SEQUENCE COMPLEXITY AS A TEST FOR CRYPTOGRAPHIC SYSTEMS

A. K. Leung & S. E. Tavares

Department of Electrical Engineering
Queen's University, Kingston, Ontario, Canada

ABSTRACT

The complexity of a finite sequence as defined by Lempel and Ziv is advocated as the basis of a test for cryptographic algorithms. Assuming binary data and block enciphering, it is claimed that the difference (exclusive OR sum) between the plaintext vector and the corresponding ciphertext vector should have high complexity, with very high probability. We may refer to this as plaintext/ciphertext complexity. Similarly, we can estimate an "avalanche" or ciphertext/ciphertext complexity. This is determined by changing the plaintext by one bit and computing the complexity of the difference of the corresponding ciphertexts. These ciphertext vectors should appear to be statistically independent and thus their difference should have high complexity with very high probability. The distribution of com plexity of randomly selected binary blocks of the same length is used as a reference. If the distribution of complexity generated by the cryptographic algorithm matches well with the reference distribution, the algorithm passes the complexity test. For demonstration, the test is applied to modulo multiplication and to successive rounds (iterations) of the DES encryption algorithm. For DES, the plaintext/ciphertext complexity test is satisfied by the second round, but the avalanche complexity test takes four to five rounds before a good fit is obtained.

G.R. Blakley and D. Chaum (Eds.): Advances in Cryptology - CRYPTO '84, LNCS 196, pp. 468-474, 1985.

INTRODUCTION

A block enciphering algorithm may be regarded as a reversible transformation which maps binary n-vectors into binary n-vectors, for a given key. In modern cryptography it is usually assumed that the cryptographic algorithm is known and only the key is kept secret. In principle, a cryptographic scheme can always be broken by an exhaustive key search. However, if the key set is large, such a search becomes computationally infeasible. On the other hand, if the cryptographic algorithm is not well designed, the key may be discovered with high probability by searching a much smaller set. Thus there is a need to develop statistical tests to reveal such weaknesses. A recent and interesting test is the complexity test. We will discuss some properties of complexity in this paper and apply the test to modulo multiplication and the DES encryption algorithm.

THE COMPLEXITY CRITERION

Lempel and Ziv [1] introduced the idea of the complexity of a finite sequence and developed several of its important properties. Fischer [2,3] recognised the application of complexity to cryptographic algorithms. Spencer and Tavares [4] applied the complexity test to a layered broadcast cryptographic system and found it to be quite sensitive. Intuitively, the complexity of a sequence is a measure of the rate at which new patterns emerge as we move along the sequence. Starting at one end, say the left, we put a marker whenever a new sequence appears. The complexity is the number of distinct patterns which have been identified. To illustrate, consider the sixteen bit sequence

$$\underline{X} = 1001101110000111.$$

Inserting a marker after each new pattern, we have

$$\underline{X} = 1|0|01|101|1100|00111|$$

and thus $\underline{X}$ has a complexity of 6. Lempel and Ziv showed that, in the limit, almost all binary sequences of length n have complexity exceeding n/log n. Thus for sequences of length n, the expression $C_n = n/\log n$, may be regarded as a threshold of complexity. If we compute the complexity of a large number of randomly selected binary sequences of length n, we can determine an IDEAL distribution of complexity as shown in Fig. 1 for sequences of length n = 64. The above sequences

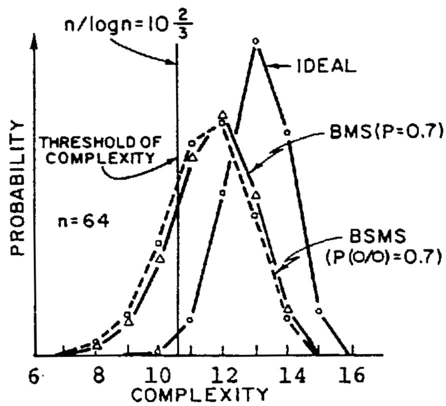

Fig. 1: Distribution of sequence complexity for 64-bit sequences from a selection of binary sources. The IDEAL curve is derived from a Binary Memoryless Source (BMS) with equiprobable symbols. The curve labelled BMS is based on a BMS with p(0)=0.7 and the dashed curve is based on a Binary Symmetric Markov Source (BSMS) with p(0/0)=p(1/1)=0.7.

