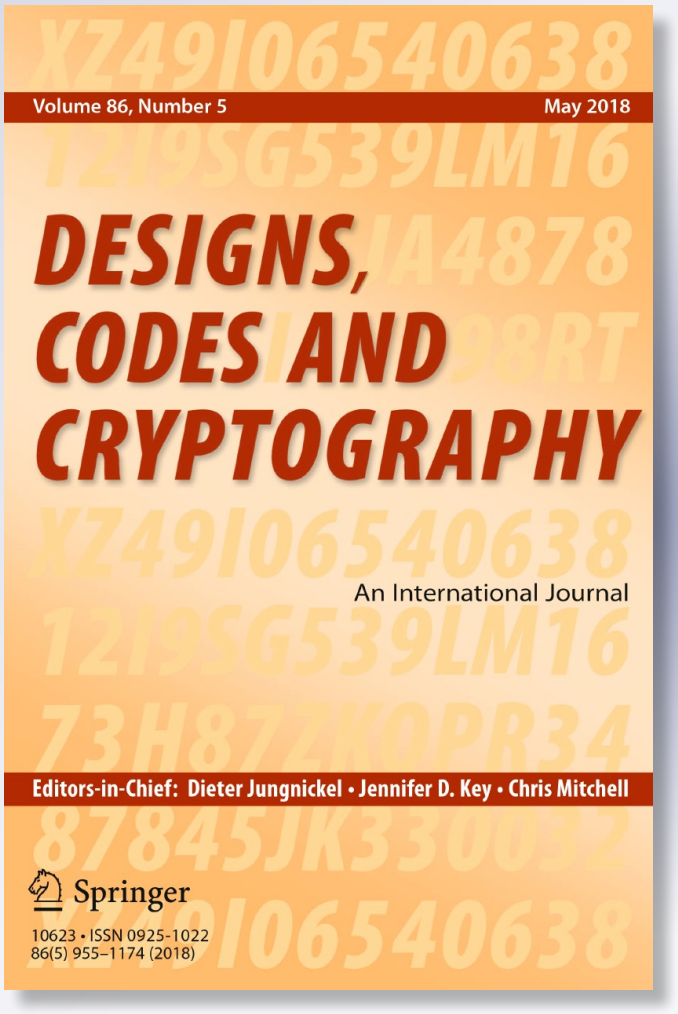
Properties of two Shannon's ciphers

Boris Ryabko

Designs, Codes and Cryptography
An International Journal

ISSN 0925-1022
Volume 86
Number 5

Des. Codes Cryptogr. (2018) 86:989-995
DOI 10.1007/s10623-017-0372-2



Springer

Des. Codes Cryptogr. (2018) 86:989–995
https://doi.org/10.1007/s10623-017-0372-2

Properties of two Shannon's ciphers

Boris Ryabko[1,2]

Received: 29 March 2017 / Revised: 25 May 2017 / Accepted: 31 May 2017 /
Published online: 22 June 2017

Abstract In 1949 Shannon published the famous paper "Communication theory of secrecy systems" where he briefly described two ciphers, but did not investigate their properties. In this note we carry out information-theoretical analysis of these ciphers. In particular, we propose estimations of the cipher equivocation and the probability of correct deciphering without key.

Keywords Shannon cipher · Cryptography · Entropy · Information theory

Mathematics Subject Classification 94A60 Cryptography · 94A15 Information theory

1 Introduction

In the famous paper [7] Shannon considered symmetric-key encryption schemes for transmitting secret messages from a sender to a receiver via an open channel which can be accessed by an adversary. It is assumed that the sender and the receiver (but not the adversary) share a key, which is a word in a certain alphabet. Before transmitting a message to the receiver, the sender encrypts it, and the receiver, having received an encrypted message (ciphertext), decrypts it to recover the plaintext.

Shannon considered the so-called running-key ciphers where the plaintext $X_1^1 \dots X_t^1$, the key sequence $X_1^2 \dots X_t^2$, and the ciphertext $Z_1 \dots Z_t$, are sequences of letters from

Communicated by C. Mitchell.

