time and places. The number of internet users becomes greatly large because of the performance enhancement of mobile terminals such as smart phone and the development of wireless communication technology. To support the real-time multimedia service such as e-business or e-learning and the traffic service of high QoS, the communication system must support ideal mobility in the next generation. Suppose all mobile terminals employ IP addresses, users would receive their private service with link layer independently and the roaming problem would be disappeared. However, the previous cellular system supports mobility in 2-level layers but it could not provide real-time application service ideally because of its bandwidth or high costs.

The IETF(Internet Engineering Task Force) proposed Mobile IPv6(MIPv6)[1] and Hierarchical Mobile IPv6(HMIPv6)[2], Fast Mobile IPv6(FMIPv6)[3] which is an extension of MIPv6. HMIPv6 is designed to minimize signaling loads from the mobile node(MN) to the correspondent node(CN) and the home agent(HA). Thus, each domain is managed by the protocol element called Mobility Anchor Point(MAP)

On the other hand, the wireless link has 10^{-2} ~ 10^{-4} high BER(Bit Error Rate) characteristic because of its path losses, fading, noises and interferences. Therefore, lots of packets would be lost even though any congestion is not occurred in the wireless link. But TCP regards that the reason of such packet losses is traffic congestion in the wired link, thereby performing unnecessary congestion control such as fast retransmissions and slow starts [4].

In this paper, we propose MF-SNOOP that loads SNOOP modules on the MAP to keep continuous TCP connection in the HMIPv6 whenever the handover happens. MF-SNOOP blocks the packet traffic pass through MAP and performs flow control by using Freeze TCP's ZWA message while handover occurs. Furthermore, the SNOOP module that is included MF-SNOOP's MAP performs continuous local retransmissions by employing the sequence number to solve SNOOP's burst error problem.

The rest of the paper is organized as follows. Section 2 provides the related work: Freeze-TCP and SNOOP. Section 3 describes the proposed scheme, MF-SNOOP. The performance evaluation of MF-SNOOP is presented in section 4 and concludes the paper in section 5.

2 Related Work

2.1 Freeze-TCP [5]

Freeze-TCP[5] was designed to prevent undesirable timeouts which lead to unnecessary slow-start and congestion avoidance phases of an already active connection. When a MN is handed over to a new wireless base station, To avoid timeout, once notified of the impending handover by the MAC layer, the MN transmits an acknowledgment carrying a zero window advertisement(ZWA) message. Upon receiving the ZWA, the remote FH freezes all data transmissions and enters a zero window probing mode, during it attempts sending one-byte probes at intervals specified by its persistence timer, until the MN is reconnected. In addition, once the MN is reconnected at the new base station, it transmits a triple acknowledgment(TR-ACK) with a non-zero window advertisement(NZWA), enables the remote FH to resume its data transmission. It is important to observe that Freeze-TCP does not require any modifications whatsoever in the remote FH; the latter is merely expected to proceed as in normal TCP upon receiving a zero window acknowledgment. No changes are called for in the base stations either. Therefore, Freeze-TCP can be naturally adapted to our mobile networking context, with the only exception being that the ZWA/NZWA advertisements are initiated by the AR, rather than the MN itself. Fig 1 is shown Freeze-TCP operation.

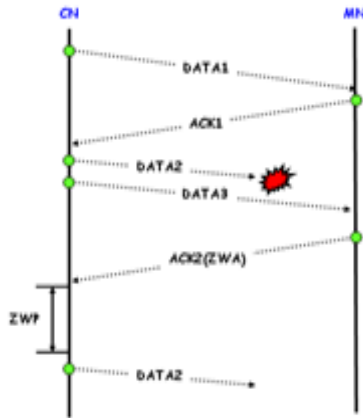


Fig. 1. Freeze-TCP

2.2 SNOOP [6]

The SNOOP protocol provides a reliable solution while maintaining the end-to-end semantics of the transport layer connection. It's a TCP-specific approach in TCP-aware module, a SNOOP agent in base station. The agent monitors every packet that passes through the BS in both direction and maintains a cache of TCP segment which sent to MH, but have not yet been acknowledged.

For data transfer from a FH to a MN, the SNOOP agent caches unacknowledged TCP segments, forwarding to MN, and monitors the corresponding ACKs.

SNOOP in Fig2 operate as follow.

1. Retransmission of lost packets locally by using local timers or TCP duplicate acknowledgements to identify packet loss, instead of waiting for the FH to do so.
2. Processing the duplicate ACKs on their way from the MH back to the FH, thus avoiding fast retransmit and congestion control at the later.

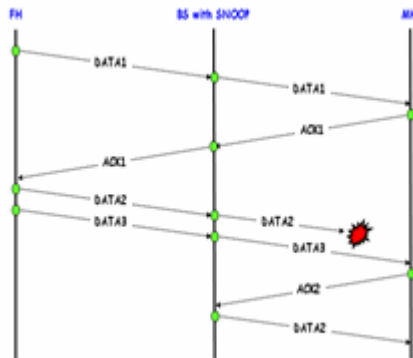


Fig. 2. Local retransmission of SNOOP

The main advantage of the SNOOP protocol is that lost packets are retransmitted locally, thereby avoiding unnecessary fast retransmission and congestion control invocation by the original sender. But if the original sender does not receive any acknowledgment from the receiver during local recovery period, then it may cause a retransmission timeout or idle time in the sender side due to limited window size.

3 Proposed Scheme

3.1 Design Concept

The existing SNOOP can manage fast recovery of error by local retransmission, but when a burst error occurs, it is not good performance of error recovery.

In MF-SNOOP, A single error on wireless-link is solved the problem by local retransmission such as SNOOP. When a burst error occurs, MF-SNOOP performs retransmission after freeze BS and FH by using M/F-SNOOP in BS. If retransmission is failed, it is attempted retransmission by suspending freeze-time

The buffering accomplishment time at FH and BS set by follow formula.

FH buffering time = $FM\_RTT \times \text{the number of burst error packets}$

BS buffering time = $SM\_RTT \times \text{the number of burst error packets}$

FM-RTT : Round Trip Time(RTT) from FH to MH

SM-RTT : Round Trip Time(RTT) from BS to MH

3.2 MF-SNOOP

Existing Freeze-TCP does not send data from FH after makes the window size during hand over. This scheme make seamless TCP connection during handover. But this scheme have problem that it could not receive data during handover and have to request retransmission to FH after abandon transmitted data. So it's not good for real-time multimedia service, since too high data packet loss In addition, SNOOP splits wired-link and wireless-link based on BS(Base Station). When it occurs packet loss in wireless-link, SNOOP makes improvement of TCP performance, since FH not run fast retransmission and congestion control by re-transmitting the local in wireless-link. However, this scheme has such risk of timeout of FH because the retransmission time increases when burst error happens.

To solve these problems, MF-SNOOP load improved SNOOP module on MAP and maintain connectivity about occurring 3 layer handover. In addition, it is proposed for real-time multimedia service, because MN receives transmission immediately without destroying before transmitted packets after handover. Enhanced SNOOP module performs local multiple retransmission to solve previous SNOOP's problem for remove burst error.

If MF-SNOOP sends Zero Window Advertisement(ZWA) message as soon as predict MN's mobility using L2 trigger before starting handover. The MAP which received ZWA message starts buffering while doesn't send data from FH to MN. Buffering is recognized using 1 bit 'B' flag of ZWA message.

After MN moves to new domain, MN acquires new CoA and send BU message to MAP. MAP stops buffering and sends packet to MN.

The SNOOP module in MAP of MF-SNOO is performs local multiple retransmission using sequence number to solve the problem that previous SNOOP burst error. When error occurs, it prevent sending the data from MAP to MN. To solve previous SNOOP's performance reduction from burst error, MF-SNOOP retransmits as same method, previous SNOOP,

Previous SNOOP must receive duplicated ACK for local retransmission, but MF-SNOOP performs retransmission on MAP after MF-SNOOP sends all of lost packet's sequence number using option field and send packet acknowledgement of burst error from MN . MF-SNOOP sends lost packets together so retransmission is faster.

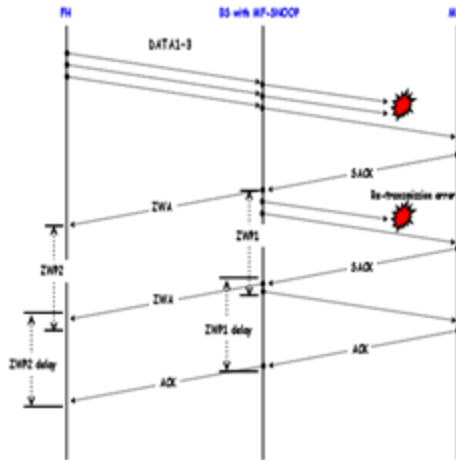


Fig 3. Signaling of MF-SNOOP in case burst error

Fig 3 is signaling of SNOOP and F-SNOOP when burst error occurrence in wireless link.

If the packets which sequence number 2 and 3 are lost, SNOOP received duplicated ACK2. And then, SNOOP re-transmit number 2 packet. We can see a persist timeout that when SNOOP received duplicated ACK3 after transmitting number 2 packet.

Network efficient decreases by accumulated retransmission timeout since burst error occurs. But MF-SNOOP send to ACK2(2,3) messages including 2,3 packets were lost in MN to the MF-SNOOP. MF-SNOOP which receives the ACK2(2,3) re-transmits the MN with 2,3 packets. Simultaneously, MN sends the ZWA to the MF-SNOOP. MF-SNOOP performs buffering during re-transmit to MN.

If MN retransmits all lost packets in buffering time, In other words, If MAP receive ACK message that transmit to FH, MF-SNOOP remove buffering and transmit the packets to MN normally.

4 Simulation Performance

4.1 Simulation Platform

We evaluate the performance of the proposed scheme using the Network Simulator(ns-2) [10] distribution version ns-allinone2.1b6. The network topology for most of the simulations is illustrated in Fig 4, which is a generic HMIPv6 network. The wireless coverage area of the access point is approximately 100 meters, and the router advertisement is once per second. We set the link layer handover delay to 200ms in the simulation. MN in the simulation moves linearly from one access router to another at a constant speed of 5m/s. Also Simulation Traffic parameters for Enhanced SNOOP are table 1.

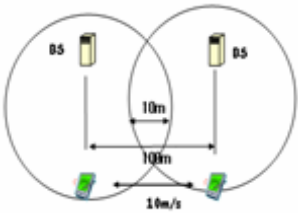


Fig 4. The network topology

Table 1. TCP parameters

Parameter	value
TCP Initial RTO	2.0(s)
TCP Maximum RTO	64(s)
TCP Minimum RTO	0.5(s)
TCP MSS	1500 Bytes
TCP Maximum ACK Delay	0.01(s)
TCP Windows Size	30 segment
PER(Packet Error Rate)	0.5% ~ 2%

4.2 Result

Fig 5 represent the simulated result of TCP window size variation for SNOOP and MF-SNOOP while transmit data. Handover occur once per 20 seconds and total simulated time is 80 seconds. We can see that TCP windows size of SNOOP reduce remarkably in the handover occur and recovery after handover finish. but, TCP

window size of MF-SNOOP is stable continually. In case of MF-SNOOP, MAP intercepts data by ZWA message and buffer data when handover occur. During handover, TCP window size variation is related to connectivity of TCP. If handover occur, TCP connectivity of SNOOP is cut off. But, TCP connectivity of MF-SNOOP is stable independent of TCP window size variation.

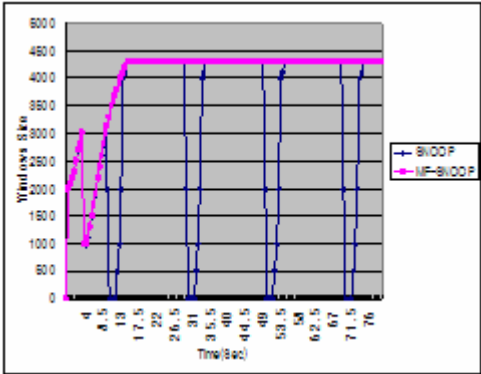


Fig. 5. TCP Window Size Variation

Fig 6 is the result that simulate transmission rate in environment that wireless link bandwidth is 144Kbps. PER is 0.5%~2%. We can see that MF-SNOOP is improved average 14% compare to SNOOP. Specially, Recording where the PER will be high, the gap increase. High PER mean possibility of burst error is high. The reason is because of the control of burst error. SNOOP performs local retransmission but MP-SNOOP performs multiple local retransmission. Thus, as burst error is increased, that is, Recording where the PER will be high, the effect of multiple local retransmission increase.

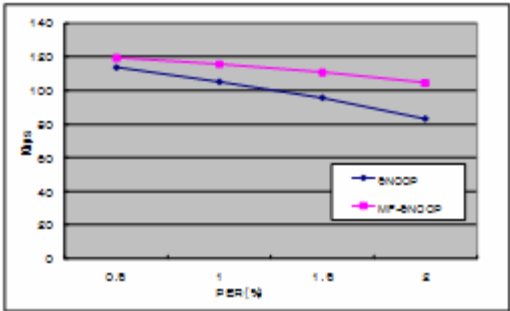


Fig. 6. Transmission Rate as PER We evaluate the perform

5 Conclusions

The wireless link has 10^{-2} ~ 10^{-4} high BER characteristic because of its path losses, fading, noises and interferences. Therefore, many packet losses occur without any congestion on wireless-link. SNOOP that proposed to solve the problem is improved performance by using local retransmission. But, when burst error is occurred, FH becomes timeout by flow control. Therefore, many packet losses occur without any congestion in wireless-link. SNOOP that proposed to solve the problem is improved performance by using local retransmission. but, if burst error is occurred, FH becomes timeout from time to time, and Freeze-TCP that proposed to maintain TCP connectivity prevent undesirable timeouts which lead to unnecessary slow-start and congestion avoidance. but during handover, MN isn't able to receive the packet. Thus after dropping received packets during handover, MN must receive packets again and must require retransmission to FH. Because of these problems, MN difficult to receive real-time multimedia service.

In this paper, we propose MF-SNOOP that loads Enhanced SNOOP module on MAP. MF-SNOOP maintains TCP connectivity when handover happens in HMIPv6. MAP including Enhanced SNOOP module performs buffering received packets from FH during handover. Furthermore Enhanced SNOOP module that is performs multiple local retransmission to solve burst error problem by using sequence number of lost packets.

Simulation was window size variation and transmission rate as PER by using NS-2. MF-SNOOP was maintained TCP connectivity during handover and transmission rate as PER variation. Particularly, the result was shown improved performance average 14% about transmission rate when compared MF-SNOOP with SNOOP.

References

- [1] Johnson, D., Perkins, C.: Mobility Support in IPv6, IETF draft, draft-ietf-mobileip-ipv6-15.txt (July 2001)
- [2] Soliman, H., Castellucia, C., Elmalki, K., Bellier, L.: Hierarchical MIPv6 Mobility Management (HMIPv6), internet draft (July 2001), draft-ietf-mobileip-hmipv6-0.5.txt, work in progress
- [3] Koodli, R.: Fast Handovers for Mobile IPv6, IETF draft, draft-ietf-mipshop-fast-mipv6-0.1.txt (January 30, 2004)
- [4] Stevens, W.R.: TCP Slow Start, Congestion Avoidance, Fast Retransmission, and Fast Recovery Algorithms, IETF, RFC 2001 (January 1997)
- [5] Goff, T., Moronski, J., Gupta, V.: Freeze-TCP: A true end-to-end TCP enhancement mechanism for mobile environments (1995)
- [6] Balakishnan, H., Seshan, S., Kartz, R.H.: Improving reliable transport and handoff performance in cellular wireless networks. ACM Wireless Networks 1 (December 1995)
- [7] Bakre, A., Badrinath, B.R.: I-TCP: Indirect TCP for Mobile Hosts. In: Proceedings of the 15th International Conference of Distributed Computing Systems (June 1995)
- [8] Brown, K., Sigh, S.: M-TCP: TCP for mobile cellular networks, ACM Computer Communication Review 127(5) (October 1997)
- [9] Sinha, P., et al.: WTCP: A reliable transport protocol for wireless wide-area networks. In: Proc. of ACM Mobicom 1999 (August 1999)
- [10] NS-2, <http://www.isi.edu/nsnam/ns>

The Influences of On-line Fashion Community Network Features on the Acceptance of Fashion Information

Kieun Song¹, Sunjin Hwang¹, Yunsik Kim¹, and Youngsik Kwak²

¹ Sungkyunkwan University

² Gyengnam National University of Science and Technology
songcci@gmail.com

Abstract. The purpose of this study is to determine the major influential factors, including network, in the acceptance of fashion information in the on-line fashion community. Communities that actively exchange fashion information among their members were selected for social network analysis and hypothesis verification. In the results, it was determined that the network features of the fashion community were concentrated on a few information activists that appeared to influence most of the information acceptors. Moreover, in the results of verifying the hypothesis through surveys on the influential factors for verbal acceptance, it was found that the informational characteristic variables, individual characteristic variables as well as the network characteristic variables, have a significant influence on the verbal acceptance which verified the importance of the network characteristic variables in the study on the word of mouth advertising on the internet.

Keywords: e-WOM, fashion information, on-line fashion community, social network.

1 Introduction

Many companies have created their own mini-webpages or blogs to advertise their products and services by increasing the marketing effect through word of mouth marketing, and fashion companies were no exception. Especially, fashion product is one of the product groups in which word of mouth marketing, such as referrals, highly affect consumer decision-making [1]. Therefore, many fashion companies and distributors have tried to increase word of mouth marketing opportunities between general consumers, when launching new product lines or brands, by holding fashion shows or using celebrity events.

In order for fashion companies to advertise their products, the effective online word of mouth marketing strategies and tactics can induce voluntary word of mouth advertising. One of the marketing communication tools that can be utilized here is the fashion community network marketing. Word of mouth marketing via the internet through a well established network has excellent word of mouth marketing effect with a fast delivery speed to broad areas. Especially, when considering that fashion information spreads through the internet fashion communities, a study on the effects

of network for fashion communities can be expected to have high marketing performances.

2 Literature Review

The internet societies that establish relationship network by means of computers and networking are social relationships that are established to accomplish common goals such as social network management or information exchange. Recently such social network services have become popular. The word of mouth advertising between consumers that is formed along with such relationships has also been considered as very important [2].

2.1 e-WOM Network

An internet communication environment that is mediated via a computer is called a CMC (computer mediated communication) environment. Brickart and Schindler stated that e-WOM (electronic Word of Mouth) is exchanging information such as product information, product reviews, or recommendations between consumers through the internet [3]. Many definitions about e-WOM are also based on CMC. The internet features have changed communities from groups to networks [4]. A network is a tangible and intangible group consisting of nodes and links, and therefore individual relationships, characteristics, and patterns of the social connections can be figured out through the connection network of the internet [5]. Especially, email, instant messaging, and online chat communities are, representatively, featured with weak ties. Granovetter emphasized the importance of 'the power of the weak tie' of the network [6]. The weakly tied relationship decreases the cost of relationship formation between people while maintaining the informational flow between them [7].

Therefore, e-WOM marketing can be considered as a social exchange activity that involves the sharing of information through the internet-based community network. Since fashion attends to follow trends, fashion products also utilize the social communication fashion which means it spreads from a fashion leader to a follower group [8]. Therefore, the sharing of fashion information that spreads through the network, and the formation of sympathies, are some of the most important factors that determine the success of the consumer-oriented fashion market in the fashion industry of the 21st Century. For such reasons, the studies on the acceptance of WOM that considered the relational feature of the internet network are required.

2.2 Influential Factors of e-WOM on Fashion Information

In order to determine the network relationship and the information exchange features of e-WOM, the influential factors of the multi-dimensional features need to be considered. Most of the previous studies found that the informational characteristics and individual characteristics of e-WOM are influential factors for the acceptance of e-WOM. Furthermore, they studied the structural characteristics of the network.

Informational Characteristics. Informational characteristics are recognized to be very important since the information exchanged on the internet consists of consumer

opinions or interests, and is uploaded by the consumers on the internet bulletin boards as text, pictures, simple replies, or scraps [9]. For such reasons, studies on the influences of e-WOM have focused on the informational characteristics [10]. The informational characteristics consist of usefulness of information, interest, and reliability.

Consumers are highly interested in information that is highly informative, highly connected, or that is equipped with good, timely mannered usefulness, and they tend to behave or act through active communication [11]. Meanwhile, it was found that the effect of the rational information process and emotional information process on WOM information is different. While the rational information process is made in a considerate and efficient way focusing on problem-solving, the emotional information process is made in the emotional way focusing on whether or not it can provide pleasure and satisfaction [12-13]. Fashion products are emotion-oriented products; and therefore, it can be assumed that the interest factor is also an important influential factor inducing the emotional information process. Also, since communication is made in anonymity and through text on the internet, reliability has been considered as a very important influential factor on e-WOM studies.

H₁. When the (a) usefulness, (b) interest, and (c) reliability of the information cognized by and internet fashion community participant is higher, more positive influences on the acceptance of WOM are made.

Individual Characteristics. Individual characteristic variables that affect e-WOM are needed for cognition, innovation, and demographic characteristics such as gender and age. It is very possible that the high need for cognition may be formed through a deep consideration process; and therefore, the relation with intension of behavior is higher than the consumers who had low need for cognition [13]. Fashion consumers who have high need for cognition are very interested in and search for the fashion information on the internet also they seemed to share the information actively due to the their strong desire to structuralize it. Therefore, it is assumed that when their need for cognition is high, they will form positive acceptance attitude on e-WOM fashion information.

Consumers who have high innovative characteristic are called opinion leaders or market mavens and are known to be influential over other consumers in the decision-making and product distribution processes [14]. Moreover, it was proved from a series of studies on consumers' innovative characteristics that they have an influence on the acceptance of new products or convergence products [15].

H₂. When an internet fashion community participant has a higher individual (a) need for cognition, and (b) innovation, more positive influences on the acceptance of WOM are made.

Demographic characteristics such as age and gender are related to internet user interests. According to Lee et al. younger consumers are more likely accept information from the internet. Also the consumers with higher knowledge on fashion are more likely to search for product information before they purchase a product [16]. Generally, it was known that more females than males and younger people are more interested in fashion. Therefore, age and gender characteristics are expected to have a

more positive influence on the acceptance of WOM fashion information on the internet.

H₂. Among the fashion community participants on the internet, (c) women and (d) younger participants will have a more positive influence on the acceptance of WOM than men and older participants.

Network Characteristics. Among the studies with the social network viewpoint in terms of the information exchange relationship within consumers' community, Kim et al. studied information distribution according to the structural characteristics of the network and compared fashion networks formed through offline friendships and social networks formed through friendships, and found that fashion networks are more concentrated and centered around a few members compared to the friendship network [17]. The results of this study showed that the network that is connected by means of an informational relationship has different characteristics than a friendship building network, which presented the necessity to conduct an independent study on fashion information networks. Network characteristics are activity, connectivity, and power.

The members who actively participate in the network structure are highly connected to other members and are situated in the center, so they tend to occupy a strategically important place and can be very influential. Therefore, they will not be situated in the duplicated relationship with other members in the activity level, and their connectivity of efficiently forming the network may positively influence on the acceptance of WOM. Moreover, the power, which means the level of connectivity with important members, shows the power dependency relationship. If the power is low, it is very possible that one will actively engage in the information activities in order to acquire high-quality information. From empirical studies, it was found that low power increases the market performance. Therefore, the positive acceptance of WOM can be expected with low power through the active information pursuing activities.

H₃. When the fashion community participants have high (a) activity (b) connection, and low (c) power through the network, they can have a more positive influence on the acceptance of WOM.

3 Methodology

Data Collection. Fashion information data was collected from June 2010 to September 2010 from a fashion community that provides internet SNS(Social Network Service) and the fashion community members were asked to participate in an internet survey from September 16, 2010 to October 5, 2010. We received 167 usable responses. In order to analyze the influence of the fashion community characteristics on the acceptance of e-WOM based on literature reviews, the social network analysis and survey analysis were also conducted. For the analysis methods, UCINET6.0 was used to analyze the network patterns of the fashion community and SPSS 17.0 was used to perform the multiple regression analysis.

Measures. The refined scales were the result of exploratory factor analyses. An exploratory factor analyses was performed with principal components extraction and varimax rotation on all scales used to examine the effect of acceptance e-WOM. The result showed the anticipated factor structure; items loaded highly on the constructs they were intended to measure. Correlations among measures appear in Table1. And, the appendix shows both the final scale items and reliability.

