ISSN 0032-9460, Problems of Information Transmission, 2015, Vol. 51, No. 1, pp. 82–86. © Pleiades Publishing, Inc., 2015.
Original Russian Text © B.Ya. Ryabko, 2015, published in Problemy Peredachi Informatsii, 2015, Vol. 51, No. 1, pp. 90–95.

SOURCE CODING

The Vernam Cipher Is Robust to Small Deviations from Randomness[1]

B. Ya. Ryabko

Siberian State University of Telecommunications and Information Sciences, Novosibirsk, Russia
Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
e-mail: rbya@yandex.ru

Received April 7, 2014; in final form, November 25, 2014

Abstract—The Vernam cipher, or one-time pad, plays an important role in cryptography because it is perfectly secure. In this cipher a key is a sequence of equiprobable independently generated symbols. We show that under small disturbance of these properties the obtained cipher is close to the Vernam cipher in the case where the enciphered plaintext and the key are generated by stationary ergodic sources.

DOI: 10.1134/S0032946015010093

1. INTRODUCTION

We consider the classical problem of transmitting secret messages from a sender (Alice) to a receiver (Bob) via an open channel which can be accessed by an adversary (Eve). It is assumed that Alice and Bob (but not Eve) know a so-called key, which is a word in a certain alphabet. Before transmitting a message to Bob, Alice encrypts it, and Bob, having received an encrypted message (ciphertext), decrypts it to recover the plaintext.

We consider the so-called running-key ciphers where the plaintext $X_1 \dots X_t$, key sequence $Y_1 \dots Y_t$, and ciphertext $Z_1 \dots Z_t$ belong to the same alphabet $A = \{0, 1, \dots, n-1\}$, where $n \geq 2$. We assume that enciphering and deciphering are given by the rules

$$Z_i = c(X_i, Y_i), \quad i = 1, \dots, t,$$
$$X_i = d(Z_i, Y_i), \quad i = 1, \dots, t,$$

so that $d(e(X_i, Y_i), Y_i) = X_i$. The functions c and d are called the coder and decoder, respectively. Often, the encoder and decoder are defined as

$$Z_i = (X_i + Y_i) \mod n, \qquad X_i = (Z_i - Y_i) \mod n, \tag{1}$$

i.e., $c(X_i, Y_i) = (X_i + Y_i) \bmod n$ and $d(Z_i, Y_i) = (Z_i - Y_i) \bmod n$. In the case of $n = 2$ relation (1) can be represented as

$$Z_i = (X_i \oplus Y_i), \qquad X_i = (Z_i \oplus Y_i), \tag{2}$$

where $a \oplus b = (a + b) \bmod 2$.

A running-key cipher (1) is called a Vernam cipher (or a one-time pad) if symbols of a key sequence are equiprobable and independent, i.e., for any word $k_1 \dots k_t$, $k_i \in A$, we have $P(Y_1 \dots Y_t =$

[1] Supported in part by the Russian Foundation for Basic Research, project no. 12-07-00125.

$k_1 \dots k_t) = n^{-t}$. This cipher plays an important rule in cryptography because it is perfectly secure [1]. This means that

$$P(X_1 \dots X_t) = P(X_1 \dots X_t \,|\, Z_1 \dots Z_t)$$

for all $X_1 \dots X_t \in A^t$ and $Z_1 \dots Z_t \in A^t$; i.e., informally, the uncertainty on the plaintext $X_1 \dots X_t$ does not change if the ciphertext $Z_1 \dots Z_t$ becomes known. Using the Shannon entropy, this property can be represented as

$$h(X_1 \dots X_t) = h(X_1 \dots X_t \,|\, Z_1 \dots Z_t),$$

where $h(X_1 \dots X_t)$ and $h(X_1 \dots X_t \,|\, Z_1 \dots Z_t)$ are the unconditional and conditional entropies of the plaintext $X_1 \dots X_t$.

If a plaintext is generated by a stationary ergodic source, then the perfect secrecy of the Vernam cipher has a simple interpretation explaining the sense of this notion. As follows from the famous Shannon–McMillan–Breiman theorem, all set of messages $X_1 \dots X_t$ for large t can be divided into two parts: $2^{h(X_1 \dots X_t)}$ messages whose probabilities are close in magnitude and in total amount to almost 1, and the set of the other messages with the total probability close to 0. Eve knows that the ciphertext almost surely belongs to the first subset, but all messages in this subset have close probabilities and the number of such messages grows exponentially (as 2^{ht}, where h is the entropy of the source). This is why Eve cannot determine the plaintext.

