Mathematical Problems of Computer Science 35, 70 76, 2011.

The Shannon Cipher System With a Guessing Wiretapper Eavesdroping Through a Noisy Channel

Evgueni A. Haroutunian and Tigran M. Margaryan

Institute for Informatics and Automation Problems of NAS of RA
e-mail: eghishe@sci.am

Abstract

In this paper we study the processes in the Shannon cipher system with discrete memoryless source and a guessing wiretapper. The wiretapper observes a cryptogram of M-vector ciphered messages passed through the noisy channel and tries to guess the secret plaintext with length N. The security of encryption system is measured by the average number of guesses needed for the wiretapper to uncover the plaintext. The problem was suggested by Arikan and Merhav as a generalization of their result for noiseless channel to wairtapper.

1. Introduction

The cryptographic system shown in Fig.1 is the Shannon cipher system with a noisy channel to wiretapper. An encrypted vector of messages of a discrete memoryless stationary source must be transmitted via a public noiseless channel to a legitimate receiver. The key-vector is communicated to encrypter and to decrypter by a special secure channel protected against wiretappers. After ciphering the vector of source messages by a key-vector, the cryptogram is sent over a public channel to a legitimate receiver. who can recover the original message using the cryptogram and the same key-vector. Not knowing the key, the wiretapper that eavesdrops on the noisy channel aims to decrypt the source messages using also known to him the source statistics and the encryption function.

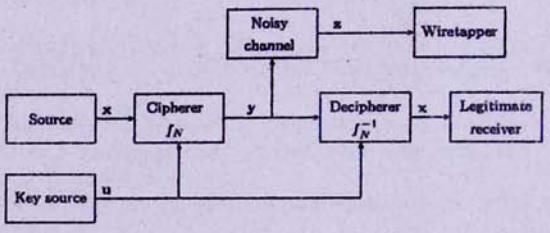

Fig. 1. The Shannon cipher system with a noisy channel to the wiretapper.

The wiretapper makes sequential guesses (suppositions), each time applying a testing mechanism by which he can know whether the estimate is successful or not and stops when the answer is affirmative. In this paper, we aim at characterizing the expectation of the number of guesses that the wiretapper may have to submit before succeeding.

The guessing problem was first considered by Massey [13], then by Arikan [1], [2] and recently by Malone and Sullivan [12]. The guessing subject to fidelity criterion was studied by Arikan and Merhav in [3], [4], for guessing subject to distortion and reliability requirements by Haroutunian and Ghazaryan in [7], [8], [9], for the Shannon cipher system with exact reconstruction by wiretapper by Merhav and Arikan in [14]. The security of the cipher system we measure by the expected number of guesses needed for reconstruction of the source messages. That approach was used also by Merhav and Arikan in [14], earlier by Hellman in [11] and by Sgarro in [15], [16]. The Shannon cipher system with a guessing wiretapper and correlated source outputs was examined by Hayashi and Yamamoto in [10].

The importance of this case of extension of the Shannon cipher system with guessing wiretapper can be explained by the fact that practically it is more probable that the wiretapper can observe a noisy version of the cryptogram. The problem was suggested by Merhav and Arikan in [14].

2. Definitions and Formulation of the Result

The discrete memoryless source is defined as a sequence $\{X_i\}_{i=1}^{\infty}$ of discrete, independent, identically distributed (i.i.d.) random variables (RVs) taking values in the finite set $\mathcal{X}$ of messages x of the source. Let $P^* = \{P^*(x), x \in \mathcal{X}\}$ be the source messages generating probability distribution (PD) which is supposed to be known also to the wiretapper. Let $\mathbf{X} = (X_1, X_2, \dots, X_N)$ be a random N-vector. Since we study the memoryless source, the probability of the vector $\mathbf{x} = (x_1, x_2, \dots, x_N)$, a realization of the random N-vector $\mathbf{X}$, is

$$P^{*N}(\mathbf{x}) = \prod_{n=1}^{N} P^*(x_n).$$

The key-source is described by a sequence $\{U_i\}_{i=1}^{\infty}$ of binary i.i.d. RVs, which take values from the set $\mathcal{U} = \{0, 1\}$. The distribution $Q^* = \{1/2, 1/2\}$ is the PD of the key bits. The key-vector $\mathbf{u} = (u_1, u_2, \dots, u_K)$ is the vector of K bits and $Q^{*K}(\mathbf{u}) = 2^{-K}$. The key-vector of K binary RVs $\mathbf{U} = (U_1, U_2, \dots, U_K)$ is independent of the vector $\mathbf{X}$.

