The simple ideal cipher system

Boris Ryabko

February 19, 2001

[1] Prof. and Head of Department of appl. math and cybernetics

Siberian State University of Telecommunication and Computer Science

Head of Laboratory of data protection

Institute of Computational Technology Siberian Branch of Russian Academy of Science

Address : Siberian State University of Telecommunication and Computer Science

Kirov str.86

Novosibirsk, 630102, Russia

Tel: 007 383 2284938

Fax: 007 383 2669343

e-mail: ryabko @ neic.nsk.su

URL: http://www.ict.nsc.ru/ ryabko/

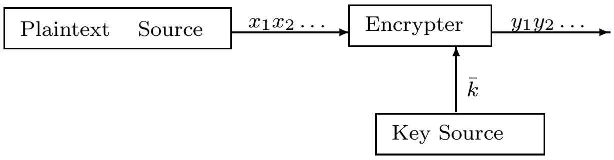
Summary. We address the problem of how to construct ideal cipher systems when the length of a key is much less than the length of an encrypted message. We suggest a new secret key cipher system in which firstly the message is transformed into two parts in such a way that the biggest part consists of independent and equiprobable letters. Secondly the relatively small second part is enciphered wholly by the Vernam cipher whereas only few bits from the biggest part are enciphered. This transformation is based on the fast version of the Elias construction of an unbiased random sequence.

The time required for encoding and decoding and the memory size of the encoder and decoder are presented as functions of the ratio of the key length and the message length. The suggested scheme can be applied to sources with unknown statistics.

Keywords: ideal cipher system, fast algorithms, Shannon entropy.

1 Introduction

We consider a common definition of a secret key cipher system as diagrammed below:



As always we assume that the secret key is statistically independent on the plaintext sequence and the letters of the key are equiprobable. For the sake of simplification we also assume that the plaintext letters, key letters and ciphertext letters are generated by Bernoulli sources and take values in the alphabet $A = \{0, 1\}$, but the suggested method is easily generalized for the case of any finite source alphabet and for Markov sources in such a way as it has been described by M.Blume [2].

In the pioneering paper [6] C.Shannon has shown that there exist systems with so-called perfect secrecy. Informally, it means that a cryptanalyst who knows an encrypted message, obtains no further information to enable him to decide which message was transmitted. Clearly, perfect secrecy is highly desirable but it is shown by Shannon that, roughly speaking, the length of a key sequence has to be equal to the length of an encrypted message for systems with perfect secrecy. Frequently the length of the key should be much less than the length of encrypted messages. In this case it is impossible to construct a system with perfect secrecy [6] but it is possible to construct so-called ideal cipher systems [6]. In such systems the uncertainty of the key makes it impossible for the cryptanalyst, among a subset of the messages, to decide exactly which message is the actual one sent.

Using this notation we can say that a simply realisable ideal system is suggested for the case when the length of a key is much less than the length of an encrypted message. It should be noted that the complexity of methods will be assessed by the memory size (S) (in bits) required to store the programs of an encoder and a decoder, as well as by the average time (T) required for encoding and decoding of a single letter measured by the number of binary operations on single- bit words when they are implemented on a computer with random access memory (it is a model of a common computer; see the definition in [1]).

Let us give some new definitions. Let A^n and A^∞ be the sets of all finite words with the length n ($n \geq 1$) and one-side-infinite words, respectively, in the alphabet A, and let $A^* = \bigcup_{n=1}^{\infty} A^n$. Let there be a plaintext source which generates letters from a finite alphabet A and a key source which generates independent and equiprobable letters from the alphabet $\{0,1\}$. A cipher α is defined as a pair of such functions α^{en} and α^{de} that α^{en} assigns to each pair of words $(x, \bar{k})$, $x \in A^\infty$, $\bar{k} \in \{0,1\}^\infty$, a sequence $y \in \{0,1\}^\infty$ in such a way that $\alpha^{de}(y, \bar{k}) = x$, where x is a plaintext , $\bar{k}$ is a sequence of letters of the key and y is an encrypted message.

We address the problem when the length of the key is (much) less than the length of an encrypted message. The formal model can be described as follows: there is a (small) number $\gamma \in (0,1)$ and, when an encoder encrypts n first letters of the plaintext, it must use not more than γn letters of the key sequence $\bar{k}, n > 0$.