Table 1. Correlations

		Information			Individual			Network			
		1	2	3	4	5	6	7	8	9	10
Information	Usefulness	1									
	Interest	-.030	1								
	Reliability	-.018	.010	1							
Individual	NFC	.308**	.139	.257**	1						
	Innovation	.153*	.051	.330**	.016	1					
	Gender	.074	.058	.020	.044	.083	1				
	Age	-.009	-.077	-.135	-.040	-.109	-.219**	1			
Network	Activity	.095	.111	.075	.176*	.054	.062	.014	1		
	Connectivity	-.117	-.050	-.195*	-.261**	-.112	-.064	.045	-.617**	1	
	Power	.039	-.099	-.016	.107	-.014	-.011	.016	.213**	-.118	1

*p<0.5, **p<0.01

4 Results

4.1 Network Characteristics of the Fashion Community

Social Network Analysis. In order to understand the influence of e-WOM, a network analysis was conducted in order to determine how the information is distributed in a fashion community. Analysis units were ‘fashion’, ‘luxury brand’, and ‘beauty’ from SNS fashion information community and two levels of subcategories of those menus. By combining these three menus, the fashion information network was established using 1,447 nodes. (See Fig.1)

Among them, the total 509 members were interactively biconnected and shared information; however, the members who formed the consumers’ power through the mutual information exchange consisted of around 30%. When these results were analyzed in the entire network aspect, it was found that the members who actively participate in the information flow and acceptance in the fashion community are not proportional. Also it was found that the network characteristics which represent the members’ relationship to evaluate the influence in the community are very critical factors even in a large fashion community with high number of members.

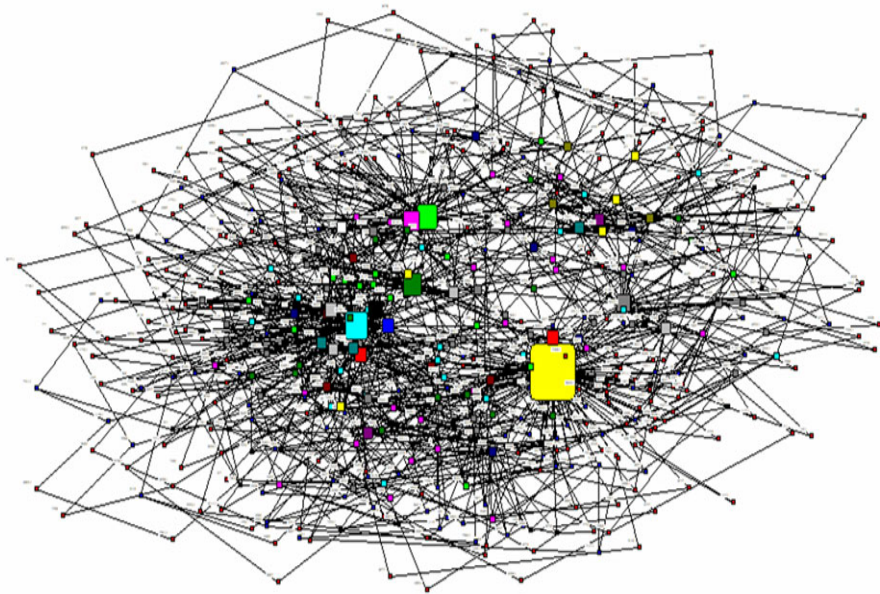


Fig. 1. Fashion Community's connections network

In the connection link characteristics of the fashion members' nodes, one node that had the most connections was connected with 104 nodes and exchanging information with approximately 20% of all of the members and the next node with the high number of connection was connected with 50 nodes; however, the nodes that had the least number of connections, a total of two, numbered as many as 215, which showed that most members were connected to a few nodes to obtain fashion information.

It showed the structural characteristics of the network in which the entire members of the fashion community network depend on a few fashion information activists who

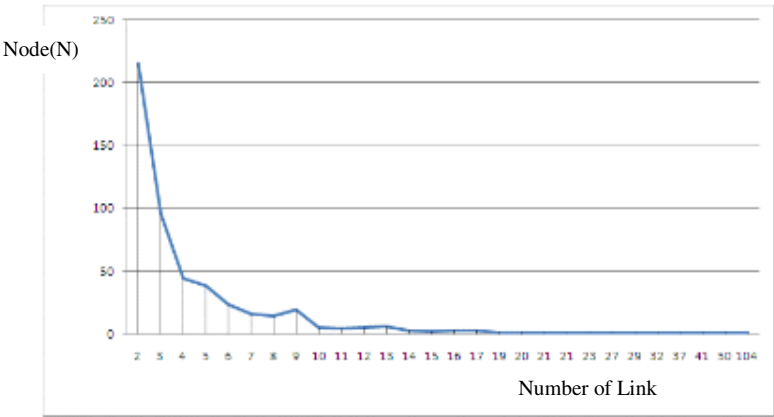


Fig. 2. Number of linked nodes

have high influence on the fashion information. Such a concentration of information demonstrates ‘the rich get richer and the poor get poorer’ phenomenon, as Barabasi mentioned [18], and also shows that such structural characteristic of the network influence the entire fashion community. In order to verify the influence of e-WOM information on the fashion community based on the studies on the network characteristics in the information distribution, the network characteristics factors were added to the existing informational and individual characteristic factors and verified their influence on the integrated acceptance of e- WOM.

4.2 Hypothesis Testing

Hierarchical Multiple Regression Analysis. Hypotheses (H₁-H₃) were verified through hierarchical multiple regression analysis. Examination of variance inflation factors (VIFs) among explanatory variables indicated that the highest VIF was 1.71, suggesting that there are not multicollinearity problems [19]. Hierarchical multi regression analyses were conducted to examine how the network features have an effect on the acceptance of e-WOM fashion information (See Table 2).

Table 2. Hierarchical multi regression analysis

Variables		Model(1)			Model(2)		
		Coefficients	Standardized coefficients	t-value	Coefficients	Standardized coefficients	t-value
Information	Usefulness	.272	.368	5.668***	.266	.359	5.647***
	Interest	.149	.196	3.220***	.133	.176	2.906***
	reliability	.190	.256	3.813***	.172	.232	3.491***
Individual	NFC	.112	.148	2.201*	.099	.130	1.931*
	Innovation	.104	.136	2.082*	.098	.128	2.000*
	Gender	-.004	-.003	-.047	-.018	-.012	-.198
	Age	-.025	-.174	-2.800**	-.025	-.175	-2.871**
Network	Activity				10.847	.041	.534
	Connectivity				-.312	-.138	-1.798*
	Power				-3.792	-.123	-2.028*
R ²			.431			.466	
Adjusted R ²			.406			.432	
F			17.232***			13.612***	

*p<0.1, **p<0.05, ***p<0.01

As we show in Table3, by including network variables, the explanatory power was improved from $R^2 = .431$ to $R^2 = .466$. The change in R^2 was a significant increase in the coefficient of determination.

Table 3. Change Statistics

Model	R^2 Change	F Change	df 1	df 2	Sig. F Change
Including network variables (1)	0.431 ^a	17.232	7	159	.000
Including network variables model's explanatory power (2)	0.035 ^b	3.369	3	156	.000

^a R^2 Change is compared with Null model

^bModel (2)-(1)

The effects of informative features on the acceptance of e-WOM. It was found that the usefulness of the information was positively related to acceptance of e-WOM ($b=.266$, $t=5.647$, $p<0.001$), which supported H_1 -a. It was predicted that the information usefulness would have a positive relationship to acceptance of e-WOM. It was found that the interest ($b=.133$, $t=2.906$, $p<0.001$) and reliability ($b=.172$, $t=3.491$, $p<0.001$) in information was positively related to acceptance of e-WOM, which supported H_1 -b and H_1 -c. Therefore, it was also predicted that the information interest and reliability would have a positive relationship to acceptance of e-WOM.

The effects of individual features on the acceptance of e-WOM. It was found that the Need for Cognition was positively related to acceptance of e-WOM ($b=.099$, $t=1.931$, $p<0.1$), which supported H_2 -a. It was predicted that NFC would have a positive relationship to acceptance of e-WOM. Also, it was found that innovation was positively related to acceptance of e-WOM ($b=.098$, $t=2.000$, $p<0.1$), which supported H_2 -b. However, H_2 -c, in which the positive impact of gender on acceptance of e-WOM was not predicted, was not supported ($b=-.018$, $t=-.198$, $p=.844$), and H_2 -d, which was predicted with a positive impact of age on acceptance of e-WOM was supported ($b=-.025$, $t=-2.871$, $p<0.01$).

The effects of network features on the acceptance e-WOM. It was found that activity was not related to acceptance of e-WOM ($b=10.847$, $t=.534$, $p=.594$), which doesn't support H_3 -a; however, it was found that the connection was positively related to acceptance of e-WOM ($b=-.312$, $t=-1.798$, $p<0.1$), which supported H_3 -b, and that H_3 -c, which was predicted with a negative impact of power on acceptance of e-WOM, was supported ($b=-3.792$, $t=-2.028$, $p<.01$).

The relative importance of impact variables. In the examination of factors affecting acceptance of e-WOM, the effects of informational, individual, and network characteristics were found. In order to understand the influence of e-WOM, the relative importance of impact variables was examined, and the result for informational characteristics was found to be more important than others. Especially the usefulness of information was the most important. The relative importance of impact variables is displayed in Table 4.

Table 4. Comparing with impact variables on acceptance of e-WOM

	Impact variables(Standardized coefficients)
3-Upper variables	Information usefulness(.359), Information reliability(.232), Information interest(.176)
3-Under variables	NFC(.130), Innovation(.128), Network Power(.123)

5 Conclusion and Suggestions

In the results of this study, the network features in the fashion community were found to be significant as explanation variables affecting the acceptance of e-WOM fashion information, and in the explanatory factors for e-WOM acceptance, the usefulness of information, interest, and reliability, which were informational characteristics; the need for cognition, innovation, and age, which were individual characteristics; and connectivity and power, which were network characteristics; were all shown to affect e-WOM acceptance.

It was found that after consumers are exposed to fashion information, they need to receive useful or reliable e-WOM information that enables them to change their attitudes and behavior. Thus, from the marketing standpoint, the quality of fashion information should not be overlooked in order to improve the organizational performance through e-WOM marketing.

Meanwhile, the variables that were assumed to be effective were found to be ineffective in the distribution of information. It was found that gender, among demographic variables, and activity, among network variables, were found to be not statistically significant. For this reason, the men involved in the fashion community were usually more highly interested in fashion than other men.

In addition, the network activity didn't have a significant impact, and activity was measured to be the central activity. The nodes that were related to star nodes played an important role in the acceptance of e-WOM. As a result, this activity was not supported; however, the effects of connectivity and power appeared to be significant.

The results this study suggest effective marketing management using network structure, examination of influence variables in order to establish verbal communication and maximize marketing by the internet fashion information acceptance.

References

1. Ahn, K.H., Hwang, S.J., Jung, C.J.: Fashion Marketing, Suhaksa, Seoul (2010)

2. No, G.Y.: New media of communication and interactivity. Institute of Information and Communications Policy (2008)

3. Brickart, B., Schinlder, R.M.: Internet forums as influential sources of consumer information. Journal of Interactive Marketing 15(5), 31–52 (2001)

4. Wellman, B., Boase, J., Chen, W.: The Networked Nature of Community: On and Off the Internet. IT and Society 1(1), 151–165 (2002)

5. Wasserman, S., Raust, K.: *Social Network Analysis*. Cambridge University Press, Cambridge (1994)
6. Granovetter, M.: The Strength of Weak Ties. *The American Journal of sociology* 78(6), 1360–1380 (1973)
7. Constant, D., Sproull, L., Kiesler, S.: The Kindness of Strangers: The usefulness of electronic weak ties for technical advice. *Organization science* (1996)
8. Kaiser, S.B.: *The Social Psychology of Clothing*. Fairchild, NY (1997)
9. Elliott, K.M.: Understanding consumer-to-consumer influence on the web. Doctoral dissertation. Duke University, Durham (2002)
10. Chevalier, J.A., Mayzlin, D.: The effect of word of mouth on sales: online book reviews. NBER Working paper 10148, 1–30 (2003)
11. Song, Y.T.: The effects of preannouncement on word of mouth diffusion in online community. Seoul national university (2007)
12. Sen, S., Lerman, D.: Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing* 21(4), 76–94 (2007)
13. Lee, H.S., Ahn, K.H., Ha, Y.W.: Consumer Behavior, Bubmons, Seoul (2008)
14. Geissler, G.L., Edison, S.W.: Market Mavens' Attitude toward. General Technology: Implications for Marketing Communications, *Journal of Marketing Communication* 11(2), 73–94 (2005)
15. Rogers, E.M.: *Diffusion of Innovation* 4th. Free Press, NY (2003)
16. Lee, H.Y., Chung, E.H., Lee, J.H.: Web2.0: Social diffusion of digital content. Institute of Information and Communications Policy (2007)
17. Kim, H.S., Rhee, E.Y., Yee, J.Y.: Comparing fashion process networks and friendship networks in small groups of Adolescents. *Journal of Fashion Marketing and Management* 12(4), 545–564 (2008)
18. Barabasi, A.: *Linked: How Everything Is connected to Everything Else and what it Means for business science and everyday life*, Plum, NY (2002)
19. Lee, H.S., Lim, J.H.: *Marketing Research*, Bubmoonsa, Seoul (2009)
20. Wilton, P.C., Myers, J.G.: Task, Expectancy and Information Assessment effects in Information Utilization Process. *Journal of Consumer Research* 12, 469–485 (1986)
21. Lee, E.Y.: Two factor model of on line word of mouth adoption and diffusion. Seoul national university (2004)
22. Cacioppo, J.T., Petty, R.E.: The Need for Cognition. *Journal of Personality and Social Psychology* 42, 116–131 (1982)
23. Rochrich, G.: Consumer Innovativeness Concepts and Measurements. *Journal of Business Research* 57, 671–677 (2004)

Appendix: Measures

		Items ^a	α	Prior Researcher
Information	Usefulness	1. Necessary information	.89	Wilson and Myers(1986),[20] song(2007), Lee(2004)[21]
		2. Satisfied my concern		
		3. Useful information		
	Interest	1. Unique and interesting information	.88	
		2. Stimulate my curiosity		
		3. Aroused my interest in fashion		
	Reliability	1. Honest information	.83	
		2. Believable information		
		3. Reliable information about fashion products		
Individual	NFC	1. I want to know more	.72	Cacioppo and Petty(1982)[22]
		2. I like to important and intelligent things		
		3. I try to come up with solutions to problems without fail		
	Innovation	1. I usually know more than others about new fashion products	.88	Roehrich(2004)[23]
		2. I am the first to buy new fashion products		
		3. Every year I buy new fashion products		
Acceptance of e-WOM		4. People ask me about new fashion trends	.80 ^b	Song(2007)
		1. Because of fashion information, improved brand images.		
		2. Because of fashion information, improved buying intention		

^a Each item was measured on a 5-point scale.

^b Correlation between two measured items about acceptance of e-WOM.

Exact Closed-form Outage Probability Analysis Using PDF Approach for Fixed-DF and Adaptive-DF Relay Systems over Rayleigh Fading Channels^{*}

Jeanyoung Jang and Kyunbyoung Ko

Dept. of Instrumentation and Control Engineering, Chungju National Univ.,
72 Daehangno, Chungju-si, Chungbuk, 380-702 Korea
Tel.: +82-43-841-5463; Fax : +82-43-841-5320
kbko@cjnu.ac.kr

Abstract. In this paper, we derive the closed-form outage probability for decode-and-forward (DF) relay systems over Rayleigh fading channels. At first, our focus is on fixed-DF (FDF) relay schemes in which the probability density function (PDF) is derived based on error-event at relay nodes. It is confirmed that an erroneous detection and re-transmission at relay nodes affects both the received signal-to-noise ratio (SNR) and the outage probability. In addition, the developed analytical method is extended to adaptive-DF (ADF) schemes and the exact outage probability expression is derived. Simulation results are finally presented to validate the analysis.

Keywords: DF relay systems, Rayleigh fading channels, FDF, ADF.

1 Introduction

Cooperative diversity has recently been widely discussed in wireless networks. Two main relaying protocols are usually used for cooperative diversity schemes: amplify-and-forward (AF) and decode-and-forward (DF). In the former, the relay retransmits the receiving signal after amplifying it, whereas in the latter, the relay detects the received signal and then retransmits a regenerated signal [1][2]. At the destination, the receiver can employ a variety of diversity combining techniques to benefit from the multiple signal replicas available from the relays and the source. A third option is to have relays forward only those correctly decoded messages, and can be considered as Adaptive-DF (ADF) schemes [3]. Use of ADF schemes presumes incorporation of, e.g., cyclic redundancy check (CRC) codes from a higher layer in order to detect errors. The advantages of general cooperative diversity schemes come at the expense of the spectral efficiency since the source and all the relays must transmit on orthogonal channels (i.e., different time slots or frequency bands) in order to avoid interfering with each other as well [2].

^{*} This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0002650).

Recent researches are widely carried out for relay selection schemes assuming that the best relay node can be selected with additional information [4][5][6][7]. For the practically attractive DF relay strategy, the authors in [3] derived a high-performance low-complexity coherent demodulator at the destination in the form of a weighted combiner. However, most papers used a moment generating function (MGF) approach [1][6][8][9]. Even if it can give exact results, a numerical integral is also necessary to obtain the analytical results [8]. Also, to the best of our knowledge, no one proposed a general probability density function (PDF) approach without the numerical integral for DF relay systems and the exact closed-form outage probability expressions, which can explain how an erroneous detection at each relay affects both the received signal-to-noise ratio (SNR) and the outage probability.

At first, we focus on fixed-DF(FDF) relay networks without selectively transmission in order to generally derive a PDF approach based on error-events at relay nodes. Then, it is modified to cover ADF schemes so that the developed analytical method can be considered as another general solution for DF relay systems. Specifically, the exact closed-form expressions are derived for the outage probability over independent and not necessarily identically distributed (INID) Rayleigh fading channels. In addition, for ADF schemes over independent and identically distributed (IID) Rayleigh fading channels, the outage probability is presented as a well-known simple tractable form.

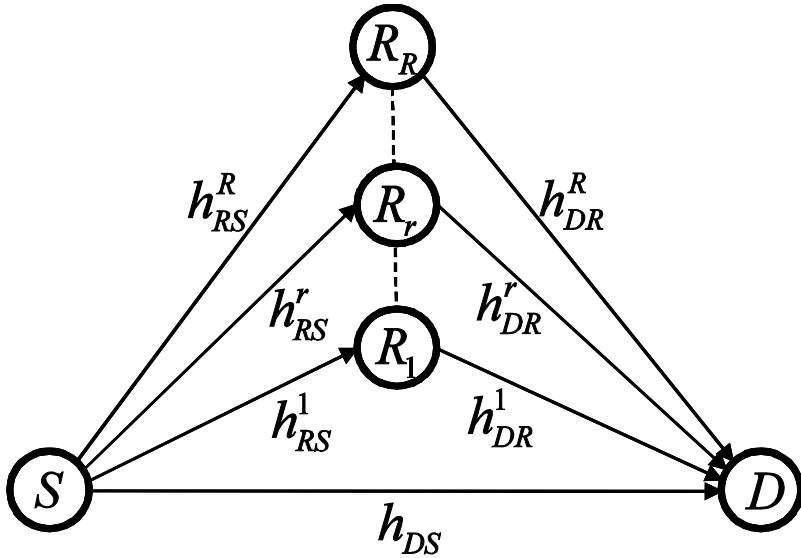


Fig. 1. Block diagram of DF relay systems (source(S), destination(D), relay(R))

2 DF Relay Systems

Fig. 1 shows the block diagram of DF relay systems with a source(S), a destination(D), and a relay(R). The number of relays is R . It is assumed that S and R transmit over

orthogonal time slots [2]. For the r th relay, let h_{DS} , h_{RS}^r , and h_{DR}^r be the channel gains of S-D, S-R, and R-D link channels, respectively. In this letter, the wireless channels between any pair of nodes in DF relay systems are assumed quasi-static INID Rayleigh fading and also corrupted by additive white Gaussian complex noise terms of n_{DS} , n_{RS}^r , and n_{DR}^r . Without loss of generality, we assume that noise terms have zero mean and equal variance of σ^2 ($= E[|n_{DS}|^2] = E[|n_{RS}^r|^2] = E[|n_{DR}^r|^2]$). The received signals for S-D, S-R, and R-D links are presented respectively as follows:

$$\begin{aligned} y_{DS} &= h_{DS}s + n_{DS} \\ y_{RS}^r &= h_{RS}^r s + n_{RS}^r \\ y_{DR}^r &= h_{DR}^r \hat{s}_r + n_{DR}^r \end{aligned} \quad (1)$$

From here, s is a binary phase shift keying (BPSK) symbol with $E[|s|^2] = 1$ and $\hat{s}_r$ is the regenerated symbol at the r th relay node. Therefore, the received instantaneous signal-to-noise ratio (SNR)s can be written as follows:

$$\begin{aligned} \gamma_0 &= |h_{DS}|^2 / \sigma^2 \\ \gamma_r &= |h_{DR}^r|^2 / \sigma^2 \\ \gamma_{R+r} &= |h_{RS}^r|^2 / \sigma^2 \end{aligned} \quad (2)$$

At the destination node, a maximal ratio combining (MRC) scheme can be applied in order to combine signals from S-D and R-D links

3 Outage Probability Analysis of FDF Relay Systems

In order to generally derive the analytical method based on error-events at relays, let us define the p th error event vector E^p as [8]

$$E^p = [e_1^p \cdots e_r^p \cdots e_R^p] \quad (3)$$

with $p \in \{1, 2, \dots, 2^R\}$ and the total number of error-events is 2^R . Generally, we can define that E^1 is all-zero vector, E^{2^R} is all-one vector, and so on. Note that for the p th error-event, $e_r^p = 0$ means the correct detection at the r th relay and $\hat{s}_r = s$ with the probability of

$$P_b(\bar{\gamma}_{R+r}) = \frac{1}{2} \left[1 - \sqrt{\frac{\bar{\gamma}_{R+r}}{1 + \bar{\gamma}_{R+r}}} \right] \quad (4)$$

which is the averaged BER of the r th S-R link, $\bar{\gamma}_{R+r} (= E[\gamma_{R+r}])$ [10]. Also, $e_r^p = 1$ leads to $\hat{s}_r = -s$ with the probability of $1 - P_b(\bar{\gamma}_{R+r})$. Furthermore, the probability of the p th error-event at DF relay systems is presented as

$$Pr^p = \prod_{r=1}^R (1 - P_b(\bar{\gamma}_{R+r}))^{\bar{e}_r^p} (P_b(\bar{\gamma}_{R+r}))^{e_r^p} \quad (5)$$

with $\bar{e}_r^p = (e_r^p + 1) \bmod 2$.