Let $f_N : \mathcal{X}^N \times \mathcal{U}^K \to \mathcal{Y}^M$ be an encryption function where $\mathcal{Y}$ is the cryptogram alphabet which is not necessarily the same as the source alphabet. This function is assumed to be invertible providing the key is given, i.e. there exists the decryption function $f_N^{-1} : \mathcal{Y}^M \times \mathcal{U}^K \to \mathcal{X}^N$. It is also assumed that f_N is such that M/N is constant and is equal to λ.

We denote the PD of RV Y taking values in $\mathcal{Y}$ by $S = \{S(y), y \in \mathcal{Y}\}$, where S depends on P^* and f_N.

The wiretapper's channel is a discrete memoryless channel (DMC) with input alphabet $\mathcal{Y}$, output alphabet $\mathcal{Z}$ and whith a stochastic matrix of transition probabilities $W^* = \{W^*(z|y), y \in \mathcal{Y}, z \in \mathcal{Z}\}$. The model for M actions of the channel is described by the stochastic matrix $W^{*M} : \mathcal{Y}^M \to \mathcal{Z}^M$, an element of which $W^{*M}(\mathbf{z}|\mathbf{y})$ is a conditional probability of receiving the vector $\mathbf{z} \in \mathcal{Z}^M$, when vector $\mathbf{y} \in \mathcal{Y}^M$ is transmitted. So for all

$\mathbf{y} \in \mathcal{Y}^M$ and $\mathbf{z} \in \mathcal{Z}^M$

$$W^{*M}(\mathbf{z}|\mathbf{y}) = \prod_{m=1}^{M} W^*(z_m|y_m).$$

We denote the joint PD of RVs Y and Z by

$$S \circ W^* = \{S(y) \circ W^*(z.y) = S(y)W^*(z|y), y \in \mathcal{Y}, z \in \mathcal{Z}\}$$

and PD of RV Z by

$$SW^* = \{SW^*(z) = \sum_{y \in \mathcal{Y}} S(y)W^*(z|y), z \in \mathcal{Z}\}.$$

The conditional probability W of $y \in \mathcal{Y}$ for given $z \in \mathcal{Z}$ is the following

$$W = S \circ W^*/SW^* = \{S(y) \circ W^*(z,y)/SW^*(z), y \in \mathcal{Y}, z \in \mathcal{Z}\}.$$

We will apply the method of types (see [5], [6]) in the proof of the theorem. Let us begin with the formulation of some basic concepts, notations and relations of this method. The type P of vector $\mathbf{x} = (x_1, \ldots, x_N) \in \mathcal{X}^N$ is a PD $P = \{P(x) = N(x|\mathbf{x})/N, x \in \mathcal{X}\}$, where $N(x|\mathbf{x})$ is the number of repetitions of the symbol x among $x_1, \ldots, x_N$. The set of vectors $\mathbf{x}$ of type P is denoted by $T_P^N(X)$. The set of all PD on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$ and the subset of $\mathcal{P}(\mathcal{X})$ consisting of the possible types of sequences $\mathbf{x} \in \mathcal{X}^N$ is denoted by $\mathcal{P}_N(\mathcal{X})$.

We denote entropy of RV X with PD P and, respectively, divergence of PD P^* from P as follows

$$H_P(X) \triangleq -\sum_{x \in \mathcal{X}} P(x) \log P(x),$$

$$D(P\|P^*) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{P^*(x)}.$$

The type of vector $\mathbf{z}$ is denoted by Q. and the set of vectors $\mathbf{z}$ of type Q is denoted by $T_Q^M(Z)$.

