Recent Developments

in Mathematics and Informatics

Contemporary Mathematics and Computer Science

Vol. 2

Wydawnictwo KUL

Lublin 2016

Editor: A. M. Zapała

Reviewers:

W. Rzymowski, Lublin Technical University

A. Chichurin, John Paul II Catholic University of Lublin

Wydawnictwo KUL, Lublin 2016

ISBN: 978-83-8061-345-4

Part II

Computer Science

Chapter 11

Theoretical Model of a Multi-Key Steganography System

Pavel Urbanovich[1,2], Nadzeya Shutko[2]

Abstract. *We describe a theoretical model of multi-key steganography systems applied for secret transmission of information hidden in text, graphics or sound files. To increase resistance of the system for unauthorized access apart from keys that deposite information we propose to use additional cryptography keys.*

2010 Mathematics Subject Classification: 94 A 40, 68 P 25, 68 P 20, 94 A 60.
Key words and phrases: text message, multi-key steganography system, key depositing information, cryptography key.

---

[1]Address: The John Paul II Catholic University of Lublin, Faculty of Mathematics, Informatics and Landscape Architecture, ul. Konstantynów 1H, 20-708 Lublin, Poland.
E-mail: pav.urb@yandex.by

[2]Address: Belarusian State Technological University, Department of Information Systems & Technologies, ul. Sverdlov 13 a, Minsk 220006, Belarus.
E-mail: shutko.bstu@mail.ru

11.1 The main features of steganographic systems

As it is known, the mathematical foundations of modern cryptographic methods of information protection were laid by C. Shannon [6]-[7]. These works have given a great impetus to the development of other methods for secure transmission or storage of information. Steganography is a technique which plays an important role among these methods.

Contrary to cryptography whose purpose is hiding data by encrypting them, the purpose of steganography is to disguise the fact of the transfer of confidential messages. It is so because through steganography system of protection the greatest degree of resistance is achieved to intentional attacks in order to destroy or identify hidden information. The steganographic system (called in short – *steganosystem*) is a set of tools and techniques that are used to form a secret channel of information transfer [2], [3], [10].

The steganosystem forms the channel that carries the filled container. This channel is treated as being exposed to the influence from the violators.

In the steganographic system embedding messages can be performed using the key or without its use. To increase the *steganoresistance* of the system the key can be used as a verification tool. It can also have an impact on the distribution of bits of the message within the container and on the order of forming a sequence of embedded bits of messages.

In this paper we present the formal logical description of steganography systems based on some analogy with cryptosystems.

11.2 General concept of the model

We will define the abstract steganographic system as a set of transformations of a suitable space (which includes the set of possible messages, $\mathcal{M}$) to another space (the set of possible *steganomessages*, $\mathcal{S}$) and vice versa.

Here are the main characteristics of the models developed in [4], [8]-[12]:
1) the processes of embedding or retrieval of information, which are based on the corresponding initial algorithms, from a formal point of view are defined by the types of the embedded/extracted information, the container and a selection of specific container elements or groups of these elements used to accommodate the relevant message components; such basic algorithms used for the textual stenography can for example be: methods of Line-Shift Coding, Word-Shift Coding and others, cf. [1], [5];
2) the foundation of the methods derived in this article and the corresponding techniques of hiding information form essential space-geometric and

color characteristics of basic elements of text containers (or else – fonts);
3) an important distinctive feature of the theoretical model developed here is the identification of the selected steganographic methods (based on the modification of the specific space-geometric or color parameters of text symbols) with essential information transmitted by means of the steganographic process; in our opinion, for the unauthorized user the information remains secret;
4) we will consider other hidden parameters of the steganographic process as additional keys that deposit information.

To complete the picture and precise requirements of the model, the assumptions will be formulated on the basis of the simulation results, in order to clarify the physical nature of the main elements of the model.

11.3 Formal description of theoretical model of a steganographic system

We shall build the model using the following notation and regulations described in [12]. Let

$\mathcal{M} = \{M_1, M_2, ..., M_n\}$ be a finite set of messages that should be hidden in various containers; in the context of solved tasks within the framework of our study, the considered messages are text documents;

$\mathcal{C} = \{C_1, C_2, ..., C_p\}$ denotes a finite set of all admissible containers (cache files or text cache documents) – in general $p > n$;

$\mathcal{K} = \{K_1, K_2, ..., K_z\}$ is a set of keys, by which we will generally understand methods and algorithms of message deposition in containers, or other operations preliminarily transforming messages or selecting the elements in containers for such a deposition.

An arbitrary concealed message $M_i \in \mathcal{M}$ can be hidden in the container $C_j \in \mathcal{C}$ using a suitable key of the set $\mathcal{K}$, $K_m \in \mathcal{K}$. The result of such transformation is a full container (or steganomessage) S_q, belonging to the set of all containers or steganomessages $\mathcal{S} = \{S_1, S_2, ..., S_r\}$, including full as well as empty containers. To avoid additional complications we assume that very long messages are divided into short portions that can be inserted in individual containers; in other words, all the elements $M_1, M_2, ..., M_n \in \mathcal{M}$ are so small that after an application of some keys they can be contained in different containers.