We consider a limiting entropy

$$H(x|_1^m / y|_1^\infty) = \lim_{t \to \infty} H(x|_1^m / y|_1^t)$$

where m, t are integers, $U|_a^b = U_a U_{a+1} \dots U_b$ for a word U and $a \leq b$, $H(\cdot/\cdot)$ is the conditional entropy (see the definitions, for ex, in[4]). We will consider the plaintext source with nonzero entropy only. (If the entropy of the plaintext source is equal to zero there is no need to transmit the plaintext.) In that case the Shannon definition may be formulated as follows: the system is ideal if

$$\lim_{m\to\infty}(H(x|_1^m/y|_1^\infty)/m) \geq \pi \tag{1}$$

where π is a positive constant. Informally it means that if somebody knows the (infinite) encrypted text y but does not know x and $\bar{k}$ then his uncertainty about the $x|_1^m$ is approximately equal to πm. In other words, a codebreaker will have around $2^{\pi m}$ possible highly probable variants of deciphering for $x_1, ..., x_m$. Of course, the more π , the better the cipher system is.

It easy to see that π cannot be larger than the plaintext entropy. On the other hand, it is obvious that if γm letters of the key were used for encryption of m first letters of the plaintext, then π cannot be larger than γ,$\gamma < 1$. So we obtain the following inequality:

$$\pi \leq \min\{\gamma, h\},$$

where h is the Shannon entropy of the plainext source. (For example, if the plaintext is generated by Bernoulli source, its entropy is defined by the equality $h = -\sum_{a\in A} p(a)\log p(a)$, where $p(a)$ is the probability of $a, a \in A$. The general difinition may be found in [4].)

Using the given definition we can say that we suggest such a cipher system that π in (1) is not less than $\min\{\gamma/2, h\}$. The time of encryption and decryption of one letter of the plaintext and the memory size of the encoder and decoder are equal to $O(\log^3(1/\gamma)\log\log(1/\gamma))$ bit operations and $O((1/\gamma)\log(1/\gamma))$ bits , correspondingly.

We can see that, if the entropy of the plaintext h is quite large, the suggested system uses only a half key digits ($\gamma/2$ instead of γ). The possibility to modify the suggested system in such a way that the key digits are used more efficiently is considered too. But in this case the complexity of the system increases.

The rest of the paper is organised as follows. The second part contains a description of the system as well as the main properties of the system are described in the third part.

2 Description of the cipher system

For a given $\gamma \in (0,1)$ let n be a minimal integer such that the inequality

$$\lceil 2\log(n+1)\rceil/n \leq \gamma/2 \tag{2}$$

is valid. (Here and below $\log n = \log_2 n$). It is easily to see that

$$n < 4\log(1/\gamma)/\gamma + O(1), \tag{3}$$

when $\gamma \to 0$.

The description of the suggested cipher system may be divided into two parts as follows: firstly, a generated sequence of letters is transformed into two subsequences and, secondly, both subsequences are encrypted by different methods.

The first part plays a key role. It is based on the method of P.Elias [3] and the fast algorithm of enumeration from [5].

Let us give some new definitions in order to describe the method of transformation. Let S_n^i be the set of all binary words of the length n with i ones, $(n \geq i \geq 0)$ and let for every $x \in S_n^i$ $code(x)$ be lexicographical number of the word x in the set S_n^i which is written in the binary number system, the length of $code(x)$ equals $\lceil \log(\binom{n}{i})\rceil$ bits. (Here and below $|x|$ is the length of x if x is a word and the number of elements if x is a set.) For example, $S_4^2 = \{0011, 0101, 0110, 1001, 1010, 1100\}$ and $code(0011) = 000$, $code(1100) = 101$.

A generated plaintext can be written in the form of a sequence of blocks of the length n, where n is defined above, see(2). Every block $\bar{x}$ is encoded by the sequence of three words $u(\bar{x})v(\bar{x})w(\bar{x})$. Here $u(\bar{x})$ is the number of units in the block $\bar{x}$ and the length of $u(\bar{x})$ is equal to $\lceil \log(n+1)\rceil$ bits. In order to describe $v(\bar{x})$ and $w(\bar{x})$ we define

$$m_k = \lfloor \log(\binom{n}{k})\rfloor (= \log |S_n^k|)$$

where k is the number of units in $\bar{x}$. Let $\alpha_{m_k}\alpha_{m_1-1}\ldots\alpha_0$ be a binary notation of $|S_n^k|$ and $\alpha_{m_k} = 1, \alpha_{j_1} = 1, \ldots, \alpha_{j_{s-1}} = 1$ as well as the other $\alpha_{j_k} = 0$. (In other words, s is the number of units in the word $\alpha_{m_k}\alpha_{m_k-1}\ldots\alpha_0$.) Let $\beta(\bar{x}) = \beta_{m_k}\beta_{m_k-1}\ldots\beta_0$ be the binary notation of the lexicographical number of $\bar{x}$ and let the following inequalities be valid:

$$\begin{aligned} &\alpha_{m_k}\alpha_{m_k-1}\ldots\alpha_{j_r}000\ldots0 \le \beta(\bar{x}) \\ &< \alpha_{m_k}\alpha_{m_k-1}\ldots\alpha_{j_{r+1}}00\ldots0 \end{aligned} \tag{4}$$

for a certain r. (Obviously such r exists.) The word $w(\bar{x})$ is defined as follows

$$w(\bar{x}) = \begin{cases} \beta_{j_r-1}\beta_{j_r-2}\ldots\beta_0, \text{if } j_r - 1 \ge 0 \\ \Lambda, \qquad\qquad\qquad \text{if } j_r - 1 < 0 \end{cases} \tag{5}$$

where Λ is an empty word and $j_0 = m_k$ by definition. Now we can describe the word $v(\bar{x})$. This word contains the binary notations of the integer r for which (4) is valid. By definition, r belongs to the set $\{j_0, j_1, \ldots, j_{s-1}\}$. (Let us recall that it is the set of such indexes that $\alpha_{j_k} = 1$ for each j_k, where $\alpha_{m_k}\alpha_{m_k-1}\ldots\alpha_0$ is the binary notation of the word

m_k). So, it is enough to keep $\lceil \log s \rceil$ binary digits in order to encode and decode each possible r and , by definition, the length of the word $v(\bar{x})$ is equal to $\lceil \log s \rceil$ bits. Let us note that

$$|v(\bar{x})| < \lceil \log(n+1) \rceil \qquad (6)$$

because by definition $|v(\bar{x})| \leq \lceil \log s \rceil = \lceil \log(m_k + 1) \rceil$, $m_k = \lfloor \log(\begin{matrix} n \\ k \end{matrix}) \rfloor$ and $(\begin{matrix} n \\ k \end{matrix}) < 2^n$ for every $k = 0, 1, \ldots, n$.

First, let us explain that it is possible to find $\bar{x}$ if $u(\bar{x})v(\bar{x})w(\bar{x})$ are known. Indeed, the first $\lceil \log(n+1) \rceil$ bits give a possibility to find a number of units in an encoded block. It gives a possibility to find the number $|S_n^k| = \alpha_{m_k}\alpha_{m_k-1}\ldots\alpha_0$ and therefore, the numbers $m_k = \log |S_n^i|$, s (the number of units in the binary notations of m_k) and $\lceil \log s \rceil$. The next $\lceil \log s \rceil$ bits contains information about the length of the word $w(\cdot)$. After that it is possible to find digits $\beta_{j_r-1}\beta_{j_r-2}\ldots\beta_0$. The word $\alpha_{m_k}\ldots\alpha_{j_r}\beta_{j_r-1}\beta_{j_r-2}\ldots\beta_0$ is the code of the lexicographical number of the encoded block $\bar{x}$.

Let us consider an example. Let the block length $n = 4$ and the word x be 0101. Obviously, $\bar{x}$ belongs to the set $S_4^2 = \{0011, 0101, 0110, 1001, 1010, 1100\}$ and its lexicographical number $\beta(\bar{x}) = 001$. It is obvious that $m_k = \lfloor \log |S_n^k| \rfloor = 2$ and $|S_n^k| = |S_4^2| = (110)_2$, $s = 2$. According to (1) we check the inequalitie

$$000 \leq 001 \ < 100$$

Hence, $j_r = 2$, $r = 0$ and we can find $u(0101) = 010, v(0101) = 0, w(0101) = \beta_1\beta_0 = 01$,see (2). In order to make clear the main goal of the suggested transformation we consider codes for all words from the set S_4^2.

$$uvw(0011) = 010\ 0\ 00$$

$$uvw(0101) = 010\ 0\ 01$$

$$uvw(0110) = 010\ 0\ 10$$

$$uvw(1001) = 010\ 0\ 11$$

$$uvw(1010) = 010\ 1\ 0$$

$$uvw(1100) = 010\ 1\ 1$$

The main property of the considered transformation may be formulated as follows: the symbols of $w(\bar{x})$ are independent and equiprobable. It is easily seen from this example. Indeed, we can see that all the words from S_4^2 are equiprobable and there are four words with w parts $00, 01, 10, 11$. So, in this set the letters $0, 1$ are equiprobable and independent. Similarly, there are two aquiprobable words with $w = 0$ and $w = 1$ and, obviously, in this set the letters $0, 1$ are eqiprobable and independent. The same property is fulfilled for each S_n^k. It is important to note that the average length of w grows as nh, whereas the length of u and v grows as $O(\log n)$ if n increases. The formal proof of both properties is given in [3].

