Evaluating the effectiveness of information security based on the calculation of information entropy

Akmaral Imanbayeva1,2*, Yerlan Tursynbek1, Rabiga Syzdykova3 and Aigerim Mukhamedova1

1IETP Al-Farabi Kazakh National University, Almaty, Kazakhstan
2Research Center “KazAlfaTech Ltd”, Almaty, Kazakhstan
3Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan

*akmaral@physics.kz

Abstract. Recently, the emergence and increasingly widespread of wireless networks has generated considerable interest in the information-theoretical approach to ensuring secure communications. The basic principle of information-theoretical security requires a combination of cryptographic methods with channel encoding techniques that use the randomness of communication channels to ensure that messages sent cannot be intercepted or decrypted by a third party maliciously eavesdropping on the wireless medium. This paper discusses the information-entropy method for assessing security. We show that information-theoretical security means that any algorithm has a negligible probability of violating the security property. This is the same as unconditional security: it does not rely on any computational assumptions and is not limited to probabilistic violators.

1. Introduction

An information-theoretical approach to ensuring secure communications has opened up a promising new direction for solving the problems of wireless network security. Wyner [1], Csiszar & Korner [2] pioneered this method in the 1970s. They showed that confidential data can be transmitted safely without using an encryption key. Agreement of the secret key (including generation and distribution) using theoretical analysis of information was later proposed in the works of Maurer [3-4] and in the work of Alswede & Chizar [5]. The showed that two or more principals (Alice and Bob) can agree on a key (for example, for subsequent encryption) is kept secret from others (Eve). The impact of these works was limited in the 1980s and 1990s, in part due to the lack of practical listening codes but mainly because strictly positive secrecy in the classical listening channel setup requires that the legitimate sender and receiver have some advantage over the intruder in the terms of channel quality. Diffie and Hellman [6] published the basic principles of public key cryptography that were to be adopted by almost all modern security schemes in 1976. Public key algorithms are simple in terms of key management but they require significant computational resources [7]. Private key algorithms are computationally efficient and provide higher data throughput than public key algorithms. Although they have key management problems such as secure storage and distribution of keys [8-11].

The basic principle of information-theoretical security requires a combination of cryptographic methods with channel encoding methods that use the randomness of communication channels to
ensure that sent messages cannot be intercepted or decrypted by a third party maliciously eavesdropping on a wireless carrier [12].

A security model consisting of three categories is considered in this papers: confidentiality, integrity, accessibility. Where confidentiality is the state of information in which the subjects having the right it exercise access to it; integrity is the avoidance of unauthorized modification of information; accessibility is avoiding temporary or permanent hiding of information from users who have received access right. The methodological basis for assessing in a telecommunications network can be the so-called intelligent methods of data analysis. Bayesian approach is the most common of these [13] but we used an information-theoretical approach.

2. Methodology

We used one of the new methods for analyzing the complexity of various objects and processes [14-15]. Usually the definition of a complex concept is formed through a list of its main properties. Information $I(X)$ of the statistical realization of some physical quantity $x$ is a positive quantity and it is determined in the presence of nonequilibrium $I(X) \neq I(x_0)$, if $X \neq x_0$. If $P(X)$ is the probability of occurrence of the quantity $x$ then the expression for the amount of information describes these properties:

$$I(X) = - \ln P(X). \quad (1)$$

Information is determined through the difference between the unconditional and conditional entropies in the presence of some condition $y$:

$$I(x/y) = S(X) - S(X/Y). \quad (2)$$

This formula is used to estimate the bandwidth of communication channels. The original signal $X$ is encrypted into message $Y$ with a key $K$ (the key is shared by transmitter Alice and receiver Bob) according to Shannon scheme [12]. An attacker Eve (who knows the family of encryption functions (keys) and the probability of choosing the keys) can intercept the ciphertext $Y$. The system is considered completely safe if the posterior probabilities of the signal $X$ of a given $Y$ are equal to the prior probabilities $X$ for all $Y$, i.e. $P(X|Y) = P(X)$. The number of different keys must be at least as large as the number of signals to achieve reliable secrecy. Shannon’s fundamental pursuit of perfect secrecy suggests that the entropy of the source cannot be greater than the entropy of the secret key originally shared by the sender and the rightful recipient. The value $S(Y) = S(X) + S(K)$ determines the uniqueness distance or the minimum length of the ciphertext that guarantees the recovery of the key used for encryption.