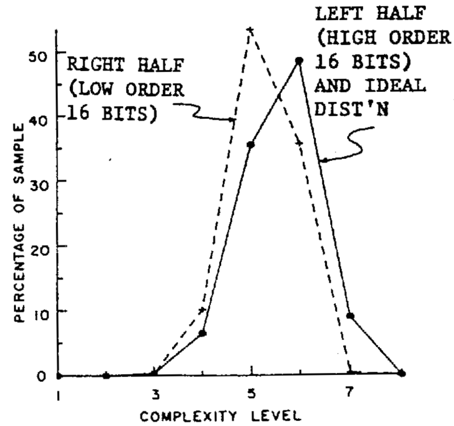

Fig. 3: Distribution of complexity for right half (low order 16 bits) and left half (high order 16 bits) for 32-bit multiplications, modulo 2^{32}.

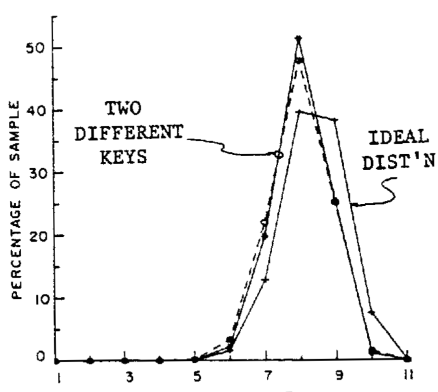

Fig. 2: Plaintext/ciphertext distribution of complexity for 32-bit modulo multiplication, modulo 2^{32}. It can be seen that the distributions fall short of the ideal distribution.

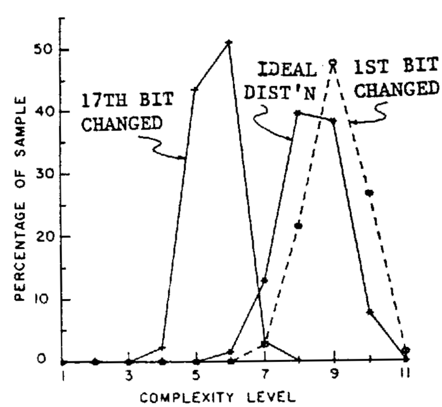

Fig. 4: Distribution of avalanche complexity for 32-bit multiplication, modulo 2^{32}. The distribution depends on which plaintext bit is complemented to generate the avalanche effect.

could also be generated by selecting 64-bit blocks from a Binary Memoryless Source (BMS) with equiprobable symbols. Lempel and Ziv [1] showed that the distribution of complexity is related to the entropy of the source generating the sequences and this is illustrated by the other two curves in Fig. 1. The curve labelled BMS is generated by a Binary Memoryless Source with p(0) = 0.7 and the curve labelled BSMS is generated by a Binary Symmetric Markov Source with p(0/0) = p(1/1) = 0.7. These two information sources have the same entropy (.881 bits/symbol) but different structure and it is seen that they are quite close together, but distinct from the ideal distribution. The threshold of complexity is given by $C_n = 64/\log_2 64 = 10\ 2/3$.

In an ideal block cryptographic system the plaintext vector $\underline{P}$ and the corresponding ciphatext $\underline{C}$ should appear to be independent of each other. Let

$$\underline{S} = \underline{P} \oplus \underline{C}$$

where $\oplus$ means the exclusive OR sum of the two binary n-vectors, term by term. Then, for a well designed cryptographic algorithm,

$$C(\underline{S}) \geq n/\log n$$

with high probability, where $C(\underline{S})$ is the complexity of the sequence $\underline{S}$ (of length n). If we pick a large number of plaintext sequences $\underline{P}$ at random and compute $C(\underline{S})$ in each instance, then the distribution of $C(\underline{S})$ should appear as indicated by the 'IDEAL' curve in Fig. 1, where n = 64 in this instance.

The complexity $C(\underline{S})$ defined above may be referred to as plaintext/ciphertext complexity, since $\underline{S}$ is the difference of $\underline{P}$ and $\underline{C}$. In a similar manner, we can define a ciphertext/ciphertext or 'avalanche' complexity as follows. Let the plaintext vector $\underline{P}$ generage the ciphertext $\underline{C}$, and $\underline{P}'$ generate $\underline{C}'$ where $\underline{P}'$ is obtained from $\underline{P}$ by complementing a bit in a designated location. Determine the n-vector

$$\underline{U} = \underline{C} \oplus \underline{C}'$$

where $\underline{U}$ is a measure of the difference between the ciphertexts and thus is also a measure of the avalanche effect. In an ideal cryptographic algorithm, $\underline{C}$ and $\underline{C}'$ should appear to be statistically independent and thus $\underline{U}$ should appear to be randomly selected from the set of all binary n-tuples. Letting $C(\underline{U})$ represent the complexity of $\underline{U}$, it should also be true that

$$C(\underline{U}) \geq n/\log n$$

with very high probability. The distribution of avalanche complexity for a specified plaintext bit position can be estimated by selecting plaintext vectors at random and complementing the designated bit. The complexity of $\underline{U}$, $C(\underline{U})$, is determined in each case. If the cryptographic algorithm is well designed, the distribution generated in this way should match very closely with the ideal distribution generated by the set of all binary vectors of length n. Note that the avalanche complexity distribution may be a function of the bit location that is complemented. Such variations would reveal cryptographic weaknesses. Avalanche complexity can also be defined by complementing key bits instead of plaintext bits. It should also be noted that the avalanche effect can be generalized by complementing a specified combination of bit positions.

THE COMPLEXITY TEST APPLIED TO MODULO MULTIPLICATION

The operations $A*B \bmod 2^n$ and $A*B \bmod 2^n-1$, where A and B are binary n-vectors, are helpful for illustrating the complexity test. The operation * between A and B is binary multiplication, and reduction mod 2^n is easily implemented since overflow high order digits fall off the end. However, due to the fact that the carries propagate from right to left and the overflow drops off the end, the mixing effect is not uniform. To examine this more closely, we applied the complexity test to the operation $A*B \bmod 2^n$, for n = 32. One of the parameters, say B, is kept fixed and may be regarded as the key (B must be an odd integer for invertibility). The other, A, is a random 32-bit binary vector which is selected many times. The plaintext/ciphertext complexity test is performed for each choice of A and a distribution of complexity is obtained. This is shown in Fig. 2 and gives the average complexity averaged over the 32 bits. To exhibit the non-uniformity, we can perform the complexity test on the left half (high-order 16 bits) and right half (low-order 16 bits) separately. It can be seen from Fig. 3 that for the same choice of B (the "key"), the left half is more complex than the right half.

The avalanche complexity test was also performed for the operations $A*B \bmod 2^n$ and $A*B \bmod 2^n-1$. It can be seen from Fig. 4 that the complexity distribution for the modulus 2^n differs quite substantially from the ideal distribution, but the fit is much better for the modulus 2^n-1. This can be seen by comparing Fig. 4 and Fig. 5, where Fig. 5 gives the avalanche complexity for modulus $2^{32}-1$. After

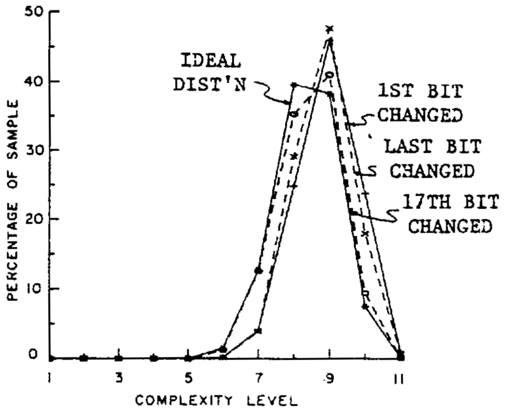

Fig. 5: Distributions of avalanche complexity for 32-bit modulo multiplication, modulo $2^{32}-1$. The curves are much closer to the ideal distribution than for modulo 2^{32}.

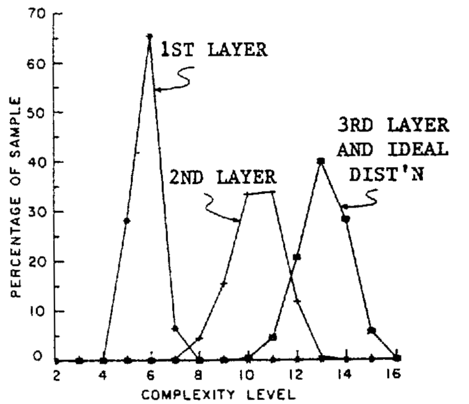

Fig. 7: Avalanche complexity for successive layers of DES produced by complementing the 32nd bit of plaintext. The curves for four or more layers are very close to the ideal distribution.

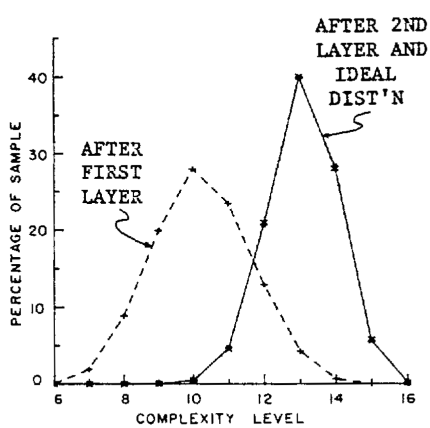

Fig. 6: Plaintext/ciphertext complexity for successive rounds (layers) of DES. From the 2nd layer on, the curves are indistinguishable from the ideal distribution.

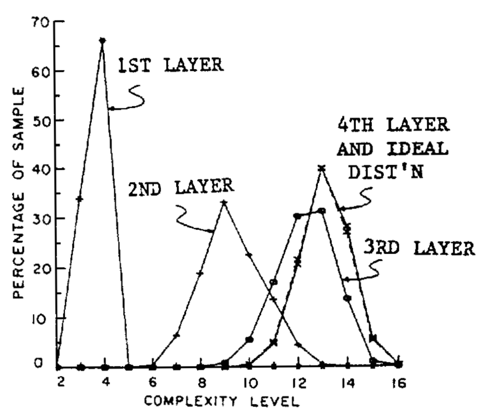

Fig. 8: Avalanche complexity for successive layers of DES produced by complementing the first bit of key. The curves are very close to the ideal distribution by the fourth layer.

a little reflection this should not be too surprising. For operations mod 2^n-1, the carries propagate around the end cyclically and the effect of the carry is much more uniform.

APPLYING THE COMPLEXITY TEST TO DES

It would be expected that the DES encryption algorithm should do well under the complexity tests, and this was found to be the case. What is also of interest is to observe how rapidly the DES algorithm approaches the ideal complexity distribution as we include more of the 16 rounds or iterations. (The initial and final permutations are ignored.) As shown in Fig. 6, the plaintext/ciphertext complexity converges to the ideal after the second iteration. However, the avalanche complexity requires four to five iterations before there is a good fit. This indicates that the avalanche complexity test is more demanding than the plaintext/ciphertext complexity test. The avalanche complexity test is performed by complementing a plaintext bit and a key bit as shown in Figs. 7 and 8, respectively.

ACKNOWLEDGEMENT

The authors wish to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada in the form of a Strategic Grant on Communications, Grant No. G0666.

REFERENCES

[1] A. Lempel and J. Ziv, "On the complexity of finite sequences", IEEE Trans. on Inform. Theory, Vol. IT-22, Jan. 1976, pp. 75-81.
[2] E. Fischer, "A Theoretical Measure of Cryptographic Performance", Cryptologia, Vol. 5, Jan. 1981, pp. 59-62.
[3] E. Fischer, "Measuring cryptographic performance with production processes", Cryptologia, Vol. 5, Jul. 1981, pp. 158-162.
[4] M.E. Spencer and S.E. Tavares, "Layered Broadcast cryptographic systems", Advances in Cryptology: Proceedings of Crypto '83, Plenum Press 1984, pp. 157-170.