ASSOCIATIVE STEGANOGRAPHY. DURABILITY OF ASSOCIATIVE PROTECTION OF INFORMATION

I. S. Vershinin,[1, *] **R. F. Gibadullin,**[1, **]
S. V. Pystogov,[1, ***] **and V. A. Raikhlin**[1, ****]

(Submitted by A. A. Editor-name)

[1]*Institute of Computer Technology and Information Security, Kazan National Research Technical University named after A.N. Tupolev–KAI, Karl Marx st. 10, Kazan, Tatarstan, 420111 Russia*

Received October 15, 2019

Abstract—The case of analysis of associatively protected cartographic scenes is considered. Protection of objects and their coordinates is achieved masking binary matrices of their code symbols. The set of inverse mask matrices is the recognition key. This allows such protection to be attributed to associative steganography. A message is considered to be unconditionally steadfast if it is statistically indistinguishable from a random sequence. Therefore, the study of its steadfastness is carried out using statistical tests of randomness NIST. If a pseudo-random sequence successfully passes the test of all 15 tests, then it is considered random («white»). If there is a failure at least on one test, then it is considered «black». But in the case of the application of the basic masking algorithm, this cannot help the disclosure of the stegomessage. The effect of masking redundancy introduced with the aim of improving noise immunity on the durability of mapping objects to the effects of various attacks is considered. It is established that associative steganography retains the property of provable (computational) stability in this case as well. Recommendations for its use to protect the text characteristics of objects are given.

2010 Mathematical Subject Classification: 68P25, 68P30

Keywords and phrases: *associative steganography, basic masking algorithm, NIST testing, steganographic durability, excessive masking, cryptographic durability, text protection*

1. INTRODUCTION AND OBJECTIVES OF THE ARTICLE

Systematization of elements of the theory and practice of associative steganography was carried out in [1–3]. Below are some of the provisions of this theory, familiarity with which is useful for understanding main content of the article.

Associative steganography is a method of protecting information with the following features:

1. Considers a finite set of conversations represented by a k-bit decimal code. The decimal digits of the message codes are styled to binary matrix patterns equal sizes $m \times n$, m = 2n–1. (Fig. 1 – example symbol representations for the n=5). Transfer (storage) any of the messages of given set is considered a priori equally probable. The received message is recognized (different) by comparing each matrix on the full set of patterns.

2. Many such matrices is subjected to masking. For each matrix, the matrix of mask of the same size is created, which keeps bits in the pattern. These bits are necessary for further identification of the patterns (the mask determines the elements of the matrices to which arbitrary values from 0, 1 can be assigned in the process of further randomization). The process of generating masks is random. The generated set of masks is the key. Regardless of n, the developed masking algorithm

* E-mail: vershinin_igor@rambler.ru
** E-mail: landwatersun@mail.ru
*** E-mail: sergey.pystogov@gmail.com
**** E-mail: no-form@evm.kstu-kai.ru

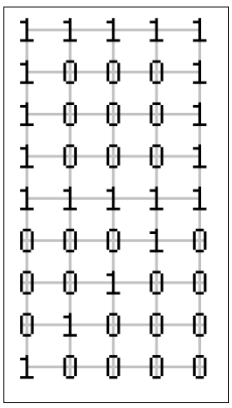

Figure 1. Example of symbol representations for the n = 5

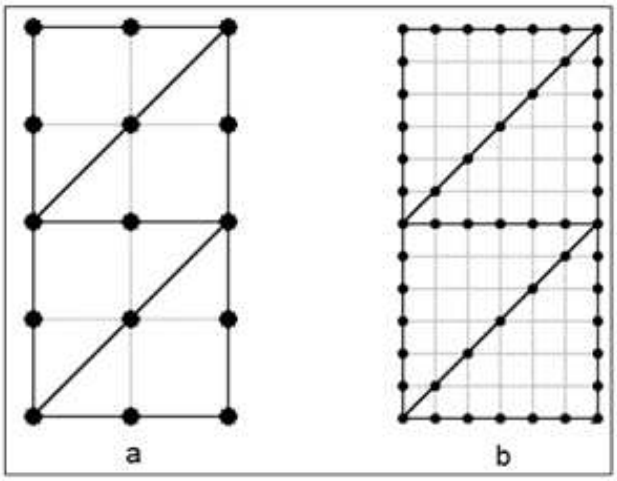

Figure 2. Distribution of significant bits along the contours of patterns

leaves in each digit from 1 to 8 significant bits from (9n–12) essential elements of these matrices (shown by dots in Fig. 2).

Many such bits are randomly distributed along the contour of selected symbols depending on the type and character of the implemented option of masking. The number of «free» elements and potential durability of protection increases with increasing n (Fig. 2a: n=3; Fig. 2b: n=7).