The joint type of $\mathbf{y} \in \mathcal{Y}^M$ and $\mathbf{z} \in \mathcal{Z}^M$ is the PD $\{M(y, z|\mathbf{y}, \mathbf{z})/M, y \in \mathcal{Y}, z \in \mathcal{Z}\}$, where $M(y, z|\mathbf{y}, \mathbf{z})$ is the number of occurrences of symbols pair (y, z) in the pair of vectors $(\mathbf{y}, \mathbf{z})$. In other words, joint type is the type of the sequence $(\mathbf{y}, \mathbf{z}) = ((y_1, z_1), (y_2, z_2), \cdots, (y_M, z_M))$ from $(\mathcal{Y} \times \mathcal{Z})^M$.

We say that conditional type of $\mathbf{y}$ for given $\mathbf{z}$ is PD $V = \{V(y|z), z \in \mathcal{Z}, y \in \mathcal{Y}\}$ if $M(z, y|\mathbf{z}, \mathbf{y}) = M(z|\mathbf{z})V(y|z)$ for all $z \in \mathcal{Z}, y \in \mathcal{Y}$. The set of all sequences $\mathbf{y} \in \mathcal{Y}^M$ of conditional type V for given $\mathbf{z} \in T_Q^M(Z)$ is denoted by $T_{Q,V}^M(Y|\mathbf{z})$ and called V-shell of $\mathbf{z}$. $\mathcal{V}_M(\mathcal{Y}, Q)$ is the set of all possible V-shells for $\mathbf{z}$ of type Q.

We denote conditional entropy of RV Y for given RV Z with PD Q and conditional PD V of Y for given Z by

$$H_{Q,V}(Y|Z) \triangleq -\sum_{z \in \mathcal{Z}, y \in \mathcal{Y}} QV(y) \log V(y|z),$$

and for conditional divergence of PD $Q \circ V$ from PD $Q \circ W$ on $\mathcal{Z} \times \mathcal{Y}$ by

$$D(Q \circ V\|Q \circ W) = D(V\|W|Q) \triangleq \sum_{z \in \mathcal{Z}, y \in \mathcal{Y}} Q(z)V(y|z) \log \frac{V(y|z)}{W(y|z)}.$$

The source generates vector $\mathbf{x}$ (with length N) and the key source generates vector $\mathbf{u}$ (with length K). The encoder f_N receiving $\mathbf{x}$ and $\mathbf{u}$ generates the cryptogram $\mathbf{y}$ (of length M) and the vector $\mathbf{y}$ is sent to a legitimate receiver. The wiretapper eavesdrops through the noisy channel and receives a vector $\mathbf{z}$ (of length M).

We assume that the wiretapper knows the encoding function f_N (hence also the decryption function f_N^{-1}) and has a strategy $g = \{\mathbf{x}_1(\mathbf{z}), \mathbf{x}_2(\mathbf{z}), \cdots\}$ to guess the correct secret $\mathbf{x}$ as fast as possible. Let $G_{f,g}^N(X|Z)$ be a number of guesses needed for the wiretapper to uncover the secret $\mathbf{x}$ by the strategy g^N.

Definition 1: *The key rate R_K of the key source is defined by* $R_K \triangleq K/N$.

Definition 2: *The guessing rate $R(R_K, W^*, P^*)$ of this system is defined by*

$$R(R_K, W^*, P^*) \triangleq \lim_{N\to\infty} \sup_{f_N} \inf_{g_N} \frac{1}{N} \log E[G_{f,g}^N(X|Z)]$$

where $E[G_{f,g}^N(X|Z)]$ is the expectation of $G_{f,g}^N(X|Z)$.

In the following theorem upper and lower bounds for the guessing rate are presented.

Theorem 1: *For given PD P^*, conditional PD W^*, and every key rate R_K,*

$$R(R_K, W^*, P^*) \le \max_{S} \max_{P,Q,V} [\min\{H_P(X), \lambda H_{Q,V}(Y|Z) + R_K\} - D(P\|P^*) - \lambda D(V\|\frac{S \circ W^*}{SW^*}|Q)],$$

$$R(R_K, W^*, P^*) \ge \max_{P} [\min\{R_K, H_P(X)\} - D(P\|P^*)].$$

Corollary: *When the wiretapper's channel is noiseless we arrive at the result of Merhav and Arikan from* [14]:

$$R(R_K, P^*) = \max_{P} [\min\{R_K, H_P(X)\} - D(P\|P^*)].$$

3. Proof of Theorem 1

We will use the following inequalities, concerting types ([4], [5]).