In the present paper we consider the case where a key sequence is generated by a stationary ergodic source and slightly differs from equiprobable and independent symbols. We show that properties of such a cipher remain to be close to the Vernam cipher in a certain sense. More precisely, we show that in this case the set of messages with close probabilities with the total probability close to 1 grows as $2^{t(h-r)}$ where, as above, h is the source entropy and r is the redundancy of a key source, equal to $\log n - h_{\text{key}}$, where n is the number of symbols in the alphabet and h_{key} is the entropy of the key source. If the redundancy approaches zero, the number of elements in the high-probability set approaches 2^{ht}, i.e., the Vernam cipher. This is what allows us to claim that the Vernam cipher is robust to small deviations. Note that here we informally estimated the level of deviation by the redundancy of a key sequence: the less the latter, the less the deviations. However, below we also give a more formal justification: in this case the redundancy coincides with the Kullback–Leibler distance, widely used in information theory for quantitative estimation of differences between probability distributions.

It is worth mentioning that Shannon in his pioneering paper [1] noted that "from the point of view of the cryptanalyst, a secrecy system is almost identical with a noisy communication system." Therefore, from the mathematical point of view, the problems of message deciphering and filtering of random processes are very close. The approach developed in the present paper is close to the methods used in [2], where the problem of filtering of stationary processes was considered.

2. PRELIMINARIES

We consider the case where a plaintext $X = X_1 X_2 \dots$ and a key sequence $Y_1 Y_2 \dots$ are independently generated by stationary ergodic processes with the same finite alphabet $A = \{0, 1, \dots, n-1\}$, $n \geq 2$, and $Z = Z_1 Z_2 \dots$ is given by (1).

The mth-order Shannon entropy and the limit Shannon entropy are defined as

$$h_m(X) = -\frac{1}{m+1} \sum_{u \in A^{m+1}} P_X(u) \log P_X(u), \qquad h(X) = \lim_{m \to \infty} h_m(X), \tag{3}$$

where $m \geq 0$ and $P_X(u)$ is the probability that $X_1X_2\dots X_{|u|} = u$ (this limit always exists; see [3,4]). Introduce also the conditional Shannon entropy

$$h_m(X \,|\, Z) = h_m(X, Z) - h_m(Z), \qquad h(X \,|\, Z) = \lim_{m\to\infty} h_m(X \,|\, Z). \tag{4}$$

For two probability distributions, P and Q, defined on an alphabet $A = \{a_1, \dots, a_n\}$, $n \geq 2$, the Kullback–Leibler divergence (see [3]) is given by

$$D(P \,\|\, Q) = \sum_{a\in A} P(a) \log(P(a)/Q(a)).$$

For two stationary processes $U = U_1U_2\dots$ and $V = V_1V_2\dots$, the mth-order Kullback–Leibler divergence and the limit divergence are defined by

$$D(U \,\|\, V)_m = -\frac{1}{m} \sum_{w\in A^m} P(U_1\dots U_m = w) \log(P(U_1\dots U_m = w)/P(V_1\dots V_m = w)),$$
$$D(U, V) = \lim_{m\to\infty} D(U \,\|\, V)_m.$$

It is easily seen from these equalities and from (3) that if V generates independent and equiprobable symbols of an n-symbol alphabet A, then

$$D(U \,\|\, V) = \log n - h(U).$$

In what follows we denote this quantity, widely known in information theory and referred to as the redundancy of the process U, by r_U:

$$r_U = \log n - h(U). \tag{5}$$

In what follows we will estimate the "closeness" of processes by the Kullback–Leibler divergence, which in our case coincides with the redundancy.

In information theory, the following theorem is well known [3, 4]

Theorem 1 (Shannon–McMillan–Breiman). *Let $U = U_1U_2U_3\dots$ be a stationary ergodic process. Then $\forall\,\varepsilon > 0$, $\forall\,\delta > 0$, for almost all $U_1U_2U_3\dots$ there exists n' such that for $n > n'$ we have*

$$P\left\{\left|-\frac{1}{n}\log P(U_1\dots U_n) - h(U)\right| < \varepsilon\right\} \geq 1 - \delta, \tag{6}$$

where $P(U_1\dots U_n)$ is the probability of $U_1\dots U_n$.