This paper was presented in part at XV International Symposium "Problems of Redundancy in Information and Control Systems" September 26–29, 2016, St. Petersburg, Russia.

✉ Boris Ryabko
boris@ryabko.net

1 Institute of Computational Technologies SB RAS, Novosibirsk, Russian Federation

2 Novosibirsk State University, Novosibirsk, Russian Federation

the same alphabet $A = \{0, 1, \dots, n-1\}$, where $n \geq 2$. Encryption and decryption are given by the rules $Z_i = c(X_i^1, X_i^2)$, $i = 1, \dots, t$, $X_i^1 = d(Z_i, X_i^2)$, $i = 1, \dots, t$, so that $d(e(X_i^1, X_i^2), X_i^2) = X_i^1$. The situation where the adversary tries to find the the plaintext $X_1^1 \dots X_t^1$ having the ciphertext $Z_1 \dots Z_t$, but not having the key sequence is considered. This model describes practically interesting cases including the so-called onetime pad, the case where the plaintext and the key sequence are texts in one language and some others, see [7].

We consider a model where it is supposed that the adversary is computationally unconstrained. (As a rule, the secrecy in this model is called either information-theoretic or everlasting).

In [7] Shannon introduced two following characteristics of a cipher, which he called equivocations: the key equivocation $H(X_1^2 \dots X_t^2 | Z_1 \dots Z_t)$ and the message equivocation $H(X_1^1 \dots X_t^1 | Z_1 \dots Z_t)$, where $H(U|V)$ is the conditional entropy of the variable U conditioned on the variable V, see for definition, for example, [2]. The properties of the equivocations and other characteristics of running-key ciphers were investigated in [4,5,7] and some other papers (see for review [1]). In particular, in [5] the so-called the average probability of correct decryptment (without key) was suggested and estimated for a certain model of the running-key cipher. In [6] the message equivocation was applied to show that the adversary has approximately $2^{H(X_1^1 \dots X_t^2 | Z_1 \dots Z_t)}$ possible decryptments whose probabilities are almost equal.

Shannon in [7] considered a running-key cipher where the key sequence (as well as the plaintext) is a text in English, and noted the following:

> The running key cipher can be easily improved to lead to ciphering systems which could not be solved without the key. If one uses in place of one English text, about d different texts as key, adding them all to the message, a sufficient amount of key has been introduced to produce a high positive equivocation. Another method would be to use, say, every 10th letter of the text as key. The intermediate letters are omitted and cannot be used at any other point of the message. This has much the same effect, since these spaced letters are nearly independent.

Note, that Diffie and Hellman in [3] claim that a running-key cipher can be strengthened by successively enciphering plaintext under two or more distinct running keys: "Since English is about 75% redundant…, four encipherments would be secure against all attacks". This is quite close to the result obtained in this paper, although Diffie and Hellman gave only this informal argumentation.

It is worth noting, that nowadays both ciphers can be used easily, because a lot of texts in English and other languages are available in Internet and other computer nets.

Our goal is to carry out the information-theoretical analysis of both ciphers. More precisely, we estimate the equivocation for both ciphers. Then, we extend the method of Lu [5] for estimation the average probability of correct decryptment (without key) to the case where letters of the key sequence can be dependent (in [5] only the case of independent and equiprobable key letters was considered). Then we apply the obtained estimations to both Shannon ciphers.

2 Information-theoretic analysis of Shannon ciphers

For this purpose we first generalize the running key ciphers in order to investigate the first Shannon cipher. Namely, let there be s sources $X^1, X^2, \dots, X^s$, $s \geq 2$, and any X^i generates

letters from the alphabet $A = \{0, 1, \ldots, n-1\}$. Suppose that X^1 is the plaintext, whereas $X^2, \ldots, X^s$ are key sequences. The ciphertext Z is obtained as follows

$$Z_i = \left(X_i^1 + X_i^2 + X_i^3 + \cdots + X_i^s\right) \mod n\,. \tag{1}$$

The deciphering is obvious.

