The Information Leakage through a Randomly Generated Function

Lennart Brynielsson
TSA
S-107 86 Stockholm, Sweden

Abstract: If a randomly filled memory is used to combine some ML-shift registers then the obtained mutual information between the output and a set of inputs will be a random variable. Its distribution is demonstrated to be approximately proportional to a χ^2-distribution.

In cryptographic systems several pseudo-random sources are often combined by means of a combining function. This is sometimes accomplished by a memory filled with random data, for instance derived from parts of the key. In order to evaluate the feasibility of a correlation attack, the statistical behaivor of such a combination function must be examined. [1]

If the arguments of the function are assumed to be independent and uniformly distributed random variables, then the correlation, or more generally the mutual information, between the output and some of the arguments can be computed. This gives, per output symbol, the maximal entropy loss of the source that controls those arguments. [1,2]

Without loss of generality we consider a function $w = F(u,v)$ of two variables, $u \in Z_m$ and $v \in Z_n$, with values $w \in Z_N$. Let the arguments U and V be independent and uniformly distributed random variables and consider the mutual information $I(U,W)$ between U and the output W. (In most applications m, n and N are powers of two). If the function values are independently drawn according to a distribution on Z_N then this information I will be a random variable. We here claim that I, if natural logarithms are used, has the approximate expectation:

$$E[I] \approx \frac{(m-1)(N-1)}{2mn}$$

Moreover, asymptotically for large n, the variable $2mn \cdot I$ is χ^2-distributed with $(m-1)(N-1)$ degrees of freedom.

Let X_{ij} be the frequency of j:s among the function values with $u=i$ and let $\bar{X}_j$ be the average frequency.

$$X_{ij} = \#\{ v ; F(i,v) = j \} \qquad \bar{X}_j = \frac{1}{m} \sum_{i=1}^{m} X_{ij}$$

The mutual information can now be expressed:

$$I(U,W) = H(W) - H_U(W) = H(W) - \frac{1}{m} \sum_{i=1}^{m} H(W \mid U=i) =$$

$$= \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{N} \frac{X_{ij}}{n} \log \frac{X_{ij}}{n} - \sum_{j=1}^{N} \frac{\bar{X}_j}{n} \log \frac{\bar{X}_j}{n} =$$

$$= \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{N} X_{ij} \log(X_{ij}/\bar{X}_j) \approx \frac{1}{2mn} \sum_{i=1}^{m} \sum_{j=1}^{N} \frac{(X_{ij} - \bar{X}_j)^2}{\bar{X}_j}$$

where we for each j have used the series expansion around $\bar{X}_j$. The last sum is however known from the statistical analysis of the homogenity of contingency tables. It is known to be approximately χ^2-distributed with $(m-1)(N-1)$ degrees of freedom. [3]

References:

[1] Siegenthaler T, "Methoden für den Entwurf von Stream Cipher Systemen", Ph.D Dissertation, ETH Zürich, 1987.

[2] Brynielsson L, "Radgeräte und ihre kritische Länge", Kryptologie Aufbauseminar, J. Kepler Universität Linz, 1984.

[3] Cramer H, "Mathematical Methods of Statistics", Princeton University Press, 1951.