3. The purpose of masking is to satisfy, firstly (in theory), the criterion of perfect secrecy according to C. Shannon [4] (equality of prior transmission probability and posteriori probability of disclosure as a result of complete busting of the keys when receiving) in its logical interpretation (replacing the term «equiprobably» with the term «equitable») and, secondly, the requirements of high interference sustainability of the message transmission. Both suppose each matrix is immersed by its mask to stegocontainer, initially filled with a binary pseudo–random sequence — gamma, the length of which is always much more than the number of stored bits. The dimensions of the matrices and gamma are selected from the condition of satisfying given set of messages to the specified criterion.

4. The prerequisite for the durability of protection according to Shannon is to perform randomization in such way that without knowing the key in each hidden object it will be possible to recognize any of the elements of the used dictionary with any coordinates on the set of randomly generated sets of masks. In the case of scene analysis [5], the fulfillment of this condition can be ensured only if the set of involved object codes and gradations of their coordinates for given k are complete. In other words, the power of this set is G = 10k. If the real number of object types and

gradation of coordinates is less than G, then «empty» objects and coordinates are entered that are «eliminated» in the recognition process.

The search time for the suitable randomization («carrier» gamma) depends on the used pseudo-random sequence generator, the values of n and k. When choosing a pseudo-random sequence generator «Mersenne Twister» [6], n = 60, k = 3, complete coverage with one attempt on sets stegocontainers with various sets of masks occurs with a probability of 0.999. The computational (provable) durability of modern crypto and stegosystems is associated with the impossibility of identifying (as a result of various attacks) the true key in a reasonable time even using supercomputers. Such durability for associative stegosystems with the specified choice is established.

The principal feature of stegosystems is the difficulty in establishing the very fact of the transfer of messages, associated with the concept of steadfastness. It is unconditional if difficulties are transformed into practical impossibility. Steganography is characterized by a significant excess of the length of the carrier L over the volume of the message. Guaranteed to detect the fact of inserting into a container is possible if their volume grows faster $\sqrt{L}$ [7]. In this case, this is indeed the case if the growth of L is associated with an increase in n [3]. But with the constancy of n and the increase in the number of containers in the stegomessage, both grow linearly. One of the objectives of this article is to identify the conditions under which there is an unconditional durability. Another task is associated with stego- and cryptanalysis in the case of excessive masking, when several sets of masks are simultaneously used to hide messages in order to improve noise immunity during storage and transmission of information.

2. STEADFASTNESS OF ASSOCIATIVE PROTECTION

The message is statistically indistinguishable from a random sequence. Therefore, the study of steganographic durability is carried out further using the NIST statistical tests of randomness [8–11], which was developed in 1999 as part of the AES (Advanced Encryption Standard) project. The NIST package includes 15 statistical tests that are designed to test the randomness hypothesis for binary sequences of arbitrary length generated by a random (or pseudo-random) number generator. All tests are aimed at identifying various randomness defects (see Appendix 1). The basic principle of testing is to test the null hypothesis H_0, which consists in the fact that the test sequence is random. An alternative hypothesis for H_a is the hypothesis that the sequence being tested is not random. According to the results of the application of each test, the null hypothesis is either accepted or rejected. The decision on whether a given sequence of zeros and ones will be random or not is made based on the totality of the results of all tests.

The order of testing a separate binary sequence S:
1. The null hypothesis H_0 is put forward – the assumption that this binary sequence S is random.
2. The sequence S calculates the statistics of the test c(S), which is constructed in such a way that its large values indicate any defect in the randomness of the sequence.
3. With the use of the additional function f errors and test statistics calculated the probability of the hypothesis H_0 justice:

$$P = f(c(S)), P \in [0; 1].$$

4. The probability value P is compared with the significance level α. If $P \geq \alpha$, then the hypothesis H_0 is accepted. Otherwise, the hypothesis of H_a is considered fair.

The following are the results of series of experiments with the recommended choice of $\alpha = 0.01$ [8] for all 15 tests. If the pseudo-random sequence passes the test of all NIST tests, it is considered random («white»). If there is failure at least on one test, then it is considered «black». The segment of a pseudo-random sequence «Mersenne Twister» with the length of 594 KB is taken as the «carrier» gamma. With k = 3, n = 60 and analysis of cartographic scenes in terms of «objects-coordinates», this corresponds to the limiting case of storing/transmitting stegomessages about 10^3 objects. The probabilities P^{WG} and P^{WS} of generating «white» gammas and receiving «white» stegomessages after bit inserts by mask into one of the «carrier» gammas are evaluated with one or another of its choices.

Series 1. 100 iterations were carried out for generating N = 1000 gamma segments and calculating the probability P^{WG} = M/N, where M is the number of gamma segments that passed NIST test.

