Definition of Chipher Key on Plaintexts and Chipher Texts by the Method of Equivalent Keys

Alexander Babash[1][0000-0001-7578-6923], Valery Sizov[0000-0002-4844-4714]

and Andrey Mikrukov[0000-0002-8206-677X]

Plekhanov Russian University of Economics
Moscow, Russian Federation
[1]babash@yandex.ru

Abstract. One of the important tasks related to the implementation of the Digital Economy program is to improve cybersecurity when creating digital platforms, as distributed information and communication systems of the subjects of a single digital market. When building distributed information security subsystems of digital platforms, an urgent task is to increase the cryptographic strength of the cryptographic mechanisms used to encrypt short texts. The paper deals with the problem of encrypting short texts with ciphers with a large number of keys, from which the equivalent keys appear in the cipher, which leads to a significant reduction in the cryptographic strength of ciphers. The concept of weak key equivalence in the C. Shannon cipher model is introduced. Methods for determining the key from the open and encrypted texts with the calculation of the parameters of their complexity are proposed. The methods are applicable to both symmetric ciphers and asymmetric ciphers. The following situations are considered: 1) representatives of classes of equivalent keys are known; 2) the capacities of the classes of equivalent keys and representatives of these classes are known; 3) only the capacities of the classes of equivalent keys are known; 4) the number of classes of equivalent keys is known. A part of encryption devices (encoders) is built using a serial connection of the control unit with an encryption unit, where the actual control unit performs the role of a pseudo-random number generator. The keys of such an encoder are the keys of the pseudo-random number generator. The output sequence of the pseudo-random number generator is the control sequence of the encryption unit. Often the encryption unit uses the gamming cipher. In this case, the equivalence of the keys of such an encoder is equivalent to the equivalence of the keys of the pseudo-random number generator. The results obtained below allow us to apply the method of equivalent keys developed in the article to ciphers that have equivalent keys in a pseudo-random number generator.

Keywords: Equivalent Keys, Cipher Encoding Algebra, Plaintext, Ciphertext.

1 Introduction

For many ciphers, decoding methods other than the “brute force” technique, sometimes called the Monte Carlo method, or the total method, the key testing method [1-3], have not yet been found. At the same time, for many of them, the absence of

equivalent keys has not been proven [4–11]. Moreover, when encrypting short texts with ciphers with a large number of keys, the presence of equivalent keys in the cipher follows from quantitative considerations. The concept is introduced in the paper - weak equivalence of keys relative to a given plaintext. The presence of equivalent keys in the cipher or weak equivalent keys leads to the possibility of grouping keys into classes of keys with subsequent testing of representatives of such classes. Such a situation, as a rule, significantly reduces the cryptographic strength of ciphers. This idea lies in the cipher key identification methods developed below. Finally, the solutions of two problems of interest for decoding of cipher are given:

1) what is the largest k, at which the probability that all the keys $\chi_1, \chi_2, ..., \chi_k$ pairwise are not equivalent is less than the given probability P;

2) what is the minimum k for the average number of equivalent key pairs from the set $\chi_1, \chi_2, ..., \chi_k$ greater than 1.

The decoding methods described in the paper are given with estimation of their complexity parameters.

2 Basic Concepts and Notation

Let's denote the cipher encoding algebra by $A = (X, K, Y, f)$. Here: X – a set of plaintexts; K – a number of keys; Y – a set of ciphertexts (cryptograms); f - encoding function $f(x,\chi)=y$, $x \in X$, $\chi \in K$, $y \in Y$.

Definition 1. Keys $\chi,\chi` $ are called equivalent if $f(x,\chi)=f(x,\chi`)$ for any $x \in X$.

Definition 2. The keys $\chi,\chi` \in K$ are called equivalent with respect to the subset $X` \subseteq X$ if $f(x,\chi)=f(x,\chi`)$ for any $x \in X$.

The binary relation $\sigma(X`)$ introduced in this definition for a set of keys $K$ is a binary equivalence relation (the properties of reflexivity, symmetry and transitivity are fulfilled). Therefore, the entire set of keys $K$ is divided into equivalence classes of the binary relation $\sigma(X`)$. We denote this partition by $R(\sigma(X`))= \bigcup_{j=1}^{L(X`)} K_j^{X`}$.