$$|\mathcal{P}_N(\mathcal{X})| < (N+1)^{|\mathcal{X}|}. \tag{1}$$

$$|\mathcal{V}_M(\mathcal{Y}, P_{\mathbf{z}})| < (M+1)^{|\mathcal{Z}||\mathcal{Y}|}. \tag{2}$$

For any type $P \in \mathcal{P}_N(\mathcal{X})$

$$|\mathcal{T}_P^N(X)| \le \exp\{N H_P(X)\}, \tag{3}$$

and for any PD P^*

$$P^{*N}(\mathcal{T}_P^N(X)) \le \exp\{-N D(P\|P^*)\}. \tag{4}$$

For any type Q , conditional type V and $\mathbf{z} \in \mathcal{T}_Q^M(Z)$

$$|\mathcal{T}_{Q,V}^M(Y|\mathbf{z})| \le \exp\{M H_{Q,V}(Y|Z)\}. \tag{5}$$

and for any conditional PD W

$$W^M(\mathcal{T}_{Q,V}^M(Y|\mathbf{z})|\mathbf{z}) = \exp\{-M D(V\|W|Q)\}. \tag{6}$$

Let the vector $\mathbf{x}$ generated by the source has type P ($\mathbf{x} \in \mathcal{T}_P^N(X)$), wiretapper receives vector $\mathbf{z}$ of type Q ($\mathbf{z} \in \mathcal{T}_Q^M(Z)$) and let our cryptogram belong to V-shell of vector $\mathbf{z}$ ($\mathbf{y} \in \mathcal{T}_{Q,V}^M(Y|\mathbf{z})$). To build some strategy for wiretapper, we consider the following two strategies g_1^N and g_2^N.

Strategy g_1^N: $\mathcal{X}^N$ can be represented as a union of vectors of various types (these types we arrange in ascending order of entropy: $H_{P_1}(X) \le H_{P_2}(X) \le \cdots\cdots$)

$$\mathcal{X}^N = \bigcup_{i=1,2,\cdots} \mathcal{T}_{P_i}^N(X).$$

The wiretapper ignores the cryptogram $\mathbf{z}$ and sequentially guesses in ascending order of entropy up to finding of $\mathbf{x}$. The message $\mathbf{x}$ belongs to $\mathcal{T}_P^N(X)$ and, therefore, it is clear that in this strategy g_1^N ($g_1^N = \{\mathbf{x}_1, \mathbf{x}_2, \cdots\}$) the number of guesses is bounded with (1) and (3) in the following way

$$G_{f,g_1}^N(\mathbf{x}|\mathbf{z}) \le \sum_{P_i: H_{P_i}(X) \le H_P(X)} |\mathcal{T}_{P_i}^N(X)| \le (N+1)^{|\mathcal{X}|} \exp\{N H_P(X)\}$$

$$\le \exp\{N H_P(X) + o(N)\} \tag{7}$$

Strategy g_2^N: $\mathcal{Y}^M$ can be represented as a union of vectors of various conditional types for given vector $\mathbf{z} \in \mathcal{T}_Q^M(Z)$ (these conditional types we arrange in ascending order of conditional entropy: $H_{Q,V_1}(Y|Z) \le H_{Q,V_2}(Y|Z) \le \cdots$)

$$\mathcal{Y}^M = \bigcup_{l=1,2,\cdots} \mathcal{T}_{Q,V_l}^M(Y|\mathbf{z}).$$

In this strategy, the wiretapper aims to find message $\mathbf{x}$ sequentially applying different keys on cryptograms $\mathbf{y}$ in ascending order of conditional entropy for given vector $\mathbf{z}$. To find vector $\mathbf{x}$ wiretapper finds the key $\mathbf{u}$ and the cryptogram $\mathbf{y}$ which belongs to V-shell of vector $\mathbf{z}$ ($\mathbf{y} \in \mathcal{T}_{Q,V}^M(Y|\mathbf{z})$) so in this strategy g_2^N
$g_2^N = \{f^{-1}(\mathbf{y}_1, \mathbf{u}_1), f^{-1}(\mathbf{y}_1, \mathbf{u}_2) \cdots f^{-1}(\mathbf{y}_1, \mathbf{u}_{\exp\{K\}}), f^{-1}(\mathbf{y}_2, \mathbf{u}_1), f^{-1}(\mathbf{y}_2, \mathbf{u}_2) \cdots\}$
the number of guesses is bounded with help of (2) and (5) as follows