For stationary ergodic processes W,V and $t \geq 1$ the t-order entropy, conditional entropy and entropy per letter are given as follows:

$$\begin{aligned} H_t(W) &= -t^{-1} \sum_{u \in A^{t-1}} P_W(u) \sum_{v \in A} P_W(v|u) \log_2 P_W(v|u) \\ H_t(W|V) &= H_t(W,V) - H_t(V)\,, \\ h_t(W) &= t^{-1} H_t(W)\,,\ h_t(W|V) = t^{-1} H_t(W|V)\,, \end{aligned} \tag{2}$$

see [2, Theorem 16.8.1 of the 2nd edition, p. 645]. The memory, or connectivity, of texts in English and other human languages is quite large, that is why estimation of their entropy is a rather complicated task, as it can be seen from the example of the entropy of English, see [8]. It is clear that the direct estimation of the entropy of the sum of two or more English texts is much more complicated. That is why we need to use other methods of estimation. The following lemma can be used for this purpose.

Lemma 1 *Let $X^1, X^2, \ldots, X^s$, $s \geq 2$, be s-dimensional stationary ergodic source and $X^1, X^2, \ldots, X^s$ be independent. If the cipher (1) is applied, then for any $t \geq 1$*

$$\begin{aligned} &h_t(X^1/Z) + h_t(X^2/Z) + \cdots + h_t(X^{s-1}/Z) \\ &\quad \geq h_t(X^1) + h_t(X^2) + \cdots + h_t(X^s) - \log_2 n \end{aligned} \tag{3}$$

and

$$\begin{aligned} &\frac{s-1}{s}\left(h_t(X^1/Z) + h_t(X^2/Z) + \cdots + h_t(X^s/Z)\right) \\ &\quad \geq h_t(X^1) + h_t(X^2) + \ldots + h_t(X^s) - \log_2 n \end{aligned} \tag{4}$$

The proof of this lemma is given in the Appendix.

Definition 1 Denote

$$\Lambda_t = \frac{1}{s-1}\left(h_t\left(X^1\right) + h_t\left(X^2\right) + \cdots + h_t\left(X^s\right) - \log_2 n\right)\,. \tag{5}$$

Note that, if X^i, $i = 1, \ldots, s$, have the same distribution, then

$$\Lambda_t = \frac{1}{s-1}\left(s\, h_t\left(X^1\right) - \log_2 n\right)\,. \tag{6}$$

The following definition is due to Lu [5].

Definition 2 Let $M = M(Z_1 \ldots Z_t) = X_1^{1*} \ldots X_t^{1*}$ be a certain function over A^t. Define $p_b = (1/t)\sum_{i=1}^t P\left(X_i^{1*} = X_i^1\right)$.

Note that, if M is a method of deciphering $Z_1 \ldots Z_t$ without the key, then p_b is the average probability of deciphering a single letter correctly. Obviously, the smaller p_b, the better the cipher.

The following theorem establishes some properties of both Shannon ciphers.

Theorem 1 *Let $X^1, X^2, \ldots, X^s$, $s \geq 2$, be s-dimensional stationary ergodic process and $X^1, X^2, \ldots, X^s$ be independent. Suppose that the cipher (1) is applied to the words of the length t, $t \geq 1$. If either $s > 2$ and all X^i have the same probability distribution (the first cipher) or $s = 2$ (the second cipher), then*

(i) *The equivocation of this cipher is low-bounded as follows:*

$$h_t(X^1|Z) \geq \Lambda_t. \tag{7}$$

(ii) *The following inequality for the average probability p_b is valid*

$$(1 - p_b)\log_2(n-1) + h_1(p_b) \geq \Lambda_t\,,$$

(iii) *For any $\delta > 0, \varepsilon > 0$ there exists t^* such that for $t > t^*$ there exists a set Ψ_Z of texts of length t for which $(P(\Psi_Z)) > 1 - \delta$, for any $V_1 \ldots V_t \in \Psi_Z$, $U_1 \ldots U_t \in \Psi_Z$*

$$1/t|\log_2 P(V_1 \ldots V_t|Z_1 \ldots Z_t) - \log_2 P(U_1 \ldots U_t|Z_1 \ldots Z_t)| < \varepsilon \tag{8}$$

and

$$\lim_{t\to\infty} \inf \frac{1}{t}\log_2 |\Psi_Z| \geq \Lambda_t. \tag{9}$$

The proof is given in Appendix.