First of all we consider a single set of keys $\mathcal{K}$. Further argumentation will be built on the basis of the fundamental concepts, formulated in the following definitions.

Definition 11.1. A suitable transformation F (in general – a relation) defined on $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$ with values in $\mathcal{S}$ (but in fact – a relation acting on $\mathcal{M} \times \mathcal{C} \times \mathcal{K} \times \mathcal{S}$) will be identified with deposition or insertion of messages $M_i \in \mathcal{M}$ in containers $C_j \in \mathcal{C}$ by means of keys $K_m \in \mathcal{K}$, which demands the use of an appropriate algorithm concerning deposition and space (geometric or some other) parameters of a container C_j of the set $\mathcal{C}$:

$$F : \mathcal{M} \times \mathcal{C} \times \mathcal{K} \to \mathcal{S}. \tag{11.1}$$

Recall that here we use the term transformation in a slightly another meaning than it is usual in mathematics, thus F need not be a function defined on the whole Cartesian product $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$, but to each triple (M, C, K) which is the first part of the relation F, written in the form $(M, C, K) \in \mathcal{D}F$, it assigns certain element $S \in \mathcal{S}$, hence we also write $F(M, C, K) = S$. A single transformation of the set of transformations

$$\mathcal{F} = \{F_1, F_2, ..., F_l\} \tag{11.2}$$

can be represented graphically as shown in Fig. 11.1. Each specific transformation F_w, where $w = 1, 2, ..., l$, of the set $\mathcal{F}$ corresponds to a particular algorithm or method in which informations M_i are placed in deposition containers C_j, using specific keys K_m. Relation (11.1) formally describes the procedure leading to the deposition of messages in containers based on selected methods.

To display such an interaction of the system components, for better comprehension it is schematically shown in Fig. 11.1.

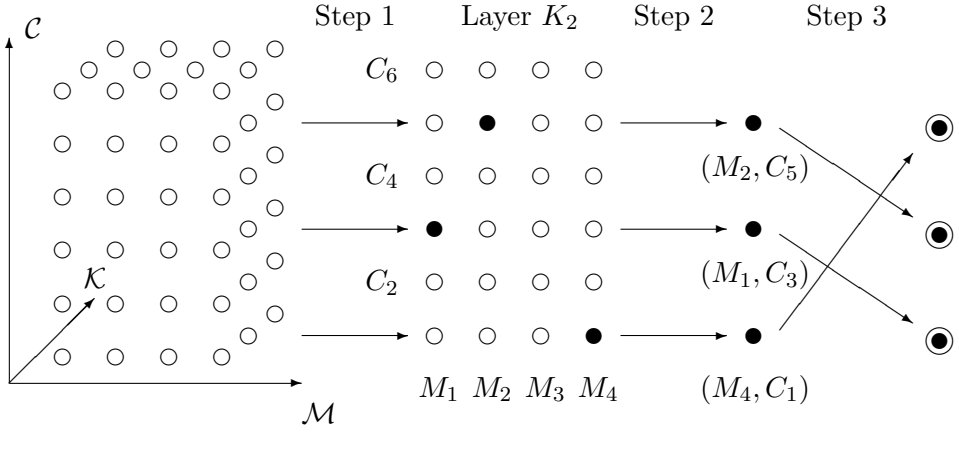

Figure 11.1: Graphic representation of some transformation $F \in \mathcal{F}$ on the basis of a fixed key $K_m = K_2 \in \mathcal{K}$; for another fixed key $K_m \in \mathcal{K}$ the same transformation F is acting in a similar manner.

On the left side of the above picture the Cartesian product $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$ is shown, where $\mathcal{M}$ contains 4 messages, $\mathcal{C}$ consists of 6 containers, and $\mathcal{K}$ is the set of 3 keys, but only the first layer corresponding to key K_1 is complete, and top-right edges of layers corresponding to keys K_2 and K_3 are visible; for further transformation of messages M_1, M_2, M_4 key K_2 is chosen; moreover, the message M_1 will be inserted into the container C_3, M_2 will be deposited in the container C_5, while M_4 is assigned to the container C_1, as is presented in the central part of the picture. This part of the picture shows that at this stage of the procedure one should select certain relation, i.e. a subset of the Cartesian product $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$, denoted by black balls. It can be seen that for a fixed key, say K_2, the picture is neither a graph of a function acting on $\mathcal{M}$ with values in $\mathcal{C}$ (for M_3 the value is not assigned), nor a function acting on $\mathcal{C}$ and taking values in $\mathcal{M}$ (for C_2, C_4 and C_6 the values are not assigned). In the next step only the triples (M_4, C_1, K_2), (M_1, C_3, K_2) and (M_2, C_5, K_2) are considered (for some other key, say K_j, $j \in \{1, 3\}$, a triple (M_3, C_i, K_j), $i \in \{1, 2, 3, 4, 5\}$, should be taken into account, but it is not presented on the figure). At this point of the steganography process the other elements of the product $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$ are excluded from further considerations, and in this way the whole set $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$ is reduced merely to the set of necessary elements. The selected elements after embedding messages into containers are ready for sending as steganomessages, which can be seen on the right side of the picture. The last stage of the described procedure is obviously a function equal to a simple permutation. The permutation presented in the picture is tantamount to ordering of the messages according to the increasing sequence of their indices, namely M_1, M_2, M_4.