For the p th error-event vector, the combined instantaneous SNR of γ_{FDF}^p can be written as

$$\gamma_{FDF}^p = \sum_{r=0}^R (-1)^{e_r^p} \gamma_r. \quad (6)$$

Note that we can define $e_0^p = 0$ for S-D link. In addition, the joint probability density function (PDF) can be derived with respect to the region of γ_{FDF}^p as

$$f(\gamma_{FDF}^p) = \begin{cases} \sum_{r=0, e_r^p=0}^R \frac{\pi_{r,FDF}^p}{\bar{\gamma}_r} \exp\left(\frac{-\gamma_{FDF}^p}{\bar{\gamma}_r}\right), & \gamma_{FDF}^p \geq 0 \\ \sum_{r=0, e_r^p=1}^R \frac{\pi_{r,FDF}^p}{\bar{\gamma}_r} \exp\left(\frac{+\gamma_{FDF}^p}{\bar{\gamma}_r}\right), & \gamma_{FDF}^p < 0 \end{cases} \quad (7)$$

with

$$\pi_{r,FDF}^p = \prod_{i=0, i \neq r}^R \frac{(-1)^{e_r^p} \bar{\gamma}_r}{(-1)^{e_r^p} \bar{\gamma}_r - (-1)^{e_i^p} \bar{\gamma}_i}. \quad (8)$$

Then, the outage probability can be expressed with respect to the γ_{th} as

$$P_{out,FDF}^p(\gamma_{th}) = Pr[\gamma_{FDF}^p \leq \gamma_{th}] = \begin{cases} \sum_{r=0}^R \pi_{r,FDF}^p \left(1 - \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_r}\right)\right), & \gamma_{FDF}^p \geq 0 \\ \sum_{r=0}^R \pi_{r,FDF}^p, & \text{else.} \end{cases} \quad (9)$$

Note that an erroneous detection at each relay (i.e., $e_r^p = 1$) can cause a negative effect on the received SNR (as shown as (6)) and generate the PDF term related with $\gamma_{FDF}^p < 0$. Consequently, it does not guarantee a diversity gain by combining the given R-D link. When we can consider all the possible error-events, the outage probability for combining R-D and S-D links can be shown as

$$P_{out,FDF} = \sum_{p=1}^2 Pr^p P_{out,FDF}^p(\gamma_{th}). \quad (10)$$

4 Outage Probability Analysis of ADF Relay Systems

In ADF relay systems, the r th relay is only to transmit the regenerated symbol of $\hat{s}_r$ when messages are correctly decoded. Note that $\hat{s}_r$ can be two values, which are $\hat{s}_r = 0$ with the probability of $P_b(\bar{\gamma}_{R+r})$ or $\hat{s}_r = s$ with the probability of $1 - P_b(\bar{\gamma}_{R+r})$. For ADF schemes, eq. (6) can be modified as

$$\gamma_{ADF}^p = \sum_{r=0}^R \overline{e_r^p} \gamma_r. \quad (11)$$

This equation means that when there is a detection-error at the r th relay node for the p th event vector (i.e., $e_r^p = 1$), no-transmission gives $\overline{e_r^p} \gamma_r = 0$. Therefore, in ADF schemes, $\overline{e_r^p}$ can be regarded as the transmission indicator for the r th relay of the p th error-event. Similarly with (7), the PDF of γ_{ADF}^p can be presented as

$$f(\gamma_{ADF}^p) = \sum_{r=0}^R \frac{\pi_{r,ADF}^p}{\bar{\gamma}_r} \exp\left(\frac{-\overline{e_r^p} \gamma_{ADF}^p}{\bar{\gamma}_r}\right), \gamma_{ADF}^p \geq 0 \quad (12)$$

with

$$\pi_{r,ADF}^p = \prod_{i=0, i \neq r}^R \frac{\overline{e_r^p} \bar{\gamma}_r}{e_r^p \bar{\gamma}_r - e_i^p \bar{\gamma}_i}. \quad (13)$$

Then, the outage probability can be expressed with respect to γ_{th} as

$$P_{out,ADF}^p(\gamma_{th}) = Pr[\gamma_{FDF}^p \leq \gamma_{th}] = \sum_{r=0}^R \pi_{r,ADF}^p \left(1 - \exp\left(-\frac{\overline{e_r^p} \gamma_{th}}{\bar{\gamma}_r}\right)\right). \quad (14)$$

By taking into account for all the possible error-events, the outage probability is presented as

$$P_{out,ADF} = \sum_{p=1}^{2^R} Pr^p P_{out,ADF}^p(\gamma_{th}). \quad (15)$$

Comparing (11) with (6), we observe that γ_{ADF}^p is always non-negative. Consequently, contrary to FDF, it is confirmed that R-D link combining in ADF can give the diversity gain.

5 Numerical and Simulation Results

In this section, we show numerical results of outage probability and verify their accuracy by comparing with simulation results. We assumed for the channel conditions

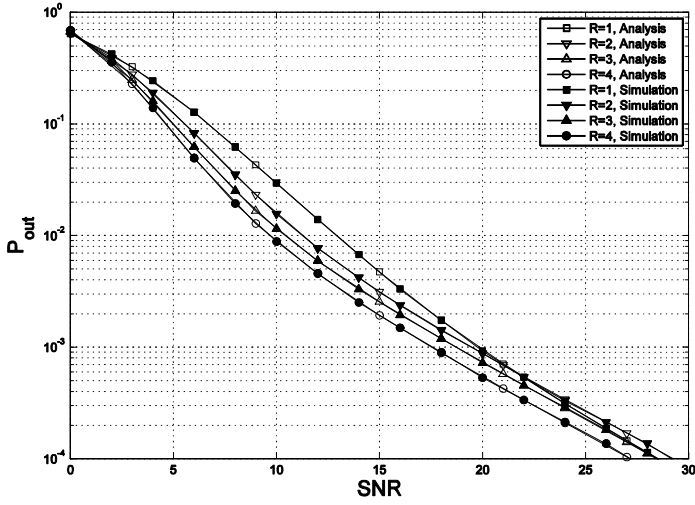


Fig. 2. Outage probability versus SNR for FDF relay systems with respect to different number of relay nodes

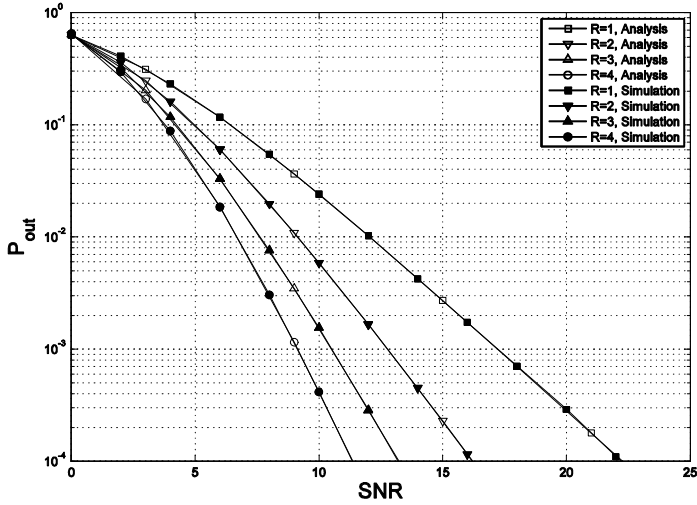


Fig. 3. Outage probability versus SNR for SDF relay systems with respect to different number of relay nodes

that $\bar{\gamma}_0 = \bar{\gamma}_{R+r}$, $\bar{\gamma}_{r+1} = \bar{\gamma}_r e^{-1/(R+1)}$, $SNR = \sum_{r=0}^R \bar{\gamma}_r$, and $R \in \{1, 2, 3, 4\}$. Fig. 2 shows the outage probability versus SNR for the FDF relay systems (where all relays transmit the regenerated symbol regardless of an erroneous detection at each relay node). Fig. 3 shows the outage probability as a function of SNR for ADF relay schemes (where each relay transmits the regenerated symbol only if the correctly detection is carried out at

each relay). For FDF relay systems as shown in Fig. 2, the increment of the number of relays generates the outage probability performance improvement but the diversity gain cannot be fully obtained. It is caused by the fact that an erroneous transmission at each relay node can give the negative effect on the received SNR as written in (6). From Fig. 3, we can see that ADF scheme gives that full diversity gain. It can be clearly seen from two figures that simulation curves math with our analytical ones.

6 Conclusions

We have developed the PDF approach based on error-events at relay nodes in order to propose analytical method as a general tool for DF relay systems over Rayleigh fading channels. For both FDF and ADF relay schemes, the outage probabilities have been derived as the exact closed-forms without numerical integrals. Moreover, they have been compared with simulations and their accuracy is verified. Therefore, we can conclude that our closed-form outage probability expressions are easily tractable form, and can be used as a tool to verify effects of an erroneous detection and transmission at each relay node on the combined SNR, the outage probability, and cooperative diversity gain.

References

1. Mazen, O., Hasna, Alouini, M.-S.: End-to-End Performance of Transmission Systems With Relays Over Rayleigh-Fading Channels. *IEEE Trans. on Wireless Commun.*, 1126–1131 (2003)
2. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. on Info. Theory*, 3062–3080 (2004)
3. Wang, T., Cano, A., Giannakis, G.B., Laneman, J.N.: High-Performance Cooperative Demodulation With Decode-and-Forward Relays. *IEEE Trans. on Commun.* 1427–1438 (2007)
4. Bletsas, A., Khisti, A., Reed, D.P., Lippman, A.: A Simple Cooperative Diversity Method Based on Network Path Selection. *IEEE Journal on selected areas in Commun.* 659–672 (2006)
5. Kim, J.-B., Kim, D.: Exact and Closed-Form Outage Probability of Opportunistic Single Relay Selection in Decode-and-Forward Relaying. *IEICE Trans. on Commun.* 4085–4088 (2008)
6. Ikki, S., Ahmed, M.H.: Performance of cooperative diversity using Equal Gain Combining (EGC) over Nakagami-m fading channels. *IEEE Trans. on Wireless Commun.* 557–562 (2009)
7. Yang, C., Wang, W., Chen, S., Peng, M.: Outage Performance of Opportunistic Decode-and-Forward Cooperation with Imperfect Channel State Information. *IEICE Trans. on Commun.* 3083–3092 (2010)
8. Lee, Y., Tsai, M., Sou, S.: Performance of decode-and-forward cooperative communications with multiple dual-hop relays over nakagami-m fading channels. *IEEE Trans. on Wireless Commun.* 2853–2859 (2009)
9. Anghel, P.A., Kaveh, M.: Exact symbol error probability of a Cooperative network in a Rayleigh-fading environment. *IEEE Trans. on Wireless Commun.* 1416–1421 (2004)
10. Proakis, J.G.: *Digital Communication*, 3rd edn. McGraw Hill, New York (1995)

From Trading Volume to Trading Number-Based Pricing at Home Trading System on Korean Stock Market

Youngsik Kwak¹, Yunkyung Lee², Jaeweon Hong³, Wanwoo Cho⁴,
Ho Jang⁵, and Daehyun Park⁶

¹ Associate Professor, Gyeongnam National University of Science and Technology, #150,
Chilam-dong, Jinju, Korea, +82-55-751-3454

yskwak@gntech.ac.kr

² Senior Researcher, Korea Culture & Tourism Institute, #827, Bangwha-3dong,
Gangseo-gu, Seoul, Korea, +82-2-2669-6938

yunky@kcti.re.kr

³ Assistant Professor, Dongseo University, #47, Jurye2-dong, Sasang-gu, Pusan,
Korea, +82-51-320-1639

jwhong@gdsu.dongseo.ac.kr

⁴ Director, Daewoo Securities Ltd., #34-3, Yoido-dong, Seoul, Korea., +82-2-768-2111

cww@bestez.com

⁵ Director, Benet Ltd., #219-25, Jayang-dong, Seoul, Korea, +82-2-737-5475

hjang@benet.co.kr

⁶ Guest Professor, Kyung Hee University, 1 Hoegi-dong, Seoul, Korea, +82-10-8881-0041

dhpark@naver.com

Abstract. The new n-block tariff can outperforms, in terms of profit, two-part tariff, all unit discount price schedule, and uniform pricing for a given service and product. This research objectives are to develop new pricing unit and to determine the optimal price break points for n-block tariff on the new pricing unit. Although the merits of developing new pricing unit and non-linear pricing are well documented, the attempt to practice the new pricing unit development and non-linear pricing in online market has been relatively rare. The researchers found that transaction log file analysis using mixture model can be the feasible methodology for developing the new pricing unit and determining the optimal break points number of n-block tariff. The researchers empirically demonstrate the feasibility and the superiority of the mixture model by applying it to the log file on Home Trading System (HTS) for futures and option transaction at a stock company in Korea. The empirical results showed that the stock company had an opportunity to set new pricing unit from trading volume-based pricing to trading number-based pricing a given time horizon.

1 Introduction

Nonlinear pricing refers to a pricing system whereby the price per unit of goods and services varies according to the quantity purchased by the customer [1]. There are research results which show that nonlinear pricing, which has been used partially, can increase the profit and sales volume more than linear pricing, which has been widely used for existing

online service [2~3]. This study focuses on n-part tariff among nonlinear pricing that has been reported by existing studies to create relatively more profit in the service sector than in the merchandise sector.

The purpose of this study is to suggest mixture modeling as a methodology to help with the normative decision making of n-block among nonlinear pricing and to empirically test it to online market data. For this, the study seeks to show the series of processes entailed in segmenting the customers using a mixture model for the online log file of the Home Trading System of a stock company where online service pricing is being actively set and formulating the optimal number of break points of n-block tariff and the optimal pricing unit.

We expect that this study will improve the capacity to create profit on the part of the online service provider at the practical level and will contribute to optimal block number decision making and to developing a new pricing unit when setting the n-block tariff online at the academic level.

2 Nonlinear Pricing and Mixture Model

1. Decision Making Required for Setting n-Block Tariff

The issue of decision making for setting the optimal n-block tariff can be divided into four categories. These are optimal block number decision making, optimal break points decision making, optimal price level decision making, and pricing unit decision making [4]. The first decision making issue entails deciding the optimal number of block, which requires deciding indeed how many n-block tariffs will be suggested to the customer.

2. Methodology for Decision Making for n-Block Tariff

2.1 Exploring Objects of Pricing Unit

The mixture model originally starts from the fact that one distribution is made up of several sub distributions which we do not know about. First, let us assume that one specimen is made up of n number of data and that each data is made up of k number of variables (formula 1).

$$\mathbf{y}_n = (\mathbf{y}_{nk}) \quad (1)$$

Where k = Number of variables used for market segmentation

n = Number of sample

$\mathbf{y}_{nk}$ = A vector made up of k number of variables and n number of sample

Then, as the conditional distribution function of vector $\mathbf{y}_n$, $\mathbf{y}_{nk}$ becomes the value that has been calculated at market segment s . $\mathbf{y}_n$ is θ_s indicated as the vector of a given density function which has a parameter that we do not know about. This has the normal function of $f_s(\mathbf{y}_n/\theta_s)$. This probability density function can have various functions of normal distribution, poisson, binomial, negative binomial, etc. By having a number of probability density, it can explore all variables which are the object of pricing unit regardless of the scale of variable used by the researcher.

2.2. Process of Verifying Optimal Breaking Points

We do not know how many sub distributions there are hidden under the original distribution but believe that the power of explanation will increase whenever the number of market segment is increased. If the power of explanation does not increase significantly when the number of market segment is increased, or if the work of segmentation is no longer required because the increase in power of explanation is small when compared to the parameter used to increase the number of market segment, then the optimal number of submarkets can be judged [9-11]. In other words, if the increase in model fit is insignificant despite an increase in market segment, then it can be said that until such a time, the number of market segment remains a meaningful number. The model's goodness-of-fit to measure this will be measured by Akaike Information Criterion or Bayesian Information Criterion.

$$BCI = -2(LL - p) / N \quad (2)$$

Where LL= log-likelihood
 p= number of parameters
 N= number of observations

Generally, if the number of market segment of n number is optimal due to a mixture model, then one can find n-1 number of breaking points. Based on this, n number of block tariff can be applied for n-block tariff.

2.3 Process of Verifying Optimal Pricing Unit

Also, we must explain the characteristics of market segment based on the variables displaying a significant difference among market segments. We must formulate a pricing unit which can induce difference among market segments and price response. Such a significant test is conducted using the Wald test, which verifies whether there is a statistically meaningful difference of each variable among market segments. This method has already been applied by Soyoung Kim (2003) and Wooksang Han (2006) to performance audience and movie audience, respectively [12-14]. As such, among the variables of difference among market segments deduced from this Wald test, the variable that can induce a price response can become a candidate for pricing unit.

3 Research Process

The data used in this study are the transaction log file of all people who engaged in transaction during seven months in the late 2000s from the online future and option transaction among the home trading system of one stock company. This file includes the date and time of online transaction per customer, transaction account, date and time of order, average transaction amount, transaction volume, total transaction number, and the number of months where transaction took place. It includes data related to the transaction of a total of 910 persons. The total transaction number is 9264. Moreover, at the time of extracting these data, the stock company was imposing a fee according to the criteria of '1 time transaction amount'.

4 Research Result

1. Optimal Market Segment Number Decision Making and Optimal Block Number of n-Block Tariff Decision Making

<Table 1> shows the changes to model fit per market segment of data used in this study via Log-likelihood value and BIC. As shown in <Table 1>, the power of explanation continued to increase according to the increase from 1 to 11 of the number of market segments (BIC value decreased). The power of explanation instead decreased when there was again an increase in the number of market segments from 11 to 12 (BIC value increased). This means that the number of market segments at 11 explained most clearly the heterogeneity of the specimen, and when it was increased to 12, the power of explanation decreased. Thus, we verified that 11 was the appropriate number of market segments.

Since this study found that 11 market segments is an appropriate number of market segments, it was revealed that a total of 11 block tariffs must be applied after finding 10 breaking points between them.

Table 1. Statistics for deducing optimal market segment

Number of market segments	Log-likelihood	BIC	Increase and decrease in BIC
1	-88184.1	176428	
2	-82833.7	165794	-10633
3	-81011.4	162217	-3577
4	-80252.1	160766	-1451
5	-79673.8	159676	-1089
6	-79260.3	158917	-760
7	-79097.2	158658	-259
8	-78851.7	158234	-424
9	-78709	158016	-218
10	-78604	157873	-143
11	-78165.1	157062	-811
12	-78433.2	157666	603

2. Pricing Unit Decision Making

As a result of applying the mixture model to 11 market segments, there were significant differences among the total transaction volume, 1 time average transaction amount, total transaction number, and number of months where transaction took place (Wald test result, $p<.001$). The stock company was already imposing a fee according to the criteria of ‘1

time transaction amount’ from among these. Hence, the ‘1 time average transaction amount’ cannot become a new optimal market segment variable.

If we list the results according to the sequence of the difference in the average per market segment, they are a multiple of 277 (transaction volume), a multiple of 274 (transaction number), a multiple of 30 (1 time transaction amount), a multiple of 11 (total number of transaction months). This sequence can be interpreted as a low variable among the variables with the highest heterogeneity of customers per market segment. As such, the ‘transaction volume’ showing the best heterogeneity market segment can become the solution for decision making on optimal market segment variable selection. In the case of using a variable called, ‘transaction volume’ during normal operation at this time, the transaction time must be specified. Likewise, in respect to the ‘transaction number’, the term of transaction must be decided upon ahead of time in the case of using new pricing criteria. As a results, these two variables can become new pricing criteria of this stock company.

Table 2. Result of market segmentation

Category	Size of market segment	Transaction volume	Average of 1 time transaction amount	Total transaction number	Transaction term (month)
Cluster1	17.0%	21,673,182.1	420,864.1	40.5	2.4
Cluster2	15.2%	20,848,120.0	982,755.6	19.2	1.0
Cluster3	12.2%	93,653,437.6	264,411.3	318.1	6.6
Cluster4	11.4%	305,101,110.2	771,238.1	368.4	5.4
Cluster5	10.4%	136,361,611.3	1,790,479.0	73.1	2.6
Cluster6	7.5%	1,325,297,059.5	1,345,830.1	984.6	6.3
Cluster7	6.3%	5,785,968,154.9	3,841,167.0	1650.3	7.3
Cluster8	5.7%	1,547,419,381.9	5,903,573.9	266.6	3.3
Cluster9	5.5%	602,649,055.1	362,894.5	1564.1	7.8
Cluster10	5.3%	3,898,103,543.8	803,645.9	5281.8	10.8
Cluster11	3.5%	340,601,726.9	8,030,859.1	58.0	1.8
Wald value		2055.6	1749.0	2015.5	4174.7
p value		.000	.000	.000	.000

5 Results and Discussion

This study sought to solve the problem of decision making on implementing n-block tariff among nonlinear pricing based on the form of online financial transaction of domestic consumers in Korea.

For this, the study has executed market segmentation by using the mixture model for log file data, which is an actual online transaction data for future option of a certain domestic

stock company, and sought to establish decision making criteria for setting the online price by marketers by applying nonlinear pricing and n-block tariff according to the result of market segmentation.

As a result, first, in regard to decision making related to, "How many blocks of price tariff should be suggest?", it was confirmed in the case of this specimen that designing n-block tariff according to 11 market segments had the possibility of bringing about the greatest profit.

Second, in respect to decision making related to, "Based on what variable criteria should n-block tariff be set?", this study has empirically deduced that the 'transaction volume' and 'transaction number' can become new pricing criteria.

The implications of this study are as follows. First, there is significance associated with deducing results by reflecting actual purchase behavior by using the Log file data which are online financial transaction data and not virtual purchase data such as intent to purchase, preference, etc. Second, an operational guideline has been provided for online pricing process by furnishing criteria for methodology related to what service should be differentiated based on several market segments. Third, by expanding its field to research on online pricing, which has relatively remained as a vacuum among financial sector, the study has contributed to establishing a balanced academic sphere which forms the foundation for financial companies to practice online price strategy. Yet, this study has sought to solve the technical decision making problem of online pricing, but there is unnecessary difficulty when generalizing it as general decision making criteria related to financial sector by using the data of a certain domestic financial company.

References

1. Yoo, P.: Theory of Pricing. Pakyoungsa (1991)
2. Tacke, G.: Nichtlineare Preisbildung: Theorie, Meassung and Anwendung. Gabler (1988)
3. Yoo, P., Park, Y.: The Study on the Service Pricing: Focused on the Non-Linear Pricing for the Maritime. Korean Journal of Management Review 26(4), 567–596 (1997)
4. Baek, S., Kwak, Y.: The Pricing Strategy for the Performance of Medical Service - Based on the Segmentation for the N-block Tariff Pricing of Medical Examination. Journal of Health Policy and Administration 12(4), 84–97 (2003)
5. Lee, Y., Hong, J., Kwak, Y.: Pricing Strategy that Finds Hidden Profit. Benet, Seoul (2004)
6. Park, Y.: A Study of Service Pricing Strategy: With a Focus on Nonlinear Pricing of Shipping Service, Sungkyunkwan University Ph. D. Dissertation (1995)
7. Wilson, R.: Nonlinear Pricing. Oxford University Press, Oxford (1993)
8. Nagle, H.: The Strategy and Tactics of Pricing, Englewood Cliffs (1995)

Design and Implementation of the Application Service System Based on Sensor Networks

Yoonsik Kwak¹ and Donghee Park^{2,*}

¹ Professor, Chungju National University, Computer Engineering,
#123, Kumdan-ri, Chungju, Korea
Tel.: +82-43-841-5345
yskwak@cjnu.ac.kr

² Professor, Chungju National University, Information & communication Engineering,
#123, Kumdan-ri, Chungju, Korea
Tel.: +82-43-841-5364
dhpark@cjnu.ac.kr

Abstract. In this paper, we will take about the design and implementation of application service system based on sensor networks. As the components of application service system, we will take about middleware and its application services. There are 4 catalogues of module that consist of middleware, on the other hand, query processing module to deal with the requests from application service system, control module to control and manage meta data, module to connect with sensor nodes, and API solution to support user's service. Based on the implemented middleware, in the high level of system, it is implemented service of system management, control service, and query processing service corresponding to happened events, and management and store service for sensing data.

Keywords: sensor networks, application service system, middleware.

1 Introduction

As advancing computer and information & communication technology, it is increasing interest for ubiquitous society. In addition to these technical trends, there have been actively studied and attended to an ubiquitous sensor networks.[1] Through interaction between persons and objects, beyond communication between persons, we can acquire a lot of valuable information for persons and can be used it to our everyday life. Then it may be a means to extract effective and productive outcome in real world.

Also, sensor network system has been actively studied.[1] As a result, we have been widely used them in our usual life. For examples, there are environmental monitoring, healthcare, etc.[1][2] A study scope in this field is classified into two catalogues: collection of sensor node configured sensor networks, middleware to support connection of user's application services and hardware layer, and then application service system.

* Corresponding author.

As above presented, defining for middleware of sensor network system is difficult to describe in simplicity. In this field of sensor network system, it was caused by ambiguous difference between application service system and middleware, also miniaturization of operating system and disappearing boundary of hardware system. In this paper, we defined middleware that is something to support application services, and collection of component easily to use of the hardware platform for sensor networks.

Components of middleware that is composed of application service system are differentiated into query processing module to deal with the request from application service system, control module to control and manage meta data, module to connect with sensor nodes and API solution to support user's service. Based on the configured middleware, in the high level of system, it is implemented services of system management, control, and query processing service corresponding to happened events, and management and store service for sensing data.

2 Requirements for System

Specified application service system based on sensor networks in this paper was designed and implemented. Roles are to measure the particular range of temperature and humidity in area to be deployed sensors. Also, throughout storing and management services for sensing data, extracting and creating various context information, it can provide intelligent services with users. Using the meta data to be information about sensor nodes and hardware system, it can provide control and management services for system. Figure 1 is shown the block diagram of implemented system.

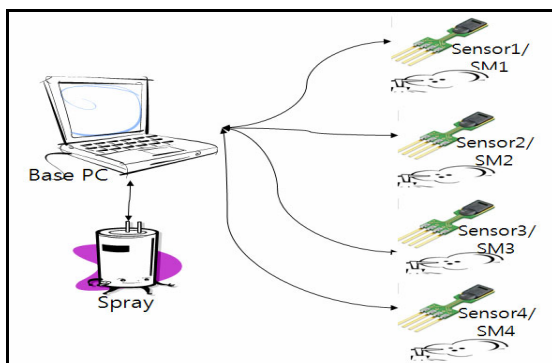


Fig. 1. Block diagram of system

In the Figure 1, Base PC is main computer that perform various application services. Spray is an actuator, Sensor nodes consist of devices that can measure a range of temperature and humidity data in sensor fields.