It is obvious that the equivalence of keys $\chi,\chi` \in K$ with respect to $X`$ implies also their equivalence with respect to any $X``$ subset of a $X`$ set. It results that any equivalence class $K_j^{X`}$ is contained entirely in a certain equivalence class $K_{j`}^{X``}$ with respect to a subset $X``$ of the set $X`$. Each class $K_{j`}^{X``}$ consists of the combination of some classes $K_j^{X`}$. In particular, $L(X``) \le L(X`)$, and classes $K_j^{X`}$ are "smaller" than classes $K_{j`}^{X``}$.

3 Formulation of the Problem

Find solutions of the equation $f(x,\chi)=y$ with respect to $\chi \in K$, i.e. the problem of determining the key χ by agiven plaintext x and a cipher text y. In the terminology adopted above, this task consists in finding the key up to *equivalence with respect to* a set consisting of a single element x.

Let's denote by $K_1, K_2, \ldots, K_L$ equivalence classes with respect to the element $x \in X$. Further, for brevity, we will call these classes simply the equivalence classes of keys, although their more meaningful name, in our opinion, would be "weak equivalence classes of keys."

4 Problem Solutions with Various Additional Assumptions

1. The representatives $\chi_1, \chi_2, \ldots, \chi_L$ of classes $K_1, K_2, \ldots, K_L$ of equivalent keys are known.

In this case, testing is carried out without the return of representatives until the first success (until receiving a representative of the equivalence class in which the key is located). That is, $f(x,\chi)=y^{\chi}$ is estimated for each test key χ, and y^{χ} is compared with the given y. The testing process ends when the equality $y^{\chi}=y$ is obtained. The performance of T in testing such a method coincides with the performance of the total method with $r=|K|=L$ and zero errors of the statistical criterion:

$$T = \frac{L+1}{2}$$

The reliability method is $\pi=1$.

2. Capacities of classes of equivalent keys and representatives of these classes are known.

Let's arrange $\chi_{i(1)}, \chi_{i(2)}, \ldots, \chi_{i(L)}$ the representatives known to us in accordance with the capacities of the classes of equivalent keys:

$$|K_{i(1)}| \geq |K_{i(2)}| \geq \ldots \geq |K_{i(L)}|$$

and try them out according to this order $\chi_{i(1)}, \chi_{i(2)}, \ldots, \chi_{i(L)}, r \leq L$. The algorithm stops its operation if the key sought is found (up to equivalence) or r tests are performed.

If the cipher key was chosen randomly and equiprobably from K, then the probability of choosing a key from the class Kj is equal to $\frac{|K_j|}{|K|}$. Therefore, the average number T_r tested in the implementation of the key algorithm is

$$T_r = \sum_{j=1}^{r-1} j \frac{|K_j|}{|K|} + r \sum_{j=r}^{L} \frac{|K_j|}{|K|},$$

and the reliability of the method is

$$\pi = \sum_{j=1}^{r} \frac{|K_j|}{|K|}.$$

When $r = L$ we have

$$T_L = \sum_{j=1}^{L} j \frac{|K_j|}{|K|}, \pi = 1.$$

3. Only the capacities $|K_1|, |K_2|, \ldots, |K_L|$ of the equivalent key classes are known.

Let's conduct testing without returning the $\chi \in K$ keys until a true key is obtained, up to equivalence.

If the cipher key was chosen randomly and equiprobably from K, then the probability of choosing the key χ from the class K_j is equal to $\frac{|K_j|}{|K|}$. Let's denote by $T(j)$ the average number of tests of the algorithm, provided that the key sought is $\chi \in K_j$. Then

$$T(j) = \frac{|K|+1}{|K_j|+1}$$

and the total average number of algorithm tests is

$$T = \sum_{j=1}^{L} T(j) \frac{|K_j|}{|K|} = \frac{|K|+1}{|K|} \sum_{j=1}^{L} \frac{|K_j|}{|K_j|+1}.$$

The reliability method is $\pi = 1$.

Let's note that if in this method testing is carried out with return, then the average number of tests of the algorithm will be equal to

$$\sum_{j=1}^{L} \frac{|K|}{|K_j|} \frac{|K_j|}{|K|} = L.$$

Consequently, under the conditions of the third problem, always $T < L$.

