On the Condition of Covering Completeness in Associative Steganography

V A Raikhlin¹, I S Vershinin¹, R F Gibadullin¹

¹Department of Computer Systems, Kazan National Research Technical University named after A. N. Tupolev – KAI, 10, K.Marx St., Kazan, Tatarstan 420111, Russia

E-mail: landwatersun@mail.ru

Abstract. Previously, the completeness of the coverage was considered by the authors as an interpretation of a kind of logical interpretation of the criterion of perfect secrecy by K. Shannon – one of the necessary conditions that the concept of associative steganography should satisfy. The article shows that in reality, the completeness of coverage is a completely independent condition for ensuring the required level of data protection when analyzing scenes. It should not be associated with the criterion of K. Shannon.

1. Introduction

Steganography has recently been taking serious attention of specialists in information protection [1-7]. Well-known methods of steganographic transformation do not provide unconditional stegostability (in practice, they do not completely hide the fact of information transfer). There are several theoretical considerations of the so-called perfect stegosystems [8,9]. But the issues of their practical implementation have not been investigated. The issue of noise immunity remains open when storing and transmitting hidden data over open communication channels without the use of additional noise-resistant coding.

Associative steganography is designed to eliminate these gaps. It is focused on data protection when analyzing scenes. Scene analysis means an enlarged description of images in terms of "objects – coordinate" [10]. In contrast to the known methods of steganographic transformation, the associative approach can provide almost unconditional stegostability, and in comparison with the known ciphers, higher noise immunity when storing and transmitting information over open communication channels.

Elements of the theory and practice of associative steganography are developed in the works [11-14] and several others. There is not much point in repeating yourself. And yet, some points should be noted for a better understanding of what follows.

In this case, the k-bit decimal encoding of the names of objects and their coordinates is used with special characters. The initial information on the scene is structured as a table with many records of the form: <Object code><Code of X coordinate><Code of Y coordinate>. Each decimal digit is represented by its binary matrix-etalon with the dimensions m×n, m=2ⁿ-1. The sizes of all etalons are the same. Many such matrices with a power of 10 are masked. The process of generating masks is random. For each matrix, its matrix of masks of the same size is created, which saves in the standard the bits that are essential for its further identification. A set of masks is a recognition key.

The masked bits are randomized. As a result, the original binary matrix-etalons are transformed into ternary matrices, the elements of which are {0,1, •}. Recognition of the decimal digits of each code is
performed by comparing a plurality of ternary standards according to the positions of the stored bits. That is the essence of associativity in this case.

Each code (object or coordinates) as a result of matrix binarization of decimal digits with subsequent masking and randomization is transformed into a k-section stegocontainer, which is formed by the following process. First, a so-called empty container is created with a length of \( L = k \times (9^n - 12) \) according to the number of significant bits of binary standards of decimal digits. It is filled with a segment of a pseudo-random sequence (PRS). Then the code bits accidentally saved by masking are inserted into it by positions. Regardless of \( n \), the average number of such bits \( 5k \ll L \), which is typical for steganography.

The completeness of the coverage is a necessary condition that must be satisfied by the concept of associative steganography. It determines the choice of the PRS generator and the sizes of binary code character matrices to obtain the desired randomization on the first attempt.

2. Justification of the condition of completeness of coverage

For a long time, the question was discussed with many specialists: what are we dealing with in this case with steganography or cryptography? The essence of the discussion is as follows. Cryptography is usually associated with serious mathematical transformations of messages with practical preservation of their length. Useful information for associative protection is randomly (by positions) small bits of the original message stored by masking, embedded in a relatively large PRS carrier, which is typical for steganography. But the representation of decimal digits by binary matrices (which determine the size of the carrier) with subsequent masking and randomization is a kind of encryption. Although it is different from the accepted cryptography, still ... . In our opinion, associative steganography embodies a kind of symbiosis of the concepts of steganography and cryptography.

The condition of completeness of coverage is associated precisely with the cryptographic aspect of the approach and is an important condition for the effective organization of the associated data protection of the scene. It is formulated as follows.