$$\begin{aligned} G_{f,g_2}^N(\mathbf{x}|\mathbf{z}) &\le \sum_{V_l: H_{Q,V_l}(Y|Z) \le H_{Q,V}(Y|Z)} |\mathcal{T}_{Q,V_l}^M(Y|\mathbf{z})| \exp\{K\} \\ &\le (M+1)^{|\mathcal{Z}||\mathcal{Y}|} \exp\{M H_{Q,V}(Y|Z)\} \exp\{N R_K\} \\ &\le \exp\{M H_{Q,V}(Y|Z) + o(M) + N R_K\} \\ &\le \exp\{N(\lambda H_{Q,V}(Y|Z) + R_K) + o(N)\}. \end{aligned} \tag{8}$$

Based on strategy g_1^N and strategy g_2^N, we define a new strategy g_3^N as follows
$g_3^N = (\mathbf{x}_1, f^{-1}(\mathbf{y}_1, \mathbf{u}_1), \mathbf{x}_2, f^{-1}(\mathbf{y}_1, \mathbf{u}_2) \cdots \mathbf{x}_{\exp\{K\}}, f^{-1}(\mathbf{y}_2, \mathbf{u}_1), \mathbf{x}_{\exp\{K\}+1}, f^{-1}(\mathbf{y}_2, \mathbf{u}_2) \cdots)$.
Then, the number of guesses in the strategy g_3^N is not more than twice the smaller number of guesses in g_1^N and g_2^N. Therefore, we have (see (7),(8))

$$G_{f,g_3}^N(\mathbf{x}|\mathbf{z}) \le 2 \min\{\exp\{N H_P(X) + o(N)\}, \exp\{N(\lambda H_{Q,V}(Y|Z) + R_K) + o(N)\}\}$$

$$\le \exp\{N \min\{H_P(X), \lambda H_{Q,V}(Y|Z) + R_K\} + o(N)\}. \tag{9}$$

The expectation $G^N_{f_N g_N}X|Z)$ is bounded by (see (1).(2).(4).(6).(9))

$$
\begin{aligned}
& E[G^N_{f_N g_N}(X|Z)] = \sum_{\mathbf{x}\in\mathcal{X}^N,\mathbf{y}\in\mathcal{Y}^M} P^{*N}(\mathbf{x})S^M(\mathbf{y})G^{*N}_{f_N g}(\mathbf{x}|\mathbf{z}) \\
\leq & \sum_{P\in\mathcal{P}_N(\mathcal{X}),V\in\mathcal{V}_M(\mathcal{Y},Q)} \exp\{-ND(P\|P^*)\}\exp\{-MD(V\|W|Q)\}G^{*N}_{f_N g}(\mathbf{x}|\mathbf{z}) \\
\leq & \max_{P,Q,V}[(N+1)^{|\mathcal{X}|}(M+1)^{|\mathcal{Z}||\mathcal{Y}|}\exp\{N(-D(P\|P^*)-\lambda D(V\|W|Q))\}] \qquad (10) \\
\times & \exp\{N\min\{H_P(X),\lambda H_{Q,V}(Y|Z)+R_K\}+o(N)\}] \\
\leq & \exp\{N\max_{P,Q,V}[\min\{H_P(X),\lambda H_{Q,V}(Y|Z)+R_K\}-D(P\|P^*)-\lambda D(V\|W|Q)]+o(N)\}
\end{aligned}
$$

Since our strategy is valid for any function f_N, from inequality (10) we obtain the upper bound for the guessing rate

$$
R(R_K,W^*,P^*)=\lim_{N\to\infty}\sup_{f_N}\inf_{g_N}\frac{1}{N}\log E[G^N_{f_N g}(X|Z)]\leq\lim_{N\to\infty}\sup\frac{1}{N}\log E[G^N_{f_N g_N}(X|Z)]
$$

$$
\leq\max_{S}\max_{P,Q,V}[\min\{H_P(X),\lambda H_{Q,V}(Y|Z)+R_K\}-D(P||P^*)-\lambda D(V||(S\circ W^*/SW^*)|Q)].
$$