This theorem shows that if Λ_t is large then the message equivocation is also large and the probability to find letters of the plaintext without the key must be small. Besides, the third statement shows that the adversary has a large set of possible plaintexts whose probabilities are close and the total probability is close to 1. So, he has around $2^{\Lambda_t t}$ possible plaintext whose total probability is close to one and the probability of all plaintexts are close. The next part contains applications of this theorem to texts on different languages.

3 Shannon ciphers

Let us come back to the Shannon's methods described above. In [8] Shannon estimated the entropy of printed English. In particular, he showed that the entropy of the first order is approximately 4.14 for texts without spaces and 4.03 for texts with spaces. He also estimated the limit entropy to be approximately 1 bit.

It is important to note that later many researcher estimated limit entropy of human languages and nowadays are quite precise estimations of entropy for many of them. A comprehensive review and estimations of entropy for English, French, Russian, Korean, Chinese, and Japanese cab be found in [9]. In particular, the estimation of the limit entropy of English is also close to 1 bit per letter.

Now we can investigate the first Shannon's cipher. He suggested to use a sum of d English texts as a key, i.e., use (1) with $s = d + 1$, and where X^i, $i = 1, \ldots, s$, are texts in printed English. Having taken into account that all X^i are identically distributed, we immediately obtain from (6) the following:

$$\Lambda_t = \frac{1}{s-1}\left(s\, h_t(X^1) - \log_2 n\right)$$

Taking into account that $h_t(X^1) \approx h_\infty(X^1) \approx 1$ and the estimation $\log_2 26 \approx 4.7$, we obtain the following approximation

$$\Lambda_t = \frac{1}{s-1}\left(s\, h_t\left(X^1\right) - \log_2 n\right) = \frac{1}{s-1}(s - 4.7)\,.$$

So, we can see that Λ_t is positive if $s - 1 \geq 4$. Moreover, Theorem 1 shows that the first cipher of Shannon cannot be deciphered without the key if four or more different texts are added to the message (i.e., used as a key).

Let us consider the second cipher of Shannon. Here $s = 2$, the sequence X^1 is a text in printed English and letters of X^2 are generated independently with probabilities equal to the frequencies of occurrence of letters in English. From (1) we obtain

$$\Lambda_t = h_t(X^1) + h_t(X^2) - \log_2 n\,.$$

Having taken into account that $h_t(X^1) \approx h_\infty(X^1) \approx 1$, $h_t(X^2) \approx 4.14$ (see [8]) and $\log_2 26 \approx 4.7$, we can see that $\Lambda_t = 1 + 4.14 - 4.7 = 0.44$. So, Λ_t is positive and Theorem 1 shows that the first cipher of Shannon cannot be deciphered without key.

Acknowledgements This research was supported by Russian Foundation for Basic Research (Grant No. 15-29-07932).

Appendix

Proof of Lemma The following chain of equalities and inequalities is valid:

$$\begin{aligned}
&h_t(X^1) + h_t(X^2) + \cdots + h_t(X^s) = h_t(X^1, X^2, \ldots, X^s)\\
&\quad = h_t(X^1, X^2, \ldots, X^s, Z) = h_t(Z) + h_t(X^1, X^2, \ldots, X^s/Z)\\
&\quad = h_t(Z) + h_t(X^1/Z) + h_t(X^2/X^1, Z) + h_t(X^3/X^1, X^2, Z)\\
&\qquad + \cdots + h_t(X^s/X^1, X^2, \ldots, X^{s-1}, Z)\\
&\quad = h_t(Z) + h_t(X^1/Z) + h_t(X^2/X^1, Z) + h_t(X^3/X^1, X^2, Z)\\
&\qquad + \cdots + h_t(X^{s-1}/X^1, X^2, \ldots, X^{s-2}, Z)\\
&\quad \leq h_t(Z) + h_t(X^1/Z) + h_t(X^2/Z) + h_t(X^3/Z) + \cdots + h_t(X^{s-1}/Z).
\end{aligned}$$

The proof is based on well-known properties of the Shannon entropy which can be found, for example, in [2]. More precisely, the first equation follows from the independence of $X^1, X^2, \ldots, X^s$, whereas the second equation is valid because Z is a function of $X^1, X^2, \ldots, X^s$, see (1). The third equation is a well-known property of the entropy. Having taken into account that X^s is determined if $X^2, \ldots, X^{s-1}, Z$ are known, we obtain the last equation. The inequality also follows from the properties of the Shannon entropy [2]. Thus,

$$\begin{aligned}
&h_t(X^1) + h_t(X^2) + \cdots + h_t(X^s)\\
&\quad \leq h_t(Z) + h_t(X^1/Z) + h_t(X^2/Z) + h_t(X^3/Z) + \cdots + h_t(X^{s-1}/Z). \qquad (10)
\end{aligned}$$

Taking into account that for any process U over alphabet $A = \{0, \ldots, n-1\}$

$$h_t(Z) \leq \log_2 n\,,$$

we obtain (3) from (10). In order to prove (4) we note that analogously to (10), we can obtain the following:

$$\begin{aligned}
&h_t(X^1) + h_t(X^2) + \cdots + h_t(X^s)\\
&\quad \leq \sum_{i=1}^{j-1} h_t(X^i/Z) + \sum_{i=j+1}^{s} h_t(X^i/Z)
\end{aligned}$$

for any $1 \leq j \leq$. From this inequality we obtain (4). □

Proof of Theorem For the first cipher $s > 2$ and all $X^i, i = 1, \dots, s$ have the same probability distribution. Having taken into account that $h_t(X^i) = h_t(X^1)$ for $i = 1, \dots, s$, from (4) and (6) we obtain (7). For the second cipher $s = 2$ and (7) follows from (3).

In order to prove ii), denote

$$H\left(X_j^i | Z_1 \dots Z_t\right) = -\sum_{X_j^1 \in A} P\left\{X_j^1 | Z_1 \dots Z_t\right\} \log P\left\{X_j^1 | Z_1 \dots Z_t\right\},$$

Let us consider any method G of encryption of $Z_1 \dots Z_k$ without key such that

$$\hat{X}_1^1 \hat{X}_2^1 \dots \hat{X}_t^1 = G(Z_1 \dots Z_k)$$

and define

$$p_j^* = P\left\{\hat{X}_j^1 = X_j^1\right\}, \; p^* = t^{-1} \sum_{j=1}^{t} p_j^*.$$

From Fano inequality (see [2,5]) we obtain

$$p_j^* \log(n-1) + \hat{h}\left(p_j^*\right) \geq H\left(X_j^1 | Z_1 \dots Z_t\right),$$

where $\hat{h}(p_j^*)$ is the following entropy:

$$\hat{h}\left(p_j^*\right) = -\left(p_j^* \log p_j^* + \left(1 - p_j^*\right) \log\left(1 - p_j^*\right)\right).$$

From the last inequality we obtain

$$t^{-1} \sum_{j=1}^{t} \left(p_j^* \log(n-1) + \hat{h}\left(p_j^*\right)\right) \geq t^{-1} \sum_{j=1}^{t} H\left(X_j^1 | Z_1 \dots Z_t\right).$$

Having taken into account convexity of entropy, from this inequality and the definition $p^* = t^{-1} \sum_{j=1}^{t} p_j^*$ we obtain

$$p^* \log(n-1) + \hat{h}(p^*) \geq t^{-1} \sum_{j=1}^{t} H\left(X_j^1 | Z_1 \dots Z_t\right).$$

From this and well known inequality for the entropy $H(u, v) \leq H(u) + H(v)$ we obtain

$$p^* \log(n-1) + \hat{h}(p^*)) \geq t^{-1} H\left(X_1^1 \dots X_t^1 | Z_1 \dots Z_t\right).$$

Taking into account the Definition (2) and the statement (i), we obtain (ii).