For all codes of objects/coordinates of the scene obtained by masking binary matrices of code symbols and randomizing them, when recognizing each code, as a result of not necessarily a complete search of keys, the full set of codes possible for this scene should "appear".

Then, if the number of possible object codes/coordinates of the analyzed scene is \( T/G \) and \( M \) objects are concentrated on it (including "empty" ones that are introduced to increase the level of protection), then the code of each of the \( M \) objects can be "any of \( T \) possible", and the coordinate codes can be "any of \( G \) possible". In this case, \( T = G = 10^6 \).

How much can the satisfaction of this condition help to increase the real resistance of protection? Queries to the scene database are selective: which objects and by which coordinates are present in a certain part of the scene. And there are usually a lot of objects on such a site. The number of stegocontainers for this section that form a message is three times greater: each container contains an encrypted k-bit decimal code – only the name of the object or only one of its coordinates. The true codes of all objects of communication and their coordinates in the aggregate will appear only on the correct key.

The total number of all possible code sequences in this case of \( R = (10^6)^M \). For \( k = 3, M = 30 \) (usually the value of \( M \) is much larger), we obtain \( R = 10^{270} \), which exceeds the number of possible keys (for example, for the recommended value of \( n = 60 \), the upper estimate for the number of keys is \( 10^{29} \) [12]). Then the set of valid code sequences will narrow down.

In general, the manifestation of the sequences of entities encoded by them will distinguish among them a non-unique subset of a posteriori truthful messages according to the criterion of semantic (not numerical) correspondence to the nature of the scene (as an example: a whale swims in the sea, and a cat sits on a tree, but not vice versa). In this case, it is theoretically impossible to unambiguously identify the true key by a complete search. Of course, we should not forget about the effect of possible attacks that can reduce the level of protection to a provable level of invincible computational complexity.
Here is an example from the field of cartography. Test map: a plot of 300×300 km² of the Republic of Chuvashia (Fig. 1). It contains 1035 point objects of 4 different types (forest, vegetation, garden, bus stop). Provided by the geodetic company "Zenit", Kazan. Gradation parameters: the step of the global coordinate grid is 300 m, the step of the local coordinate grid is 0.3 m, the number of coordinate gradations in the local and global areas of the map is 1000.

![Figure 1. Test map.](image)

The results of visualization of the test map layer using the MapInfo geoinformation system after hiding it and recognizing it on the true key and one of the false keys are shown in Figures 2 and 3, respectively. The pictures are different, but equally plausible if it is not known what is in the area of interest to us.

![Figure 2. Results of visualization of the map opened on the true key.](image)
Figure 3. Results of visualization of the map opened on the false key.

In some special cases, for example, when analyzing protected text messages, a coherent text can appear only on the true key. Of course, it is impossible to assert the absence of a plurality of plausibility in this case, too, because of the significance of the R. But always a positive answer to the question of how much the criterion of completeness of the coating remains valid here is indirectly given by the fact that the method is unconditionally quilted when this condition is satisfied [14, 15].

3. Was the connection with the K. Shannon criterion justified?
At the early stages of the study, we were engaged in the recognition of masked k-bit decimal codes as such. Oddly enough, the idea of the completeness of coverage arose in association with the criterion of perfect secrecy by K. Shannon [16]. This criterion is determined by the condition:

For all elements of the set of messages, they are a posteriori probabilities according to the results of a complete search of the keys should be equal to the a priori probabilities, regardless of the magnitude of these latter. In this case, the interception of the message does not give any information to the unauthorized user.

But in the simplest case, $\gamma = 3$, after representing decimal code digits with binary matrices, applying masking and randomization mechanisms to them, the a posteriori probabilities of recognizing a complete set of codes in one stegocontainer turned out to be different. This led to the idea of replacing the term "equal probability" in the K. Shannon criterion with the term "equal likelihood" in the sense that any of the $\gamma^k$ codes obtained as a result of a complete search is a sequence of bits without any features. This was called the logical interpretation of the criterion. By inertia, the name was preserved in the works [11-14] as a justification for the completeness of the coverage.