Before designing the application system based on sensor networks, we should need to considerations about the requirements for system. In the most case, application system of sensor networks was designed and implemented for specified purposes. It is usual. We divide requirements for system into hardware and software parts and are presented in this chapter.

2.1 Hardware Platform

A hardware platform refers to a set of devices: sensor nodes and base system. An application service system is composed of sensor nodes and base system that will be presented about entire hardware in this chapter.

First, it is a part of sensor nodes, an eyes and ears of sensor, we must select a type of sensor. It basically is components that are deployed in inaccessible terrines or disaster relief operation. And then monitoring, measuring, processing to sensing data is performed on. On the other hand, it is used sensor nodes to convert physical phenomena into electrical signal in real world. It uses physical characteristics of sensor that are corresponding to the variation of environmental phenomena. In this system, selected devices could measure a range of temperature and humidity in sensing field.

Second, it is a part of base system, where is performed monitoring, and storing, and processing for sensing data from sensor nodes. Base system is composed of processor, memory, other component that is designed to satisfy these requirements.

Third, prior conditions which should run over limited source that must be satisfied. To realize the prior conditions, it must possess reasonable energy and minimum energy consumption mechanism. It is the reason that sensor nodes basically run unattended.

2.2 Software Platform

An application service system must provide user to various services and administer the hardware system which consist of a lot of sensor nodes. To realize these purposes, it should be necessary to be prepared various functions. In this chapter, we will take about it.

First, it is function of query processing. Catalogues of query are differentiated into snapshot, continuous, events, and spatiotemporal query. A snapshot query corresponds to user's requests from base system. Its requests are happened in real time. For example, there is request to output the current status information of sensor nodes. A continuous query corresponds to user's request which is happened at an interval of time or spatial. Function of event query processing refers to means that is provided for corresponding to happened events. When specified event is happened, the service corresponding events is performed. For last, there is a spatiotemporal query. As advancement, intelligence, and context awareness of sensor network system, it should be necessary to accomplish this goal.

Second, it should be necessary to acquire, process, and manage sensing data. Various data are acquired from sensor nodes. Depending on the features of system, an acquired data can be managed, stored, processed by a set of methods. On the other hand, the methods are differentiated into 3 catalogues: local storage, external storage, and clustered storage. In case of the first method based on local storage, where uses a local storage for monitoring, storing, and managing acquired data. It uses memory inside sensor nodes. In case of clustered storage, after selecting head sensor node in sensor fields, it uses head sensor node's memory to store sensing data. At last, in case of external storage, it is necessary to transmit sensing data to base system. All sensing data is accumulated at memory of base system. Then an amount of traffic rate may be occurred. For above various methods, we must select method depending on the

features of designed system. In this case of proposed system, sensing data is composed of a set of temperature and humidity data is transmitted to base system in a period of time. Furthermore an amount of traffic rate relatively is smaller than other system.

Third, it should be necessary to administrate Meta data. As described by figure 1, designed system consists of sensor nodes, base system, and actuator. Meta data about hardware components of implemented system may be used to provide various services with users. Where, Meta data refer to information about hardware components is differentiated into static and dynamic type. Static information refers to ID of sensor node, the number of sensor node, and catalogue of sensors and actuators. Dynamic information refers to a status data of sensor nodes and power changes in time. To use the Meta data, after initial configuration for system, it should be necessary to monitor given system a period of time.

Fourth, it is necessary to create and administrate context information. According to sensor network application system and services is advancing, it is necessary to context information. Creating and administrating context information refers to make an export system connected with existed DB system. Throughout extracting sensing data from DB and connecting existed DB, we can create new context information.

Fifth, it is necessary to acquire and administrate location, time, and other information. As I mentioned, in order to create and manage the described context information, temperature, humidity, time, position, and proximity information should be taken and managed.

3 Design of Service System

Proposed application service system was composed of sensing, meta, and context information services. We will take about these.

3.1 Sensing Information Services

It is use SHT7X devices to acquire an amount of temperature and humidity data in given environment. Figure 2 is data sheet of SHT7X. It can measure a range of 0~99°C temperature and 0~99% humidity. Also using the real time clock and GPS module, we can acquire and store time and location information.

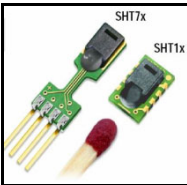
	Part Number	Humidity accuracy [%RH]	Temperature accuracy [K] @ 25 °C	Package
	SHT10	±4.5	±0.5	SMD (LCC)
	SHT11	±3.0	±0.4	SMD (LCC)
	SHT15	±2.0	±0.3	SMD (LCC)
	SHT71	±3.0	±0.4	4-pin single-in-line
	SHT75	±1.8	±0.3	4-pin single-in-line

Fig. 2. SHT7X and Data Sheet

3.2 Meta Information Services

Meta data used in the application service system was static and dynamic data. Where, static Meta data- ID of sensor node to provide sensor’s identification and location information of deployed sensor node, status information of valve of actuator 1(water tank) to check “open” or “close”, status information of valve of actuator 2(heater) to check “on” or “off”- was included. Also dynamic information consists of temperature of water in water tank, water level information, temperature and humidity information of place where sensor nodes is deployed, operation status information of actuator3(fog spray) and actuator4(hot/cool spray).

3.3 Context Information Services

According to advancement of sensor networks, a sophisticated system should need to make context awareness. Using the context information, an application service system can provide various and intelligent services to users. Proposed system also can be able to use for context information. As provided intelligent services, functions of operating mode configuration- an intelligence mode, a humidity control mode, a timer control mode, a sequential sterilization mode- can be supplied. Throughout calculating a heat index, it can be implemented to control proposed system. Figure 3 show the implemented panel to be shown various status of the system control.

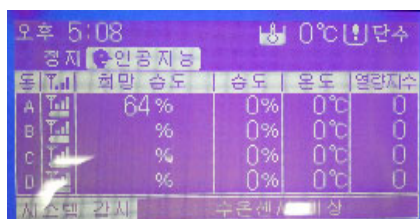


Fig. 3. Implemented Panel

4 Conclusion

Proposed sensor network system is related with the design of system to control temperature and humidity in specified environment. To achieve these purposes, sensor module is composed of many sensing devices-temperature, humidity, time, location. Based on the acquired information from these devices, we can implement the application service system that can provide functions of query processing, acquiring and processing of sensing data, acquiring and managing of meta data, and processing of a context information to users. As the result, we can effectively maintain and administer specified environment.

Acknowledgements

This work(Grants No.09-03) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2010.

References

1. Akyildiz, I.F., Su, W.: A survey on Sensor Networks. *IEEE Communication Magazine*, 102–114 (August 2002)
2. Culler, D., Estrin, D., Srivastava, M.: Overview of Sensor Networks. *Computer*, 41–49 (August 2004)
3. Heinzelman, W., Murphy, A., Carvalho, H., Perillo, M.: Middleware to Support Sensor Network Applications. *IEEE Network Magazine Special Issue* (January 2004)
4. Hadim, S., Mohamed, N.: Middleware Challenges and Approaches for Wireless Sensor Networks. *IEEE Distributed System Online* 7(3) (2006)
5. Yoneki, E., Bacon, J.: A survey of Wireless Sensor Network Technologies: Research Trends and Middleware's Role. Technical Report Published by University of Cambridge (2005)
6. Kwak, Y.S., et al.: Design and Implementation of Sensor Node Hardware Platform Based on Sensor Network Environments. *The Journal of Korea Navigation Institute* 14(2), 227–232 (2010)
7. Kwak, Y.S., et al.: Design and Implementation of the Control System of Automatic Spray Based on Sensor Network Environments. *The Journal of Korea Navigation Institute* 15(1), 91–96 (2011)

Software RAID 5 for OLTP Applications with Frequent Small Writes

Kijeong Khil¹, Dongho Kwak¹, Seokil Song¹,
Yunsik Kwak¹, and Seungkook Cheong²

¹Dept. of Computer Engineering, Chungju National University, Chungbuk, Korea

²Dept. of Electronics and Telecommunications Research Institute(ETRI),
Yuseong, Daejeon, Korea
setoutroad@gmail.com, kb2clsrn@nate.com,
sisong@cjnu.ac.kr, yskwak@cjnu.ac.kr, skjeong@etri.re.kr

Abstract. In this paper, we propose a DDR-SSD conscious software RAID system for OLTP applications. The proposed software RAID employs a logging based method to improve the performance of software RAID 5. Small writes that are frequently occurred in OLTP applications are the main factor to degrade the performance of RAID systems. Several methods for a hard disk based software RAID system are proposed to solve the small write problems. It is straightforward to apply the existing methods to a DDR-SSD based software RAID system. However, DDR-SSDs have unique characteristics which are different from hard disks. If the unique characteristics are taken into account carefully when designing the software RAID system, its overall performance will be improved. In this paper, we propose a differential-logging method considering the DDR-SSD to solve the small write problem. We show that our proposed method outperforms the existing software RAID through simulations.

Keywords: DDR-SSD, RAID, Small Write.

1 Introduction

RAID (redundant array of independent disks) is a popular technology to achieve performance, reliability or both by using many inexpensive hard disks. RAID has various implementations and configurations in terms of performance, reliability, and price. Hardware RAID implements RAID functions on the system board or by inserting an add-on board into a PCI bus slot. In addition, hardware RAID generally includes an acceleration circuit for performance enhancement, escalating the price. On the other hand, software RAID implements RAID functions in the kernel of the operating system or a device driver, and software RAID is provided by most commercial operating systems. Since software RAID does not require dedicated hardware, reliability and performance can be improved at a relative low cost.

Various RAID architectures can be configured, such as RAID 0, 1, 2, 3, 4, 5, 6, 0+1, 1+0, and 5+0, according to how logic blocks are stored in physical disks [1]. RAID 2, 3, and 4 are rarely used; RAID 0, 1, and 5 and hybrid forms of these

configurations are mainly used in commercial products. RAID 5 partitions data according to predefined block units and calculates parity among the blocks. Parity is stored in a separate parity block, which is dispersed into storage devices to distribute the workload. Upon disk failure in RAID 5, data can be restored using the parity block. RAID 5 provides a relatively high performance for read requests using its enhanced high parallelism. However, RAID 5 increases the workload because write operations require access to the parity block, as well as the data block, and modifying the existing parity block. Since software RAID system does not have separate hardware, additional workload caused by parity in RAID 5 affects the overall performance even more so than hardware RAID. In particular, frequent small writes in enterprise applications trigger RMW (read-modify-write) operations, further undermining the performance of the storage system.

In order to recover data during disk failure in RAID 5, the parity block has to be maintained with consistency. Therefore, if some blocks of the stripe are modified, parity must be re-calculated and recorded in the parity block. If the entire stripe has been modified or newly written, the parity block that had been calculated does not need to be read again for parity calculation. However, in case of a small scale change, the pre-modification block and the parity block must be read from the storage device, and the XOR (exclusive OR) logical operation has to be performed on the new block, pre-modification block, and the parity block. Then the new block and the parity block are written on the disk. In summary, even if there is change in a single block of the stripe, four input/output operations and two XOR operations must be performed. This modification process is referred to as RMW (read-modify-write). Enterprise applications often require small writes that trigger frequent RMW, severely undermining the overall performance of the storage system, as shown in Table 1,

A number of technologies have been developed to resolve this problem. Existing technologies for considering small writes can be categorized into cache management techniques [3,4] and parity logging techniques [5,6]. These technologies either propose means for improving small writes in RAID based on hard disk, or require separate hardware in hardware RAID. The novel approach to improving RMW performance in RAID 5 proposed in this paper is based on DDR-SSD software RAID. DDR-SSD is SSD based on DRAM, and offers I/O performance up to 100 times greater than that of hard disk. Moreover, there is virtually no performance difference between sequential I/O and random I/O [7]. Taking into full consideration these properties of DDR-SSD, this paper proposes an RMW technique for RAID 5 using differential logging. As a modified version of the conventional parity logging algorithm, the proposed RMW technique reduces the costs associated with I/O and parity logging, and prevents data loss during system failure.

This paper is organized as follows: Chapter 2 summarizes related work, and Chapter 3 explains the proposed RMW technique based on differential logging. Chapter 4 provides the results of performance evaluation, and the conclusion of the paper is given in Chapter 5.

2 Related Work

As mentioned in the previous chapter, conventional technologies addressing small writes can be categorized into cache management techniques and parity logging techniques. The RMW technique proposed in this paper is based on parity logging. Accordingly, let us briefly examine conventional parity logging schemes.

The parity logging technique proposed in [5] reduces the number of I/O operations caused by small writes. When a small write occurs, a parity log is created and multiple parity logs are compiled and converted into a few large writes, reducing the number of I/O operations. Parity logging requires an additional disk for recording logs. Whereas RAID 5 requires $N+1$ disks (N data disks and one parity disk), parity logging requires $N+2$ disks (N data disks, one parity disk, and one log disk). When a small write occurs in parity logging, the entire block is read, and XOR is performed with the new block to create the parity block, which is referred to as the parity log. The parity log is temporarily registered in the log buffer of the main memory device, and transferred to the log disk after the buffer size reaches a specified level. As with the parity block of RAID 5, the parity log stored in the log disk can be distributed over multiple disks.

[6] proposes parity log compression, which involves reducing the size of the parity block (log) using a specific compression technique. For stability reasons in parity logging, logs are stored in a buffer located in nonvolatile memory, which is very expensive. Therefore, it is difficult to use a sufficient amount of memory. Parity log compression allows a more efficient use of the nonvolatile memory and reduces the number of I/O operations.

3 Proposed RAID 5 RMW Based on Differential Logging

This paper proposes a differential logging technique for improving small write processing performance in software RAID 5 based on the properties of DDR-SSD. The proposed differential logging technique prevents block losses caused by failure in storage devices, as well as loss of latest data during system failure.

In the differential logging technique proposed in this paper, when creating a parity log, XOR is performed only on the modified sectors of the pre-modification block and the modified block, rather than on entire blocks. The log created in this manner shall be referred to as a "differential log" in this paper. The primary difference between the parity log proposed in [6] and the differential log is the log size. While the size of a parity log is equal to the block size, the size of a differential log is equal to the number of modified sectors. Furthermore, the cost associated with XOR operation for creating the logs is reduced. In disk-based software RAID, the cost of XOR operation was minimal compared to that of disk I/O. In software RAID based on DDR-SSD, however, the cost of XOR operation cannot be dismissed because it is approximately same as the I/O cost of DDR-SSD. Therefore, in addition to reducing the number of I/O operations, decreasing the costs associated with XOR operation becomes an important factor for improving the overall performance in software RAID based on DDR-SSD. Since differential logs are created only for modified sectors, header information is required to explain the differential logs. The format of a complete differential log that includes header information is *diff-log (disk_id, block_id, start_sector,*

end_sector, type, log). A *type* is assigned to distinguish a normal differential log from a dummy log, which will be explained later.

The architecture of the proposed software RAID systems is shown in Figure 1. The stripe manager receives an I/O request from the file system and creates a stripe according to the number of storage devices and the RAID level. Since the differential logging technique proposed this paper focuses on improving the performance of small writes, we shall assume hereafter that I/O requests are small write requests. The stripe created according to a small write request is transferred to the differential log manager, which reads the old version of the modified block from the I/O manager. XOR is performed on the modified sectors of the modified block to create a differential log, which is recorded in the log region of the disk where the parity block of the corresponding stripe is stored. The differential log is then transferred to the differential log buffer manager so that it can be managed in the differential log buffer.

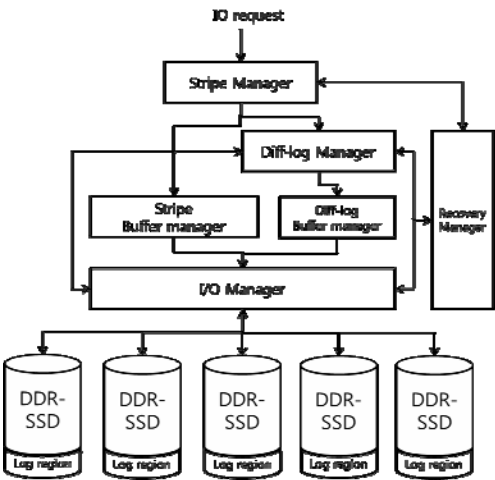


Fig. 1. Structure of software RAID using differential logging technique

The stripe manager sends the created stripe to the stripe buffer manager, which stores and manages the stripe in the stripe buffer. The stripe buffer uses the LRU (least recently used) policy to manage the stripe buffer. When flushing a particular disk block, the stripe buffer reads the corresponding block and related differential logs from the differential log buffer manager. Then the parity block is read from the storage device to create a new parity block. The newly created parity block is recorded in the storage device and the stripe is flushed. At the same time, the stripe buffer creates a dummy log in the differential log region of the disk that stores the parity block, indicating that the disk block to which the corresponding parity block is flushed has been updated. The differential log buffer manager checks the block where modification has occurred to generate the differential log sent by the differential log manager. The differential log buffer manager also verifies whether the differential log buffer contains another differential log created in relation to the corresponding block. If a previously created differential log exists, the manager finds the sectors that overlap with the new differential log, performs XOR, and combines non-overlapping sectors

to create a single differential log. The combined differential log is again recorded in the differential log buffer, and the old differential log can be deleted.

The restoration manager is activated when the stripe is lost due to system failure or when there is a problem with a storage device in RAID. The stripe buffer is lost during a reboot caused by system failure, and some stripes will have old data rather than latest data. Since the differential log buffer is also lost, the log region of each storage device must be read and transferred to the main memory in order to restore lost data. The log region of each storage device is sequentially read in the reverse-chronological order (from the latest differential log), and the differential log that corresponds to each disk block is appended. If a dummy log for a disk block is found, the process of reading and appending differential logs is no longer performed for that particular disk block. After completing the process for every differential log, the combined differential log and the disk block stored in the storage device (old disk block) are read to perform XOR operation, restoring the latest disk block.

Disk blocks lost due to faulty storage device are restored using the typical restoration process performed in RAID 5. XOR is performed on data blocks and parity blocks read from undamaged disks to restore lost blocks. In the differential logging technique of this invention, the parity block for the data blocks of a single stripe in a storage device always maintains consistent parity. Parity update for the modified data block takes place when the corresponding block is flushed (modification of the block is reflected in the storage device). Therefore, from the storage device's perspective, the data block is updated simultaneously with the parity block. When the log has to be flushed because the log region of each storage device has become full, the parity block is updated and the modified blocks are flushed at the same time, maintaining consistency between the data blocks and the parity block of the storage device. Accordingly, the stripe stored in the disk (including the parity block) is read and XOR is performed to restore the blocks of the damaged disk.

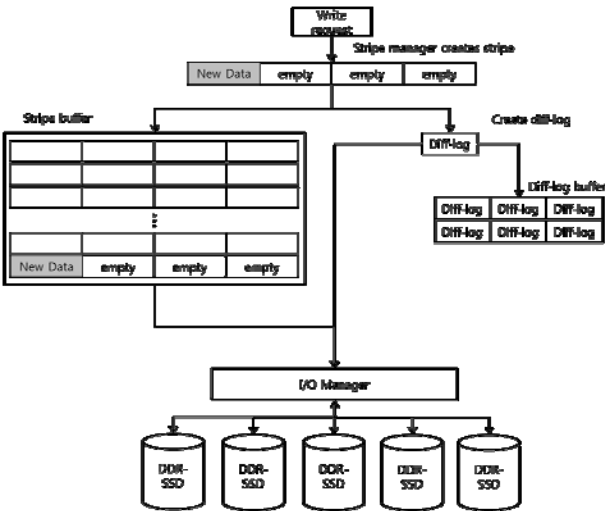


Fig. 2. Process of a small write request

Figure 2 depicts how the differential logging technique processes a small write request. A small write request is created into a stripe and stored in the stripe buffer. At the same time, the differential log manager creates a differential log, which is stored and managed in the log region and the differential log buffer of the disk.

Figure 3 explains the difference between conventional software RAID, parity logging technique and the differential logging technique. In the case of typical write processing, differential logging performs two I/O and one XOR operations, whereas the conventional scheme performs four I/O and two XOR operations. Moreover, the single XOR operation is performed among the modified sectors of the block rather than on the entire block.

Original RMW	Parity logging RMW	Diff-logging RMW
1. Read old block (OB) 2. Read old parity (OP) 3. Calculate new parity (NP) $NP = OB \text{ XOR } NB$ XOR OP 4. Write new block (NB) 5. Write new parity (NP) Commit	1. Read OB 2. Calculate parity log $\text{parity log} = OB \text{ XOR } NB$ 3. Write parity log to NVRAM 4. Write NB 5. When NVRAM is full, write all parity log in NVRAM to HDD Commit	1. Read OB 2. Calculate diff-log $\text{diff-log} = OB_{\text{user data}} \text{ XOR } NB_{\text{user data}}$ 3. Write diff-log to log region Commit

Fig. 3. Comparison between conventional RMW vs. parity logging RMW vs. differential logging RMW

4 Performance Evaluation

In order to evaluate the performance of RMW based on the differential logging technique proposed in this paper, a simulator was implemented on the Linux platform using the C language. Comparison was made with the software RAID solution that Linux provides by default. For proper comparison, the simulator for software RAID was also implemented on the Linux platform using the C language. We used Financial1 data from SPC for the experiment. Other parameters used in simulation are shown in Table 2.

Table 1. Simulation parameters

Data Set	SPC Financial 1 Trace
Number of Disks	8
Page Size	4K ~ 32K
Read Time (us)	4K(3.7), 8K(4.8), 16K(7.1)
Write Time (us)	4K(13.8), 8K(15.8), 16K(19.6)
Log Region	1M ~ 128M/Disk
Stripe Buffer	(128 ~ 1024) X Stripe Size
XOR Time (us)	4K XOR 4K -> 11us

The read and write times of DDR-SSD shown in Table 2 were acquired from [8]. Figure 6 displays the IOPS results when the page sizes were modified to 4k - 32k and the log sizes were 1M - 128M. It can be seen from Figure 4 that the proposed technique yields substantially higher performance than the software RAID scheme provided by Linux. As for the log region, while performance improved for large log regions, the differences were not significant. Figure 5 displays the IOPS results obtained by the two methods as page sizes were varied. It can be seen from Figure 7 that the proposed technique yields higher performance compared to the software RAID scheme provided by Linux. For pages larger than 8k, performance improvement was not significant. As the page size increased and the size of the log region became smaller, performance decreased.

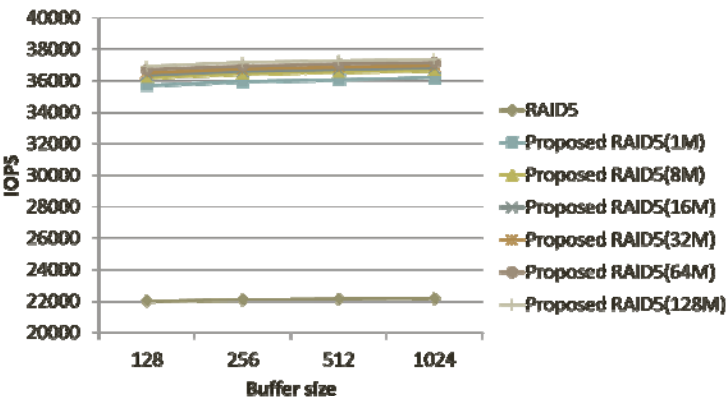


Fig. 4. IOPS with various buffer sizes

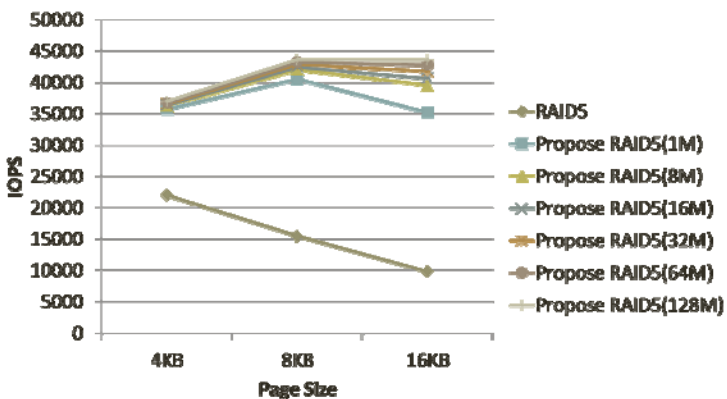


Fig. 5. IOPS with various page sizes

5 Conclusion

This paper proposed a differential logging based RMW technique for improving small write performance of software RAID based on DDR-SSD. The proposed technique restores data from disk failure, as well as lost data due to system failure. Since the I/O performance of DDR-SSD is substantially higher than that of hard disk, the differential logging technique is proposed to minimize parity calculation and prevent the cost of parity calculation from significantly affecting the overall performance. Results of performance evaluation indicate that with only 1 – 8M of additional disk space, performance of RMW can be greatly enhanced compared to the software RAID scheme provided by Linux. In future research, we intend to design and implement a differential logging algorithm based on Linux and examine its performance in applications.