From the above considerations it follows that each transformation $F \in \mathcal{F}$ is a composition of 3 operations: 1^0 a relation defined on the Cartesian product $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$, which determines certain subset of this product, 2^0 a function, whose domain is the subset selected in step 1^0, (the mentioned function with the domain restricted to common part of this subset and certain layer K_m can be identified with the chosen key $K_m \in \mathcal{K}$), and 3^0 a mapping that is in a fact a permutation of elements of set $\mathcal{S}$ (thus the last step may be even omitted).

Definition 11.2. The transformation F^*, acting on $\mathcal{S} \times \mathcal{K}^*$ and taking values in $\mathcal{M} \times \mathcal{C}$, (in fact with values in $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$, but since the key used for embedding of information is inessetial, we here write in short $\mathcal{M} \times \mathcal{C}$ instead of $\mathcal{M} \times \mathcal{C} \times \mathcal{K}$), where $\mathcal{K}^* = \{K_1^*, K_2^*, ..., K_z^*\}$, and in general $K_m^* \neq K_m$, $K_m \in \mathcal{K}$, $K_m^* \in \mathcal{K}^*$ for $m = 1, 2, ..., z$, will be identified with the recovery of

hidden messages $M_i \in M$ from steganomessages $S_q \in \mathcal{S}$:

$$F^* : \mathcal{S} \times \mathcal{K}^* \to \mathcal{M} \times \mathcal{C}. \tag{11.3}$$

We use here an analogous notation as above in the case of transformation F.

The set $\mathcal{F}^*$ of reverse transformations, similarly as $\mathcal{F}$, consists of l elements:

$$\mathcal{F}^* = \{F_1^*, F_2^*, ..., F_l^*\}, \tag{11.4}$$

where each specific backward transformation $F_w^* \in \mathcal{F}^*$ corresponds to the forward transformation $F_w \in \mathcal{F}$, $w = 1, 2, ..., l$, and is determined by some combination of messages $M_i \in \mathcal{M}$, containers $C_j \in \mathcal{C}$ and appropriate keys $K_m^* \in \mathcal{K}^*$ suitable to previously applied keys $K_m \in \mathcal{K}$.

Thus, the expression (11.3) defines the inverse to the transformation (11.1), and although it is not an inverse mapping in the mathematical sense, the transformation F^* to some element S_q of the set $\mathcal{S}$ and a fixed element K_m^* of the set $\mathcal{K}^*$ assigns the appropriate element M_i of set $\mathcal{M}$ and an element C_j of set $\mathcal{C}$.

The expression (11.3) formally describes the procedure of extraction of hidden messages from the containers by means of appropriate methods corresponding to their deposition into the containers. Consequently, each concrete transformation F_w^*, $w = 1, 2, ..., l$, of the family $\mathcal{F}^*$ corresponds to particular algorithms or methods of embedding some information messages M_i into containers C_j using specific keys K_m. A graphic representation of this procedure is demonstrated in Fig. 11.2.

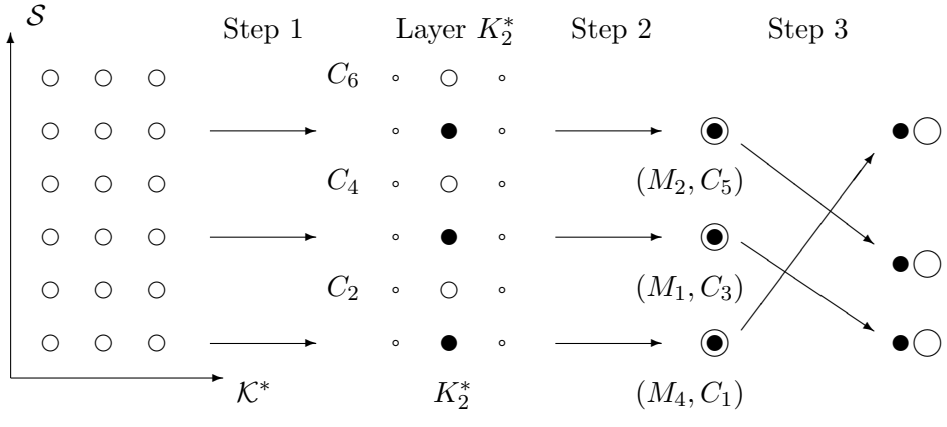

Figure 11.2: Graphic representation of a fixed transformation $F^* \in \mathcal{F}^*$ on the basis of key $K_m^* = K_2^* \in \mathcal{K}^*$.

On the left side of the above picture the Cartesian product $\mathcal{S} \times \mathcal{K}^*$ is shown; set $\mathcal{S}$ contains 6 elements (recall that we consider here filled as well as empty containers), and set $\mathcal{K}^*$ consists of 3 keys. For a fixed key, e.g. $K_2^* \in \mathcal{K}^*$, the layer corresponding to this key is determined. Acting by key K_2^* on all the containers one can find filled containers with information that was deposited by key K_2. In the same manner the other keys $K_m^* \in \mathcal{K}^*$ are acting on the whole set $\mathcal{S}$ of all steganomessages and in this way some relation on $\mathcal{S} \times \mathcal{K}^*$ is defined. Small white balls in the central part of the picture denote pairs (S_q, K_m^*) which are not taken into consideration when key K_2^* is active. Next filled containers are separated from the set of all containers. In the last step, the selected pairs (M_4, C_1), (M_1, C_3) and (M_2, C_5) are treated by the key K_2^* and the messages M_4, M_1 and M_2 are decoded, and then the decoded messages together with containers are ordered according to the increasing indices of messages.

Analyzing the above description and Fig. 11.2, we arrive at the following conclusion – similarly as in the input part of the steganography system, the transformation used in the receiver part of this system is a composition of 3 operations: in the first step of the discussed procedure an arbitrary key $K_m^* \in \mathcal{K}^*$ is fixed, which determines a layer corresponding to this key, say key K_2^* and layer K_2^* as is shown above; next, in step 2^0, acting by the chosen key K_2^* on all the elements of layer K_2^*, full containers with messages embedded by key $K_2 \in \mathcal{K}$ are separated from the set of all containers; finally, in step 3^0, the embedded messages are extracted from containers by means of key K_2^* and ordered together with their containers according to increasing indices of messages.

Definition 11.3. The *steganographic system* is an ordered structure Σ, consisting of 6 connected elements:

$$\Sigma = (\mathcal{M}, \mathcal{C}, \mathcal{K}, \mathcal{S}, \mathcal{F}, \mathcal{F}^*), \tag{11.5}$$

where $\mathcal{M}$ denotes the set of messages, $\mathcal{C}$ is the set of containers, $\mathcal{K}$ is the set of keys, $\mathcal{S}$ denotes the set of steganomessages (filled and empty containers), $\mathcal{F}$ and $\mathcal{F}^*$ are transformations (forward and reverse resp.).

11.4 Alternative description of a steganographic system

The model of the steganography system proposed in Section 11.3 is not unique. It is possible to describe also such systems in a slightly different manner. We explain it briefly.

Let $\mathcal{M} = \{M_1, M_2, ..., M_n\}$, $\mathcal{C} = \{C_1, C_2, ..., C_p\}$ and $\mathcal{S} = \{S_1, S_2, ..., S_r\}$ denote, as above, the set of messages, set of containers and set of steganomessages, respectively. To simplify the considerations, assume that the set of keys $\mathcal{K}$ contains only 1 key K, i.e. $\mathcal{K} = \{K\}$. In this case the operation of embedding of messages into containers can be identified with certain mapping $G : \mathcal{M} \cup \mathcal{C} \to \mathcal{S}$. Clearly, the mapping G here cannot be quite arbitrary. To make the requirements concerning G more precise, without loss of generality we reduce the set of steganomessages taking $\mathcal{S} = G(\mathcal{M} \cup \mathcal{C})$. Since each container can contain at most 1 message, G must satisfy the following condition:

$$\bigwedge_{S \in \mathcal{S}} \quad \mathrm{card}(G^{-1}(S) \cap \mathcal{M}) \in \{0, 1\} \ \wedge \ \mathrm{card}(G^{-1}(S) \cap \mathcal{C}) = 1.$$

More precisely, condition $\mathrm{card}(G^{-1}(S) \cap \mathcal{M}) = 1$ means that the message $M = G^{-1}(S) \cap \mathcal{M}$ is embedded (by means of a fixed key K) into the container $C = G^{-1}(S) \cap \mathcal{C}$, while $\mathrm{card}(G^{-1}(S) \cap \mathcal{M}) = 0$ denotes the situation in which the container $C = G^{-1}(S) \cap \mathcal{C}$ remains empty. Furthermore, equation $\mathrm{card}\, G(M) = j$, where $2 \leq j \leq p$, expresses the fact that the same message M is embedded into j various containers. The mapping G is illustrated in Fig. 11.3.

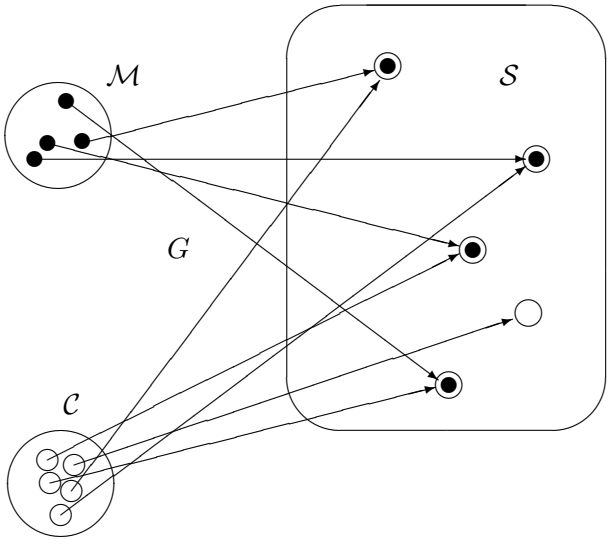

Figure 11.3: Graphic representation of the mapping $G : \mathcal{M} \cup \mathcal{C} \to \mathcal{S}$.

The reverse transformation G^* is now a relation defined on the Cartesian product $\mathcal{S} \times (\mathcal{M} \cup \mathcal{C})$. Obviously, G^* must also satisfy the appropriate conditions. To formulate these restrictions, for every $S \in \mathcal{S} = G(\mathcal{M} \cup \mathcal{C})$ we put

$$S\,G^*\mathcal{M} = \{M \in \mathcal{M} : S\,G^*M\} \quad \text{and} \quad S\,G^*\mathcal{C} = \{C \in \mathcal{C} : S\,G^*C\}.$$

Then G^* may be characterized by the requirement

$$\bigwedge_{S \in \mathcal{S}} \quad \operatorname{card}(S\,G^*\mathcal{M}) \in \{0,1\} \ \wedge \ \operatorname{card}(S\,G^*\mathcal{C}) = 1.$$

Now if $\operatorname{card}(S\,G^*\mathcal{M}) = 1$, then S contained a message M such that $S\,G^*M$ in the container satisfying condition $S\,G^*C$. On the other hand, if $\operatorname{card}(S\,G^*\mathcal{M}) = 0$, then S did not contain any hidden message. Moreover, $\operatorname{card} G^{-1}(M) = j \geq 2$ denotes that the message M was embedded into j containers. The relation G^* is presented in Fig. 11.4.

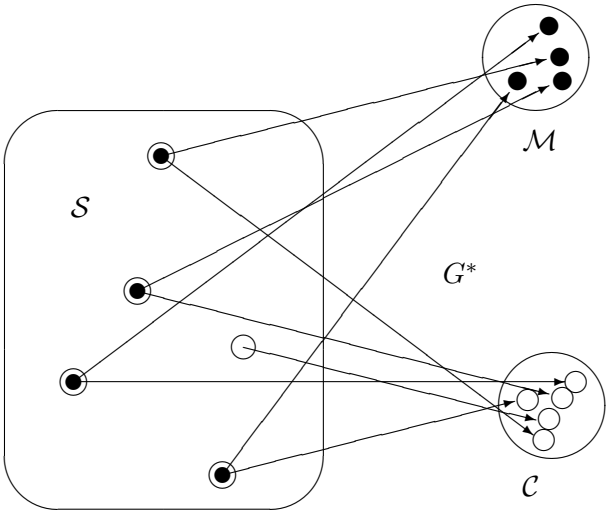

Figure 11.4: Graphic representation of the relation G^* on the set $\mathcal{S} \times (\mathcal{M} \cup \mathcal{C})$.

Using the just introduced terminology the other features of a steganographic system can be formulated as well, but we do not discuss it in greater detail. To describe in this manner a multi-key steganography system one may consider layers for single keys as in Section 11.3.

11.5 Collisions in steganographic systems

We shall discuss now some problems arising in steganography systems. It should be pointed out that situations described in this section occur in practice very rarely, but they cannot be excluded entirely from theoretical considerations.

The first situation which is rather inconvenient is the case when two or several messages are embedded into the same container by means of the same or various keys. Then it may happen that the second (or another message) during the process of embedding, erase and destroy (partially or even entirely) the first portion of information. Therefore one should prevent the system from such functioning. There is no doubt that the relation chosen in the initial step of the steganography procedure cannot be quite arbitrary, thus first of all we must precise the conditions which the mentioned relation should fulfill.

Let F denote the relation on $\mathcal{M} \times \mathcal{C} \times \mathcal{K} \times \mathcal{S}$ used at the initial stage of the steganography process.

Definition 11.4. Relation F acting on $\mathcal{M} \times \mathcal{C} \times \mathcal{K} \times \mathcal{S}$ is called *unambigious with respect to the set* $\mathcal{C}$, if

$$(M_i, C_j, K_m) \in \mathcal{D}F \wedge (M_{i'}, C_j, K_{m'}) \in \mathcal{D}F \quad \Rightarrow \quad M_i = M_{i'} \wedge K_m = K_{m'} .$$

Evidently, to avoid the just mentioned drawback, the relation used in the input part of the steganography system must satisfy the above condition.

It is also evident that the correct steganographic system must not contain internal inconsistencies. Some errors encountered in steganographic systems are described briefly below.

Let $\mathcal{S}$ denote the set of all steganomessages, including filled and also empty containers.

Definition 11.5. *Collision of a steganographic system*

$$\Sigma = (\mathcal{M}, \mathcal{C}, \mathcal{K}, \mathcal{S}, \mathcal{F}, \mathcal{F}^*)$$

(or *intersection*) is called a situation in which various steganomessages are the same, but they contain two various messages (including empty) with different information, which can be written in symbols

$$F_w \left(M_i, C_j, K_m\right) = F_{w'} \left(M_{i'}, C_{j'}, K_{m'}\right),$$

where $\mathcal{D}F_w \ni (M_i, C_j, K_m) \neq \left(M_{i'}, C_{j'}, K_{m'}\right) \in \mathcal{D}F_{w'}$ for certain indices $1 \leq i, i' \leq n$, $1 \leq j, j' \leq p$, $1 \leq m, m' \leq z$, and $1 \leq w, w' \leq l$.

In other words, two various triples (M_i, C_j, K_m) and $(M_{i'}, C_{j'}, K_{m'})$ are the first parts of relations F_w and $F_{w'}$ resp., and after steganographic transformations form identical steganomesages. Another formulas in this section should be interpreted in a similar way.

Recall that we consider here together filled as well as empty containers, thus the last definition regards also situations in which may happen that some sensible information is retrieved from the container which is in fact empty and does not contain any message. The similar inconvenient case may occur when merely filled containers are taken into account. We formalize it in the next definition.

Let $\mathcal{S}' \subseteq \mathcal{S}$, $\mathcal{S}' = \{S_1, S_2, ..., S_z\}$ denote the subset of all steganomessages being filled containers, where z is the number of messages in set $\mathcal{M}$.

Definition 11.6. *Subcollision* (or *subintersection*) *of a reduced steganographic system*

$$\Sigma' = (\mathcal{M}, \mathcal{C}, \mathcal{K}, \mathcal{S}', \mathcal{F}, \mathcal{F}^*)$$

is called a situation in which various steganomessages containing some hidden information are the same, but they contain various messages, i.e. we have

$$F_w\left(M_i, C_j, K_m\right) \cap \mathcal{S}' = F_{w'}\left(M_{i'}, C_{j'}, K_{m'}\right) \cap \mathcal{S}' \neq \emptyset,$$

although $\mathcal{D}F_w \ni (M_i, C_j, K_m) \neq (M_{i'}, C_{j'}, K_{m'}) \in \mathcal{D}F_{w'}$ for certain indices $1 \leq i, i' \leq n$, $1 \leq j, j' \leq p$, $1 \leq m, m' \leq z$, and $1 \leq w, w' \leq l$.

Some errors which occur during extraction of messages with information may be caused also by methods of transformations used in steganographic systems. The next definition describes one of such situations.

Definition 11.7. *Collision of a steganographic transformation* (or *intersection*) is called a situation in which various triples

$$\left(M_i, C_j, K_m\right), \left(M_{i'}, C_{j'}, K_{m'}\right) \in \mathcal{M} \times \mathcal{C} \times \mathcal{K}$$

after some transformation $F \in \mathcal{F}$ form as a result the same steganomessage, that can be formally written as follows:

$$F\left(M_i, C_j, K_m\right) = F\left(M_{i'}, C_{j'}, K_{m'}\right),$$

where $\mathcal{D}F \ni (M_i, C_j, K_m) \neq (M_{i'}, C_{j'}, K_{m'}) \in \mathcal{D}F$ for certain indices $1 \leq i, i' \leq n$, $1 \leq j, j' \leq p$, $1 \leq m, m' \leq z$.

We do not discuss here in greater detail all the possible inconsistencies in steganographic systems, but we have to mention that some situations which at the first glance seem to be inadmissible, are in fact allowed. For instance, the same message, such as electronic signature, may be indeed embedded into many various containers (text, image, sound files, etc.).

11.6 The role of keys

As was already mentioned, set $\mathcal{K}$ consists of a finite number of keys. Returning to the discussion of the number of keys, it should be pointed out that the number of various keys has a straightforward effect on the camouflage of transmitted information.

Formally, the set $\mathcal{K}$, which is a part of the system Σ, can be regarded as a sum of a finite number of disjoint subsets. That number is dependent on how many factors affect the resistance for cracking of the created system. By such factors (following the well-known postulate of Kerkhoffs) we understand the information on keys available for outsiders, in other words the number of keys. If the secret of embedding information in the container is caused only by an algorithm or a way of implementation of such an operation, the steganographic system should be treated only as a system with a simple set of keys, i.e. $\mathcal{K}$ does not contain subsets. A single key in such systems is equivalent to the algorithm or method of embedding/extraction information, and for such steganographic systems the correct formal description is given by (11.5). However, for additional protection from hacking or other influence on the deposition of information, usually additional preventive means are applied. Such remedies may include:

1) symmetric or asymmetric cryptographic keys for encryption and decryption of information messages contained in the set $\mathcal{M}$, used in the process of embedding/extraction of its elements in containers; 2) algorithms or methods that enable error-correcting coding of information messages $M_i \in \mathcal{M}$, in order to detect and/or correct errors in the information that occurred during transmission or storage of information in the container; and other similar remedies. Each of these cases 1) – 2) can be identified with the use of a key added to the basic algorithm of embedding/extraction of messages M_i in containers. Thus, we can speak about the class of steganosystems with additional keys.

Definition 11.8. By an *additional key* K_i^a of a steganography system we mean a concrete secret value of a set of cryptographic parameters or any other algorithm used for cryptographic encryption/decryption during error-correcting coding/decoding of the message, or other additional operations used in the process of deposition/extraction of messages M_i in containers; the set of additional keys

$$\mathcal{K}_a = \{K_1^a, K_2^a, ..., K_l^a\}$$

can be treated as additional means of increasing the resistance of a steganographic system.

Typically, as a separate key pseudo-random or other labels can be applied that determine the position or location of the elements in a container and that are modified when a particular element (the symbol of an embedded message M_i) is placed in this container (for example, such a label can be a position of a pixel in a matrix, formatting of the image of text, or a single character in the document-container (modifying its) with modified color, brightness, and other characteristics, in the process of embedding/extraction of the secret message). The value of such information in providing total resistance of steganographic system to cracking is comparable with the value of the additional key determined by the type of set $\mathcal{K}_a$.

However, based on the physical principles of functioning of steganosystems, the key applied for deposition of information that relates to the principles of selecting the location for placing a particular piece (bit, byte, symbol, etc.) of any message $M_i \in \mathcal{M}$, $i = 1, 2, ..., n$ in document-container $C_j \in \mathcal{C}$, $j = 1, 2, ..., p$ should be assigned to set $\mathcal{K}$. Because in practice the generation and the use of additional keys is not connected with the base algorithm of deposition/extraction of some message M_i in the container C_j using any key K_m, in such a case the discussed method of transmission of information is considered as a two-key steganographic system.

Definition 11.9. The *two-key steganography system* Σ_2 is an ordered structure consisting of 7 connected elements: the set of messages $\mathcal{M}$, containers $\mathcal{C}$, keys $\mathcal{K}$, additional keys $\mathcal{K}_a$, steganomessages (filled and empty containers, or merely filled containers) $\mathcal{S}$, and sets of transformations (forward $\mathcal{F}$ and reverse $\mathcal{F}^*$):

$$\Sigma_2 = (\mathcal{M}, \mathcal{C}, \mathcal{K}, \mathcal{K}_a, \mathcal{S}, \mathcal{F}, \mathcal{F}^*). \tag{11.6}$$

In accordance with (11.6), steganography conversion (embedding and retrieval of information) for two-key steganographic systems is described in a general form by the relations:

$$\mathcal{F} \ni F : \mathcal{M} \times \mathcal{C} \times \mathcal{K} \times \mathcal{K}_a \to \mathcal{S}, \tag{11.7}$$

$$\mathcal{F}^* \ni F^* : \mathcal{S} \times \mathcal{K}^* \times \mathcal{K}_a^* \to \mathcal{M} \times \mathcal{C} \left(\times \mathcal{K} \times \mathcal{K}_a\right). \tag{11.8}$$

The transformations defined by (11.1), (11.3) and (11.7)-(11.8) should be referred to the number of functional (or more precisely – relational) dependencies, because between each element of the set specified by an appropriate relation on the left-hand side (from an arrow) and on the right side there must be the one to one correspondence. However, since practically the receiver is not interested in which keys were used to code the message, we

often write $\mathcal{M} \times \mathcal{C}$ instead of $\mathcal{M} \times \mathcal{C} \times \mathcal{K} \times \mathcal{K}_a$ on the right side of (11.8). In addition, note that in all the steganographic systems the primary purpose of which is the transmission of information and the conditions (information content) characterizing the container are not substantially important, the expressions (11.3) and (11.8) can be written in the canonical form:

$$\mathcal{F}^* \ni F^* : \mathcal{S} \times \mathcal{K}^* \to \mathcal{M}, \tag{11.9}$$

$$\mathcal{F}^* \ni F^* : \mathcal{S} \times \mathcal{K}^* \times \mathcal{K}_a^* \to \mathcal{M}. \tag{11.10}$$

However, for systems related to the problems of protection of intellectual property rights (in case of document containers), the only valuable expressions are (11.3) and (11.8). This follows from the logic of the task: the inverse transformation should give an equivalent result and fully restore the original elements of the system – the secret (authorial) information $M_i \in \mathcal{M}$ and the container document $C_j \in \mathcal{C}$, the authorship of which must be confirmed by a message M_i.