Legitimate questions: 1) If such an interpretation is fair, then how can we accurately assess its plausibility? For some small-sized problems, a comparative digital assessment of the degree of plausibility of a particular event is possible [17]. But not in this case. 2) Was it necessary to introduce a logical interpretation of the K. Shannon criterion into associative steganography, especially since the meaning of a priori implausibility is far from clear? The answer is unambiguous: it is not necessary, because the concept of completeness of coverage, as shown earlier, is self-sufficient.
4. Conclusion
Achieving the completeness of the coverage is necessary for associative steganography. The article justifies this provision without involving the concept of the logical interpretation of the K. Shannon criterion introduced in previous articles. There was no need for such an interpretation.

Gratitude: the authors are grateful to Professor A.V. Babash for his critical comments, which made us think about the consistency of the logical interpretation of the K. Shannon criterion and the need to use it to justify the completeness of the coverage.

5. References
[1] Gupta S and Jain R 2015 An Innovative Method of Text Steganography in Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP) (Solan, India) pp 60-64
[2] Mandal, Kunal & Koley, Dr. Santanu & Dhar, Sudipto 2016 A Mathematical Model for Secret Message Passing Using Steganography 10.1109/ICCIC.2016.7919527
[3] Seo J O, Manoharan S and Mahanti A 2016 Network steganography and steganalysis - a concise review 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) pp 368-371 doi: 10.1109/iCATCCCT.2016.7912025
[4] Wang Z and Zhang X 2019 Secure Cover Selection for Steganography in IEEE Access vol 7 pp 57857-57867 doi: 10.1109/ACCESS.2019.2914226
[5] Yang Z, Huang Y and Zhang Y 2019 A Fast and Efficient Text Steganalysis Method in IEEE Signal Processing Letters vol 26 4 pp 627-631 doi: 10.1109/LSP.2019.2902095
[6] Korzhik V, Duy N 2019 Cuong and G. Morales-Luna, "Cipher Modification Against Steganalysis Based on NIST Tests 2019 24th Conference of Open Innovations Association (FRUCT) pp 179-186 doi: 10.23919/FRUCT.2019.8711958
[7] Yeung Y, Lu W, Xue Y, Huang J and Shi Y 2020 Secure Binary Image Steganography With Distortion Measurement Based on Prediction in IEEE Transactions on Circuits and Systems for Video Technology vol 30 5 pp 1423-1434 doi: 10.1109/TCSVT.2019.2903432
[8] Cachin C 1998 An Information-Theoretic Model for Steganography IACR Cryptol. ePrint Arch. 28
[9] Ryabko Boris & Ryabko Daniil 2007 Information-Theoretic Approach to Steganographic Systems pp 2461-2464 10.1109/ISIT.2007.4557588
[10] Duda R & Hart P 1973 Pattern classification and scene analysis A Wiley-Interscience publication
[11] Raikhlin V A, Vershinin I S, Gibadullin R F & Pystogov S V 2013 Reliable recognition of masked binary matrices Connection to information security in map systems Lobachevskii Journal of Mathematics 34(4) 319-325
[12] Raikhlin V A, Gibadullin R F, Vershinin I S and Pystogov S V 2016 Reliable recognition of masked cartographic scenes during transmission over the network 2016 International Siberian Conference on Control and Communications (SIBCON) pp 1-5 doi: 10.1109/SIBCON.2016.7491657
[13] Raikhlin V A, Vershinin I S & Gibadullin R F 2019 The Elements of Associative Steganography Theory MoscowUniv.Comput.Math.Cybern. 43 40–46
[14] Vershinin I S, Gibadullin R F, Pystogov S V et al 2020 Associative Steganography Durability of Associative Protection of Information Lobachevskii J Math 41 440–450
[15] Vershinin I S, Gibadullin R F, Pystogov S V et al 2021 Associative Steganography of Text Messages MoscowUniv.Comput.Math.Cybern. 45 1–11
[16] Shannon C E 1949 Communication theory of secrecy systems in The Bell System Technical Journal vol 28 4 pp 656-715 doi: 10.1002/j.1538-7305.1949.tb00928.x
[17] Raikhlin V A 2015 Remarks on temporal-fuzzy clustering Nelinei’ny i’ mir vol 13 8 pp 48-53