For the obtained general set of probability values the expected value $\overline{P^{WG}}$ and standard deviation $\sigma^{P^{WG}}$ are calculated.

$$\overline{P^{WG}} = \frac{1}{N}\sum_{i=1}^{N} P_i^{WG} = 0,44951;$$

$$\sigma^{P^{WG}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(P_i^{WG} - \overline{P^{WG}})^2} = 0,015425608.$$

Series 2. 100 iterations on the choice of any gamma segment that passed the NIST test were carried out with this segment of N = 1000 experiments in which stegoinserts were performed on different sets of masks (keys), and probability calculations $P^{WG\to WS}$, where L is the number of stegosegments that passed NIST test. It is counted:

$$\overline{P^{WG\to WS}} = \frac{1}{N}\sum_{i=1}^{N} P_i^{WG\to WS} = 0,49778;$$

$$\sigma^{P^{WG\to WS}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(P_i^{WG\to WS} - \overline{P^{WG\to WS}})^2} = 0,204357608.$$

Series 3. It was carried out similarly to series 2, but with the NIST test that did not pass. Results of the series are

$$\overline{P^{BG\to WS}} = \frac{1}{N}\sum_{i=1}^{N} P_i^{BG\to WS} = 0,22007;$$

$$\sigma^{P^{BG\to WS}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(P_i^{BG\to WS} - \overline{P^{BG\to WS}})^2} = 0,218848144.$$

Series 4. 100 iterations were carried out on the generation of a gamma segment, conducting with this segment (without prior sampling) N = 1000 experiments in which stego-inserts were performed on different sets of masks (keys), and probability calculations $P^{\forall G\to WS} = L/N$ where L is the number of stegosegments that passed NIST test. In this case, we have:

$$\overline{P^{\forall G\to WS}} = \frac{1}{N}\sum_{i=1}^{N} P_i^{\forall G\to WS} = 0,34014;$$

$$\sigma^{P^{\forall G\to WS}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(P_i^{\forall G\to WS} - \overline{P^{\forall G\to WS}})^2} = 0,245837222.$$

The aggregated results of the experiments are presented in table 1.

Graphs of these functions are plotted on Figures 3 and 4 for the specified general populations. They clearly illustrate the significant differences in their statistical properties.

A «white» message is equivalent to the action of random sequence and therefore unconditionally steadfastly by condition. As established, with a single masking, the maximum likelihood of receiving a steganographically stable message occurs when choosing the «white» gamma as the carrier. Therefore, this choice is preferable. But if the message is «black» then finding out if it is not «black» sequence of «empty» containers without stegoinclusions is problematic. For this, it would be necessary to compare the results of stegoanalysis with the statistics for the corresponding «black»

Table 1. Aggregated results of experiments

General population	Estimation of mathematical expectation	Standard deviation estimate
Probability of obtaining «white» gamma	0,44951	0,015425608
Probability of getting «white» stego based on «white» gamma	0,49778	0,204357608
Probability of getting «white» stego based on «black» gamma	0,22007	0,218848144
Probability of obtaining «white» stego on the gamma without prior sampling	0,34014	0,245837222

