Lecture Two

Randomness

7. CRYPT'X Randomness Tests

Designer and users of encryption algorithms used in cipher systems need a systematic approach in examining their cipher prior to use, to ensure that they are safe from cryptanalytic attack. In this manner we will introduce a new package of randomness instead of the mentioned five tests. CRYPT-X is a microcomputer package that is intended to be used to test either large binary strings that are to be used as keystream in stream ciphers or block cipher algorithms.

7.1 Statistical Tests Distributions

The package uses a number of statistical distributions including the standard normal and the chi-squared distribution.

The **standard normal distribution** is used to compare a sample measure obtained from the cipher with expected measure of hypothesized distribution. The test statistic for the standard normal distribution is $z=(x-\mu)/\sigma$, where x is the sample measure and μ and σ are the expected measure and variance of the hypothesized distribution.

The **chi-squared distribution** is used to compare the goodness-of-fit of the observed frequencies of a sample measure to the corresponding expected frequencies of the hypothesized distribution.

The test statistic is:

$$\chi^2=\sum_{i=1}^{n}\frac{(o_i-e_i)^2}{e_i} \qquad \ldots(7.1)$$

Where o_i and e_i are the respective observed and expected frequencies of all possibilities of the measure. This result in the number of degree of freedom ν being one less than the resulting number of e_i values obtained i.e. ν=n-1.

Algorithms calculating the tail-area probabilities of the standard normal and chi-squared distributions have been incorporated into the package. If the tail-area probability α of the statistic is extremely small (α<0.001 for large sample) then it would be interpreted that the cipher does not satisfy the test applied.

7.2 Stream Cipher Tests

The security of a stream cipher depends on the keystream appearing random. In most cases the keystream is formed by a deterministic generator which produces a periodic sequence. If the stream deviated significantly from randomness in some fashion a cryptanalyst may be able to use some the decrease in entropy caused by this deviation. The method applied in this package examines the hypothesis that the string was based on Bernoulli trials, for which

$$\Pr(z(t)=1) = \Pr(z(t)=0) = \frac{1}{2} \qquad \ldots(7.2)$$

Tests employed in the package to examine this hypothesis are the **Frequency** tests on the original stream and the 1st and 2nd **Binary Derivative** stream, **Change Point** test, **Subblock** (**Poker**) test and **Run** test. The details of these tests are as follows:

1. **Frequency Test**: the aim of this test to determine how the proportion of ones in the sample stream of length n bits fit into the hypothesized distribution where the proportion of one's is 0.5. Using n_1 as the number of ones, the standard normal test statistic is:

$$z = 2\sqrt{n}\left(\frac{n_1}{n} - 0.5\right) \qquad \ldots(7.3)$$

2. **Binary Derivative Test**: it's a new stream found by the modulo-two addition of successive bits in the stream. In forming the 1st binary derivative we are looking at the overlapping 2-tuples 00, 01, 10, 11 in the original stream. The proportion of ones in the 1st binary derivative gives the proportion of the total of 01 and 10 patterns in the original stream.
3. **Change Point**: at each bit position t in the stream the proportion of ones to the point is compared to the proportion of ones in the remaining stream. The bit where the maximum change occurs is called the change point. This test determines whether this change is significant for a binomial distribution where the proportion of ones in the stream is expected to be 0.5.
4. **Subblock Test: (poker)** its partitions the stream into F hands of length m bits. For a stream of size n, where $F/2^m \geq 5$, the total number of hands is $\lfloor n/m \rfloor$, where $\lfloor\ \rfloor$ denotes the integer value. The aim of this test is to show that there is an equal number of each of 2^m possible hands. If f_i denotes the frequency of hand pattern i, then the test statistic used is:

$$\chi^2=(\frac{2^m}{F})\sum_{i=0}^{2^m-1} f_i^2 - F \qquad \ldots(7.4)$$

   This compared with chi-squared distribution with degree of freedom equal to 2^m-1.
5. **Runs Test**: this test counts the number of runs of one's (blocks) and runs of zeros (gaps) for each possible run length. For random data there should be an equal number of blocks and gaps. The expected number of blocks (gaps) of length i is:

$$e_i=\frac{\frac{n-i-1}{2}+2}{2^{i+1}} \qquad \ldots(7.5)$$

A chi-squared test is performed on the bit stream to test for the goodness-of-fit of the number of blocks and gaps to this distribution. The chi-squared statistic is calculated as:

$$\chi^2=\sum_{i=1}^{k}\frac{(B_i-e_i)^2}{e_i}+\sum_{i=1}^{k}\frac{(G_i-e_i)^2}{e_i} \qquad \ldots(7.6)$$

Where B_i is the number of blocks of length i, G_i is the number of blocks of length i, and Σ denotes the summation over all possible run k of length i such that $e_i \geq 5$. This is compared with 2k-2 degree of freedom.