It is easy to see that the most difficult part of encoding is the calculation of the lexicographical number $\beta(\bar{x})$ of $\bar{x}$ in the S_n^k. Respectively, the most difficult part of decoding is the calculation of $\bar{x}$ on the basis of its lexicographical number $\beta(\bar{x})$.

Let us describe the second part of the suggested method. Briefly, this part may be described as follows: at least $\lfloor \gamma n \rfloor$ first letters of the sequence $u(.)v(.)w(.)$ are encrypted by the Vernam cipher, whereas all other letters of this sequences are not changed. Hence, the words $u(.)v(.)$ and at least $\lfloor \gamma n/2 \rfloor$ first letters of the word $w(.)$ will be encrypted. That is why a codebreaker will have at least $2^{\lfloor \gamma n/2 \rfloor}$ equiprobable alternate versions of deciphering an initial block.

The suggested scheme is a little bit more complicated because it is possible that the length of the sequence $u(.)v(.)w(.)$ is shorter than $\lfloor \gamma n \rfloor$. In this case it is natural to use extra bits of the key for the encryption of the following words.

In order to give a formal description of this part of the method we define an auxiliary value $R(m)$ as the number of digits of the key which were used for encryption of the first m blocks

$$u(1)v(1)w(1)u(2)v(2)w(2)\dots u(m)v(m)w(m), m > 1.$$

It is obvious that $R(m) \leq \lfloor m\gamma n \rfloor$ bits. It is convenient to denote

$$u(i)v(i)w(i) = L^i$$

and let $L^i = L^i_1 L^i_2 \dots L^i_K$ where $K = |L^i|$.

The letters of each L^i are encrypted according to following equality

$$encrypted(L^i_j) = \begin{cases} L^i_j \bigoplus \bar{k}_{R(i-1)+j}, \text{if } \ R(i-1)+j \leq \lfloor i\gamma n \rfloor \\ L^i_j, \qquad\qquad\quad \text{if } \ R(i-1)+j > \lfloor i\gamma n \rfloor \end{cases} \tag{7}$$

where $\bar{k} = \bar{k}_1 \bar{k}_2 \bar{k}_3 \dots$ is the key and $a \bigoplus b = a + b (\text{mod} 2)$.

3 The properties of the cipher system

The suggested transformation of the original plaintext plays the key role. That is why we describe the properties of the transformation first.

Theorem 1 *Let a Bernoulli source generate letters from the alphabet* $A = \{0, 1\}$ *with (unknown) probabilities* p *and* q*, respectively. Let a generated sequence be transformed as it is described above and* $n, n \geq 2$*, be a parameter of the transformation. Then the following holds:*

i) The symbols of the sequence $w(x|_1^n)w(x|_{n+1}^{2n})w(x|_{2n+1}^{3n})\ldots$ *are independent and equiprobable.*

ii) $E(w(x|_{rn+1}^{(r+1)n})) > nh - 2\log(n+1)$ *where* $h = -(p \log p + q \log q)$ *is the entropy of the source,* $E(\cdot)$ *is an expectation.*

iii) the transformation requires the memory size $O(n \log n)$ *bits and has the time of encoding and decoding* $O(\log^3 n \log\log n)$ *bit operations per letter as* $n \to \infty$*.*

Proof. We will omit the proof of i) because it is given in P.Elias paper [3] and the main idea is quite obvious.

In order to prove ii) we first note that every block $\bar{x}$ may be decoded using three encoded words $u(\bar{x})v(\bar{x})w(\bar{x})$. Due to the classical Shannon theorem we know that the length of encoded words is more than or equal to the entropy:

$$E(|u(\bar{x})|) + E(|v(\bar{x})|) + E(|w(\bar{x})|) \geq nh$$

By definition $|u(\bar{x})| = \lceil \log(n+1) \rceil$ bits and $|v(\bar{x})| < \lceil \log(n+1) \rceil$, see (6). From the last three inequalities we obtain the claim ii).

Let us estimate the complexity of the method. In order to find $u(\bar{x})$ it is enough to look through the word $\bar{x}$ and calculate the number of units. It takes $O(n \log n)$ bit operations or $O(\log n)$ bit operations per letter. In order to find $v(\bar{x})$ and $w(\bar{x})$ it is necessary to calculate $\binom{n}{i}$ and the lexicographical number of the word $\bar{x}$ in S_n^i where i is the number of units. We suggest to use the method from [5] for this purpose. The time of calculation and the memory size are equal to $O(\log^3 n \log\log n)$ bit operations per letter and $O(n \log n)$ bits, respectively. Using estimations from [5] it is easy to obtain the same estimation for the complexity of decoding. Theorem 1 is proved.

The next theorem shows that the suggested cipher is ideal.

Theorem 2 *Let a Bernoulli source generate letters from the alphabet $\{0, 1\}$ with probabilities p and q, respectively, $p > 0, q > 0$, and let the suggested cipher system be used for encrypting the source messages and $\gamma \in (0, 1)$ be a parameter of the system. Then*

$$i) \lim_{m\to\infty} m^{-1} H(x|_1^m / y|_1^\infty) \geq \min\{\gamma/2, h\},$$

$$ii) S \leq const \log(1/\gamma)/\gamma$$

$$iii) T \leq const \log^3(1/\gamma) \log\log(1/\gamma),$$

where T and S are the time of calculation and the memory size, respectively.

Remark 1 From i) we can see that, roughly speaking, only half of the digits of the key is used, when the entropy h is quite large. It is possible to modify the suggested system in such a way that the key digits will be used more efficiently, but the complexity of the system will increase. Namely, for every $\varphi \in (0, 1)$ it is possible to construct a cipher system for which the following inequality is valid:

$$\lim_{m\to\infty} m^{-1} H(x|_1^m / y|_1^\infty) \geq \min\{\gamma\varphi, h\}. \tag{8}$$

If φ is more than $1/2$ then the modified system uses the key more efficiently. In order to describe the modified system it is enough to change the parameter n whereas all other parts are not changed. Namely, the parameter n is defined as such a minimal integer that the inequality

$$\lceil 2\log(n+1)\rceil / n \leq \gamma(1-\varphi)$$

is valid (instead of inequality (2)). As before, a plaintext is divided into blocks of the length n and $\lfloor \gamma n \rfloor$ first letters of each blocks are encrypted by the Vernam cipher. In this case, at least $n\gamma\varphi$ equiprobable letters from each block will be encrypted that results in (8). But the compexity increases as

$$S = const \log(1/(\gamma(1-\varphi))/(\gamma(1-\varphi)), T = const \log^3(1/(\gamma(1-\varphi))) \log\log(1/(\gamma(1-\varphi)))$$

that is bigger than ii) and iii) when $(1-\varphi)$ is small.

Outline of Proof. We divide the proof into two deferent parts which correspond to the following cases: i) with probability one all (might be except a finite number) letters of the transformed plaintext are encrypted , ii) with probability one infinite number of letters of the transformed plaintext are not encrypted.

In the first case the uncertainty is equal to the entropy source h because asymptotically all letters of the transformed plaintext are encrypted by the Vernam cipher. In the second case the average number of encrypted bits is γn a block. It means that at least $\gamma n/2$ equirobable bits are encrypted by the Vernam cipher. Hence, at least $2^{\gamma n/2}$ equiprobable variants of deciphering are possible (when the key is unknown) and we obtain the first claim of the theorem.

The two other claims of the theorem are obtained from the definition of the value n and the theorem 1 by direct calculation.

References

[1] *Aho A.V.,Hopcroft J.E.,Ullman J.D.* The design and Analysis of Computer Algorithms//Reading, MA: Addison- Wesley, 1976.

[2] *Blum M.* Independent Unbiased Coin Flips From a Correlated Biased Source: a Finite State Markov Chain // IEEE Symposium on Foundations of Computer Science, 1984, pp. 425-433.

[3] *Elias P.* The Efficient Construction of an Unbiased Random Sequence // The Annals Math. Statist. 1972. V. 43. 3. P. 864-870.

[4] *Gallager R.G.* Information Theory and Reliable Communication // J.Wiley,1968.

[5] *Ryabko B., Machikina E.* Fast and Efficient Construction of an Unbiased Random Sequence //IEEE Trans. Inform. Theory 46 (2000), no. 3, 1090–1093.

[6] *Shannon C.* Communication Theory of Secrecy Systems// Bell System. Techn. J.,28,4, 1949, p.656-715.