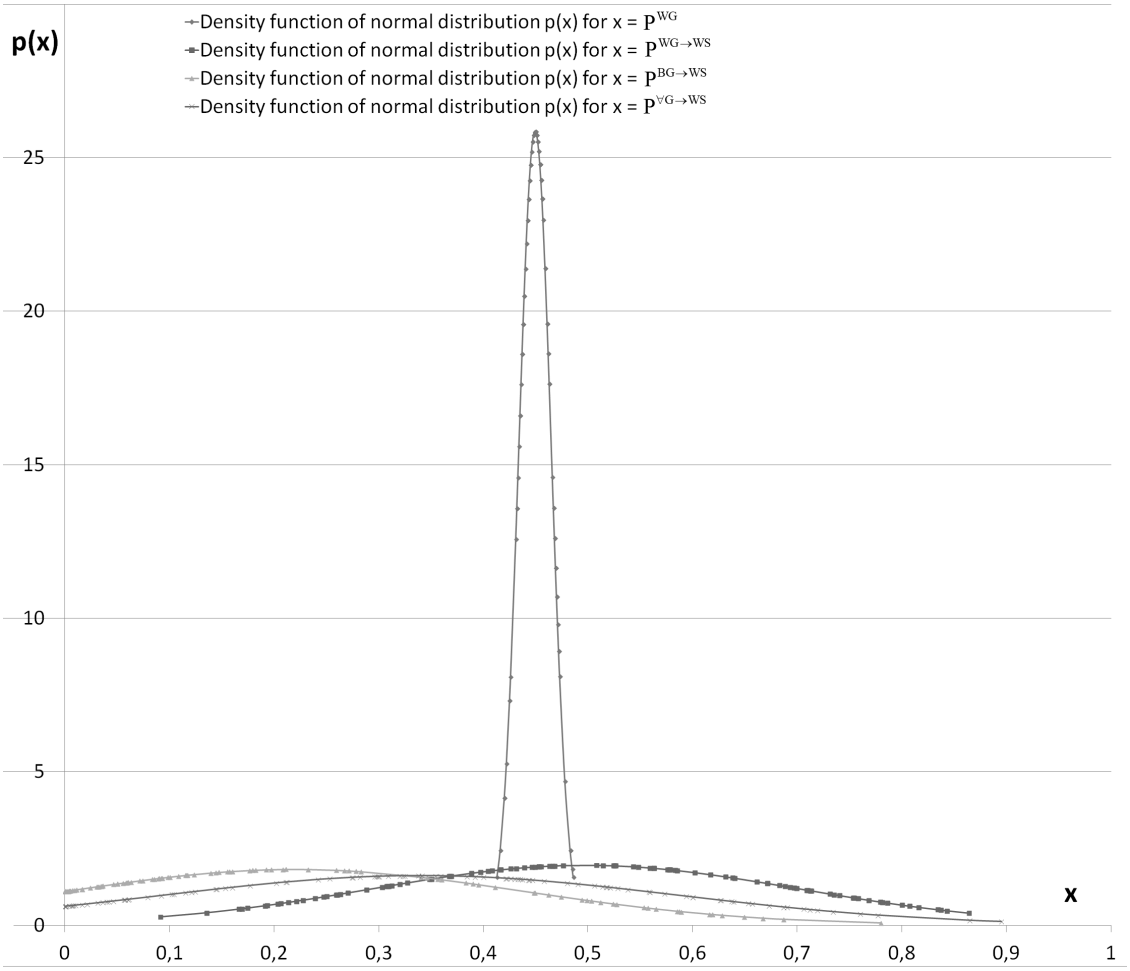

Figure 3. Graphs of probability density functions

gamma. However, the adversary can not have results similar to those shown in Table 1 and Figures 3-4 for single sequences of arbitrary length (he will need data for the case of double masking). Therefore, he will not be able to compare the behavior of the functions p(x) and F(x) for gamma and the received message, and information about the computational steadfastness of the method will always be necessary to analyze any «black» sequence.

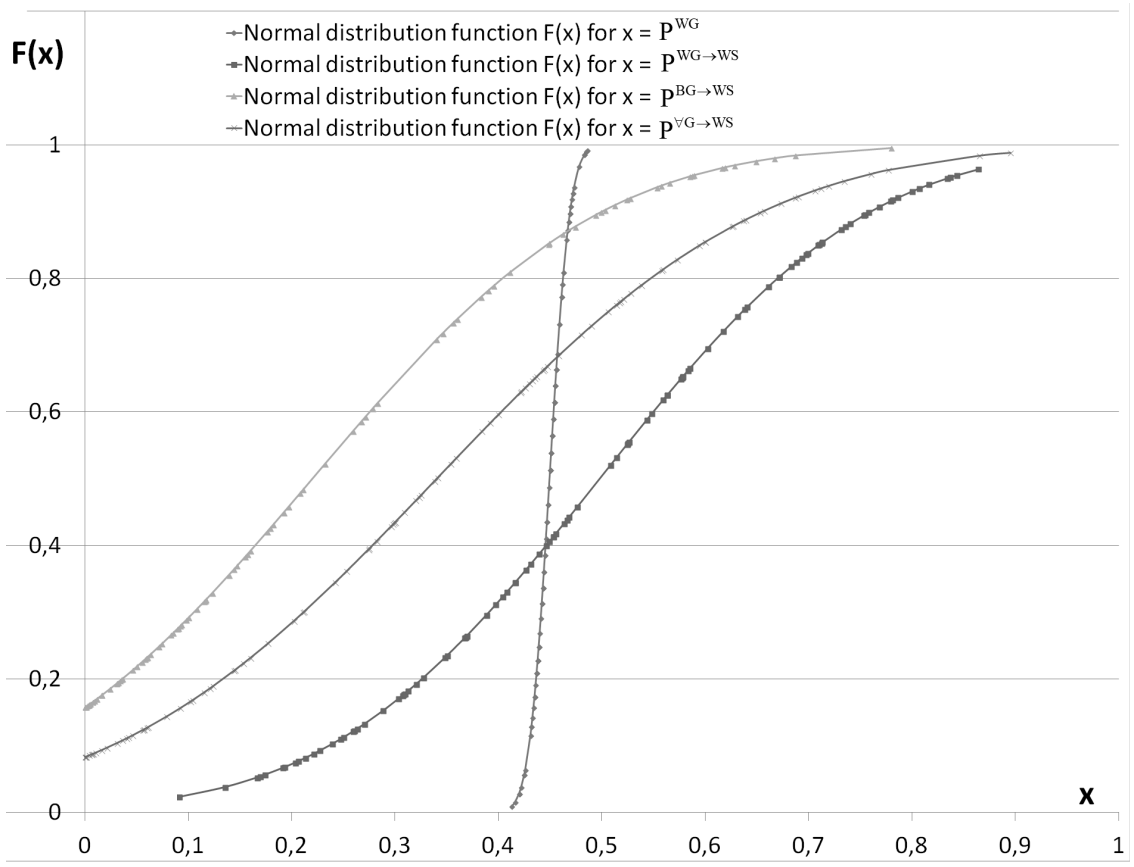

Figure 4. Distribution function graphs

In practice stegomessages lengths may be different. Under the same experimental conditions for statistical gamma and stegosegments of sizes 66, 198 and 396 KB, statistical estimates were established for populations taking values P^{WG}, $P^{WG\to WS}$. They are shown in Table 2.