Obviously,

$$1 \le T \le \frac{K+1}{2},$$

in this case, the lower bound is attained at $L = 1$, $|Kj| = |K|$, and the upper one at $L = |K|$, $|Kj| = 1$.

If the estimated capacities ratings of equivalent key classes are

$$c_j \le |K_j| \le C_j, \; j \in \overline{1, L},$$

then

$$\frac{|K|+1}{|K|} \sum_{j=1}^{L} \frac{c_j}{c_j+1} \le T \le \frac{|K|+1}{|K|} \sum_{j=1}^{L} \frac{C_j}{C_j+1}.$$

4. The number L of classes of equivalent keys is known. Carrying out the method of paragraph 3, for the complexity of the method, we obtain the estimate $T < L$.

5 Discussion

Let's draw attention to the fact that the methods outlined were based on the equivalence of keys with respect to a given $x \in X$ ("weak equivalence of keys"). Usually, the exact estimation of capacities of such equivalence classes is difficult, and therefore, the equivalence of keys with respect to the whole set X is used. In this case, it is easy to obtain lower bounds for the capacities of the classes of weak equivalences we used. The direct use of the capacities of the classes of equivalent keys with respect to the whole set X in methods 1–4 allows to estimate upper bounds of performances of the above methods of "weak equivalences".

The second circumstance to which attention should be paid is that in a number of cases other key equivalences can be used in a similar way. For example, using the mode of generating one-time keys with the help of a markant (special cipher mode for obtaining one-time keys from a long-term key).

6 Methods of using equivalent keys and the birthday paradox

Let the number L of classes of equivalent keys of the used cipher be known, all the classes having equally capacities and the number $\boldsymbol{k}$ of ciphered texts being used randomly and equally probably selected keys $\chi_1, \chi_2, ..., \chi_k$.

In various tasks of cryptographic practice, the solution of the following problems is of interest.

1. what is the largest k, at which the probability that all the keys $\chi_1, \chi_2, ..., \chi_k$ pairs are not equivalent is less than the given probability P.
2. what is the minimum k for the average number of equivalent key pairs from the set $\chi_1, \chi_2, ..., \chi_k$ greater than 1.

The birthday paradox [11] is connected with the answer to the question: how many people should be in the room so that with high probability there are two born on the same day? The paradox is that the answer is significantly less than the number of days in a year, which seems implausible. So, we consider that keys are people, and the number L of classes of equivalent keys is the number of possible dates of birth. We believe that in the year 365 = L days and that the birthdays of k people are chosen randomly and independently from each other.

We first estimate the probability that all the birthdays of the selected k people ($k \le L$) will be different. Let the birthday of the first is already chosen. Then the birthday of the second coincides with it with a probability of $1/L$. With selected (and different) birthdays of the first and second person, the probability that the third birthday will coincide with one of the existing ones will have $2/L$, and so on. As a result, the probability P_k that k people will have different birthdays has $P_k=(1-1/L)(1-2/L)...(1-(k-1/L))$.

The factors P_k can be increased using a known inequality $1+x \le e^x$:

$$P_k \le e^{-1/L} e^{-2/L} ... e^{-(k-1)/L} = e^{-(1+2+3+...+(k-1))/L} = e^{-k(k-1)/2L}.$$

With increasing k, the probability P_k decreases. For which k is this probability strictly less than a given P? Let's solve inequality

$$e^{-k(k-1)/2L} < P.$$

We have

$$-k(k-1)/2L < \ln P, -k^2+k < 2L\ln P, k^2-k > 2L\ln(1/P),$$
$$k^2-k+1/4 > 2L\ln(1/P)+1/4,$$
$$(k-(1/2))^2 > 2L\ln(1/P)+1/4.$$

Find k_0 in which $(k_0-(1/2))^2=2L\ln(1/P)+1/4$. We have

$$k_0-(1/2)=(2L\ln(1/P)+1/4)^{(1/2)}.$$

Whence,

$$k_0=(1/2)(1+(1+8L\ln(1/P))^{1/2}.$$

In this connection, when k is smaller than k_0, the probability P_k is less than the given P. Therefore, when k is not less than k_0, the probability P_k is not less than the given P. Assuming, for example, L = 365 (669) is the number of different birthdays, P = 0,5,

we find that for $k \geq 23$ ($k \geq 31$) the probability that among k people there will be two born on one day no less than $P = 0.5$. In other words, if the cipher has 365 (669) classes of equivalent keys and there is a set of $k \geq 23$ ($k \geq 31$) cipher telegrams, then among them with a probability of at least 0.5 there will be a pair of cipher telegrams encrypted on equivalent keys.