3. MAIN RESULT

Theorem 2. *Let a plaintext $X = X_1X_2\dots$ and a key $Y = Y_1Y_2\dots$ be independent processes with an alphabet $A = \{0, 1, \dots, n-1\}$, $n \geq 2$, such that the process $(X, Y) = (X_1, Y_1), (X_2, Y_2), \dots$ is stationary and ergodic, and let $Z = Z_1, Z_2, \dots$ be given by (1). Then with probability 1 for any $\varepsilon > 0$ and $\delta > 0$ there exists an integer n' such that for any $t > n'$ and $Z_1^t = Z_1, Z_2, \dots Z_t$ there exists a set $\Psi(Z_1^t)$ of texts of length t for which*

(i) $P(\Psi(Z_1^t)) > 1 - \delta$;

(ii) *For any $X^1 = X_1^1\dots X_t^1$ and $X^2 = X_1^2\dots X_t^2$ in $\Psi(Z_1^t)$, we have*

$$\frac{1}{t}\left|\log P(X^1 \,|\, Z_1^t) - \log P(X^2 \,|\, Z_1^t)\right| < \varepsilon;$$

(iii) $\liminf\limits_{t\to\infty} \frac{1}{t} \log|\Psi(Z_1^t)| \geq h(X \,|\, Z)$.

Proof. Since the process (X, Z) is a deterministic function of the process (X, Y), the ergodicity and stationarity of (X, Y) implies ergodicity and stationarity of (X, Z).

Let us construct the set $\Psi(Z_1^t)$. To this end, we will apply the Shannon–McMillan–Breiman theorem, but first let us note that

$$h(X \,|\, Z) = h(X, Z) - h(Z)$$

and

$$\log P(X_1 \dots X_n \,|\, Z_1 \dots Z_n) = \log P(X_1 \dots X_n; Z_1 \dots Z_n) - \log P(Z_1 \dots Z_n).$$

Applying (6) to $\log P(X_1 \dots X_n; Z_1 \dots Z_n)$ and $h(X, Z)$ and to $\log P(Z_1 \dots Z_n)$ and $h(Z)$, the Shannon–McMillan–Breiman theorem can be written as follows: $\forall\, \varepsilon > 0$, $\forall\, \delta > 0$, for almost all $(X_1, Z_1), (X_2, Z_2), \dots$ there exists n' such that for $n > n'$ we have

$$P\left\{\left|-\frac{1}{n}\log P(X_1 \dots X_n \,|\, Z_1 \dots Z_n) - h(X \,|\, Z)\right| < \varepsilon\right\} \geq 1 - \delta. \tag{7}$$

Hence, for all $\varepsilon > 0$ and $\delta > 0$ and for almost all Z there exists n' such that for $t > n'$ we have

$$P\left\{\left|-\frac{1}{t}\log P(X_1 X_2 \dots X_t \,|\, Z_1 Z_2 \dots Z_t) - h(X \,|\, Z)\right| < \varepsilon/2\right\} \geq 1 - \delta. \tag{8}$$

Define

$$\Psi(Z_1^t) = \{X = X_1 X_2 \dots X_t : |P(X_1 X_2 \dots X_t \,|\, Z_1 Z_2 \dots Z_t) - h(X \,|\, Z)| < \varepsilon/2\}. \tag{9}$$

Property (i) of the theorem immediately follows from (8). To prove (ii), note that for $X^1 = X_1^1 \dots X_t^1$ and $X^2 = X_1^2 \dots X_t^2$ in $\Psi(Z_1^t)$ equations (8) and (9) imply

$$\begin{aligned}\frac{1}{t}&\left|\log P(X^1 \,|\, Z_1^t) - \log P(X^2 \,|\, Z_1^t)\right| \\ &\leq \frac{1}{t}\left|\log P(X^1 \,|\, Z_1^t) - h(X \,|\, Z)\right| + \frac{1}{t}\left|\log P(X^2 \,|\, Z_1^t) - h(X \,|\, Z)\right| < \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

From (9) and property (i) we obtain

$$|\Psi(Z_1^t)| > (1 - \delta)2^{t(h(X|Z)-\varepsilon)}.$$

Taking into account that this holds for all $\varepsilon > 0$, $\delta > 0$, and $t > n'$, we arrive at (iii). $\triangle$

Thus, the set $\Psi(Z_1^t)$ of all possible decipherings grows exponentially, its total probability is close to 1, and probabilities of words within it are close to each other in magnitude.

The above theorem makes it possible to estimate characteristics of a cipher with the help of the conditional entropy $h(X \,|\, Z)$. The following estimates do not require the computation of the conditional entropy but are based on quantities that are easier to compute.

Corollary. *For almost all $Z_1 Z_2 \dots$ we have*

$$\liminf_{t\to\infty} \frac{1}{t}\log|\Psi(Z_1^t)| \geq h(X) + h(Y) - \log n, \tag{10}$$

$$h(X \,|\, Z) \geq h(Y) + h(X) - \log n. \tag{11}$$

Proof. From the well-known representation $h(X, Z) = h(X) + h(Z \,|\, X)$ (see [3, 4]) we obtain

$$h(X \,|\, Z) = h(X, Z) - h(Z) = h(Z \,|\, X) + h(X) - h(Z).$$

Taking into account that $\max h(Z) = \log n$ (see [3, 4]), where n is the number of symbols in the alphabet, from the last equality we obtain

$$h(X \,|\, Z) \geq h(Z \,|\, X) + h(X) - \log n.$$

Since X and Y are independent, from the definition (1) of the running-key cipher it is seen that $h(Z \,|\, X) = h(Y)$. From this and the preceding inequality we obtain (11). Property (iii) of Theorem 2 and inequality (11) imply (10). $\triangle$

Remark. Using the notion of redundancy introduced above, one can express the growth rate of the set $\frac{1}{t}\log|\Psi(Z_1^t)|$ as follows:

$$\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t}\log|\Psi(Z_1^t)| \geq h(X) - r_Y,\\
&\liminf_{t\to\infty} \frac{1}{t}\log|\Psi(Z_1^t)| \geq h(Y) - r_X,\\
&\liminf_{t\to\infty} \frac{1}{t}\log|\Psi(Z_1^t)| \geq \log n - (r_X + r_Y),
\end{aligned} \tag{12}$$

where $r_Y = \log n - h(Y)$ and $r_X = \log n - h(X)$ are the redundancies. Furthermore, from the definition (5) of the redundancy and from inequality (11) we obtain

$$h(X \,|\, Z) \geq h(Y) - r_X.$$

These inequalities allow us to quantitatively estimate the influence of redundancy on the security of a cipher and justify the fact well known in cryptography and information theory: reducing the redundancy improves the security of ciphers.

Now let us return to the question on the influence of deviations from randomness for the Vernam cipher. Let a plaintext be $X_1X_2\ldots$, $X_i \in \{0,1\}$, and let a key sequence $Y_1Y_2\ldots$, $Y_i \in \{0,1\}$, be generated by a source different from the Bernoulli source with $P(0) = P(1) = 0.5$. (For instance, a key $Y_1Y_2\ldots$ is generated by a Bernoulli source with symbol probabilities $P(0) = 0.5 - \tau$ and $P(1) = 0.5 + \tau$, where τ is a small number.) One can see from (12) that the cardinality of the set $\Psi(Z_1^t)$ of high-probability decipherings grows exponentially with exponent not less than $h(X) - r_Y$, where $r_Y = 1 - h(Y)$ is the redundancy. It is seen that if r_Y tends to 0, then the cardinality of the set of high-probability decipherings tends to the cardinality of the corresponding set in the Vernam cipher. Indeed, $h(Y) = 1$, and therefore $r_Y = 0$, as in the Vernam cipher. Informally, we may say that the one-time pad is robust to small deviations from randomness.

REFERENCES

1. Shannon, C.E., Communication Theory of Secrecy Systems, *Bell Syst. Tech. J.*, 1949, vol. 28, no. 4, pp. 656–715.

2. Ryabko, B. and Ryabko, D., Confidence Sets in Time-Series Filtering, in *Proc. 2011 IEEE Int. Sympos. on Information Theory (ISIT'2011), St. Petersburg, Russia, July 31 – Aug. 5, 2011*, pp. 2509–2511.

3. Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, Hoboken, NJ: Wiley, 2006, 2nd ed.

4. Gallager, R.G., *Information Theory and Reliable Communication*, New York: Wiley, 1968. Translated under the title *Teoriya informatsii i nadezhnaya svyaz'*, Moscow: Sov. Radio, 1974.