Table 2. Statistical estimates for different sizes of pseudo-random sequence

	66 KB		198 KB		198 KB	
	$\overline{P}$	σ^P	$\overline{P}$	σ^P	$\overline{P}$	σ^P
Values P^{WG}	0,046	0,006	0,232	0,013	0,375	0,016
Values $P^{WG\to WS}$	0,254	0,138	0,427	0,188	0,493	0,193

Thus, when choosing as carrier «white» gamma, good steadfastness of associative protection is retained for stegomessages with length of not less than 200 KB.

3. THE EFFECT OF REDUNDANT MASKING ON CRYPTOGRAPHIC DURABILITY OF ASSOCIATIVE PROTECTION TO VARIOUS ATTACKS

The monograph [14] summarizes the results of the original researches of cryptographic durability of associative protection of cartographic scenes in the case of single key (set of masks) for the following attacks:
– Brute force attack.
– Associations with map.

– Attack on the gamma.
– Plaintext attack.

3.1. Case of single masking

This subsection briefly provides information about the previously obtained results.

Brute Force Attack. The computational steadfastness of modern crypto- and stegosystems is associated with the impossibility of a complete brute force of keys in order to identify the true key in a reasonable time even with the use of supercomputers. The number of possible keys for the protection method plays a key role.

Stegokeys are generated by the basic masking algorithm with different sizes of matrix patterns, which are represented in the alphabet of zip codes. An upper estimation of the number of stegokeys was obtained experimentally. The results of the experiment are shown in table 3.

Table 3. Upper estimation of the number of stegokeys

n	**3**	**18**	**30**	**40**	**60**
Number of stegokeys	10^{13}	10^{23}	10^{25}	10^{27}	10^{29}

According to the table, an increase of n by 1.5 times leads to keys number increase by 2 order of magnitude. Therefore, in order to compare, for example, with the AES algorithm (128 bits key, number of keys 10^{38}), a choice of n≈450 is necessary. It will increase the size of the contours of the matrix patterns to 4038 bits. Size of the key (set of masks for 10 patterns) will be 40380 bits, which is excessive. However, in practice, it is possible to use a smaller key. For cartography it is important not only to recognize the objects of a certain area, but also to analyze the plausibility of the obtained results. This process may require significantly more time than the keys search. If we set the time for obtaining and analyzing the result of applying one key to 1 nanosecond, then a full search of keys with n = 60 will take 10^{20} seconds, i.e. about 3×10^{12} years. Thus in the case of single masking, the method is provably resistant to brute force attack.

Associations with map (attack with known object / absence of object). Possible associations with the terrain may reduce the uncertainty of the plaintext, thereby increasing the possibility of successful identification by an unauthorized user. Two cases of possible associations were considered:
1.The adversary knows one object (its type) and the coordinates of this object on the map.
2.Association as such, when the object cannot be located in a certain place on the map from the standpoint of вЂњcommon senseвЂќ.

In the first case, the adversary task is to search for keys with the dropping those that do not provide correct recognition. At the same time, on each key it is necessary to recognize the global coordinates of all clusters [1] with the selection of those of them in which a known object may be located. Next step is the recognition of objects in selected clusters.

To implement this approach, the experiment with brute force of 1.6×10^6 keys was performed. No key, other than the true one, matches the codes for each of the objects on the map (object code, local coordinates) with the codes of a known object. We believe this is sufficient reason to consider the situation in this case like the full brute force.

The second case is identical to the full brute force.

Attack on the gamma. Baseline: There are many randomized containers with three sections (letters) in each. Used letters (patterns) are represented in the alphabet of zip codes by binary matrices of the size $m \times n = 79 \times 40$ with a «significant contour» length of 348 bits. The carrier gamma is formed by the pseudo random sequence generator «Mersenne Twister». The value of the pseudo-random sequence bit, marked by the one-bit of the corresponding inverse mask matrix, is replaced by the true value of pattern matrix.

If a segment of a pseudo-random sequence used to randomize the set of containers has become known to the adversary, then to find the true key (set of masks) the gamma could be attacked with the following algorithm.