However, the information entropy $S(X)$ itself is the average value of information:

$$H(X) = \sum_i P_i(X) I_i(X) = - \sum_i P_i(X) \ln P_i(X), \quad (3)$$

where $i$ is the number of cells of the partition of the set of values $X$. The probability of realizing information $P(I)$ is

$$P(I) = e^{-l}. \quad (4)$$

The probability $f(I)$ density is

$$0 \leq P(I) \leq 1, \quad 0 \leq l \leq \infty, \quad \int_0^\infty f(I)dl = 1, \quad P(I) = \int f(I)dl, f(I) = P(I) = e^{-l}. \quad (5)$$
The probability function of information realization \( P(I) \) coincides with the probability density distribution function \( f(I) \). We define the informational entropy of the distribution of information values \( H(I) \) as the average of information:

\[
S(I) = \int_{-\infty}^{\infty} f(I) dI = (1 + I) e^{-I}. \tag{6}
\]

Entropy varies \( 1 \geq S \geq 0 \), for \( 0 \leq I \leq \infty \) i.e. entropy is normalized to unity. It is known that the entropy of a continuous set under a jumpwise change in variables is infinite [16] and therefore the integral is calculated in the sense of Lebesgue by introducing a certain measure. Information is adopted as this measure. Any continuous function \( g(x) \) at its fixed point satisfies equation

\[
g(x) = \alpha g(g(x/\alpha)), \tag{7}
\]

where \( \alpha \) is scale factor. If \( f(I) \) and \( H(I) \) are characteristic functions then fixed points of Eq.(7) are

\[
f(I) = I, \quad e^{-I} = I, \quad I = I_1 = 0.567, \tag{8}
\]

\[
S(I) = I, \quad (1 + I)e^{-I} = I, \quad I = I_2 = 0.806. \tag{9}
\]

There are various interpretations of the physical meaning of the numbers \( I_1 \) and \( I_2 \) [14].

3. Result and Discussion

We considered a distributed computing network with three areas of protection against a certain threat: 1\textsuperscript{st} group are protected systems; 2\textsuperscript{nd} group are systems with a high degree of security; 3\textsuperscript{rd} group are systems a low degree of security. There are three hypotheses \( Q \) belonging to the \( n \)-th group, where \( n = 1, 2, 3 \). It is known that 50% of systems nodes were protected, 30% of the nodes have high and 20% low protection from the general statistics of the impact of threats \( Z \). Using this data you can determine the prior probabilities of hypotheses \( S(Q_1) = 0.5; S(Q_2) = 0.3; S(Q_3) = 0.2 \). Expression for determining information entropy

\[
S(Q_n, z_n) = S(Q_n) + S(z/Q) + S(Q_n)S(z/Q). \tag{10}
\]

We will choose three indicators of system security to assess the effectiveness of information protection based on the calculation of information entropy:

- The ability of the system to ensure the confidentiality of information when exposed to the threat \( Z(z_1) \).
- The ability of the system to ensure the integrity of information \( (z_2) \).
- The ability of the system to ensure the availability of information \( (z_3) \).

The following impacts on the resources of the 1\textsuperscript{st}, 2\textsuperscript{nd} and 3\textsuperscript{rd} groups are derived threat analysis, and it is known that confidentiality was ensured from 10% to 90% of cases. Conditional entropy can then be written:

\[
S(z_1/Q_1), \quad S(z_2/Q_2), \quad S(z_3/Q_3).
\]

The calculated values of information entropies for various states of confidentiality, integrity and availability are shown respectively in Tables 1, 2, 3.

| Groups | Values | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|--------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1\textsuperscript{st} | 0.65 | 0.80 | 0.95 | 1.10 | 1.25 | 1.40 | 1.55 | 1.70 | 1.85 |
| 2\textsuperscript{nd} | 0.43 | 0.56 | 0.69 | 0.82 | 1.00 | 1.10 | 1.21 | 1.42 | 1.40 |
| 3\textsuperscript{rd} | 0.32 | 0.44 | 0.56 | 0.68 | 0.80 | 0.92 | 1.04 | 1.16 | 1.28 |
Table 2. Information entropy at different values of integrity for different groups

| Groups | Values | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.85 |
|--------|--------|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 1\textsuperscript{st} | 0.65   | 0.80| 0.95| 1.10| 1.25| 1.40| 1.55| 1.72| 1.775|
| 2\textsuperscript{nd} | 0.43   | 0.56| 0.69| 0.82| 0.95| 1.08| 1.21| 1.34| 1.47 |
| 3\textsuperscript{rd} | 0.32   | 0.44| 0.55| 0.68| 0.80| 0.82| 1.04| 1.16| 1.19 |

Table 3. Information entropy at different values of availability for different groups

| Groups | Values | 0.15 | 0.25 | 0.35 | 0.45 | 0.55 | 0.65 | 0.75 | 0.85 | 0.95 |
|--------|--------|------|------|------|------|------|------|------|------|------|
| 1\textsuperscript{st} | 0.725  | 0.87 | 0.95 | 1.125| 1.325| 1.475| 1.625| 1.77 | 1.87 |
| 2\textsuperscript{nd} | 0.500  | 0.62 | 0.755| 0.885| 1.015| 1.145| 1.275| 1.40 | 1.47 |
| 3\textsuperscript{rd} | 0.380  | 0.50 | 0.620| 0.740| 0.860| 0.980| 1.100| 1.22 | 1.28 |

The probabilities of hypotheses can increase in the process of gathering facts if the facts support them, or decrease if the facts disprove them.