In order to prove the third statement we will use the well-known Shannon–McMillan–Breiman theorem, see [2]. For conditional entropies it can be stated as follows:

$\forall \varepsilon > 0, \forall \delta > 0$, for almost all
$Z_1, Z_2, \dots$ there exists n' such that if $n > n'$ then

$$P\left\{\left|-\frac{1}{n} \log P(X_1^1 .. X_t | Z_1 .. Z_t) - h(X|Z)\right| < \varepsilon\right\} \geq 1 - \delta, \tag{11}$$

where (X^1, Z) is stationary ergodic process.

According to Shannon–McMillan–Breiman theorem for any $\varepsilon > 0, \delta > 0$ and almost all $Z_1, Z_2, \ldots$ there exists such n' that for $t > n'$

$$P\left\{\left|-\frac{1}{t}\log P(X_1X_2\ldots X_t|Z_1Z_2\ldots Z_t) - h(X|Z)\right| < \varepsilon/2\right\} \geq 1-\delta. \tag{12}$$

Let us define

$$\Psi_Z = \left\{X_1^1X_2^1\ldots X_t^1 : |P\left(X_1^1X_2^1\ldots X_t^1|Z_1\ldots Z_t\right) - h_t(X^1|Z)| < \varepsilon/2\right\}. \tag{13}$$

The equation $P(\Psi_Z) > 1-\delta$ immediately follows from (12). In order to prove (8), note that for any $X^1 = X_1^1, \ldots, X_t^1$, $\bar{X}^1 = \bar{X}_2^1, \ldots, \bar{X}_t^1$ from Ψ_Z we obtain from (12), (13)

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

From (13), (7) and the proven equation $P(\Psi(Z)) > 1-\delta$ we obtain the following: $|\Psi_Z| > (1-\delta)2^{t\,(h_t(X|Z)-\varepsilon)}$. Taking into account that it is valid for any $\varepsilon > 0, \delta > 0$ and $t > n'$, we obtain (9). Theorem is proven. □

References

1. Calmon E.P., Medard M., Varia M., Duffy K.R., Christiansen M.M., Zeger L.M.: Hiding Symbols and Functions: New Metrics and Constructions for Information-Theoretic Security. arxiv:1503.08515 (2015).
2. Cover T.M., Thomas J.A.: Elements of Information Theory. Wiley-Interscience, New York (2006).
3. Diffie W., Hellman M.E.: Privacy and authentication: an introduction to cryptography. Proc. IEEE **67**(3), 397–427 (1979).
4. Hellman M.E.: An extension of the Shannon theory approach to cryptography. IEEE Trans. Inf. Theory **23**(3), 289–294 (1977).
5. Lu S.-C.: The existence of good cryptosystems for key rates greater than the message redundancy. IEEE Trans. Inf. Theory **25**(4), 475–477 (1979).
6. Ryabko B.: The Vernam cipher is robust to small deviations from randomness. Probl. Inf. Transm. **51**(1), 82–86 (2015).
7. Shannon C.E.: Communication theory of secrecy systems. Bell Syst. Tech. J. **28**(4), 656–715 (1949).
8. Shannon C.E.: Prediction and entropy of printed English. Bell Syst. Tech. J. **30**(1), 50–64 (1951).
9. Takahira R., Tanaka-Ishii K., Debowski L.: Entropy rate estimates for natural languagea new extrapolation of compressed large-scale corpora. Entropy **18**(10), 364 (2016).