Algorithm:
1. For each randomized letter of all stegocontainers, the corresponding letter bits and the pseudo-random sequence bits are added modulo 2.
2. Identified mismatch bits are considered true bits.
3. The number of such bits and their coordinates are determined for each letter.
4. i=1. Assume that the first level contains all analyzed letters. All of them are not initially marked.
5. Among all the letters of the current i (with i = 1 only for unmarked letters), the letter with the most true bits is searched (marking by ${B^i}_{max}$). ${B^1}_{max}$ is marked. If all letters of level with i = 1 have marks, go to step 10.
6. True bits of selected letter ${B^i}_{max}$ are compared with the corresponding bits of all other letters (only unmarked, when i = 1) on current level. Marks of all unmatched letters are canceled when $i \neq 1$.
7. Each matched at the previous step letter ${B^i}_j$ of the i-th level compared with ${B^i}_{max}$ in the same way: the true bits of ${B^i}_j$ are compared to bits of ${B^i}_{max}$. The mark of the letter ${B^i}_j$ should be canceled in the case of a mismatch and when $i \neq 1$. In case of complete match the letter ${B^i}_j$ is taken to the next level (i + 1). If i = 1, then this letter is marked.
8. i = i + 1. If the i-level has more than one letter taken out on previous step, go to step 5. Otherwise, i = 1.
9. All letters marked after comparison with the current letter ${B^1}_{max}$, should be referred to one set. Go to step 5.
10. For each of the ten sets found, a mask matrix is formed based on the coordinates set of the true bits of all the letters of the set.
11. END.

Researches have shown that to achieve the correct assignment of the identical letters to one set and identify all the bits of the masks being formed is possible with the number of stegocontainers equal to 33 (99 letters). With smaller number of letters single errors of attribution to the wrong set or detection of not all bits of the masks were observed. For the case of 33 stegocontainers, 102 experiments were carried out according to the algorithm described above. Different masks and pseudo-random sequence segments for randomization were used during experiments. No errors were detected. The found number of stegocontainers (33) is the uniqueness distance [15], i.e. the average length of the stegotext necessary for unambiguous restoration of the true key. However, the time complexity of the experimental determination of the true gamma is 10^{5966} years.

Plaintext attack (with known / selected message). Let the plaintext is the sequence (1 2 3 4 5 6 7 8 9 0), represented in the zip code alphabet. The adversary knows the stegotext of this message, obtained using the basic masking algorithm and the randomization procedure. It is required to estimate the possibility of finding the stegokey (true set of masks). Research results helped to establish a set of keys, on base of them the corresponding stegotext can be obtained from the source text. The number of keys is presented in table 4.

Table 4. Number of keys, necessary for source text recognition

Pattern matrixes size n	**3**	**60**
Full set of keys	10^{13}	10^{29}
Basic masking algorithm	10^3	10^{14}
Modified masking algorithm, single masking	10^3	10^{14}
Modified masking algorithm, redundant masking	10^3	10^{14}

But the situation when the adversary has information about the stegogram with all 10 characters is unlikely. Therefore, the situation also analogous to the full search.

3.2. Redundant masking case

To improve the noise immunity of associative protection was proposed [9] to generate Q sets of masks which are disjunctively combined during immersion in stegocontainers. Recognition of the received stegomessage is carried out on all sets of masks. For i-stegosymbol a pattern is taken for the result of recognition if the number of recognitions of this pattern is $r_i \geq](Q+1)/2[$. If this condition is not met, then the recognition will be marked as failed. Provable durability of associative protection to various attacks with a single masking is established [11]. Our further task is to show that this property is preserved even with redundant masking.

Brute force attack. In case of redundant masking the probability of random finding any true key on set of five or seven used keys (minimum recommendations) increases five or seven times to 5×10^{-29} or 7×10^{-29}, respectively. Therefore, the redundant masking practically does not provide advantages for the implementation of this attack in comparison to case with one key.

COMMENT. The key that allows to correctly open a single stegomessage (one stegocontainer), in general, is not unique. This is due to the properties of randomization. However, with a large number of stegocontainers, there is a high probability of incorrect recognition of some part of them. In other words, the recognition result on false keys will be incorrect. Therefore, the given estimate of the probability of random formation of a correct key is not in doubt.

Associations with the map. As in the case of a brute force attack, an increase in the number of true keys with redundant masking does not provide practical advantages. Brute force is required. Therefore, provable durability is preserved.

Attack on gamma. To determine the distance of uniqueness with redundant masking (the case of Q = 5), another software experiment on base of above algorithm was performed. A set of 33 stegocontainers obtained by randomization with five sets of stegokeys was used as the initial information. It should be noted that in this case we are talking about checking the possibility of finding an aggregate mask for each pattern. An aggregate mask is the result of a disjunctive bitwise combining of all used mask sets for a given pattern.