As regards the lower bound, we have not get better result for it and we will use the result achieved by Merhav and Arikan in [14]. It is obvious that any lower bound on $R(R_K,W^*,P^*)$ for Shannon cipher system with a noiseless channel to the wiretapper is also a lower bound for the same system with a noisy channel. Thus,

$$
R(R_K,W^*,P^*)\leq R(R_K,P^*)\leq\max_{P}[\min\{R_K,H_P(X)\}-D(P||P^*)].
$$

References

[1] E. Arikan. "An inequality on guessing and its application to sequential decoding". *IEEE Trans. Inform. Theory*. vol. 42. no. 1, pp. 99-105, 1996.

[2] E. Arikan a "On the Average Nuber of Guesses Required to Determine the Value of a Random variable". *Transactions of the 12th Prague Conference on Information Theory, Statistical Decision Function and Random Processes*, pp. 20-23. 1994.

[3] E. Arikan and N. Merhav, "Guessing subject to distortion". *IEEE Trans. Inform. Theory*. vol. 44. no. 3. pp. 1041-1056. 1998.

[4] E. Arikan and N. Merhav,"Joint source-channel coding and guessing with application to sequential decoding". *IEEE Trans. Inform. Theory*. vol. 44. no. 5. pp. 1756-1769. 1998.

[5] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic. 1981.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley. 2006.

[7] E. A. Haroutunian and A. R. Ghazaryan. "On cipher system with a wiretapper guessing with respect to fidelity and reliability criteria". *Proceedings of the Third Conference on Computer Science and Information Technologies* (Yerevan, Armenia. 2001). pp. 215-219.

[8] E. A. Haroutunian and A. R. Ghazaryan. "On the Shannon cipher system with a wiretapper guessing subject to distortion and reliability requirements". *Proceedings of the 2002 IEEE Int. Symp. Inform. Theory* (Lausanna. Switzerland), p. 324.

[9] E. A. Haroutunian, "Realibility approach in wiretapper guessing theory". in "Aspects of Network and Information Security", NATO Science for Peace and Security. series D: Information and Communication Security, vol. 17, pp. 248 260. IOS Press, 2008.

[10] Y. Hayashi and H. Yamamoto, "Coding theorems for the Shannon cipher with a guessing wiretapper and correlated source outputs", *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2808-2817, June 2008.

[11] M. E. Hellman, "An extention of the Shannon theory approach to cryptography", *IEEE Trans. on Inform. Theory*, vol. 23, no. 3, pp. 289-299, 1977.

[12] D. Malone and W. G. Sullivan, "Guesswork and entropy", *IEEE Trans. Inform. Theory*, vol. 50. no. 3. pp. 525-526. 2004.

[13] J. L. Massey. "Guessing and entropy", *Proceedings of the 1994 IEEE International Symp. Inform. Theory* (Trondheim, Norway, 1994), p. 204.

[14] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper", *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1860-1866, 1999.

[15] A. Sgarro, "Error Probabilities for Simple Substitution Ciphers", *IEEE Trans. Inform. Theory*, vol. 29, no. 2, pp. 190-197, 1983.

[16] A. Sgarro, "Exponential-type parameters and Substitution Ciphers", *Problems of Control and Inform. Theory*, vol. 14(5), pp. 393-403, 1985.

Շենոնյան ծածկագրման համակարգում գաղտնագողի գուշակումը աղմկոտ կապուղով

Ե. Հարությունյան և Տ. Մարգարյան

Ամփոփում

Հոդվածում լուծված է Մերհավի և Արիկանի կողմից առաջադրված խնդիրը: Դիտարկված է հատնյալ համակարգը. ծանիչը բանալու օգնությամբ ծածկագրում է հաղորդագրությունը և անաղմուկ կապուղով ուղարկում է օրինական հասեատիրոջը, որին ուղարկում է նաև բանալին մեկ այլ անաղմուկ, պաշտպանված կապուղով: Գաղտնագողը, օգտվելով աղմկոտ կապուղուց, ստանալով ծածկագիրը, ձգտում է վերծանել հաղորդագրությունը հաջորդական գուշակումների միջոցով:

Հոդվածում ստացված է կռահման արագության գնահատականները: