Reliable Recognition of Masked Cartographic Scenes During Transmission over the Network

V. A. Raikhlin
Department of computer systems
KNRTU-KAI
Kazan, Russia
no-form@evm.kstu-kai.ru

I. S. Vershinin
Department of computer systems
KNRTU-KAI
Kazan, Russia
vershinin_igor@rambler.ru

R. F. Gibadullin
Department of computer systems
KNRTU-KAI
Kazan, Russia
landwatersun@mail.ru

S. V. Pystogov
Department of computer systems
KNRTU-KAI
Kazan, Russia
sergey.pystogov@gmail.com

***Abstract*—General questions compressed representation of a given set of binary matrices of the same size by associative masking. Existence and uniqueness are proved: 1) identification in a random binary matrix of one of the matrices of the set of masked patterns of the same size; 2) mistaken recognition on this set of matrices by mask immersed in a random sequence of the inversion of any number of unmasked bits. Connection of these questions with reliability of protection of cartographical data is established at action of interference misinformation in a corporate network.**

Keywords— masking algorithm, recognition, mapping

I. INTRODUCTION

General formulation of the binary matrices recognition problem can be found in [1]. For the ternary pattern (ternary matrix):

$X^t = |x^t_{pq}|; x^t_{pq} \in \{0, 1, –\}; p = \overline{1,k^t}; q = \overline{1,l^t}; t \in \{\overline{1,\gamma}\}$

(γ - the number of types of objects) to find the coordinates of the objects covered by their binary image (boolean matrix).

$A = |a_{ij}|; \; a_{ij} \in \{0, 1\}; i = \overline{1,K}; j = \overline{1,L}; K \geq k^t; L \geq l^t.$

Record x^t_{pq} = (–) means indifferent (masked) value of the corresponding element of the pattern. Set of masked elements is displayed single elements of the matrix masks:

$$M^t = |m^t_{pq}|; m^t_{pq} = \begin{cases} 0, x^t_{pq} \in \{0,1\}; \\ 1, x^t_{pq} \in \{–\}. \end{cases}$$

Further elements of the pattern X^t that cannot be masking defined unit components inverse matrix mask $\overline{M}^t = |\overline{x^t_{pq}}|$ for this pattern.

Ternary pattern is represented by two binary matrices: ideal binary pattern X^t_2 and the inverse matrix of masks $\overline{M}^t$. For example, let

$$X^t_2 = \begin{Vmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{Vmatrix},$$

$$\overline{M}^t = \begin{Vmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{Vmatrix}.$$

Then the ternary pattern

$$X^t = \begin{Vmatrix} – & 0 & 1 & – & – & – \\ – & – & – & – & – & – \\ – & – & – & – & – & 1 \\ – & – & 0 & – & 0 & – \end{Vmatrix}.$$

Recognition procedure is carried out via masked comparison of the matrix with the pattern and is described in [2, 3].

The required non-ambiguous (unique) recognition stipulates that none of the pairs of triple patterns of the set $\{X^t\}$, $t \in \{\overline{1,\gamma}\}$ are mutually covered. A sufficient condition for this would be the difference in at least one significant elements x^t_{pq} of ternary matrices X^{t1} and X^{t2}, $t1 \neq t2$. In this paper we propose masking algorithm satisfying this condition. The algorithm successively extracts subsets based on the presence of a common significant bit. Fig. 1 shows a sample zip code template representation, with alphabet 0 … 9 and size n×m = 5×3. Noticeably, only the bits of the entire circuit can

978-1-4673-8383-7/16/$31.00 ©2016 IEEE

be treated as significant bits in this case (outer contour + "inner zigzag" on Fig. 2a). Fig. 2b shows the direction of traverse of the entire circuit for case n×m = 5×3. Size of the entire circuit is equal to (9m – 12) bits.

II. PROPERTIES OF THE MASKING ALGORITHM

Definition. The random binary matrix is a matrix with randomassignment of unit/zero values to elements of each row.

Theorem 1. If, for a random binary matrix of size n × m, we apply the recognition procedure on a set of similar-sized templates using masks generated by our masking algorithm, then one and only one template from the specified set will be recognized in this matrix.

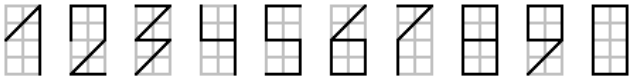

Fig. 1. Zip code template

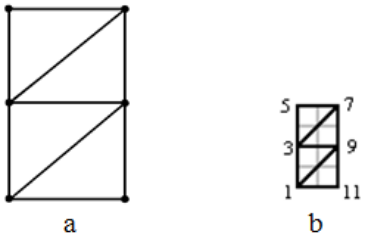

Fig. 2. Significant bits

Proof. At each step of the masking algorithm we divide into two parts (dichotomy) the previously selected subset of masked binary matrices based on the value of some bit. First step of the masking algorithm takes as an input a complete set of matrices. Therefore, all inverse masks necessarily contain one and only one common unit at a certain position

$$\alpha_{i,j}\ (i \in \{\overline{1,n}\}, j \in \{\overline{1,m}\}).$$

This unit divides the original set into two subsets, where the capacity of each part is less than γ. It remains to search iteratively on the subset for which the value of the bit at position $\alpha_{i,j}$ coincides with the observed for a random matrix.
Let us proof the convergence of this search by induction. The case γ = 2, corresponding to the only unit bit in both inverted masks, is obvious. For the case γ = 3, first step of the masking algorithm selects two subsets of capacity 1 and 2 respectively from the original set. Based on the unit bit which is common for all generated inverse masks, the random matrix will be attributed to either a singular subset or to one of the binary matrices of the second subset, based on the value of another significant bit common to both matrices of this subset. Suppose that the theorem is true for all $\gamma \le k$. Then in case of $\gamma = k+1$ the first step of the masking algorithm distinguishes two subsets, each having capacity of $\le k$. Of these two, we select a subset with the corresponding value of the dichotomy bit. By induction hypothesis, a search on any of the chosen subsets will succeed. The recognition uniqueness follows from the inability to cover any mutual pair of triple patterns. *The theorem is proved.*

Corollary 1. Using the masking algorithm for generation of mask sets ensures, that for any realization of the randomized objects set (i.e., the set containing those objects which are spread over a random sequence by applying a mask), and for any newly generated set of triple patterns of these objects, each randomized object is covered with one and only one pattern.

Corollary 2. For a given set of triple patterns and for a given ternary pattern of the matrix, spread by a mask over a random a sequence, the inversion of any number of significant bits in this matrix results in the recognition of one and only one pattern, which is different from the true pattern.

Validity of the first corollary was considered in [5]. The validity of the second corollary follows from the fact that, as a result of significant bits inversion, some random matrix will be obtained. This will result in false recognition, because significant bits of the "distorted" pattern will be inverted, causing mismatch of these bits with the corresponding bits of the true triple pattern.

III. THE INTERFERENCE MISINFORMATION

The interference misinformation means the interference, leading to distortions of the part of the information, with provided credibility. According to the consequence 2, the introduction of misinformation can be implemented by the distortion of the ternary model's substantial bits of any quantity. In order to identify this substantial bits, which have got the highest probability of the distortion as a result of the additive interference, an additional investigations of the masking algorithm's properties were provided. There was carried out the following experiment for the different measures of models m (with $n = 2m-1$) in order to derive the distribution function for the bit definitions, which are commonly used as substantial.

1. According to the masking algorithm a random set of inverse matrix masks $A_i, i = \overline{1,10}$ are generated for the models in the postcode's alphabet for a given m.
2. $j := 0$.
3. Then there is provided an additional generation of the inverse matrix masks $B_i, i = \overline{1,10}$.
4. After this an addition of the appropriate matrix A_i *and* B_i is carried out. The result of the addition is recorded in the matrix $C_i, i = \overline{1,10}$.C= A+B (for each bit)
5. $A_i := C_i, i = \overline{1,10}$.
6. $j := j + 1$. If $j < 10^5$, go to the item 3, otherwise – to the item 7.
7. End.

As the result of this experiment in the matrices $A_i, i = \overline{1,10}$ for each bit we can define the frequency (quantity) of its use as a substantial bit on set for 10^5 sets of inverse matrix masks. A graphical representation of the derived distribution functions for m = 20, 40 and 60 is shown in Fig. 3–5. Note that the general form of the distribution does not depend on the measure of m. From the given graphics follows that so

called hotspots of the total circuit for all models are frequently used as the substantial ones (Fig. 2a). Their coordinates (from the bottom left corner): 1, $2m - 1$, $4m - 3$ и $5m - 4$. The coordinates of the hotspots in the case $n \times m = 5 \times 3$ are shown in the Fig. 2b. It should be noted that hotspots with the coordinate 3 (coordinate m in general case) and 7 ($3m$-2) can not be used as a dichotomical bits, as all the models in the postcode's alphabet contain according to this coordinates the same (single) bit measures (Fig. 1). Additionally, it was estimated that among the whole multiplicity of keys (a set of inverse matrix masks) for about 61 percent of indicated multiplicity of keys, these bits are used as substantial. So this way the position of the inverse matrix masks substantial bits, most commonly used in the generation, was defined. This information can be used to carry out so called "purposeful" misinformation. The result of the interference's use in practice with the indicated above information is shown below in the example of the real map.

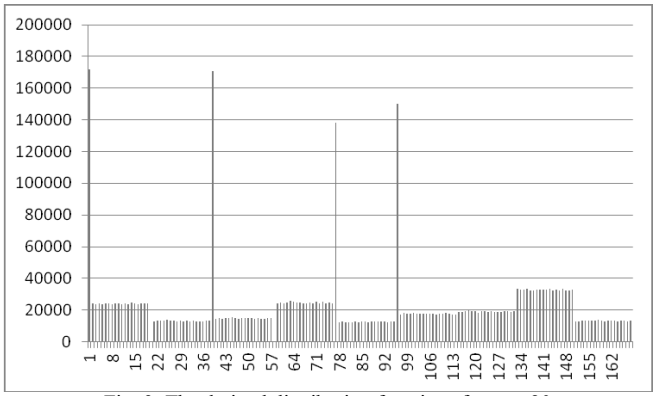

Fig. 3. The derived distribution functions for m = 20

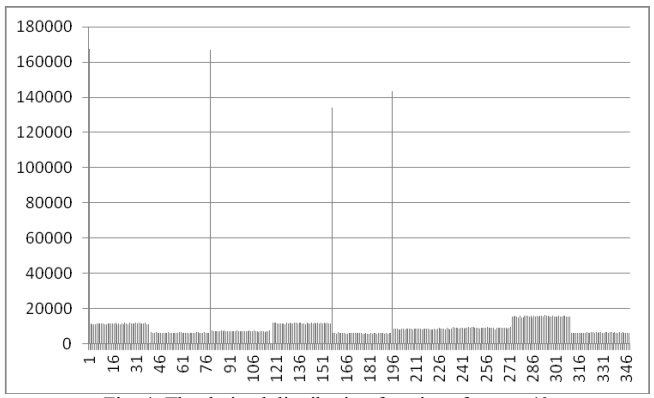

Fig. 4. The derived distribution functions for $m = 40$

With the purpose to determine the probability of distortion at least of the one substantial bit of ternary model in the process of this interference misinformation's generation the following experiment was provided.

1. A multiplicity of the ternary models (800) is generated with the use of some key (a set of masks).
2. Then a set of masks is generated, which is used as interference.
3. With every ternary model an addition is carried according to the mod2 of a randomly chosen mask from the set in point 2.
4. After this the recognition of the ternary model's distorted set on the true key is carried (according to the consequence 2, the result will be an incorrect recognition).
5. Then the number of incorrect recognitions is determined.
6. The items 1-5 are repeated 5×10^4 times.
7. The total number of ternary model's incorrect recognitions of indicated multiplicity for a given number of iterations is counted.
8. And at last relying on the data obtained we can determine the distortion's probability at least of the one substantial bit.

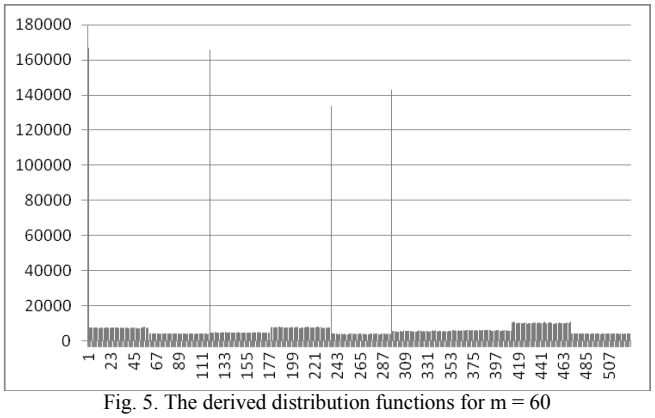

Fig. 5. The derived distribution functions for m = 60

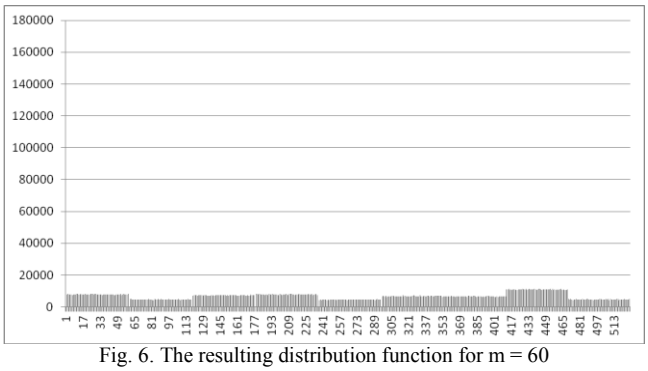

Fig. 6. The resulting distribution function for m = 60

Distortion's probabilities for different measures of m are shown in the Table 1.

TABLE 1. Distortion Probabilities

m, bit	3	10	20	30	40	60
P	0,764	0,234	0,155	0,130	0,117	0,108

To obtain a more uniform distribution function is useful when forming a random set of inverse matrix of masks prohibit identified through the research nodal bits as essential. For this a random choice of a distinctive bit in the masking algorithm to carry out additional verification does not apply if the selected bit to the "exceptions" bits. In this case, selection of distinctive bit is repeated. There has been a repetition of the above experiment using the modified masking algorithm. In Fig. 6 presents a graphical representation of the resulting distribution function for $m = 60$. Exception by modifying the masking algorithm of significant bits will reduce the power of a set of keys about an order. However, this reduction will not affect the provable (computing) persistence of the proposed

method. For example, for m = 40 the number of keys is 10^{27} without modification, the modification is $3{,}9 \cdot 10^{26}$. Taking the time to process a single key to 1 ns (brute-force attack), for the first case we get the total time of the exhaustive search $32 \cdot 10^9$ years, in the second case we get $12 \cdot 10^9$ years. Table 2 shows the probability of distortion for different values of m for the modified masking algorithm.

TABLE 2. Probability of Distortion

m, bit	3	10	20	30	40	60
P	0,875	0,174	0,081	0,054	0,039	0,026

Fig. 7. The territory of Chuvashia size of 300 km^2

Fig. 8. The territory of Chuvashia size of 300 km^2 (*by base masking algorithms*)

IV. CONNECTION WITH THE PROTECTION OF INFORMATION IN THE MAPPING SYSTEM

Today, for the storage and management of spatial data in database management systems are widely used ArcSDE server software in the ArcGIS Server [6]. ArcSDE is closely integrated with packages ArcEditor, ArcInfo, ArcView to work with multi-user cartographic databases. This solution is quite universal, but the construction of database management systems oriented job-protected database mapping can significantly improve the management of such a database by performance at the required level of strength protection. Basic principles of a secure mapping database for point, line and polygon objects using masking algorithm described in [7, 8].

Fig. 9. The territory of Chuvashia size of 300 km^2 (*by modified masking algorithms*)

Fig. 7 shows a section of area on the territory of Chuvashia size of 300 square kilometers coated with a point object. The number of types of objects is 4 ("forest", "vegetation", "sad", "stop"), the number of these objects is 1035. Each object on the map correspond to the three-digit code of the object depending on its type (000 ÷ 003), as well as a three-digit code of its coordinates in the alphabet zip codes, for m = 18, which are masked, and then defined to be using a randomization procedure. Used for masking a random set of masks is a key role of knowledge which allows authorized users to spend one distinction.

Further the procedure is performed clustering [9], which resulted in the original section of area is divided into several clusters. Clustering procedure involves the alignment number of objects in each cluster for which we introduce the so-called empty (non-essential) objects which are assigned the unused codes in the range 004 ÷ 999. As a result of clustering for the sample was obtained 5 clusters with 356 objects each.

Previously considered the case of the fixed effects of the additive interference at the nodal points, that contains all the units. The result of such exposure was the distortion of objects and their coordinates on the map. In this case all of distorted objects as a result of recognition in the class were insignificant that made it impossible to render.

Thus the condition of preserving the plausibility is not satisfied here. Except that the specified fixed interference is easily eliminated by re-inverting nodes before the procedure of recognition.

Further considered two variants of masking: using a base and modified algorithms. While the card is exposed to the additive noise of misinformation using the following method for their generation.

1. Generates a random set (or multiple sets) of inverse matrix masks using masking algorithm.
2. Carried out by sections in the introduction of interference transmission unit (container) by adding mod2 inverse matrix of mask of a given set at each of the container.

Fig. 8 and 9 shows the results of recognition cards disguised using base and modified masking algorithms respectively.
Information of the object is distorted if the result of the disturbance change in the type of an object or its coordinates x (y) or found a set of these changes. Number of false recognitions types of objects or their coordinates for the base masking algorithm is 45% and for the modified algorithm is 35%. The slight decrease in number of false recognitions using a modified masking algorithm is associated with a large number of empty (insignificant) objects which are under the influence of interference are mostly in the same class. Note that in Fig. 8 and 9 in the general kind of maps remains almost unchanged from the original map that is stored plausibility. To lead this result to a pre-given estimates (Tables 2 and 3), we assume that all objects for this example maps are essential. Then the number of false recognitions will be 66% for the basic masking algorithm, and 30% will be for the modified algorithm.

V. CONCLUSION

All this research prove the existence and uniqueness of 1) identification in a random binary matrix of one of the matrices of the set of masked patterns of the same size; 2) mistaken recognition on this set of matrices by mask immersed in a random sequence of the inversion of any number of unmasked bits.

Established the possibility of using the masking algorithm as a generator of additive noise for the introduction of misinformation in the transmitted message. To establish the fact of misinformation to the use of special techniques such as the use of cryptographic hash functions [10].

REFERENCES

[1] Raikhlin V.A. On the use of the unit two-dimensional associative search in the recognition process, in: Interuniversity collection. Problem-oriented means to improve computer systems, Kazan Aviation Institute, Kazan, 1991, pp. 38–54 (in Russian).

[2] Raikhlin V.A. Performance analysis processor matrices for recognition of binary images, Avtometriya, Novosibirsk, 1996, No. 5, pp. 97–103 (in Russian).

[3] Vershinin I.S., Gibadullin R.F., Zemtsov P.E. Parallel algorithms for protecting binary objects cartography, in: Proceedings of the Kazan Scientific Seminar "Methods of Modeling", Kazan State Technical University named after A.N. Tupolev, Kazan, 2007, No. 3, pp. 96–108 (in Russian).

[4] Raikhlin V.A., Vershinin I.S., Glebov E.E. Solution of the problem of masking stylized binary images, in: Vestnik of Kazan State Technical University named after A.N. Tupolev, Kazan, 2001, No. 1, pp. 42–47 (in Russian).

[5] Raikhlin V.A., Vershinin I.S. Elements of a two-dimensional mapping cipher cryptanalysis, in: Vestnik Kazan State Technical University named after A.N. Tupolev, Kazan, 2002, No. 4, pp. 48–54 (in Russian).

[6] Mapping platform for your organization, Esri © 2012, URL: http://www.arcgis.com.

[7] Raikhlin V.A., Vershinin I.S., Gibadullin R.F. Constructive modeling systems in application data security mapping, in: Proceedings of the Kazan Scientific Seminar "Methods of Modeling", Kazan State Technical University named after A.N. Tupolev, Kazan, 2010, No. 4, pp. 68–95 (in Russian).

[8] Gibadullin R.F., Pystogov S.V. Formation secure cartographic databases for the error localization objects in the Security Map Cluster, in: Tupolev reading: Proceedings of the 20th International Youth Conference, Kazan State Technical University named after A.N. Tupolev, Kazan, 2012, Vol. 3, pp. 97-99 (in Russian).

[9] Raikhlin V.A., Vershinin I.S. Modeling of the processes of two-dimensional associative masking distributed point objects cartography, in: Nonlinear world, Radiotechnika, Moscow, 2010, No. 5, Vol. 8, pp. 288–296 (in Russian).

[10] Alferov A.P., Zubov A., Kuzmin A.S., Cheryomushkin A. Foundations of cryptography, Gelios, Moscow, 2002.