Let us turn to the solution of the second task. At what minimum k the average number of equivalent key pairs of $\chi_1, \chi_2, \ldots, \chi_k$ is greater than 1. For each key pair (i, j) from the set $\{\chi_1, \chi_2, \ldots, \chi_k\}$, let's consider the random variable X_{ij}

$X_{ij}=1$, if χ_i and χ_j are equivalent, otherwise $X_{ij}=0$.

Since the classes of equivalent keys of the used cipher have equal capacities and the keys $\chi_1, \chi_2, \ldots, \chi_k$ are chosen randomly and equiprobably, the probability of equivalence of any key pair is $1/L$. Therefore, the mathematical expectation $M(X_{ij})$ of the random value X_{ij} ($i \neq j$) is calculated by the formula

$$M(X_{ij}) = 1 \cdot 1/L + 0 \cdot (1 - 1/L) = 1/L$$

The random value Y equal to the sum of all X_{ij} (in all $C_k^2 = \frac{k(k-1)}{2}$) has a mathematical expectation equal to the sum of all $M(X_{ij})$

$$M(Y) = \frac{1}{L} \cdot \frac{k(k-1)}{2}.$$

Let's find the value k_0 from equality

$$\frac{1}{L} \cdot \frac{k_o(k_0 - 1)}{2} = 1.$$

This value is $\frac{1+\sqrt{1+8L}}{2} \approx \sqrt{2L}$. Consequently, when $k \geq k_0$, the average value of pairs of equivalent keys will be no less than 1. So if L=365, (669), then with $k \geq 28$ ($k \geq 38$) the expected number of pairs of equivalent keys is not less than $(28 \cdot 27)/2 \cdot 365 = 1.0356$.

7 Conclusion

1. The concept of weak key equivalence in the C. Shannon cipher model is introduced. A method is proposed for decrypting both symmetric and asymmetric ciphers using the weak key equivalence parameters and calculating performances and reliabilities.
2. The proposed method can be used to determine the initial states of pseudo-random generators from known input and output sequences.
3. Due to the lack of proof of the absence of weakly equivalent keys, many ciphers have a successful chance of practical application of the above described method of decoding.

References

1. Panasenko S.P. Encryption algorithms. Special reference book BHV-Petersburg, 2009.
2. 2Schneier Bruce, Ferguson Nils. Practical cryptography, (2005).
3. Schneier B. Applied Cryptography. Protocols, Algorithms, and Source Code in 2nd ed. N.Y.: Wiley, (1996).
4. GOST 28147-89. Information processing systems. Cryptographic protection. Algorithm of cryptographic transformation. M.: State Standard of the USSR, (1989).
5. Adams C. RFC 2144: The CAST - 128 Encryption Algorithm // Entrust Technologies, (May 1997).
6. Adams C., Gilchrist J / RFC 2612: The CAST-256 Encryption Algorithm // Entrust Technologies, June (1999).
7. Advanced Encryption Standart (AES). Questions and Answers // http // csrc.nist.gov - January 28, (2002).
8. AES Round 1 Inforation //http:csrc.nist.gov - January 26,(2001).
9. Biham E., Dunkelman O., Ketller N. Rectangle Attacks on the 49th Round SHACAL-1 //http://vipe.technion,ac.il - Technion, Haifa, Israel.
10. Biham E., Biryukov A., Dunkelman A., RichardsonE., Shamir A. Observations on Skipjack: cryptanalysis of Skipjack-3XOR // http://www.cs.technion.ac.il - Technion –(1998).
11. T. Cormen, C. Lazerson, R. Rivest. Algorithms and analysis. M., MCNMO, (2002).