References

1. Chen, P., Lee, E., Gibson, G., Katz, R., Patterson, D.: RAID: High – Performance, Reliable Secondary Storage. *ACM Computing Surveys* 26 (June 1994)
2. <http://www.storageperformance.org/>
3. Kim, J.-H., Noh, S.H., Won, Y.-H.: Cache management schemes for efficient small-writes in a Software RAID. In: *KISS 1996 Fall Conference*, vol. 23(2B), pp. 857–860 (October 1996)
4. Kim, J.-H., Noh, S.H., Won, Y.-H.: A Cache Replacement Policy for a Software RAID File System Considering Small – Writes and Reference Counts. In: *KISS 1997 Spring Conference*, vol. 24(1A), pp. 123–126 (April 1997)
5. Stodolsky, D., Gibson, G., Holland, M.: Parity logging overcoming the small write problem in redundant disk arrays. *ACM SIGARCH* 21(2), 64–75 (1993)
6. Kim, G.H., Chang, E.J., Choi, H.K.: Compressed Parity Logging for Overcoming the Small Write Problem in Redundant Disk Arrays. In: *KISS 1998 Spring Conference*, vol. 25(2III), pp. 12–14 (October 1998)
7. Hwang, J., Chung, S.-K.: Technology trend of DDR based SSD storage System. *NIPA Weekly Trend Report*, vol. 1421, pp. 28–41 (2009.11.4)

The Effects of Project, Customer and Vendor Properties on the Conflict of IS Outsourcing Development

Dong-Hwan Cho¹ and Haeng-Nam Sung²

¹ Venture & Business, Gyeongnam National University of Science and Technology,
150, Chilam-dong, Jinju, Gyeongsangnam-do, 660-758, Republic of Korea

² Management Information System, Gyeongsang National University,
900, Gazwa-dong, Jinju, Gyeongsangnam-do, 660-701, Republic of Korea

Abstract. This study aims to investigate the major causes of conflict between clients and vendors in the outsourced IS development. A research model is established based on the prior research of IS outsourcing, system development and conflict. An empirical study was executed using 214 survey questionnaires of project teams composed of client and vendor with PASW 18.0. Research results using multiple regressions show that contract concreteness, requirement constancy, accordance of goals and Technology of Knowledge have negative effects on conflict.

Keywords: IT Outsourcing, System Development, Project Performance, Conflict, Conflict Management.

1 Introduction

Nowadays, new type of information system outsourcing such as Offshore Outsourcing and ASP (Application Service Provider) has emerged. For this reason, as the market volume of outsourcing increases, Information System outsourcing has become an alternative for IS management of an organization [1][2]. This study targets Information System Development Outsourcing Project, a type of Information System Outsourcing. Information System Development Outsourcing Project(hereinafter OISD) means the activity on outsourcing the whole or partial course including requirement definition, design, development, evaluation, installation for Information System development to the outside the company. This study includes the package solutions such as ERP, SCM and CRM.

Regarding conflicts degrading achievement of OISD, the published studies have included only conflicts in context of developing internal system, not explained more significant development of external system in reality. Therefore, they could not effectively take conflicts occurred from OISD into account.

The goal of this study is to explore the main causes of conflicts between vendors and clients in OISD. To achieve the goal, the empirical analysis was performed on 214 OISD project teams. The results of this study can provide practical and academic implications for the improvement of OISD achievement.

2 Theoretical Background

2.1 Information Systems Outsourcing

Information system outsourcing is to outsource the whole or partial information system functions to outside service provider. That is, the outside service provider contributes to the whole or partial physical or personnel resources of an organization which forms the base structure of an Information technology [4].

While IS outsourcing can enhance the results of a company by focusing its capability on the core function through cutting the cost, and the effective management and distribution of resources, it does not have the Information technology ability and the control of the information resources within the organization in the long term. Therefore, the outsourcing has both danger and opportunity. When a company is considering outsourcing, it needs to be prepared for them.

2.2 Conflicts in the Area of Information System

This study on the conflicts in the area of Information System was mostly conducted by Robey et al, and Barki and Hartwick. Robey and Farrow (1982) suggested user participation, effects and conflicts, and the model about the solutions [6]. Subsequently, Robey, Farrow and Frans(1989) verified the model of conflicts in the system development in the level of group[7]. Robey, Smith and Vijayasaraty(1993) figured out the links between user participation in System Development, conflicts, its resolution and project success[8]. Barki and Hartwick(1994) verified the associates' models and established the parameter role of multidimensional property and effect of conflicts, and Barki and Hartwick(2001) established the property of inter individual conflict and the role of conflict management during the system development focusing on the individual conflicts.

Since then, Yeh and Tsai(2001) shed new light on the cause of potential conflict and the role of user participation[11], Cohen et al(2004) suggested the causes of conflicts and the method of managing them targeting the test procedure, a part of System development. Recent studies reviewed the effect of conflict on the results of project and suggested the specific and various causes of conflicts [10][12].

3 Research Model and Hypotheses

3.1 Research Model

This study based on the precedent research set up the research model which can be seen in [fig.1]. In OISD, the precedent factors will be divided into vendor property (vendor's power, vendor developer's knowledge about application domain), client company property (the degree of accordance between goals and client company worker's knowledge on development process) and project property (the concreteness of a contract, the constancy of requirement). The effects of those properties will be

verified on conflict. The dotted line between conflicts and project results shows the relation verified in the earlier empirical studies [7][9][13].

The precedent factor affecting the conflicts between client company and the vendor is based on Jehn et al. (1999)[13], one of the studies about the diversity and conflicts within the work group, project team. In those studies, the diversity in the group was assorted into social category diversity, informational diversity, the value diversity, and the relations between these diversities and the conflicts are reviewed.

As the period and the volume of the projects increases, the complexity of the system development task [1][2], which leads to the high possibility of conflicts among team members. Therefore, the period and the volume of the projects are set as control variables.

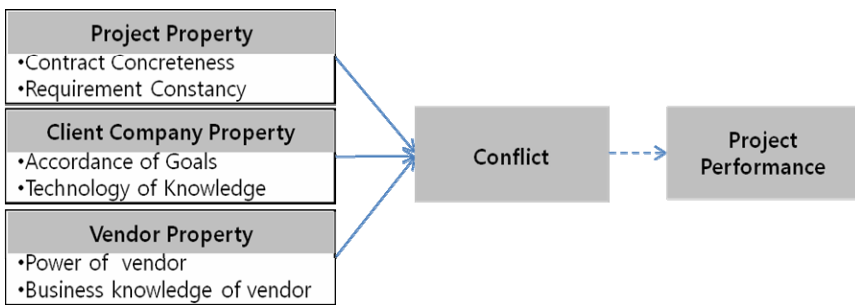


Fig. 1. Research Model

3.2 Research Hypotheses

Project Property. In outsourcing, it happens the case where the content of the posted RFP(Request For Proposals) is different from that to be developed in reality. That's because the requirements were not reflected clearly. OISD is the process where the developers of the vendor find the requirement about the system which client company users have. The changes of requirement of client company occurs [15][16].

First, when, in the beginning of the development, the requirements of users about the system were not clear. Second, when there is not enough mediation between the departments of users in different situations, or when the range or user is so comprehensive that the range includes the inside organization and the outside of it as well, the conflicts could occur between the requirements from them. The above discussion leads to the next hypotheses.

Hypothesis1 Project property affects the conflicts.

Hypothesis1-1 Contract concreteness will have a negative effect on the conflicts.

Hypothesis1-2 Requirement constancy will have a negative effect on the conflicts.

Client Company Property. The client company decides to outsource when it needs to get strategic, economic, technological benefits [4][12], and this decision is made

finally by management or steering committee. The goal of an organization or management does not always correspond with that of the team conducting outsourcing project.

If they do not fully understand the development process, this can cause communication gap, confusion, tension and complaints, linked to the lack of participation, leading to conflicts [14]. Here following hypotheses can be established.

- Hypothesis 2* Client company property affects the conflicts.
- Hypothesis2-1* The degree of the accordance of client company goal with project team goal will have a negative effect on the conflicts.
- Hypothesis2-2* The technology knowledge of the client company works will have a negative effect on the conflicts.

Vendor Property. Power or authority is the ability for a member or an organization to make another member or organization do what he or it wants to do. The exercise of power or the power relations (power asymmetry) is the major cause for conflicts.

In the case of developing from outsourcing, vendor developers will be able to have lower knowledge on organizations or business. If these vendors do not know the industry or the organization to which the client company belongs, they will have communication gap with the client company users, which will lead to misunderstanding their explanations or requirements [19]. The misunderstanding causes disagreement and conflict in the system development. Here following hypotheses can be established.

- Hypothesis 3* Vendor property affects the conflicts.
- Hypothesis3-1* The power of the vendor will have a negative effect on the conflicts.
- Hypothesis3-2* The business knowledge of vendor will have a negative effect on the conflicts.

3.3 Survey Measures

Survey design was chosen to verify the research model. The survey for this research was pretested by the work site experts before distribution and measured by 7 point Likert scale. The definition for the research variables are summarized in [Table 1].

Table 1. Operational Definition of the Constructs

Variable, Source	Operational Definition
Contract Concreteness [5]	Setting up the standards and evaluation about development process and output, the concreteness about the regulation to handle the exceptions.
Requirement Constancy [16]	Unclearness and inexactitude of the user’s requirement in the beginning of the project
Accordance of Gals [12] [9]	Accordance of the goal of project team with that of client company at the level of organization.

Table 1. (*Continued*)

Technology of knowledge [17] [9]	Understanding about general system development process and project development process
Power of Vendor [18]	Potential influence of the vendor that client company perceives, Which is the ability to lead some change in the relation of outsourcing
Business Knowledge of Vendor [9]	The knowledge of the application domain which the vendor developers retain for the system development
Conflict [10]	The experience of negative feeling between client company user and the vendor form the communication gap or interference during the project

4 Research Method and Results

4.1 Research Method

The data collection was carried out by mail, e-mail, fax and a personal visit. For sampling frame were more than 200 businesses participating in the business management course for executives in Y University in Korea, and more than 200 businesses attending IT related course of study conducted by the Federation of the Korean Industries. The 586 surveys were sent and distributed during the 9 weeks of survey period, and the final collected surveys were 242. To remove the abnormality for the date, the surveys including errors or insincere responses were excluded, 214 projects (163 companies) were used for the final analysis. PASW 18.0 of SPSS was used for the statistics analysis.

The project period for more than 6 months and less than 1 year has the highest percentage of 39.7, the volume of the project team of more than 6 people and fewer than 10 people has the highest percentage of 35.0, and the project budget of less than 100 million was the highest percentage of 21.5.

4.2 Analysis and Results

Reliability and Validity Tests of Measurement Instrument. According to [Table 2], the analysis shows this research the reliability as the value of Cronbach's α is more than 0.8, and the validity through the factor analysis of independent variable and dependent variable.

To verify the distinction validity, correlation analysis and AVE(Average Variance Extracted) were conducted.

All the square root of AVE value about the variable proved bigger than the all the coefficient of correlation, which means that there are distinction validity among variables.

Table 2. Reliability Analysis & Factor Analysis

Var.	Cronbach's α	de.1	de.2	de.3	de.4	de.5	de.6	Inde.1
PV1	.890	<u>.757</u>	.090	.076	.249	.169	.006	
PV2		<u>.717</u>	-.084	.157	.107	-.022	-.161	
PV3		<u>.728</u>	-.049	.197	.003	.196	.017	
PV4		<u>.766</u>	.027	.227	.116	.161	-.001	
PV5		<u>.790</u>	-.007	-.004	.110	.105	-.006	
PV6		<u>.777</u>	-.018	.152	.150	.084	.093	
PV7		<u>.728</u>	.059	.190	-.042	.065	.021	
TK1	.904	.051	<u>.797</u>	.033	.061	.089	.092	
TK2		-.019	<u>.857</u>	.126	-.060	.085	.066	
TK3		-.064	<u>.849</u>	.154	.013	.012	-.016	
TK4		.039	<u>.854</u>	-.003	.064	.070	.059	
TK5		-.012	<u>.866</u>	.032	.034	.124	.034	
KV1	.902	.120	.076	<u>.833</u>	.109	.070	.002	
KV2		.194	.122	<u>.835</u>	.035	.108	-.071	
KV3		.224	.072	<u>.864</u>	.041	.095	.026	
KV4		.180	.076	<u>.778</u>	.161	.128	.118	
KV5		.217	.025	<u>.693</u>	.280	.158	.165	
CC3	.848	.104	-.029	-.059	<u>.832</u>	.108	.003	
CC4		.055	-.087	.010	<u>.866</u>	.144	.047	
CC5		.113	.152	.160	<u>.774</u>	.081	-.057	
CC6		.054	.135	.240	<u>.704</u>	.082	-.015	
CC7		.259	-.042	.231	<u>.636</u>	-.003	.005	
AG2	.902	.217	.133	.140	.124	<u>.844</u>	.148	
AG3		.219	.116	.140	.217	<u>.879</u>	-.032	
AC4		.206	.166	.223	.098	<u>.810</u>	.023	
RC1	.810	-.067	.017	.032	.001	-.034	<u>.861</u>	
RC2		-.021	.082	.006	-.059	.048	<u>.896</u>	
RC3		.057	.100	.101	.041	.093	<u>.771</u>	
Cf1	.899							<u>.814</u>
Cf2								<u>.848</u>
Cf3								<u>.851</u>
Cf4								<u>.845</u>
Cf5								<u>.865</u>
Eigen-value		4.392	3.753	3.684	3.272	2.416	2.269	3.568
Pct of Var (%)		15.69	13.40	13.16	11.69	8.630	8.10	71.36
Cum Pct (%)		15.69	29.09	42.25	53.94	62.57	70.67	

PV: Power of Vendor, TK: Technology of Knowledge, KV: Biz Knowledge of Vendor, CC: Contract Concreteness, AG: Accordance of Goals, RC: Requirement Constancy

Hypothesis Test. To test the hypothesis, the multiple regression analysis was conducted. The project period and the volume were controlled [1], While the project period was not statistically meaningful, the project volume proved statistically

significant. In fact, the project volume means the complexity of the project and the one of the task, which means the bigger or the more complex the project is, the more conflicts there are among the team members.

Table 3. The Result of Hypothesis Testing

Independent variable	Dependent variable	R ²	Beta	t-value	not reject or reject
project period	conflict	.321	-.014	-.215	control
Project volume			.185	2.756***	control
Contract concreteness			-.138	-2.172**	not reject
Requirement constancy			-.270	-4.264***	not reject
Accordance of goals			-.224	-3.450***	not reject
Technology of knowledge			-.140	-2.224**	not reject
Power of vendor			-.071	-1.121	Reject
Business knowledge of vendor			-.083	-1.287	reject
* p <.1 , ** p<.05, *** p<.01					

As the result of the hypothesis testing, the two variables of the project property, both the contract concreteness and the requirement constancy are significant. When the project contract is concrete and specific, and when the requirements of the user are clear and consistent from the beginning of the project throughout the period, the conflicts decline. [Table 3] shows the summary of the result.

The two variables of the client company property, accordance of goals and technology of knowledge indicates both meaningful variables. Accordance with outsourcing purpose of client company and achieving goals of project team enables conflicts to be lower. This means conflicts can be lower if client company staff in project have higher knowledge on developing process.

The two variables of the vendor property, power of vendor and the business knowledge of vendor are not all meaningful. In a word, power or authority of vendor and application knowledge of vendor developers in project do not influence conflicts any more. In the outsourcing project, client company constantly controls the vendor and tries to explore and implement effective control mechanism [1][3].

5 Conclusions

The main conclusion and implication in this study are as follows. Firstly, it aims to clarify the clauses, which conflicts happen between customer company and vendor, degrading OISD achievement through outsourcing. This suggests precautions against conflicts in advance or manageable plan.

Second, like the precedent studies [5][15][16], the project property turned out the important factor influenced on the conflict. Despite the inherent uncertainty and complexity of information system development, the concrete contract reduces the

possibility of conflict. If the requirements of users in client company are clearly investigated, and changes on the requirements are reduced, the conflict can be lower.

The close investigation on the vendor power in various ways is essential in the future study. A profound study for the developers' knowledge together with the knowledge on the application domain and IT is required.

References

1. Rustagi, S., King, W.R., Kirsch, L.J.: Predictors of Formal Control Usage in IT Outsourcing Partnerships. *Information Systems Research* 19(2), 126–143 (2008)
2. Gopal, A., Gosain, S.: The Role of Organizational Controls and Boundary Spanning in Software Development Outsourcing: Implications for Project Performance. *Information Systems Research*, Published Online in Advance 1–23 (2009)
3. Choudhury, V., Sabherwal, R.: Portfolios of Control in Outsourced Software Development Projects. *Information Systems Research* 14(3), 291–314 (2003)
4. Grover, V., Cheon, M.J., Teng, J.T.C.: The Effect of Service Quality and Partnership on the Outsourcing of Information Systems Functions. *Journal of Management Information Systems* 12(4), 89–116 (1996)
5. Lacity, M.C., Hirschheim, R.: The Information Systems Outsourcing Bandwagon. *Sloan Management Review* 35(1), 73–86 (1993)
6. Robey, D., Farrow, D.: User Involvement in Information System Development: A Conflict Model and Empirical Test. *Management Science* 28(1), 73–85 (1982)
7. Robey, D., Farrow, D.L., Franz, C.R.: Group Process and Conflict in System Development. *Management Science* 35(10), 1172–1191 (1989)
8. Robey, D., Smith, L.A., Vijayasarathy, L.R.: Perceptions of Conflict and Success in Information Systems Development Projects. *Journal of Management Information Systems* 10(1), 123–139 (1993)
9. Barki, H., Hartwick, J.: User Participation, Conflict and Conflict Resolution: The Mediating Roles of Influence. *Information Systems Research* 5(4), 422–438 (1994)
10. Barki, H., Hartwick, J.: Interpersonal Conflict and Its Management in Information System Development. *MIS Quarterly* 25(2), 195–228 (2001)
11. Yeh, Q., Tsai, C.: Two Conflict Potentials During IS Development. *Information & Management* 39, 135–149 (2001)
12. Aladwani, A.M.: An Integrated Performance Model of Information Systems Projects. *Journal of Management Information Systems* 19(1), 185–210 (2002)
13. Wakefield, R., Leidner, D.E., Garrison, G.: A Model of Conflict, Leadership and Performance in Virtual Teams. *Information Systems Research* 19(4), 434–455 (2008)
14. Jehn, K.A., Northcraft, G.B., Neale, M.A.: Why Differences Make A Difference: A Field Study of Diversity, Conflict and Performance In Workgroups. *Administrative Science Quarterly* 44, 741–763 (1999)
15. Cohen, C.F., Birkin, S.J., Garfield, M.J., Webb, H.W.: Managing Conflict in Software Testing. *Communications of the ACM* 47(1), 76–81 (2004)
16. Nidumolu, S.: The Effect of Coordination and Uncertainty on Software Project Performance: Residual Performance Risk as an Intervening Variable. *Information Systems Research* 6(3), 191–219 (1995)

17. Wallace, L., Keil, M., Rai, A.: How Software Project Risk Affects Project Performance: An Investigation of the Dimensions of Risk and an Exploratory Model. *Decision Sciences* 35(2), 289–321 (2004)
18. French, J.R., Raven, B.: The Bases of Social Power, In *Studies In Social Power*, pp. 150–167. University of Michigan Press, Ann Arbor (1959)
19. Cross, J.: IT Outsourcing: British Petroleum's Competitive Approach. *Harvard Business Review* 73, 94–102 (1955)

A Study on the Split Algorithm of URL LIST Collected by Web Crawler

IL-Kwon Lim, Young-Hyuk Kim, Seung-Gu Kang, and Jae-Kwang Lee*

Department of Computer Engineering, Hannam Univ., Daejeon, Korea
{iklim,yhkim,sgkang}@netwk.hannam.ac.kr, jklee@hnu.kr

Abstract. The development of cloud technology and the web-based technology and serching technology is becoming important. About the web crawler technology which collects URL for serching, one of issues of the distributed crawler system is the effective URL split. Therefore, this study designed algorithm to split the URL LIST effectively collected by the web crawler and the split algorithm environment of URL LIST collected by web crawler was composed by implementing depositary in Hadoop environment.

Keywords: WebCrawler, URL split, Hadoop.

1 Introduction

Since the term as cloud computing first emerged in 2006, the appearance of smart phone and personal computing environment moved to the web-based mobile & the various phones with PC, the use of information and communication method have changed, and the existing searching engine technology is becoming important in addition to web-based cloud technology. And the technology to find information which is needed by the users through Internet as the sea of information more correctly and rapidly, accordingly, the various search engines to find and gather information were born, and the thing that collects the documents instead of the human being is the crawler[1].

The web crawler is the computer program to explore World Wide Web by the organizational and automatic method. It acts the role that conducts proper process after bringing the web page during surfing internet to the local machine and keeps it in the depositary for applying this information by the search engine in the future. This is called ants, automatic indexers, bots, worms, Document Collection robots, web spider, web robot, documents collection agents etc. [1].

The process of web crawler is called the web crawling or spidering, and the web crawling is conducted continually for maintaining the latest status of data in the various sites like search engine. The study on establishing web crawler by the Hadoop based distributed system to realize crawling is processed, and it is a problem about treatment of URL split as one of the issues of the distributed crawler system [2][3].

* Corresponding author.

The existing studies adopted method that distributing URL in each node by using domain of URL or IP address [3]. Therefore, the slotting method was designed as URL split method in this study, and the Hadoop-based data base system was proposed as the depositary of URL LIST.

2 Related Research

2.1 Single Web Crawler Architecture

As Figure 1, the single web crawler architecture downloads the web page by the starting URL and the URL was extracted from the downloaded page. The extracted URL should be checked in duplicate and un-checked URL should be stored in URL list. The important component of the single web crawler architecture is as follows, and the operation method of this components is as [Figure 1][3].

Visiting URL Queue: It is URL list that is needed to be downloaded is still not loaded.

Downloader: It is component which is processed by thread and downloads the relevant web page of URL by bringing URL from Visiting URL Queue.

URL Extractor: It is component to extract out-link from the downloaded page.

URL Duplication Eliminator: It is component which reduces the duplicated URL by checking extracted URL.

Seed URL: It is component having given URL list when the crawler started first. The reason that this URL was not inputted to the Visitation URL Queue is for preventing the duplicate check of start URL.

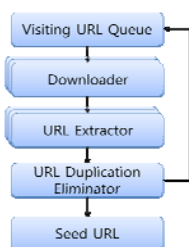


Fig. 1. Single web crawler structure

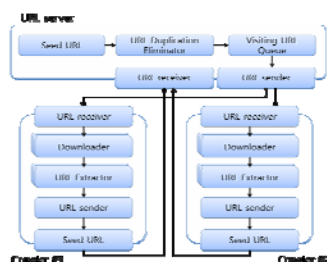


Fig. 2. Centralized web crawler structure

2.2 Distributed Web Crawler

As the amount of information on the Web grows very rapidly, the web crawler is impossible to bring data as single crawler any more and more data can be collected by distributing web crawler. There are two methods of distributed crawling, the centralized client model is the first and other method is non-centralized model of P2P method.

The centralized distributed is the method to deliver URL for each crawler’s demand after duplicate checking by the URL server. This shows the distributed web crawler structure of the centralized client server model in the [Figure 2].

Each crawler of P2Pmethod is acted like general web crawler as [Figure 3], and each crawler downloads the document and extracts the OutLine URL and removes the duplicate URL, so each crawler should operate independently [2][3].

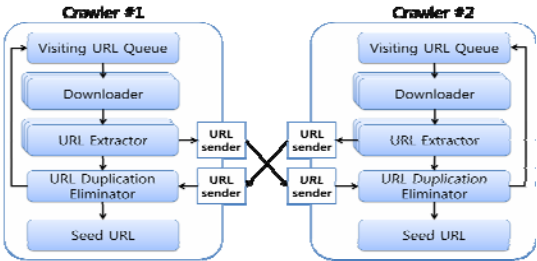


Fig. 3. P2P Drawing of the method distributed web crawler structure

2.3 Hadoop

The open source based Apache Hadoop project is the representative platform of the large sized distributed computing environment. It implemented same properties with GFS, Bigtable, MapReduce of Google platform, and it is developed & applied actively focused on Yahoo at the present. Hadoop is composed by three properties such as HDFS, HBase, MapReduce largely. As it can be seen in [Figure 4], HDFS is composed by one master node and several slave nodes, and the master node manages the file system namespace and it is composed by the single name node which controls the file access by the clariant. In each slave node, the file data is distributed by the block unit so it is stored in many data node. The data node processes the data input/output request from the client. In addition, the MapReduce is the programming model which is made for efficient support of the distributed parallel systems by Google. Through a combination of Map function and Reduce function, the data was made with both of {key, Value}, and it has the process to support the distributed parallel system operation [4][5][6].

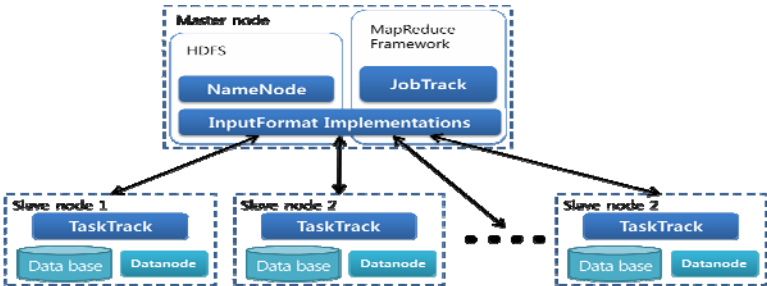


Fig. 4. HDFS structure

3 System Design

In this study, the collected URL LIST from the general distributed web crawler was stored in the Hadoop based URL collection system. The whole system structure is as following [Figure 5]. At this time, the distribution of the URL LIST is the biggest issue, the crawled URL data by each keyword was transported by multi-pressing. In addition, the collected URL from the specific seed URL is by designing into the slot structure.

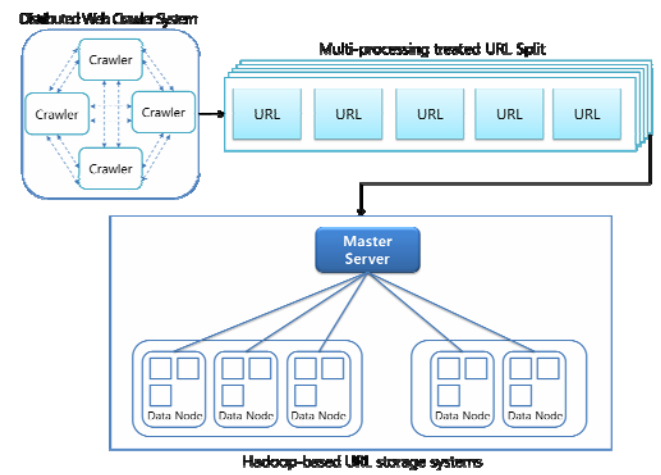


Fig. 5. The whole system structure

At this time, the web crawler sets the order of seed URL and transmits as string formation by converting into one data. The structure like this is as following [Figure 6]. The converted file is divided as [Figure 7] by the separator and it is converted in the order of designated seed URL. Thus, the collected URL is stored in the specified slot location and if the URL list is filled in the slot structure, then it is sent.

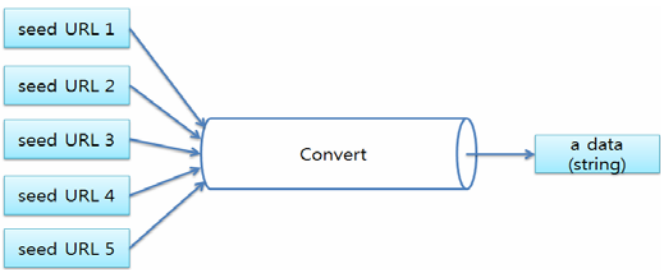


Fig. 6. Conversion of data type

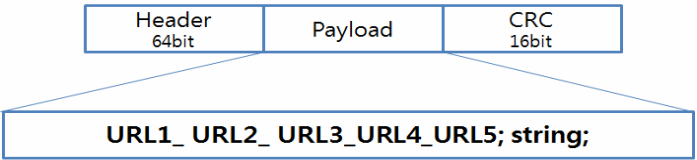


Fig. 7. Payload formations by using separator

In the transported Hadoop based distributed file system, the collected URL list in the seed URL is stored in order. It is stored by distributing in each server and the reproduction for backup in the distributed file system.

4 System Implementation

For this system experiments and implements, the socket programming for URL split is implemented by using C # and five programs are implemented at the same time so the multi-thread environment was constructed. The URL List data base for storing URL and system was composed by the split data as Hadoop. For this, the Hadoop based URL storage system as Intel Xeon Nehalem CPU was implemented in the server of RAM 4 GB, and the URL split program was implemented in Quad core, RAM 4GB desktop. The following [Figure8] shows that the virtual URL value was split and it was transported to the Hadoop based URL storage system like [Figure 9].

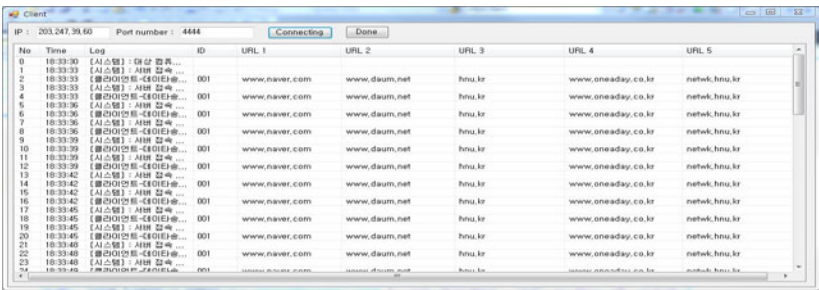


Fig. 8. Screen of transferring URL split collected by web crawler

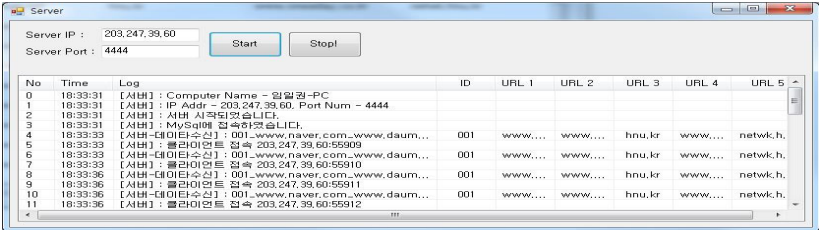


Fig. 9. The screen of sending URL split collected by web crawler

The following [Figure10] shows the implemented Hadoop based URL storage system. It shows one master and three slaves are composed by the virtualization.

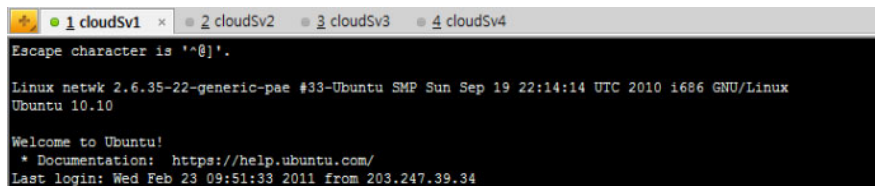


Fig. 10. The screen composed by one master and three slaves through Hadoop virtualization

The serial process of storing split URL data in the Hadoop based URL storage system is as follows and it is shown in the [Figure 11].

1. The request of file path for creation of file: file path:/foo/bar, number of replicas:3
2. Creation of file path information in memory: Creating Lock for preventing creation of other client
3. Returns the host information after selecting data node for storing file data (Slave1, Slave2, Slave3)
4. Transfer the file data & data node list: transferring the fsimage file to the Name Node
5. Storing local
 - 5.1 Storing the first replicas
 - 5.2 Storing the second replicas
 - 5.3 Storing local Complete (close())
6. Recording the memory contents in the edit file(name space registered): Merger the edits and fsimage merger after periodical download

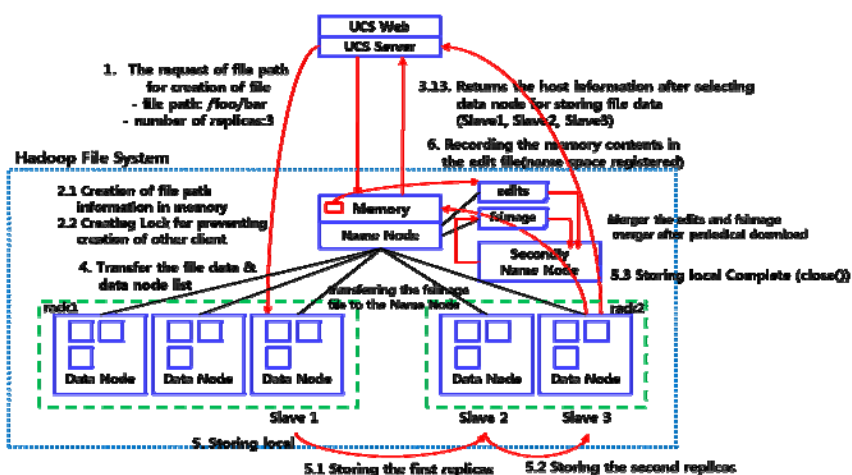
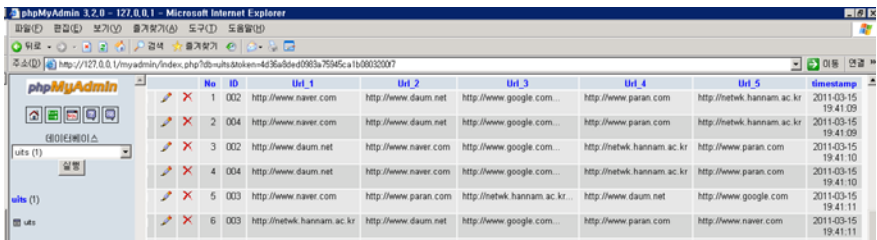


Fig. 11. Implemented Hadoop work process

The management of URL LIST by the Hadoop distributed system makes the rapid analysis and search of URL List for analysis data possible.



No	ID	url_1	url_2	url_3	url_4	url_5	timestamp
1	002	http://www.naver.com	http://www.daum.net	http://www.google.com...	http://www.paran.com	http://netwk.hannam.ac.kr	2011-03-15 19:41:09
2	004	http://www.naver.com	http://www.daum.net	http://www.google.com...	http://www.paran.com	http://netwk.hannam.ac.kr	2011-03-15 19:41:09
3	002	http://www.daum.net	http://www.naver.com	http://www.google.com...	http://netwk.hannam.ac.kr	http://www.paran.com	2011-03-15 19:41:10
4	004	http://www.daum.net	http://www.naver.com	http://www.google.com...	http://netwk.hannam.ac.kr	http://www.paran.com	2011-03-15 19:41:10
5	003	http://www.naver.com	http://netwk.hannam.ac.kr...	http://www.google.com...	http://www.daum.net	http://www.google.com	2011-03-15 19:41:11
6	003	http://netwk.hannam.ac.kr	http://www.daum.net	http://www.google.com...	http://www.paran.com	http://www.naver.com	2011-03-15 19:41:11

Fig. 12. URL LIST screen collected by split from the Multi- process environment

Table 1. Comparison analysis of single process and multi process

D v		
m	0	
m p c		
m a	00	00

5 Conclusion and Evaluation

This system sends five client programs to five slots at the same time for composing as the multi-processor method for transferring crawled data. At this time, the effectiveness was compared through the transfer experiment with single process program with 10 slots for one minute. The following [Figure 12] is the URL LIST screen collected by split from the multi-processor environment.

The implementation by the multi-processor like following [Table 1] can transfer more URL lists, and higher efficiency can be obtained by transferring meets the keyword by each process.

In addition, the existing parallel system may bring the duplicate collection results between each crawler because of each independent crawler to collect a random URL, so it is very difficult to distinguish. However, this system sets the keyword by the crawler process to minimize the redundancy of URL through filtering.

This study attached importance to the web-based technology in addition to the development of cloud technology, so the web crawling based transfer split technology was implemented. In this study, the crawling was not conducted from internet environment and the test was performed by the virtual value, so the study through the implementation of web crawling environment and several devices and multi-processor will be required in the future.

References

1. Lee, J.-S., Kim, Y.-W., Lee, P.-W.: Design and Implementation of Distributed Web Crawler Using Globus Environment. In: The Korean Institute of Information Scientists and Engineers Spring Conference, vol. 31(1) (April 2004)
2. Kang, M.-S., Choi, Y.-S.: Design Hadoop Based P2P Distributed Web Crawler. In: Korean Society for Internet Information 2010 Conference (June 2010)
3. Kang, M.-S., Choi, Y.-S.: P2P Distributed Web Crawling Architecture. In: Korean Society for Internet Information Fall Conference, vol. 6(2) (November 2005)
4. Cho, S.-H., Lee, S.-H., Kim, Y.-W.: Development of Large-Scale Spam Mail Filtering System based on Hadoop Framework. In: Korean Society for Internet Information Fall Conference, vol. 11(2) (October 2010)
5. Hadoop web page, <http://hadoop.apache.org/>
6. White, T.: Hadoop the Definitive Guide. Hanbit Media, Inc. (May 2010)
7. Cloudera web page, <http://www.cloudera.com/>
8. Kim, H.-J., Jo, J.-H., An, S.-H., Kim, B.-J.: Implementation of cloud computing technology. Acorn Inc. (December 2010)
9. Shin, E.-J., Kim, Y.-R., Heo, J.-S., Whang, K.-Y.: Implementation of a Parallel Web Crawler for the Odysseus Large-Scale Search Engine. Journal of The Korean Institute of Information Scientists and Engineers: the Actual of Computing 14(6) (August 2008)
10. Yoo, D.-H., Chung, S.-H., Kim, T.-H.: Applying TIPC Protocol for Increasing Network Performance in Hadoop-based Distributed Computing Environment. Journal of The Korean Institute of Information Scientists and Engineers: Systems and Theory 36(5) (October 2009)
11. Kim, J.-H., Lee, L.-S., Ra, I.: Hadoop-based Redistributed Sample Sort Model for Cloud Computing. Journal of Korean Institute of Information Technology 8(6) (June 2010)
12. Kim, H.-W., Han, Y.S.: Web Crawler Design for the ARANES Search Engine. In: Korean Society for Internet Information Conference, vol. 2(1) (May 2001)
13. Hong, S.-J., Park, Y.-B.: Design for WEB Crawler of Reverse RSS for a Large Quantity Contents Search. Journal of Korea Information and Communication Society 09-02 34(2) (February 2009)
14. Kim, H.-H., Kim, Y.-W., Lee, P.-W.: A Method of GridIR System Configuration over Distributed Experiment of Web Crawler. In: Korean Society for Internet Information Fall Conference, vol. 8(2) (November 2007)

A Study on Authentication System Using QR Code for Mobile Cloud Computing Environment

Dong-Sik Oh¹, Bong-Han Kim², and Jae-Kwang Lee^{1,*}

¹ Department of Computer Engineering, Hannam University,
Daejeon, Korea

² Department of Computer Information & Engineering, Cheongju University,
Cheongju, Korea

{dsoh, jklee}@netwk.hannam.ac.kr, bhkim@cju.ac.kr

Abstract. In cloud computing environment, users use authentication system for being provided the whole or the part of needed resources directly from servers, and for authentication, thousands of status bit are necessary. The problems is that thousands of status bit information should be searched under the central authority, and the searching process causes high network bandwidth, delay, and network congestion. This paper suggests authentication system appropriate for mobile cloud computing environment using QR code, a typical 2-dimensional code, and QR code is a technology suitable to display information needed for user authentication through its strengths such as high data integration, compressibility and error correction capability. Then, it suggests a technology to authorize users with totally new information vs information through new type of data set having 3 kinds of QR code by converting user information to different 3 versions of QR code and by storing each QR code in distributed server system of cloud computing environment in a circulation-loop way.

Keywords: Mobile Cloud Computing, QR code, Authentication.

1 Introduction

Recently, cloud computing, which provides IT resources as a part of service using Internet technology, is getting public attention. The reason why cloud computing is widely used is that it enables users to store programs or documents stored individually in large-scale computer to which can be accessed through Internet, and to perform necessary works by running necessary application such as web browser through various terminal units including PCs or mobile phones[1]. In cloud computing environment where users borrow and use IT resources as required, are supported with real-time scalability according to service load, and pay the expenses as used, most large-sized companies which have enough capital strength and high technical skills are constructing Private Cloud actively for the reason of security. However, small and medium-sized companies which lack investment power compared with large-sized companies tend to introduce Public Cloud with lower initial investment and operation

* Corresponding author.

costs compared with Private Cloud. Especially, it is expected that SMO (Smart Mobile Office) utilizing Public Cloud will be on the rise as killer service in order to enhance productivity and efficiency of small and medium-sized companies. SMO means environment where users can deal with businesses efficiently with no limitations of time and place by utilizing portable high-performance network devices such as smart phone and tablet PCs. Smart work is being suggested as a solution for problems across the whole industries such as low fertility, aging society, decline of labor productivity and environmental pollution, and SMO can be said to be the core of existing smart work [2].

Like this, though mobile cloud computing has an advantage of efficiency increase attained by sharing IT assets with other users and cost-reduction according to that, it can also generate some problems in data integrity and server authentication by sharing data, that is, assets, with other users [3]. In cloud computing environment, servers utilize authentication system of cloud service in order to provide the whole or the part of needed resources directly to users, and in order for the server to authorize numerous clients in cloud computing environment, thousands of status bits are necessary. Here, status bit means information of status change of links (interface) among paths from the server to same clients. Such status bits should be searched under central authority by a main verifier, and the searching process causes high network bandwidth, delay and network congestion [4]. Then, this paper suggests an authentication system introducing QR code as a plan for solving authentication and security problems which are on the rise as serious problems in general cloud computing environment, especially mobile cloud computing environment.

Thanks to QR code's fast recognition and data processing, easy method for use and wide utilization, services applying QR code continue to be spread widely. But, while currently-used QR code is being used intensively in enhancing service accessibility, this paper would propose a new method about mobile cloud authentication by suggesting a plan for utilizing QR code as a kind of authentication certificate. Technology relevant to QR code and mobile cloud authentication will be examined in the 2nd chapter, concept of cloud computing suggested in this paper and design of authentication system using QR code in the 3rd chapter, its implementation and evaluation in the 4th chapter, and conclusion for the last.

2 Relevant Studies

2.1 QR Code

QR code is 2-dimensional code developed in order to overcome the limit of storing and expressing information of bar code, 1-dimensional code, and it was invented by Denso Wave, Japan. They announced that they would exercise their patent right for common use and distributed it in 1994. QR is an abbreviation of "Quick Response", QR code is a type of matrix, enables fast decoding, and uses a method of mass storage, much record, high density, and Reed-Solomon error correction (8 bits/codewords) [5]. So, with its advantage which enables fast recognition and data processing in mobile environment, relevant services applying QR code are being vitalized currently.

As for 2-dimensional codes, there are many kinds of code introduced including QR code as shown in [table 1].

Table 1. Typical two-dimensional code

	QR Code	PDF417	DataMatrix	MaxiCode
				
Company	Denso Wave	Symbol	CI Matrix	UPS
(Country)	(Jap)	(USA)	(USA)	(USA)
Method	Matrix	Scotkudo bar code	Matrix	Matrix
Number	7,089	2,710	3,116	138
Muneral	4,296	1,850	2,355	93
Binary number	2,953	1,018	1,556	-
Chinese character	1,817	554	778	-
Major function	Mass storage Space-saving Fast data-processing	Mass storage	Space-saving	Fast data- processing
Major usage	All fields	OA	FA	Distribution
Standardization	AIM International JIS ISO	AIM International ISO	AIM International ISO	AIM Internation ISO

Among above codes, QR code has 6 distinguished merits from other codes including that it is opened [6].

2.2 Mobile Cloud Authentication Technology

1) Way for issuing authentication certificate using application [Figure 1]



Fig. 1. Authentication certificate mounted in mobile cloud

Users access to services in mobile cloud environment, and when accessing, they use SSL (Secure Sockets Layer) and OTP (One-Time Password) with ID/PW for security in network communication. After access to the service, they are issued authentication certificate from a Certificate Authority through an application installed in mobile cloud environment. When completing all process, they access to the service by using ID/PW, SSL, and OTP for using the service, and are provided with financial services by selecting installed authentication certificate and by inputting password [7].

The problems of this system is that it requires for users to input and transfer ID/PW, OTP, and authentication password in all process such as issuing initial authentication certificate, service use, and reissuance and renewal of the certificate. That is, though means for safe authentication is already applied multiply, it is difficult to implement environment for dealing with service promptly due to increase of network traffic and processing load, it is necessary to install an application preferentially, and advance arrangements are also needed for using OTP.

3 Design of QR Code Authentication System in Mobile Cloud Computing Environment

QR code authentication system under the concept of cloud computing can be divided into 3 stages.

3.1 Creation of QR Code

As shown in [Figure 2], user ID and password, and user's image are converted into QR code. QR code is converted to total 3 kinds of version, and even if it has the same input data, total different QR code is created according to set version, as shown in [Figure 3].

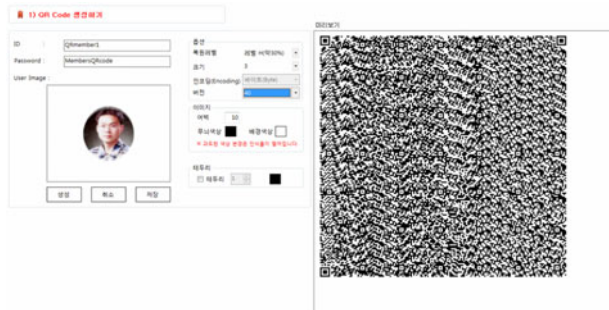


Fig. 2. QR code creation page

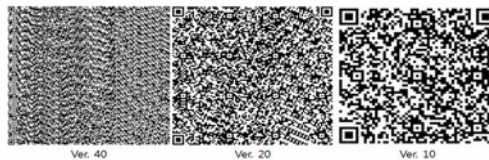


Fig. 3. Creation type of 3 different versions of QR code

3.2 Distributed Storage of QR Code Data Cell

As shown in [Figure 3], 3 kinds of QR code having different shape each other change the structure according to algorithm of each version. That is, 3 kinds of QR code are

created with different structure and shape and with the same data meaning, and are stored in distributed cloud computing servers by 1 divided cell in grid shape, as shown in [Figure 4], playing a role of authentication certificate for using services provided by many distributed servers.

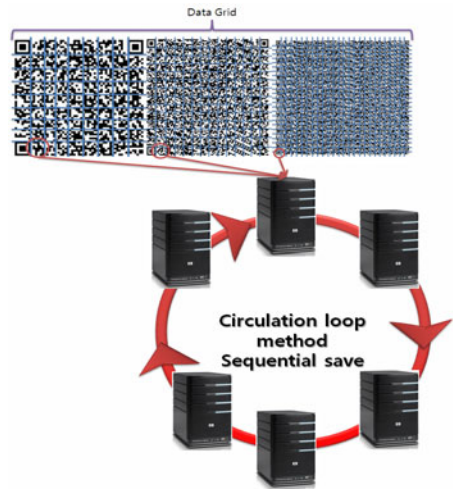


Fig. 4. Distributed server storage after making QR Code into grid shape

When they are stored sequentially with different QR structure and grid structure as shown in [Figure 4], each server is stored in the form shown in [Figure 5].

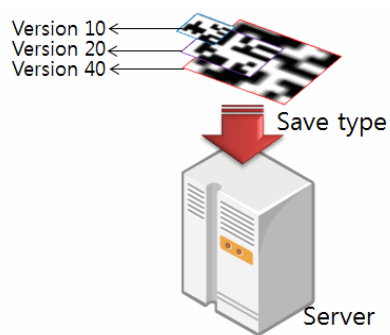


Fig. 5. Overlaid storage of QR Code Cell by each version

3.3 Environment for Mobile Cloud User Authentication

Security concern discussed in cloud computing environment is that if the initial authentication is passed, security of all other servers connected to each other through generalized SSO (Single Sign On) would be destroyed. Especially, in mobile

computing environment, it demands high security as well as user authentication by the least concise communication due to unique characteristics of mobile communication. Therefore, each cloud server has cell storage values generated by different structures after going through ‘1’ and ‘2’ designed above, and such values have totally different data from the initial QR code. For example, QR code having a totally different form with the original form as shown in [Figure 6] was generated, but totally different form of authentication system could be implemented as stored values according to each position corresponded to the initial QR code values generated by the user.