The calculated data of confidentiality ($a$), integrity ($b$) and availability ($c$) are presented respectively in Fig. 1.

Figure 1. Change in normalized information entropy at different values of confidentiality ($a$), integrity ($b$) and availability ($c$) for different groups (1\textsuperscript{st} group ‘+’, 2\textsuperscript{nd} group ‘o’, 3\textsuperscript{rd} group ‘*’).

4. Conclusion
We have assessed the security model in three categories: confidentiality, integrity and availability. The methodological basis of this approach was the information-entropy analysis of random processes. This method can be used to quantitatively substantiate the composition of resources of various levels of security using key elements of information theory and entropy. The results obtained make it possible
to formulate recommendations on the use of practical and theoretical results of work to protect information system. Similar calculations are also used as independent tasks for university students who are studying in the specialty “Radio Engineering and Telecommunications” at the Al Farabi Kazakh National university [16-17].

Informed decisions for configuring the capacity of systems and resources of different levels of security can be made by performing similar calculations for all types of security threats in accordance with the threat model.

References

[1] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, pp. 1355-1387, 1975.

[2] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, pp. 339-348, 1978.

[3] U. M. Maurer, "Provably secure key distribution based on independent channels," *Proc. of the IEEE Information Theory Workshop (ITW), Veldhoven, The Netherlands*, June 1990.

[4] U. M. Maurer, "Secret-key agreement by public discussion based on common information," *IEEE Transactions on Information Theory*, vol. 39, pp. 733-742, 1993.

[5] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. Part I: Secret sharing," *IEEE Transactions on Information Theory*, vol. 39, pp. 1121-1132, 1993.

[6] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 644–654, 1976.

[7] A. Salomaa, "Kriptografiya s otkrytym klyuchom," 1995. (in Russ.)

[8] N. Asokan and P. Ginzboorg, "Key-agreement in ad hoc networks," *Computer Communications*, vol. 23, no. 17, pp. 1627-1637, 2000.

[9] N. Satish G., Ch.V. Raghavendran, P.T.K. Mehar, Dr. Varma P. Suresh, "Secret key cryptographic algorithm," *Intern. J. of Computer Science*, vol.1, no. 1-2, 2012.

[10] R. D. Pietro, L. V. Mancini, and S. Jajodia, "Efficient and secure keys management for wireless mobile communications," *Proc. of the 2nd ACM Intern. Workshop on Principles of Mobile Computing, Toulouse, France*, pp. 66-73, 2002.

[11] B. Zhu, F. Bao, and et al, "Efficient and robust key management for large mobile ad hoc networks," *Comp. Networks*, vol. 48, pp. 657-682, 2005.

[12] Y. Liang, H. V. Poor and S. Shamai, "Information Theoretic Security," *Foundations and Trends R in Communications and Information Theory*, vol. 5, no. 4-5, pp. 355-580, 2008.

[13] I. A. Zikratov, S. V. Odegov, "Otsenka informatsionnoy bezopasnosti v oblaschnykh vychisleniyakh na osnove bayyesovskogo podkhoda," *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki*, vol. 80, no 4, pp. 121–126, 2012.

[14] Z. Zh. Zhanabayev, "Kriterii samopodobiya i samoaffinnosti dinamicheskogo khaosa, Recent Contributions to Physics*, vol. 44, pp. 58-66, 2013.

[15] Z. Zh. Zhanabayev, S. M. Mukhamedin, A. K. Imanbaeva, "Information criteria for the degree of turbulence self-organization," *Russian Physics Journal*, vol. 44, pp. 756–762, 2001.

[16] A. K., Imanbayeva, R. N., Syzdykova, A. A. Temirbayev, "Concept formulation and university teaching methodology for dynamic chaos," *Journal of Physics: Conference Series*, vol. 1136, pp. 012029, 2018.

[17] A. Imanbayeva, A. Temirbayev, R. Syzdykova, N. Saduev, "Practical experience of using problem training in the formation of a knowledge system for students in the field of information security," *EDULEARN18 Proc. Palma, Spain*, pp. 9919-9924, 2018.