IJCTA, 10(07), 2017, pp. 97-102
© International Science Press

Chosen-Plaintext Attack on Block Ciphers Based on the use of Fast Information Calculation Algorithm

Ivan Fedyanin* and Valery Korzhik*

Abstract : In this work we consider a problem of chosen-plaintext attack (CPA) on symmetric block ciphers. Our solution is based on well known Shannon inequality for mutual information between plaintext and ciphertext. Since cipher key entropy is limited this inequality gives nonzero value for mutual information between genuine plaintext and cipertext for any cipher. Numerical calculation of mutual information has received a support through relatively recent paper by A. Kraskov at al. This approach executes a computing of *k*-nearest neighbour distance. We have applied such technique to 16-bit and 32 bit block lengths, four rounds substitution-permutation Heys's ciphers. Our simulation showed that CPA problem can be reliably solved against such cipher. Unfortunately, we are faced so far with unreal computing problem for ciphers with more than 32 bit block lengths.

Keywords : Semantic security, block ciphers, Shannon inequality, *k*-nearest neighbour distance.

1. INTRODUCTION

Contemporary cryptosystems should be computationally secure. This means that the best known cryptanalytic attack is untractable with point of view computational resourses of codebreakers. But there is also more stronger requirement to modern cryptosystems – resistance to so called *chosen-plaintext attack(CPA)* or in another words semantic security[1].

Let us consider an attacker who has several plaintexts and one ciphertext. The following question arises: which of plaintexts has been encrypted by given ciphertext or none of them? If such problem cannot be solved in a feasible time and with tractable hardware, then cryptosystem is called indistinguishable under CPA or semantic secure one.

It is well known that many public-key cryptosystems (RSA among them) are not semantic secure. In fact, because public key is insecure an attacker is able to encrypt with it all plaintexts and compare the results with given ciphertext. As far as symmetric block ciphers this problem was not solved completely [2].

We consider substitution-permutation four rounds block cipher proposed by Heys [3] starting with block length 16 bits and key string having 80 bits. That means that total number of keys is 280 $\approx 1.2.10^{24}$ and hence brute force attack by key exhaustion is unrealistic.

In Fig. 1 is presented a scheme of such cipher and in Table 1 and Table 2 are presented S-box transforms and permutation mapping.

Table 1

S-box Transforms that Have the Same Structure for all S-boxes and they are Presented in Hexadecimal System

Input	0	1	2	3	4	5	6	7	8	9	10(A)	11(B)	12(C)	13(D)	14(E)	15(F)
Output	E	4	D	1	2	F	B	8	3	A	6	C	5	9	0	7

* State University of Telecommunications, Saint-Petersburg, Russia *E-mail : ivan.a.fedyanin@gmail.com, val-korzhik@yandex.ru*

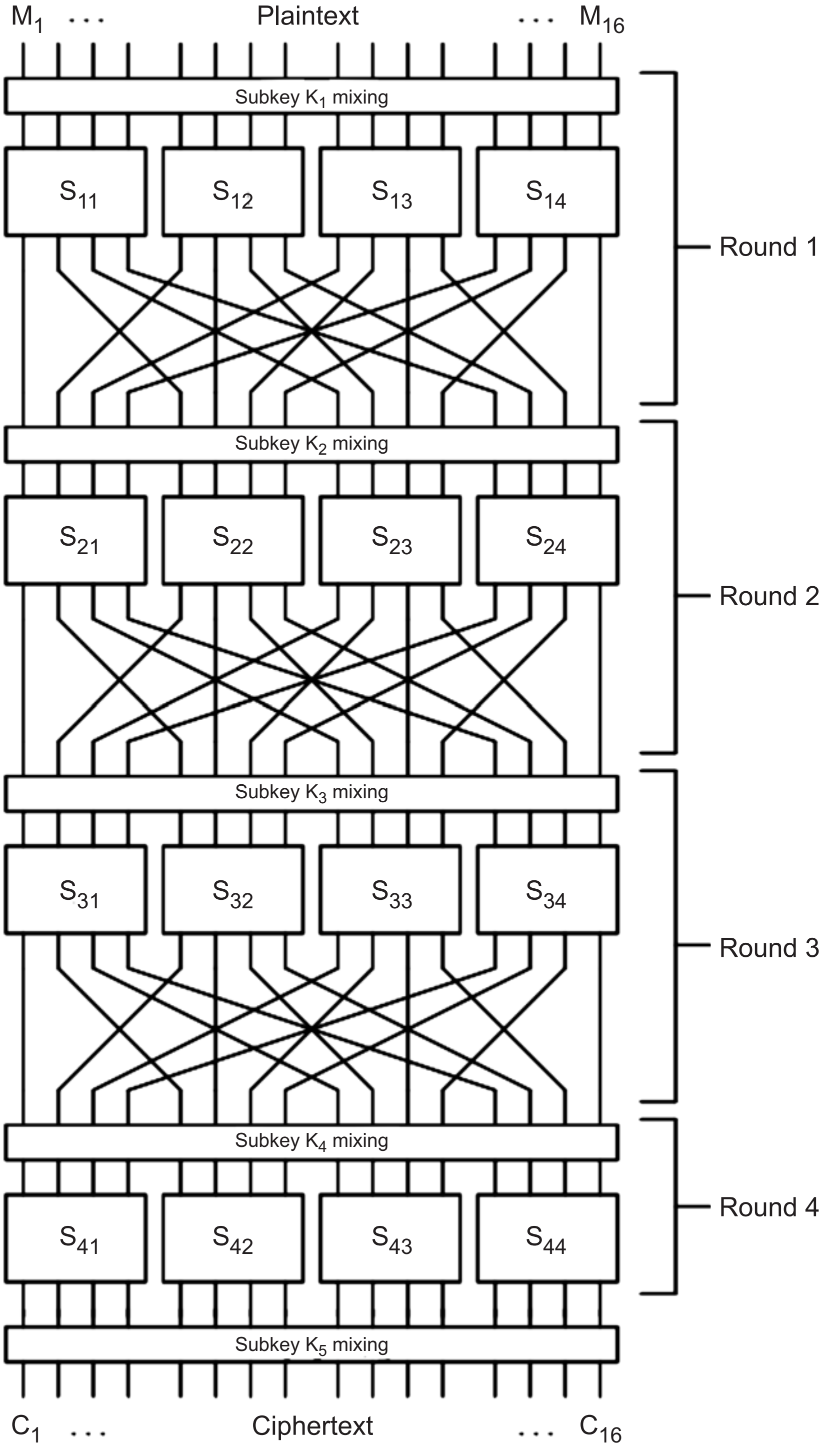

Figure 1: Substitution-permutation block cipher with block length 16 due to Heys[3]

Table 2
Permutation Mappings that Have the Same Structure for all Cipher Rounds

Input	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Output	1	5	9	13	2	6	10	14	3	7	11	15	4	8	12	16

For the case of substitution-permutation cipher with block length 32 bits we extend the 16 bit cipher with addition of the second halve of scheme to the first one keeping previous transforms in S-boxes and changing Table 2 for permutation mapping to Table 3 showed below.

Table 3

Permutation Mappings for 32-bit Block Length Cipher

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Output	1	5	9	13	17	21	25	29	2	6	10	14	18	22	26	30	3	7	11	15	19	23	27	31	4	8	12	16	20	24	28	32

The remainder of work is organised as follows: In Section 2 the CPA is described theoretically. The simulation results on CPA both for block cipher with block length 16 and 32 are given in Section 3. Finally, we summarize the results of our work in the conclusion.

2. THEORETICAL DESCRIPTION OF CPA AGAINST BLOCK CIPHERS

It is well known inequality for mutual information between plaintext and ciphertext that should be valid for any cryptosystem[4],[5]:

$$I(M^N, C^N) \geq H(M^N) - H(K^L) \tag{1}$$

where M^N is a sequence of message symbols of the length N, C^N is a sequence of ciphertext symbols of the length N (without of the generality lost we believe that these lengths are equal one to another), K^L is the binary key string of the length L .

We can transform inequality (1) by dividing both its sides on N:

$$I'(M^N, C^N) \geq H'(M^N) - H'(K^L) \tag{2}$$

where symbol “ ′ ” means that we consider a normilized to N corresponding values.

Since for computationally secure contemporary block ciphers the length of the key L is much less than the length of the message, we get asymptotically (as $N \to \infty$):

$$I'(M^N, C^N) \sim H'(M^N) > 0 \tag{3}$$

It follows from inequality (3) that if some message M^N has been in fact encrypted into ciphertext C^N with any unknown key K^L of the limited length L then for very large message length N we get nonzero mutual information $I'(M^N, C^N)$ but this value approaches to zero if M^N is not encrypted as C^N with some key. Hence we can take a decision about a choice of message that is encrypted into given ciphertext comparing the value $I'(M^N, C^N)$ with some threshold.

But the following problem appears – how it is possible to calculate mutual information $I'(M^N, C^N)$? Solution to this problem based on “binning” [6] was very hard generally but relatively recent has been published the paper [7] where it was used a method based on the notion of *k-nearest neighbour distance*. This approach can be termed as *fast mutual information calculation(FMIC)* between two N-dimension random vectors X and Y. It has been proved in [7] that FMIC can be performed by the following algorithm:

$$I(X, Y) = \Psi(1) - \langle \Psi (n_x + 1) + Y(n_y + 1)\rangle + \Psi(N) \tag{4}$$

where $X = \{x_1, x_2, ..., x_N\}$, $Y = \{y_1, y_2, ... , y_N\}$ vectors corresponding to M^N and C^N, $\Psi(x)$ is digamma function, $\Psi(x) = \Gamma(x)–1\ d\Gamma(x)dx$ that satisfies the recursion $\Psi(x + 1) = \Psi(x) + 1/x$ and $\Psi(1) = C$, where C = 0.5772156... is the Euler-Mascheroni constant. For large x, $\Psi(x) \approx \log x - 1/2x$. $n_x(i)$ is the number of points x_j whose distance from x_i is strictly less than $\varepsilon(i)/2$ and similarly for y instead of x. Here $\varepsilon(i)/2$ is the distance from $z_i = (x_i, y_i)$ to its neighbour and $\varepsilon_x(i)/2$ and $\varepsilon_y(i)/2$ are distances between the same points projected into the X and Y subspaces. Obviously, $\varepsilon(i) = \max (\varepsilon_x(i), \varepsilon_y(i)$. $\langle...\rangle$ is symbol that denotes an averaging both over all $i \in [1, ... , N]$ and over all realizations of random samples. But in our case we average only on all samples $i \in [1, ... , N]$ that is $\langle ... \rangle = \frac{1}{N}\sum_{k=1}^{N}(...)$.

In order to implement relation (4) for estimation of left side inequality (3) we map each of plaintext blocks $M_i = (m_{i1}, m_{i2}, \dots m_{in})$ into one integer X_i and each of the ciphertext blocks $C_i = (c_{i1}, c_{i2}, c_{in})$ into one integer Y_i following trivial relations, respectively:

$$\begin{aligned} X_i &= \sum_{j=0}^{n-1} x_{ij} 2^j , \\ Y_i &= \sum_{j=0}^{n-1} y_{ij} 2^j , \\ i &= 1, 2, \dots , N \end{aligned} \qquad (5)$$

where n is the block cipher length, x_{ij}, y_{ij} binary symbols of plaintext M^N and ciphertext C^N, respectively. (We assume of course that block cipher is binary and has the same length n of input and output blocks). Experimental investigation of technique described above is presented in the next Section.

3. SIMULATION RESULTS OF CPA AGAINST BLOCK CIPHERS

A. Heye's cipher with block length 16

We generate pseudo randomly two binary sequences M_I and M_{II} both of the length $n \cdot N$, where $n = 16$ is the cipher block length and N is the number of tested blocks. One of these sequences, say M_I is encrypted by Heye's block cipher that gives $n \cdot N$ ciphertext bits. (It is worth to noting that in the case of meaningful plaintext the entropy $H'(M^N)$ in (2) be lesser than for truly random binary sequence but it be still nonzero. Hence the proposed method works but we should select plaintext as close to truly random one only for simplicity reasons). Next we calculate mutual information $I(M_I, C)$ by (4) and (5), where X_i are integers corresponding to M_I and Y_i are integers corresponding to $C = f(M_I, K)$, where $f(\cdot)$ is the encryption function for Heye's 16-bit block cipher with 80-bit key chosen pseudo randomly. After that it is calculated also by (4) and (5) mutual information $I(M_{II}, C)$ between ciphertext C obtained after encryption of plaintext M_I and independent on it another plaintext M_{II}. The results of such calculations against the number of message bits N are presented in Table 4.

Table 4
Mutual Information Between Ciphertext and Plaintext Corresponding and no Corresponding to Given Ciphertext Against the Plaintext Bit Length N..

N	10^2	10^3	10^4	2×10^4	4×10^4	8×10^4	3×10^5	10^6
$I(M_I, C)$	0,3	1,2	5,52	7,057	8,77	10,3	12,65	14.24
$I(M_I, C)$	– 0,09	0,053	0,03	0,04	0,08	0,13	0,373	0.89

We can see from this Table 4 that in fact mutual information $I(M_I, C)$ for valid plaintext M_I encrypted into C increases with increasing of N and approaches to normilized entropy of truly random binary string of the length 16. Mutual information $I(M_{II}, C)$ between ciphertext (obtained for plaintext M_I) and plaintext M_{II} is close to 0. It is sufficiently to select some threshold in order to distinguish between valid and invalid plaintexts for given ciphertext already for N ≥ 103.

In Table 5 are presented results of calculation for cross correlation R(C, M) between sequence C and sequences M_I and M_{II} which show that such criteria cannot be used for a breaking of block cipher semantic security. (This is a consequence of course, a presence of nonlinear transforms in algorithm of Heye's block cipher containing into its S-boxes.)

Table 5

Cross Correlation Between Ciphertext C and Plaintexts M_I , M_{II} Against the Plaintext Bit Length N..

N	4×10^4	8×10^4	3×10^5	10^6
$R(M_I , C)$	–0,000680	–0.038	0.011	–0.000977
$R(M_I , C)$	–0.0019	–0,000556	0.003	–0.00016

B. Heye’s cipher with block length 32

We consider block cipher with the same structure as Heye’s cipher but with block length 32 and with round keys consisting from 32 bit each. S-box transforms are shown in Table 1 and permutation mapping is shown in Table 3. Experiment with such “extended cipher” was arranged similarly as for ordinary cipher described in the point A with only differences that two plaintexts M_I and M_{II} have the length 32 bits and the same length has ciphertext C. The results of simulations are presented in Table 6.

Table 6

Mutual Information Between Cipher Text and Plaintext Corresponding and not Given Cipher Text Against the Plaintext Bit Length N..

N	10^3	10^4	2×10^4	4×10^4	8×10^4	3×10^5	10^6
$I(M_I, C)$	–0.065	0.025	0.038	0.078	0.083	0.3626	0.976
$I(M_I, C)$	–0.03	-0.007	0.0025	-0.012	0.0055	0.0014	0.0017

We can see from Table 6, that despite of the fact that mutual information $I(M_I, C)$ grows much slower with increasing of N than similar value for 16-bit block cipher (see Table 4) it is still exceeds the value $I(M_{II}, C)$ where $N \geq 10^4$

This means that after a choice of appropriate threshold it is possible to distinguish “valid” plaintext from “invalid” one for given ciphertext. Thus the proposed approach can break semantic security of at least for block ciphers with limited block length $n \leq 32$.

Our experiments with DES block cipher having block length 64 bits showed that this problem is rather untractable at least with the use of ordinary PC.

4. CONCLUSION

We have proposed a novelty approach to a breaking of block cipher semantic security based on Shannon inequality for mutual information and fast calculation of mutual information using nearest neighbour distance relatively recently proposed by A. Kraskov et.al.

Our simulation results performed on PC confirmed that in fact this problem can be solved at least for block length less than 32. It is possible to increase the last value at the cost more powerful computers or by optimization of neighbour distance approach. We are going to do it in the nearest future. It is worth to noting that the use of the same technique allows to detect a presence of stegosystems [8] if it is known a message that could be embedded into some cover object and extraction algorithm is known also due to Kerchgoff assumption [9].

5. REFERENCES

1. A. Menezes, P.C. van Oorshot and S.A. Vanstone “Handbook of Applied Cryptography”, CRC Press, 1996
2. A more powerful adversary. Security against chosen-plaintext attacks. Computer Science Department, Wellesley College (http://cs.wellesley.edu/~cs310/lectures/07_CPA_slides_handouts.pdf)
3. H.M. Heys “Tutorial on Linear and Differential cryptology”, Technical Report CORR 2001-17, Centre for Applied Cryptographic Research, Department of Combinatorics and Optimization, University of Waterloo, March 2001

4. C.E. Shannon “Communication theory of secrecy system”, Bell System Technical Journal 28, P. 650-715, 1949
5. C. Henck van Tillory “Fundamentals of Cryptography”, Kluwer Acad. Publisher, 2000
6. A. Hyvariach, J. Karpunen and E. Oja, “Independent Component Analysis”, Willey, 2001
7. A. Kraskov, H. Stogbauer and P. Grassberger “Estimating mutual information”, Physical Review, E69, 066138, 2004.
8. J. Fridrich, “Steganography in Digital Media””, Cambridge Publisher, 2010
9. A. Kerchkhoffs, “La cryptographie militaire”, Journal des sciences militaires, vol. IX, pp. 5–38, Jan. 1883, pp. 161–191, Feb. 1883.