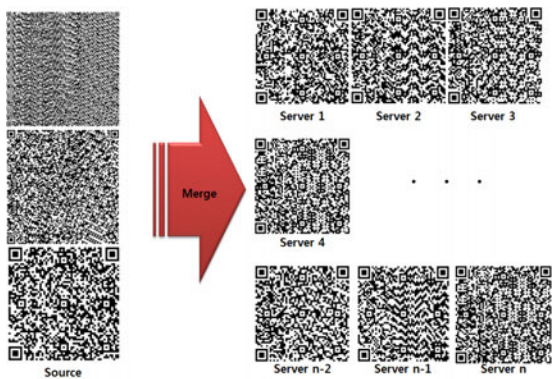


Fig. 6. Concept for QR Code authentication

4 Implementation and Evaluation

This paper prepared the following environment in [Table 3] for implementation and evaluation of the system suggested by this paper.

Table 2. System Environment

Item	Content
System	Server 1set PC 1set
OS	Windows Server 2003, Windows7
Server Spec	Intel Xeon Nehalem 5500 16GB RAM
PC Spec	Intel Quadcore 2.4GHz 4GB RAM
S/W Development Environment	Visual Studio 2008 .Net Framework 3.5
DataBase	SQLite

6 DB files were defined through SQLite in order to set 6 types of storage which plays a role as a virtual server within 1 server for implementation and evaluation of the system. The system designed in above 3rd chapter was implemented through C#, and the experiment was performed by defining PC as mobile for testing.

The experiment prepared its environment in order to generate total 20 times of client access and authentication for 1 minute by making user ID, password, authentication password and OTP value into packets except complexity and process of advance preparation pointed out in the 2nd chapter, and the result is shown in [Figure 7].

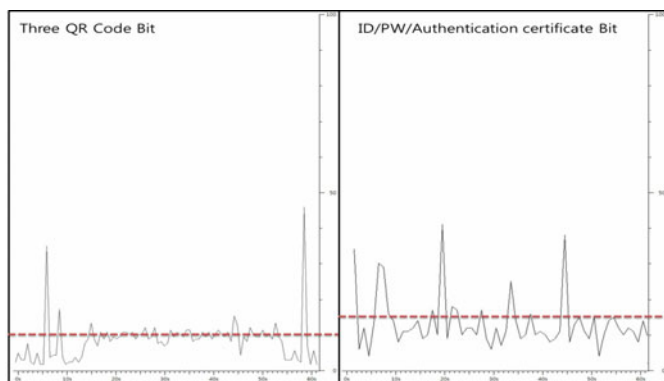


Fig. 7. Comparison with the existing systems

5 Conclusion

Recently, cloud computing, which provides IT resources as a service by using Internet technology, is getting public interest, and especially, mobile cloud computing utilizing mobile phones is receiving much attention. Therefore, while most large-sized companies having enough capital power and technical skills for application of could computing are constructing Private Cloud for the reason of security, small and medium-sized companies, which lack investment capability, tend to introduce Public Cloud requiring lower initial investment and operating expenses than that of Private Cloud. Especially, under current situation where smart mobile office utilizing Public Cloud for enhancing small and medium-sized companies' productivity and efficiency is getting attention as a killer service, the biggest problem when using Public Cloud is security and user authentication. Existing mobile cloud authentication utilized user ID and password, authentication certificate, authentication password, and OTP, and this generates much network traffic only by performing 1 time of user authentication process. Therefore, this paper suggested an authentication system generating great effect with the minimum usage of user's network traffic in mobile cloud environment by utilizing QR codes, and confirmed that the average amount of traffic was lower than that in existing systems from the experiment.

But, as this paper did not conduct analysis on weak points in security between existing technology and suggested technology, and evaluation about security itself such as time spent for interpretation when external attack occurred, it is difficult to guarantee that suggested system has higher effect on security. Therefore, future evaluation on security should be conducted.

References

1. CheolSoo, L.: Cloud computing security technology. *Journal of Information Security and Cryptology* 19(3), 14–17 (2009)
2. HyeonBong, G., HyeonDeok, L., JaeIl, L.: Policy for vitalizing Smart Mobile Office based on Cloud Computing, Korean Institute of Information Scientists and Engineers 2010 Collection of Conference Dissertations, vol. 37(2) (2010)
3. SeongGyeong, E.: Trends of cloud computing security technology. *Journal of the Korea Institute of Information Security and Cryptology* 20, 27–31 (2010)
4. YoonSoo, J., YongTae, K.: Flooding packet authentication based on double hash chain and mechanism for guaranteeing integrity. *Journal of the Korea Institute of Information Security and Cryptology* 9(1), 147–158 (2011)
5. YeongJae, M., DongOh, S., SeongHo, K., DaeHeon, Y.: Analysis on weak points in falsification in E-commerce and consideration of responding measures. *Journal of the Korea Institute of Information Security and Cryptology* 20(6), 17–27 (2010)
6. <http://www.scany.net/kr/generator/barcodeQrcode.php>
7. MoonYeong, H., Woong, G., DongByeom, L., Jin, G.: Plan for managing authentication certificate in smartphone banking using mobile cloud computing. In: Summer Academic Conference of Institute of Electronics Engineers of Korea, vol. 33(1), pp. 1873–1876 (2010)
8. Zhang, X., Schiffman, J., Gibbs, S., Kunjithapatham, A., Jeong, S.: Securing elastic applications on mobile devices for cloud computing. In: ACM workshop on Cloud computing security, CCSW 2009 (2009)
9. Sun, A., Sun, Y., Liu, C.: The QR-code reorganization in illegible snapshots taken by mobile phones. In: The 2007 International Conference Computational Science and its Applications, pp. 532–538 (2007)
10. Chen, W.-Y., Wang, J.-W.: Nested image steganography scheme using QR-barcode technique. *Society of Photo-Optical Instrumentation Engineers* 2009 (2009)

Research on the I/O Performance Advancement of a Low Speed HDD Using DDR-SSD

Seung-Kook Cheong¹, Jae-Jin Jeong², Yong-Wan Jeong³,
Dae-Sik Ko³, and Young-Hoon Lee²

¹ Electronics and Telecommunications Research Institute, Daejeon, Korea

² Circuit & System Research Lab. HANNAM University, Daejeon, Korea

³ Department of IT Engineering Lab. MOKWON University, Daejeon, Korea
skjeong@etri.re.kr, slayer86@naver.com, jyywing@naver.com,
kds@mokwon.ac.kr, yhlee@hnu.kr

Abstract. In this paper, we proposed Hot/Cold File Management Algorithm to maximize performance to environment combining a low speed I/O property HDD with a high speed I/O SSD. As a result to test by applying that algorithm, we confirm that performance get better multiples of 11.36(in case of block size=4k), multiples of 12.79(in case of block size=8k) and that performance is unrelated to a type of second disk (HDD, Flash-SSD, DRAM-SSD).

Keywords: DDR-SSD, HDD, I/O Performance.

1 Introduction

Hot/Cold File Management Algorithm is proposed to overcome memory quality of relatively low-capacity despite overwhelmed performance and to improve write method flaw of Flash-SSD.

Commonly, user (application) data is frequently used data or rarely used data. In case of data capability polarization, per cell write cycle is limited, flash memory having memory capacity limit affects the deadly problem.

Accordingly, by managing data as a way of classifying frequently used data into a hot data group, rarely used data into a cold data group, with problem that overwriting is not operated in case of changing flash memory data, in progressing write after elimination that do canceled arithmetic operations through garbage collection after nullifying a formerly existed data after writing relevant data in new area. Study on the Hot/Cold data management method, reducing abrasion of particular cell and helping efficient memory management by moving cold data to backup disk in backup disk install time to back up low-capacity flash memory has progressed.

As a relevant study, there is many papers “An Efficient Adaptive Flash Translation Layer using Hot Data Identifier for NAND Flash Memory”, “A wear-leveling improving method by periodic exchanging of cold block areas and hot block areas”, “New Flash Memory Management Method for Reliable Flash Storage Systems”, “A Prediction Scheme between Cold Data and Hot Data on Flash Memory Through Input

data Size” and “Dynamic block-allocation Flash Translation Layer according to the hot and cold data pattern”[1-5].

Also, there is many patents “Hot Data Management Based on Hit Counter from Data Servers in Parallelism”, “Method For Data Processing And Asymmetric Clustered Distributed File System Using The Same”, “Hybrid density Memory storage device”, “DEVICE DRIVER AND METHOD FOR EFFECTIVELY MANAGEMENT A FLASH MEMORY FILE SYSTEM”, “Apparatus and method for storing data in nonvolatile cache memory considering update ratio and Clustering device for flash memory and method thereof”[6].

However, above technologies and patents is hot/cold management technology of block data and are not relevant to hot/cold management technology of file unit. Therefore, in this paper, we proposed a new hot/cold file management technology using two different disks (1st disk is memory DRAM-SSD, 2nd disk is HDD or Flash-SSD).

2 Hot/Cold File Management Algorithm

2.1 The Concept of Algorithm

A fundamental conception of the algorithm is to operate as a way of that a high speed I/O property SSD saves frequently used file and a low speed I/O property HDD saves rarely used file using two different disks.

Read/Write of all files is to operate as a way that saves files to 1st disk (memory DRAM-SSD) commonly, in case of lack of capacity, files that access is oldest moves out into 2nd disk (HDD).

The algorithm concept equals Fig 1. If file creation or modification(read/write) event is occurred, file system using I-node information of relevant file forms Hash Function and decides position of hot/cold management file(I-node information management File : ii_file.ord) and modifies ii_file.ord management data and processes a ordered file service. If 1st disk capacity reaches 80 %(marginal value), files that access is oldest among the recorded file moves out into 2nd disk. At this point, 1st disk has I-node information, file is deleted.

If this type of file(move-out file) is required, required file move-in from 2nd disk to 1st disk again after renews I-node information, since then, the disk proceeds with file read service, in that case, 2nd disk has not relevant file.

2.2 I-node Management

Hot/Cold File Management uses I-node. Hash Function uses to find the File Position using I-node information of file. Algorithm finds a particular integer through Hash Function and sets a File Position using that. File Position plays a roll which indicates that relevant file is located, and Hash Table is composed with I-node Number(8Bytes), Start Point(4Bytes) and End Point(4Bytes).

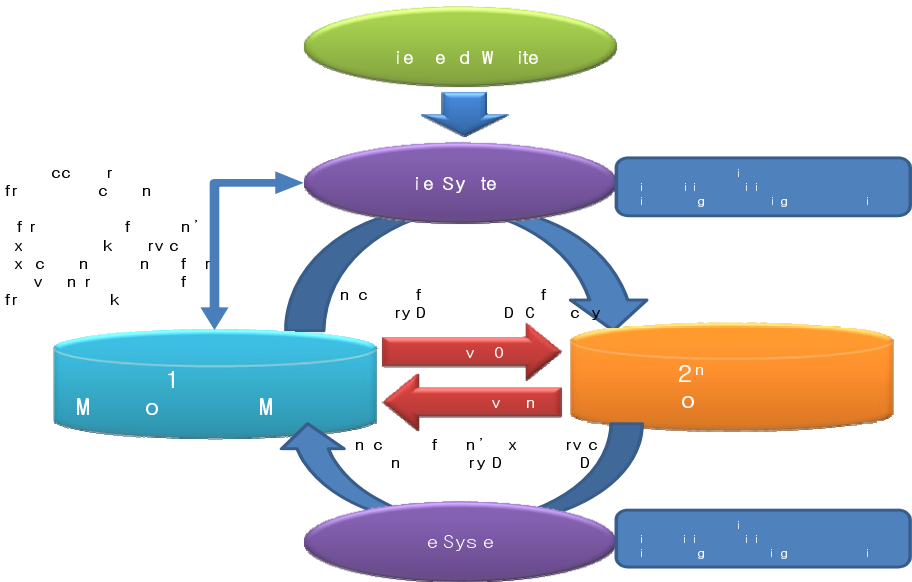


Fig. 1. The concept of algorithm

Table 1. I-Node Management Table

File Position	I-node Number(8Bytes)	Start Point(4Bytes)	End Point(4Bytes)
1	vvvvvvvv	First : 0	First : 1
2	yyyyyyyy	H(1)	H(1)
3	xxxxxxxx	H(1)	H(2)
...	...	...	...
n	zzzzzzzz	H(n-1)	H(n)

3 Test Result Analysis

3.1 Test Environment and Conditions

Test environment equals Fig 2. We used to PostMark[7] which is macro benchmark for OLTP(small size random I/O) property test.

Test conditions set up FS(File Size)=327KB, BS(Block Size)=(4k and 8k), the number of file=20,000, the number of transaction=100,000, the number of sub directory = 100, the number of test = 30, r/w rate = 1:1.

Test condition as follows: (refer to Fig.2)

- Test Case A : only Performance measurement of memory type DRAM-SSD(MD-SSD)

- Test Case B : Performance measurement of 1st Disk(memory type DRAM-SSD) with 2nd Disk(PCI-e type DRAM-SSD)
- Test Case C : Performance measurement of 1st Disk(memory type DRAM-SSD) with 2nd Disk(SCSI type SAS-HDD)
- Test Case D : only Performance measurement of PCI type DRAM-SSD
- Test Case E : only Performance measurement of PCI type Flash-SSD
- Test Case F : only Performance measurement of SCSI type SAS-HDD

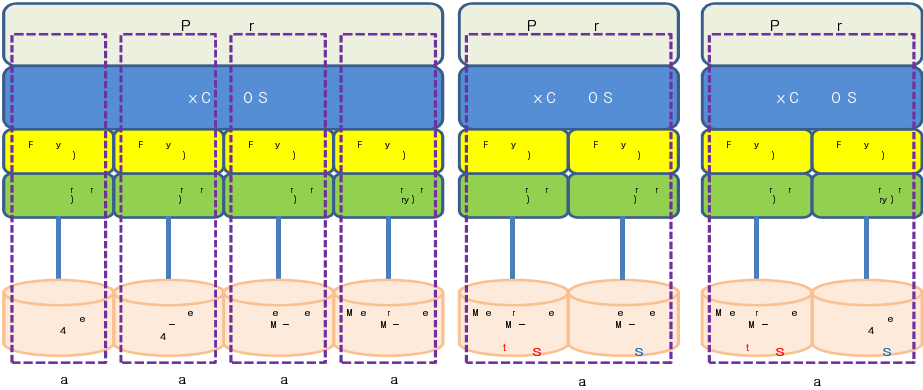


Fig. 2. Test Environment and Condition

3.2 Test Result Analysis

Table 2 is test results about case of BS(Block Size)=4K, and Table 3 is test results about case of BS(Block Size)=8K.

Having done a comparative analysis with the Table 2, MD-SSD quality is best and are faster multiples of 29 than HDD using the SAS bus, multiples of 4 than D-SSD using the PCI-e bus, multiples of 12 than Flash-SSD using PCI bus.

Also, through standard deviation (STDV) about test time, we confirm that SSD provides better service quality than HDD.

Through performance measure using proposed algorithm, we got the result that Test Case A(in case of setting up 1st disk for MD-SSD and 2nd disk for D-SSD) gets better quality by 43% than D-SSD unit quality based on PCI-e and gets better quality by 1,135% than HDD unit quality and that Test Case C(in case of setting up 1st disk for MD-SSD and 2nd disk for HDD) gets better quality by 50.5% than D-SSD unit quality based on PCI-e and gets better quality by 1,196% than HDD unit quality.

Also, we got the result that 2nd disk is not affected according to I/O Speed of 1st disk. Accordingly, 2nd disk can raise the use efficiency by using cheap HDD.

From a result of Table 2, we confirmed that (BS=8K) get better quality than (BS=4K). The ratio of BS=4K and BS=8K is as follows:

Table 2. Test Results (Block Size=4K)

Unit: Time(second)						
Number of Trial	Test Case A	Test Case B	Test Case C	Test Case TD	Test Case E	Test Case F
1	40	90	86	141	457	1,148
2	40	97	86	142	466	1,130
3	39	109	108	142	461	1,106
4	39	88	93	143	462	1,142
5	40	95	102	145	464	1,123
6	40	108	100	144	461	1,120
7	39	97	96	142	462	1,116
8	39	94	101	145	462	1,134
9	39	106	101	143	459	1,149
10	38	96	96	141	463	1,128
11	39	99	102	142	463	1,122
12	39	101	100	142	461	1,161
13	39	103	101	142	465	1,137
14	39	97	100	141	465	1,109
15	39	102	99	143	463	1,161
16	38	104	92	139	464	1,125
17	39	99	105	140	462	1,138
18	39	89	101	141	463	1,143
19	39	101	88	142	465	1,154
20	40	103	102	144	460	1,156
21	39	104	95	145	463	1,136
22	39	98	101	144	464	1,134
23	39	97	91	144	465	1,125
24	39	104	104	142	466	1,129
25	39	114	109	145	463	1,161
26	39	95	98	142	465	1,111
27	39	103	73	143	464	1,148
28	40	103	75	141	465	1,152
29	39	108	74	140	470	1,141
30	39	98	73	145	467	1,126
AVER	39.13	100.07	95.07	142.50	463.33	1,135.50
STDV	0.51	6.00	10.21	1.66	2.52	15.75

- Test Case A Ratio(4K:8K)=5.6%
- Test Case B Ratio(4K:8K)=16.1%
- Test Case C Ratio(4K:8K)=1.1%
- Test Case C Ratio(4K:8K)=2.3%
- Test Case E Ratio(4K:8K)=0.6%
- Test Case F Ratio(4K:8K)=0.4%

As a result, we got the result that BS(Block Size) does not affect proposed Hot/Cold File Management Algorithm.

4 Conclusion

In this paper, we proposed Hot/Cold File Management Algorithm to maximize performance to environment combining a low speed I/O property HDD with a high speed I/O SSD. As a result to test by applying that algorithm in combining HDD with the MD-SSD, we confirm that performance is unrelated to a type of 2nd disk and that existing HDD I/O performance sharply is improved.

In conclusion, we assure that a hybrid disk or storage combining a low speed I/O property HDD with a high speed I/O quality SSD can be formed and used. Therefore, SSD which has a high speed I/O characteristic surely needs to improve the normal massive HDD storage performance.

References

1. Yun, H.-S.: An Efficient Adaptive Flash Translation Layer using Hot Data Identifier for NAND Flash Memory. *Journal of KIISE: Information Networking* 35(1), 18–29 (2008)
2. Jang, S.-w.: A wear-leveling improving method by periodic exchanging of cold block areas and hot block areas. In: *KIMICS 2008*, pp. 175–178 (2008)
3. Kim, H.-j.: New Flash Memory Management Method for Reliable Flash Storage Systems. *Journal of KIISE: Information Networking* 27(6), 567–582 (2000)
4. Shin, H.: A Prediction Scheme between Cold Data and Hot Data on Flash Memory Through Input data Size. In: *Proceedings of KIIS Spring Conference*, vol. 20(1), pp. 38–39 (2010)
5. Cho, E.: Dynamic block-allocation Flash Translation Layer according to the hot and cold data pattern. *Journal of KIISE autumn Conference* 34(2), 245–246 (2007)
6. <http://www.kipo.go.kr/>
7. Katcher, J.: PostMark: A New File System Benchmark
8. Jeong, S.-K., Ko, D.: Technical trends of next generation storage system SSD. *Weekly journal* 1369 (2008.10.22)
9. Seung-kook, C.: Web Performance Enhancement of E-business System using the SSD. *Ussnet 2008*, Hainan China, (2008)
10. Cheong, S.-K., Ko, D.-s.: Data I/O Performance Evaluation of the SSD and HDD in Database. In: *ITCS 2010*, Cebu Philippine, (2010.8.11)

Neighborhood Evolution in MANET Experiments

Julien Franzolini^{1,2}, Frédéric Guinand¹, and Damien Olivier^{1,*}

¹ LITIS Laboratory, Rue Philip Lebon University of Le Havre, France

² Direction de l'Information sur les Risques Majeurs, DIRM-CODAH, Le Havre City
{julien.franzolini, frederic.guinand, damien.olivier}@univ-lehavre.fr
<http://www.litislab.eu>

Abstract. Many algorithms and protocols in Ad-Hoc Networks are based on the use of neighborhood information. It is usually accepted that one-hop neighbors in static networks are stable. However, based on intensive sets of experiments, it appears that wireless links both in indoors or outdoors contexts are very sensitive to conditions. During two years we have recorded sets of measures and draw today some conclusions about the stability of links between wireless-enable small communicating objects like smartphones and PDAs.

1 Introduction

An ad hoc network may be defined as an infrastructureless network of wireless-enable communicating devices. We choose the term *station* for naming such devices. Our attention is focused on the neighborhood of stations. We study the sensitivity of this neighborhood of stations for various endogenous as well as exogenous conditions. Conditions of use of the network, rates of transmission, antenna power, algorithms and protocols are called endogenous conditions while environmental conditions (position of the stations, weather, etc.) are called exogenous conditions. Depending on the ability of the stations to move, the position of the station may be declared endogenous or exogenous. While the testbed as well as the different system layers were designed and implemented for mobile ad hoc networks, the results presented in this paper only concern static topology MANET since the experiments with mobility have not been completed yet.

This work has been performed in the context of a collaboration between our laboratory and a local authority in charge of alerting populations in case of industrial disaster: the DIRM section of CODAH¹². The aim of this office is to develop and deploy all possible means for alerting populations. In addition to radio, TV and alert sirens, they are looking for new solutions for reducing the delay between the time at which the alert is launched and the time at which the information reaches most of the population. We proposed to investigate the

* PROTEC project - GRR SER (MRT), granted by Region Haute-Normandie, France.

¹ DIRM: Major Risk Information Department, CODAH: Le Havre City Community.

² PROTEC project partially funded by Haute-Normandie Region.

possibility of reducing the distance between people and the vector of information, and it appears that nowadays mobile phones are among the best candidates for that purpose. One pitfall to a SMS-based broadcast is that the probability of a network congestion during a disaster is very high. Thus, we propose to look forward in the direction of MANET [AJF⁺08].

All presented results in that paper come from real-world experiments. Neither simulations nor emulation were used for this study. As underlined by Kiess and Mauve implementing real-world MANETs requires a big effort, but this was unavoidable since *“real-world experiments are the ultimate way to prove that an algorithm of protocol works as expected”* [KM07], and it was clear for us that the proposed solution has to be reliable or at least that authorities should know the percentage or the people that may be reached. Neighborhood stability plays a central role in the efficiency of the system. Thus, we finally developed a dedicated framework and testbed for measuring neighborhood in MANETs under various conditions.

It is worth noting that very few works were dedicated to the problem we address. Many of wireless networks testbeds were mostly designed for bandwidth and protocol efficiency measures. These experiments focus mainly on local results about the delay, packet loss [KS06] [MBJ00] or link quality [ABB⁺04], but only a few [DVL03] try to understand the evolution of network topology, even in the static case, evolution caused by the instability of the connection between two nodes. One of the closest studies is due to Anastasi et al. [ABC⁺05] who analyze the dependence of communication range on several parameters such as transmission data rate or environment humidity. With respect to this work, we focus on stations’ neighborhood evolution according to changes in usage as well as environmental conditions, rather than on communication range between stations. Usually, strategies designed for neighborhood discovery mechanism (NDM) try to minimize bandwidth consumption to prevent data transfer perturbations. But in our case, we decided to analyze the impact of network congestion on NDM [KS06].

For that purpose, we have developed an experimental plan for testing the reliability of mobile ad hoc networks in real environments. In addition to the experimental protocol, we have developed a software layer, dedicated to mobile devices, for managing the configuration, the connection and the communication. The protocol and the software layer are presented in Section 2. Results obtained from real experiments are exposed in Section 3 and based on the results, some comments conclude this work.

2 Methodology

2.1 The MANET Framework

The framework consists mainly in a dynamic library, functional today on Windows Mobile from Win 2003 to 6.5 and Android platform. The objective of this layer is twofold. On the one hand it allows the station to automatically set up sets of parameters for entering an existing MANET and, on the other hand, it enables a careful observation and logging of the connections and communications of this station.

This layer interacts at different levels of the mobile device architecture:

- management of system;
- management of power of the IEEE 802.11 interface;
- settings of the wireless interface (channel, bssid, rate, mode...);
- network connection and disconnection;
- communication management (receive,send message).

An additional piece of software manages operations of message emission and reception. To identify uniquely each node, it uses its IMEI³ number. Communications are based on the UDP protocol. This choice was motivated by the fact that UDP supports broadcasting. Finally, as exchanges are based on information broadcasting, IP addresses are useless.

Two types of message are considered: HELLO messages are intended to help stations to build and maintain neighborhood information and “classical” messages that contain application-specific data (see Figure 1). Due to MTU (default Maximum Trame Unit) restriction, single UDP packet is limited to 1472 bytes. Sending larger messages is still possible thanks to a special dedicated mechanism not used in the current set of experiments.

In the framework, three thread operate simultaneously. One thread is associated to the emission of HELLO messages and to the management of the list of neighbors (let us call this thread the neighbor-thread. Another thread (emission-thread) is dedicated to the emission of application messages and a last one, the reception-thread, to the reception of all messages (HELLO and application).