Researches have shown that it is not always possible to identify all the bits of the formed aggregate masks (Table 5) with the number of stegocontainers equal to 33 (99 letters). Therefore, the introduced redundancy does not impair the previously obtained estimates of cryptographic durability for this attack.

Table 5. Aggregate masks finding

Masking	**No redundancy**	**With redundancy**
Number of experiments	100	100
Number of successful implementations	100	27
Number of erroneous implementations	0	73

Cleartext attack. According to the table 4, using of redundancy does not reduce the set of keys in comparison with the non-redundant case.

4. CONCLUSION

As a result of the research it is founded:
1. For the basic masking method, when using the «Mersenne Twister» pseudo-random sequence, regardless of the choice of gamma, the sequence of empty and stegocontainers differ significantly in the form of functions p(x) and F(x).
2. On average, when using a pseudo-random sequence generator «Mersenne Twister» is selected as the carrier of «white» and length stegomessage range of not less than 200 KB, for the basic method approximately 50% stegocontainers sequences remain «white», thereby certainly are steganographically steadfast. The choice of such carrier is preferred.
3. If for some message NIST test result is «black», that almost always occurs when excess masking

is used to improve the noise immunity of storage and transmission stegomessages, in any case it is advisable to pay attention to the results of the analysis of computational stability.
4. On the set of considered attacks, associative steganography retains the property of computable (provable) durability even in the case of excessive masking.

Appendix A: NIST TEST SUMMARY

№	**Statistical test**	**Test statistics c(S)**	**Detectable defect**
1	Frequency (monobit) test	Normalized absolute sum of sequence element values	Too many zeros or ones in the sequence
2	Frequency test inside unit	Measure of matching observed number of units inside block with theoretically expected	Localized deviations of frequency of occurrence of units in block from the ideal value of 1/2
3	Check cumulative sums	Maximum deviation of the accumulated amount of elements in a sequence from an initial reference point	Large value of ones or zeros at the beginning or end of a binary sequence
4	Series check	Total number of series for entire length of sequence	Too fast or too slow reversal during sequence generation
5	Check maximum length of series in block	Measure matching observed value of maximum length of single series with theoretically expected value	Deviation from theoretical distribution of maximum lengths of series of units
6	Checking rank of binary matrix	Measure matching observed value of different order of ranks from theoretically expected	Deviation of empirical law of value distribution of matrix ranks from theoretical, which indicates dependence of characters in the sequence
7	Spectral test based on the discrete Fourier transform	Normalized difference between the observed and expected number of frequency components that exceed 95% threshold level	Identification of periodic components (trends) in binary sequence
8	Check overlapping patterns	Measure of matching the observed number of overlapping patterns in sequence with the theoretical value	Large number of m-bit series of units in the sequence
9	Maurer's universal test	Sum of logarithm of distance between l-bit patterns	Sequence compressibility
10	Entropy test	Measure matching observed value of entropy source with the theoretically expected for random source	Uneven distribution of m-bit words in sequence (regularity of source properties)
11	Checking random deviations	Measure matching observed number of visits random walk in predetermined state inside loop with theoretically expected	Deviation from theoretical law of distribution of visits to particular state by random walk
12	Check random deviations (option)	Total number of visits to given state by random walk	Deviation from theoretical expected total number of visits in case of random walk to given state
13	Sequential test	Measure matching observed number of all variants encountered m-bit patterns with theoretically expected	Uneven distribution of m-bit words in sequence
14	Checking non-overlapping patterns	Measure matching observed number of non-periodic patterns in sequence with theoretical value	Large number of specified non-periodic patterns in the sequence

№	Statistical test	Test statistics c(S)	Detectable defect
15	Linear complexity check	Measure matching observed number of events is appearance of fixed length equivalent linear recurrence register for given block with theoretically expected	Deviation of empirical distribution of lengths of equivalent linear recurrent registers for a sequence of fixed length from theoretical distribution law for random sequence, which indicates insufficient complexity of test sequence

Appendix B: NOTE ABOUT THE APPLICABILITY OF ASSOCIATIVE STEGANOGRAPHY FOR TEXTUAL CHARACTERISTICS OF OBJECTS PROTECTION

Associative steganography is quite acceptable for the protection of text characteristics of various objects: rocket mines; oil wells; mineral deposits; shore shelf contents; medical records of patients; personal data, for example, information on the availability of particular property, etc.

The infological database schema for such associative protection application is shown in Figure 5.