Each time reception-thread receives a HELLO message, it timestamps the message and transmits it to the neighbor-thread. This one periodically updates the list of neighbors by removing neighbors for which HELLO messages have not been received for a given time, the discovery time (2 seconds for our tests), and by updating the last-reception date of neighbors for which an HELLO message has been received. Neighbor-thread periodically sends HELLO messages 1 containing its network identifier, its number of neighbors and the list of identifiers of its neighbors.

The reception-thread discriminates HELLO messages and application messages. The former are transmitted to the neighbor-thread while the later are processed. Short messages are made of only one packet while long messages are composed of several packets.

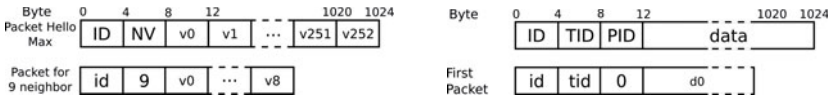


Fig. 1. Hello packet (left) Unique enclosed message (right)

³ International Mobile Equipment Identity

A packet is composed of two parts: the header and the data. The header is made of the identifier of the source (*ID*), the number of the message (*TID*) and the number of packets (*PID*) if the packet belongs to a long message.

When the size of the data to be transmitted is lower than or equal to 1012 bytes, only one packet is required for sending the message, and in the header the field *PID* is set to 0 (see figure 1).

To have ad-hoc behavior feed back, this layer can log different information:

- network signal strength (also the fact if it's connected);
- battery power;
- 1-Hop neighborhood;
- neighbor event;
- data message event (size, number and type of each message transmitted, received and send back);
- data, itself;
- GPS event.

This layer is designed to be extended to many routing protocols and offers necessary resources to implement various strategies of routing on smartphones. In this article we only use simple send mechanism with no information routing to have a referential of the behavior of real mobile devices.

2.2 Experiments Testbed and Plan

In this experimentation, we focus on static topologies in order to prevent modifications on the neighborhood detection mechanism by mobility. To test the sensitivity of neighborhood detection versus the network load, we fix all parameters which can directly have an influence on the neighborhood (position, mobility and transmission rate [ABC⁺05]). Devices used are HTC Mega (T3333) on Windows Mobile 6.5 with broadcom 802.11 DHD Network Adapter. This adapter is tuned by the application in mode ad-hoc on channel 1 at 54MBit/s rate. The network was tested in open space, 500 meters away from the first houses, so the newtork is not pertubated by other Wireless station.

For the experiments, some parameters were fixed:

- **Spatial Coordinates:** the coordinates (x,y,z) in meters of each device are stored in a MANET position file (.mpf) (it can be a dynamic file generated by GPS). The stations form a square, and the maximum distance between two stations is less than 5 meters. In that configuration, when the network is switched on, it is full-connected (see Figure 2).
- **Discovery Time:** the refresh delay of neighborhood; Here, 2000ms;
- **Send Time:** the time between two emissions of message; Variation range: 300 ms, 150 ms, 100 ms, 50 ms, 25 ms;
- **Size of Data:** the size can be set (by default message which contains the neighborhood); Variation range: 1024 bytes, 512, 256, 128, 80 bytes;
- **Type of Devices and number:** indicates the network heterogeneity; 8 HTC Mega (T3333);

- **Environment:** urban or obstacle-free place as an indication of radio propagation perturbations. Here, obstacle-free place;
- **Wireless interface setting:** includes interface type, rate, mode and channel. (802.11g, 54 MBit/s rate, ad-hoc mode, channel 1);

Before beginning the experimentation, all devices must be synchronized on the same clock, to have a usable timestamp for the log. For 1 hour of experimentation, 8 mobile devices produce about 800Mb of log distributed on each smartphone.

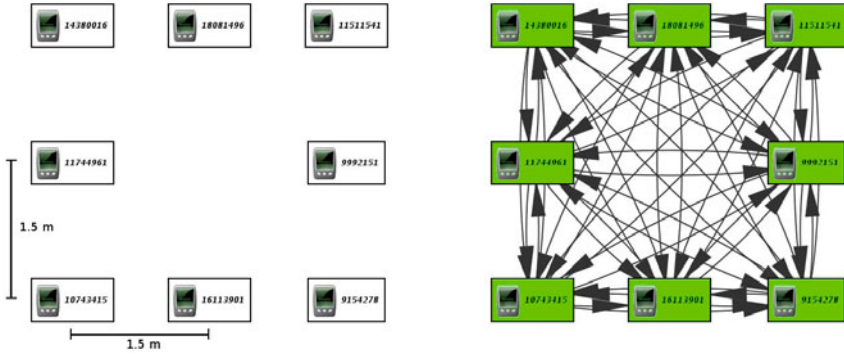


Fig. 2. Spatial Coordinates when the network is down (left side) and MANET layer started with neighborhood discovery (right side). Each edge represents an unidirectional neighborhood link.

To analyze logs, the application first sorts events according to timestamps then replay the execution step-by-step and event-by-event. The application displays topology in real-time on a dynamic graph (see Figure 4), and various information about the network (number of messages, bandwidth, message route, signal strength evolution, battery consumption ...).

3 Neighborhood Stability Results

3.1 With Fixed Data Size (80 Bytes)

Within this set of experiments we measure the sum of neighbors (SoN). This value corresponds to the sum of neighbors of each station. This corresponds to the number of arcs in the communication graph. For a full-connected network with eight stations, this value is equal to 56.

During the first test a message is sent every Δ ms. We make Δ varying from 300 to 25. From 300ms to 100ms no noticeable change occurs for the value of SoN. The value is close to 56 (left side of Figure 3). The topology of the network is close to be full-connected as illustrated on Figure 4. When Δ is equal to 50ms, the neighborhoods are less stable and the average value of SoN is close

to 54. Finally, when we increase the frequency of the emissions, for a value of $\Delta = 25\text{ms}$, the neighborhoods are no longer stable and the mean value of SoN decreases down to 26.8, which means that more than 50% of the neighbors are lost. This clearly appears on the right side of Figure 4. Moreover, after 30s of test, that can be considered as a transitional period, the value of SoN varies between 11 and 38 as illustrated on Figure 3.

Remark about the transitional period: during the first 5 seconds after switching on the network, the only exchanged messages concern neighbor detection. This explains why in the first few seconds the value of SoN tends to its maximum as the network is not congested (the situation corresponds to the right side of Figure 2). Then, the value of SoN decreases according to the neighborhood refresh mechanism.

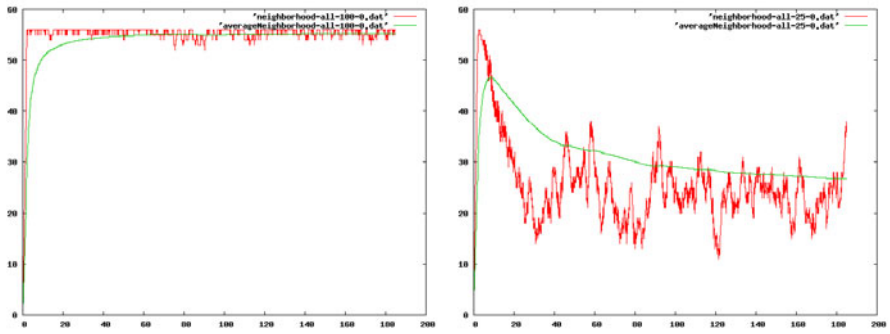


Fig. 3. Global neighborhood evolution on left for 100ms and on right 25ms. Time corresponds to x axis and global neighborhood to the y axis.

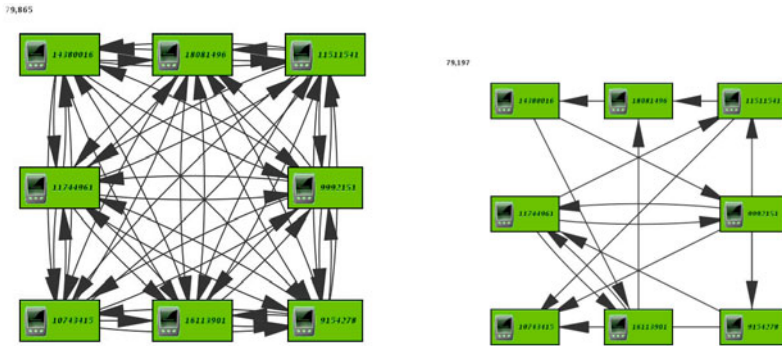


Fig. 4. State of the network for different values of Δ . When $\Delta = 100\text{ms}$, the network is not congested and is full-connected most of the time (left side). When $\Delta = 25\text{ms}$, the neighbor detection mechanism fails because of the congestion.

3.2 With Fixed Send Time (150ms)

In this second set of experiments, we fixed the period of emissions to 150ms. We focus on the same measure: the sum of neighbors (SoN). During this test, a message is sent by each station every 150ms but the size of the packet varies from 80 bytes to 1024 bytes. From 80 bytes to 256 bytes no noticeable change occur for the value of SoN. This value is close to 56 (left side of Figure 5). When the size is equal to 512 bytes, the neighborhoods are less stable and the average value of SoN is close to 50. Finally, when we increase the size of the packet up to 1024 bytes, the neighborhoods are no longer stable and the mean value of SoN decreases down to a value less than 50. This appears on the right side of Figure 5. But what is more important is that, after 30s of test, that can be considered, as previously, as a transitional period, the value of SoN varies between 56 (full-connected network) and 40.

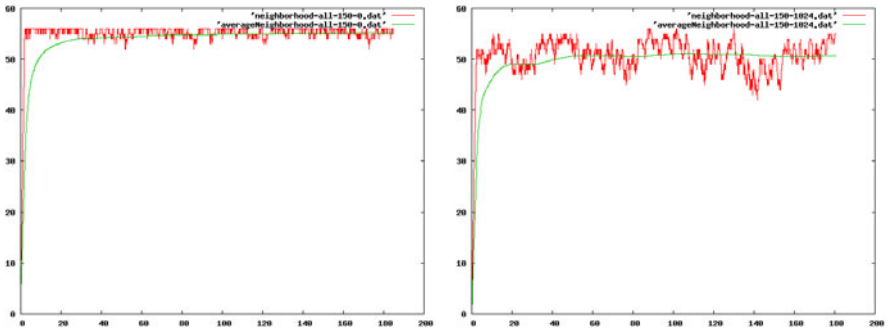


Fig. 5. Stability and variation of the neighborhoods when the size of the packets vary from 80 bytes (left side) to 1024 bytes (right side), for a fixed period of emission equal to 150ms

4 Conclusion and Outlook

The goal of our experiments is to understand what are the crucial parameters and phenomenon that allow stable neighborhoods. The analyses of the results stemmed from our experimental framework show that neighborhood may be unstable in spite of use in normal conditions. The neighborhood management is a key for most routing protocol in MANET.

To the light of our experiments, it is worth noting that changing the emission frequency seems to have an impact more important than the size of packets, and if it is possible to restrict the size with the network MTU, it is more difficult to limit the frequency of messages sending locally without a global perception of the network.

With respect to our initial goal, risk-oriented application design, some guarantees should be given to the risk actors in order for them to take the right

decisions. The design and the implementation of our testbed and experimental platform of MANETs is a step in that direction. We are now aware of the limitations of classical approaches for broadcasting information and for routing, and we can claim that designing routing protocols has to take into account neighbor detection mechanism sensitivity according to the traffic.

The two main next steps are to design a routing protocol which avoid the neighborhood instability, and to test it in mobile topology.

References

- [ABB⁺04] Aguayo, D., Bicket, J., Biswas, S., Judd, G., Morris, R.: Link-level measurements from an 802.11b mesh network. *SIGCOMM Comput. Commun. Rev.* 34(4), 121–132 (2004)
- [ABC⁺05] Anastasi, G., Borgia, E., Conti, M., Gregori, E., Passarella, A.: Understanding the real behavior of mote and 802.11 ad hoc networks: an experimental approach. *Pervasive and Mobile Computing* 1, 237–256 (2005)
- [AJF⁺08] Dutot, A., Franzolini, J., Guinand, F., Olivier, D., Mallet, P.: Information routing for risk management based on mobile ad-hoc networks. In: *Proceedings of LambdaMu, Avignon (France)*, vol. 16, pp. 7–9 (October 2008)
- [DVL03] Dhoutaut, D., Vo, Q., Lassous, I.G.: Global visualisation of experiments in ad hoc networks. Technical Report 4933, INRIA (2003)
- [KM07] Kiess, W., Mauve, M.: A survey on real-world implementations of mobile ad-hoc networks. *Ad Hoc Networks* 5(3), 324–339 (2007)
- [KS06] Kim, K.-H., Shin, K.G.: On accurate measurement of link quality in multi-hop wireless mesh networks. In: *Proceedings of the 12th annual international conference on Mobile computing and networking, MobiCom 2006*, pp. 38–49. ACM, New York (2006)
- [MBJ00] Maltz, D.A., Broch, J., Johnson, D.B.: Quantitative lessons from a full-scale multi-hop wireless ad hoc network testbed. In: *Proceedings of the IEEE Wireless Communications and Networking Conference*, pp. 992–997 (2000)

Author Index

- Abraham, Ajith II-72
Ahn, Chi-Hyun I-437
Alaa, Ghada II-80
Almulla, Sameera Abdulrahman I-75
An, Jun-Ho II-230
Arasteh, Bahman I-110
Arvidsson, Adam II-34
- Baek, Woon Kyung I-280
Bat-Erdene, Byambasuren II-214
Blanken, Malte II-1
Boo, Jun-Pil II-167
Byun, Giseob II-288
Byun, Haewon II-352, II-357
Byun, Sang-Bong I-237
- Cha, Sungdeok II-131
Chahir, Y. II-196
Chang, Jyh-Biau I-67
Chang, K.H. II-123
Chen, Po-Cheng I-67
Cheong, Seung-Kook I-475, I-508
Chiang, Tzu-Chiang II-16
Chilamkurti, Naveen I-101
Cho, Choong-Ho I-157
Cho, Dong-Hwan I-393, I-417, I-426, I-483
Cho, Hyunhun II-238, II-245
Cho, Seongsoo I-273
Cho, Wanwoo I-463
Choe, Yoonkyung II-348
Choi, Chan-Yung II-256, II-371
Choi, Dongjin I-37
Choi, Ikkyu I-21
Choi, JiYoung II-321
Choi, Ki-Young II-415
Choi, Kwansun I-140
Choi, Min I-183
Choi, Yong-Hoon I-93
Choi, Yong-Seok I-323
Choi, Young-Gyu I-308
Chu, Hai-Cheng II-123
Chun, Se-Hak I-190, I-196, I-200
Chung, Hyunji II-100
- Chung, Kwang Sik II-313, II-379
Chung, Taewon II-345
Coelho, Jorge II-44
Colleoni, Elanor II-34
Cristea, Dan II-188
- Dai, Bin I-1
Du, Guoping I-85
- Engelbert, Benedikt II-1
Etter, Michael II-34
- Fatima, Iram I-47
Filipiak, Joanna I-148
Florido, Mário II-44
Fragopoulou, Paraskevi II-9
Franzolini, Julien I-514
- Gao, Yuhang I-85
Gifu, Daniela II-188
Gorrepati, Rajani Reddy II-159
Gu, Bongen I-285
Gu, Mi-Ran I-212
Guinand, Frédéric I-514
Gulliver, Stephen R. II-54
Guo, Shaohui I-85
- Ha, Sang-Yong I-175
Hakim, R.B. Fajriya II-90
Hamad, Ahmed II-80
Han, Saeron I-140
Han, Seungjin I-123
Han, Young-Hwan I-323
Hansen, Lars Kai II-34
Hassan, Mohammad Mehedi II-397
Hecker, Artur II-24
Heo, Kwang-Seon I-308
Hong, Bong-Hwa I-230, I-240, I-273, I-299
Hong, Jaeweon I-463
Hong, Jin-Hyuk I-393
Hong, Kwang-Seok II-230
Hong, Sengphil II-224
Horak, Zdenek II-72
Hosseini, Mohammad Javad I-110

- Huh, Eui-Nam II-387, II-397, II-407
 Hussein, Dina II-80
 Hwang, Joon Ho I-200
 Hwang, Myunggwon I-37
 Hwang, Sunjin I-445
 Hyun, Myung Hwa II-313

 IL Park, Jeong I-343
 Im, Kyoung-Mi II-296, II-371
 Islam, Motaharul Md. II-407

 Jang, Ho I-463
 Jang, Hyunmi II-224
 Jang, Jeanyeung I-409, I-456
 Jang, Sung-Won I-329
 Jeon, Changwan I-140
 Jeon, Heunggu I-140
 Jeon, Hyekyoung I-123
 Jeon, Moongu I-316
 Jeong, Chang-Sung II-415
 Jeong, Chorong II-115
 Jeong, Hwa-young I-230, I-240, I-256
 Jeong, Jae-Jin I-508
 Jeong, Yong-Wan I-508
 Jeong, Young-Sik I-21, I-183
 Ji, Wonkyu I-364
 Jin, Ik Soo I-167
 Jin, Sung Ho II-379
 Jo, JaeChoon II-304
 Jo, Sung-Up II-415
 Joo, Moon G. I-280
 Jun, Seungsu I-224
 Jung, Yong-kuk I-308

 Kang, Dae-Seong I-212, I-218
 Kang, Jang-Mook I-299
 Kang, Jinwon II-345
 Kang, Seung-Gu I-204, I-492
 Khattak, Asad Masood I-47
 Khil, Kijeong I-475
 Kim, Ahsol II-238
 Kim, Bong-Han I-500
 Kim, Byung-Chul I-131
 Kim, Dae Jung II-272
 Kim, Dae-Sung I-308
 Kim, Do-Hyeun II-159, II-167
 Kim, Dong-Han I-200, II-214
 Kim, Dong-Keon I-248
 Kim, Dongsik I-140
 Kim, Du Yong I-316
 Kim, Eun-Kyung I-237
 Kim, Huy Kang I-29
 Kim, Hye Jeong II-321
 Kim, Hyongsoon I-175
 Kim, Hyoungchul I-437
 Kim, Iksoo I-265
 Kim, Jae-Jung II-224
 Kim, Jin-Sung I-417
 Kim, Jung-sik I-371, I-378
 Kim, Ki Youn II-272, II-288
 Kim, Kwangeak II-264
 Kim, Kyongjin II-224
 Kim, Minkoo I-21
 Kim, Pankoo I-37
 Kim, Se-Han I-131
 Kim, Se-Jin I-131, I-157
 Kim, Seonmin I-196
 Kim, Seunghae II-238, II-245
 Kim, Sung-Woo I-293
 Kim, Taisiya II-288
 Kim, Young-Gab II-131
 Kim, Young-Hyuk I-204, I-492
 Kim, Yunsik I-445
 Ko, Byeongkyu I-37
 Ko, Dae-Sik I-508
 Ko, Jong-won I-256
 Ko, Kyunbyoung I-401, I-409, I-456
 Kruthoff-Brüwer, Ralf II-1
 Ktari, Salma II-24
 Kudelka, Milos II-72
 Kumar, Neeraj I-101
 Kwak, Dongho I-475
 Kwak, Yoonsik I-285, I-387, I-469
 Kwak, Youngsik I-358, I-445, I-463
 Kwak, Yunsik I-475
 Kweon, Min-Su I-308
 Kwon, Mi-kyung II-329
 Kwon, Sook-Youn II-256
 Kwon, Soon T. I-280

 Lassoued, I. II-196
 Le, Thi-Lan II-175
 Lee, Bong Gyou II-264, II-272, II-280, II-288
 Lee, Byung-Bog I-157
 Lee, Chung-Hong II-62
 Lee, Eunyoung I-175
 Lee, Jae-Kwang I-204, I-492, I-500
 Lee, Jae-Yong I-131
 Lee, Jun-Eui II-337

- Lee, Jung-Hoon I-293
 Lee, Junghyun I-123
 Lee, Jun-Hyung II-407
 Lee, Keun-gi II-115
 Lee, Kilhung I-93
 Lee, Min-Ho II-367
 Lee, Saebyeok II-309
 Lee, Sang-Goog I-316
 Lee, Sangjin II-100, II-115
 Lee, Sang-Won II-337
 Lee, Seungyeol I-265
 Lee, Seungryong I-224
 Lee, Sungyoung I-47
 Lee, Sunheum I-140
 Lee, Won-bong I-350
 Lee, WonGyu II-321
 Lee, Won-hyung II-329, II-337
 Lee, Yong-Hwan I-237
 Lee, Young-Hoon I-508
 Lee, Young-Koo I-47
 Lee, Yunkyung I-358, I-463
 Lim, Changbeom II-245
 Lim, Heui-Seok II-304, II-309, II-325
 Lim, Hye-Youn I-212, I-218
 Lim, IL-Kwon I-204, I-492
 Lim, Jae-Hwan I-293
 Lim, Jae-Hyun II-256, II-296, II-371
 Lim, Jongsik I-140
 Liu, Yujing I-1

 Min, Jin-Hong II-230
 Moon, Young Hoon I-29
 Morisse, Karsten II-1

 Na, Yoonhye II-348
 Nakata, Keiichi II-54
 Nam, Kichun II-345, II-348
 Nam, Sang-Yep II-214
 Nielsen, Finn Aarup II-34
 Niu, Jianwei I-85

 Oh, Dong-Sik I-500
 Oh, Duckshin II-280
 Olivier, Damien I-514
 Overill, Richard E. II-110

 Pak, Jugeon I-335
 Panagiotakis, Costas II-9
 Papadakis, Harris II-9
 Park, Byeong-Ho I-329
 Park, Chan-Hong I-323
 Park, Daehyun I-463
 Park, Dong-Hee I-387, I-469
 Park, Dong-Hwan II-159
 Park, Doo-Soon I-101, I-140
 Park, Dowoo II-238
 Park, Hee-Dong II-167
 Park, Heejin II-348
 Park, Hyeon II-159
 Park, Jaesung I-93
 Park, Jinhyung II-238, II-245
 Park, Jisu II-313
 Park, Jong-Dae I-218
 Park, Jong Hyuk I-101, I-183, II-123
 Park, Joongoo II-238, II-245
 Park, Jungheum II-100
 Park, Keehyun I-335
 Park, Sang-Ju I-329
 Park, Sang-yong II-329
 Park, Seongwon II-264, II-280
 Park, Seung-Hun I-293
 Liu, Haipeng II-206
 Pérez, Pablo Martín I-148

 Ragab, Khaled II-148
 Rho, Seungmin I-21
 Rim, Keewook I-123
 Roe, Paul II-167
 Ryou, Jae-Cheol II-367
 Ryu, Hyun-gi I-371, I-378
 Ryu, Jee-Youl I-293
 Ryu, Kwang-Taek I-175

 Salehi, Mostafa E. I-57
 Seong, Hyeon-Kyeong I-323, I-329
 Se Park, Hwa I-343
 Shalaik, Bashir II-139
 Shieh, Ce-Kuen I-67
 Shim, Hyoung Seop I-426
 Shin, Munbong I-224
 Shin, Sung-Ho II-369
 Shon, Jingon II-313
 Shrestha, Bhanu I-273
 Sierra, José María I-148
 Sim, Han-moi II-329
 Skulimowski, Andrzej M.J. I-11
 Snasel, Vaclav II-72
 Son, Kwang Chul I-273
 Song, Biao II-387, II-397
 Song, Jinyoung II-345

- Song, Kieun I-445
 Song, Seokil I-475
 Song, Young-jae I-256
 Songvetkasem, Tirayu II-183
 Su, Jinshu I-1
 Su, Yen-Liang I-67
 Subanar, II-90
 Sung, Haeng-Nam I-483
 Sung, Won-Kyung II-367, II-369

 Tian, Yuan II-387
 Tong, Chao I-85

 Ul Haque, Asrar II-148

 Wang, Shan-Yang II-16
 Wang, Xu I-364
 Wang, Yi-Da II-123
 Wang, Xiaojie II-206
 Wiafe, Isaac II-54
 Winarko, Edi II-90
 Winstanley, Adam II-139

 Woo, Choongchae I-401, I-409
 Woo, Jongjung I-437
 Woo, Yoseop I-265
 Wu, Chih-Hung II-62

 Zhong, Yixin II-206

 Yang, Chu-Sing II-16
 Yang, YeongWook II-325
 Yazdanbakhsh, Amir I-57
 Yeun, Chan Yeob I-75
 Yoon, Hwa-Mook II-367
 Yoon, Min I-293
 Yoon, Wha-Mook II-369
 Yu, Heon-Chang II-379
 Yu, Weikuan I-285
 Yun, SungHyun II-360

 Zagrouba, E. II-196
 Zahrani, Mohammed II-148
 Zhu, Peidong I-1