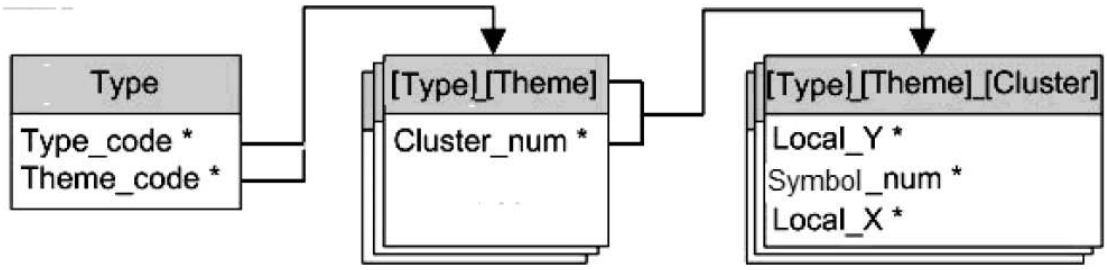

Figure 5. The infological database schema

Here:
– Type – relation containing information about all thematic layers represented by a pair of hidden codes: Type code (for example, mine) – Layer code (specific mine). For each type, the code of the layer is chosen randomly on the full set of codes of a given width.
– [Type]_[Theme] (the name is composite, contains open codes of type and layer) – set of relations, each of which describes separate thematic layer (list of cluster names representing many characteristics of this object).
– [Type]_[Theme]_[Cluster] (the name is composite, contains open codes of type, layer and cluster names) – set of relations, each of which represents one of the characteristics of this object.
– Data on the attributes marked with an asterisk are stored in masked form.

Line number code (Local_Y*)	Character code (Symbol_num*)	Position code in line (Local_X*)

Let any textual characteristic occupy no more than one typewritten page. Then:
– The number of lines of the message – no more than 30.
– The number of characters per line – no more than 60.

With k = 3, n = 60, the maximum response volume for any requested characteristic is $3 \times 3 \times (9 \times 60 - 12) \times 1.8 \times 10^3 = 8.5536 \times 10^6$ bits ≈1MB, which is acceptable. Earlier estimates of the durability and noise immunity of associative protection remain valid for stegotexts. Sequence of characters representing the requested textual characteristic can be interpreted as a linear cartographic object, which allows the development of a full-featured DBMS of cartographic scenes with associative protection [16] and with the necessary adaptation to a new application to be used.

REFERENCES

1. V. A. Raikhlin, I. S. Vershinin, R. F. Gibadullin and S. V. Pystogov, “Reliable Recognition of Masked Binary Matrices. Connection to Information Security in Map Systems”, Lobachevskii Journal of Mathematics **34** (4), 319–325 (2013).
2. V. A. Raikhlin, I. S. Vershinin, R. F. Gibadullin and S. V. Pystogov, “Reliable Recognition of Masked Cartographic Scenes During Transmission over the Network”, IEEE: 2016 International Siberian Conference on Control and Communications (SIBCON), May 2016, DOI:https://doi.org/10.1109/ICIEAM.2018.8728629.
3. V. A. Raikhlin, I. S. Vershinin and R. F. Gibadullin, “The Elements of Associative Steganography Theory”, Moscow University Computational Mathematics and Cybernetics **43** (1), 40–46 (2019).
4. C. E. Shannon, “Communication Theory of Secrecy Systems”, Bell System Technical Journal **28** (4), 656–715 (1949).
5. R. O. Duda, P. E. Hart Pattern Classification and Scene Analysis (Wiley, New York, 1973).
6. M. Matsumoto, M. Saito, H. Haramoto and T. Nishimura, “Pseudorandom Number generation: impossibility and compromise”, Journal of Universal Computer science **12** (6), 672–690 (2006).
7. D. A. Ker, “A capacity result for batch steganography”, IEEE Signal Processing Letters **14** (8D), 525–528 (2007).
8. Lawrence Bassham, Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, N. Heckert and James Dray “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications”, April 2010. https://doi.org/10.6028/NIST.SP.800-22r1a.
9. U. Maurer, “A Universal Statistical Test for Random Bit Generators”, Journal of Cryptology **5** (2), 89–105 (1992).
10. J. K. M. Sadique, Uz Zaman, R. Ghosh“Review on fifteen Statistical Tests proposed by NIST”, IJTPC 1, 18–31 (2012).
11. K. Gyarmati, “On a pseudorandom property of binary sequences”, The Ramanujan Journal (8), 289–302 (2004).
12. I. S. Vershinin, “Elaboration of the criterion of redundancy of noise-resistant information hiding within associative steganography”, Information and Security **19** (4), 511–514 (2016).
13. E. S. Wentzel, Probability theory: Textbook for universities. 6th ed. (Higher school, Moscow, 1995).
14. V. A. Raikhlin, I. S. Vershinin, R. S. Minyazev, R. F. Gibadullin, Constructive modeling of computer science systems. 6th ed. (Publishing House "Science", Kazan, 2016) [in Russian].
15. A. P. Alferov, A. Yu. Zubov, A. S. Kuzmin, A. V. Cheremushkin, Basics of cryptography. Textbook, 2nd ed. (Helios ARV, Moscow, 2002) [in Russian].
16. S. V. Pystogov, DBMS MapCluster. https://github.com/pystogov/MapCluster. Accessed 2019.