Now we turn to the discussion of the set $\mathcal{K}_a$. Due to the arguments presented above in points 1) – 2), this set can be represented as the sum of nonintersecting subsets, the number of which corresponds to the number of types of additional keys used in the system (for example, cryptographic keys may form a subset $\mathcal{K}_{a,c}$). If the set of additional keys is limited only to cryptographic keys, the mentioned set $\mathcal{K}_a$ and the subset $\mathcal{K}_{a,c}$ coincide. Having fixed the set of keys $\mathcal{K} = \{K_1, K_2, ..., K_z\}$ and the set of additional keys $\mathcal{K}_a$ as cryptographic keys $\mathcal{K}_{a,c} = \{K_1^{a,c}, K_2^{a,c}, ..., K_l^{a,c}\}$ such that for all $w = 1, 2, ..., l$ the transformation F_w according to (11.7) is uniquely defined by these keys, it can be formally rewritten as follows:

$$F_w : \mathcal{M} \xrightarrow{K_w^{a,c}} \mathcal{M}_c \nearrow \mathcal{M}_c \times \mathcal{C} \xrightarrow{K_w} \mathcal{S}, \tag{11.11}$$

and

$$F_w^* : \mathcal{S} \xrightarrow{K_w^*} \mathcal{M}_c \times \mathcal{C} \searrow \mathcal{M}_c \xrightarrow{\left(K_w^{a,c}\right)^*} \mathcal{M}, \tag{11.12}$$

where $w = \{1, 2, ..., l\}$, and $\mathcal{M}_c$ is the set of cryptomessages. In the last case each of the transformations (11.11) and (11.12) may be viewed as a compound process of several operations. However, since (11.11) and (11.12) in general represent relations, the arrows here denote rather ordering of their connections than mappings; merely some of these operations are functions. The graphic representation of the corresponding transformation, for example (11.11), is built on the basis of observation that the first part of relation (11.11) is nothing else than the map acting on the set of open messages into

a set of cryptograms, (i.e. cryptograms obtained by means of the encoding key $K_w^{a,c}$), see the below Fig. 11.5.

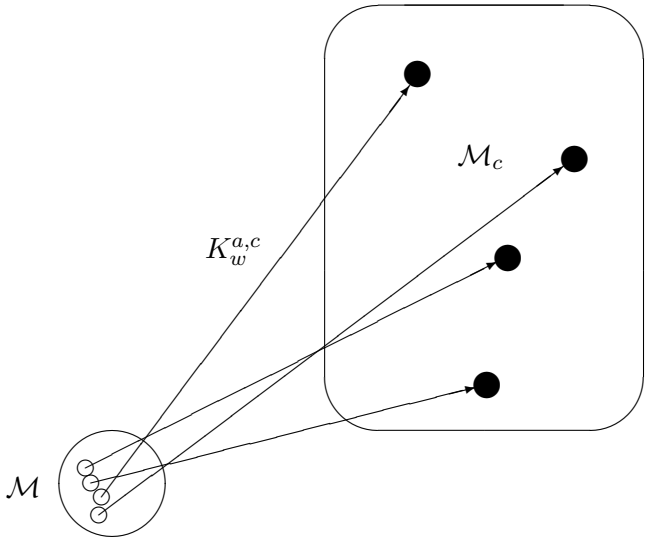

Figure 11.5: Graphic representation of the first part of map (11.11): $\mathcal{M} \overset{K_w^{a,c}}{\longrightarrow} \mathcal{M}_c$.

The last part of the transformation F_w^* given by (11.12), say $(F_w^*)_{\dashv}$, graphically looks the same, with the only difference that instead of the set of plain texts $\mathcal{M}$ we use the set of cryptograms $\mathcal{M}_c$:

$$(F_w^*)_{\dashv} : \mathcal{M}_c \overset{\left(K_w^{a,c}\right)^*}{\longrightarrow} \mathcal{M}. \tag{11.13}$$

The symbol $\nearrow$ in (11.11) denotes the operation of injection of $\mathcal{M}_c$ into the Cartesian product $\mathcal{M}_c \times \mathcal{C}$ by a suitable relation, while $\searrow$ in (11.12) is the projection of $\mathcal{M}_c \times \mathcal{C}$ onto $\mathcal{M}_c$.

One should remember about the important condition of unambiguity of reverse transformations: the number of elements contained in sets interconnected by keys of transformations must be the same. If the direct and inverse cryptographic transformations are carried out on the basis of an asymmetric system, the keys used for encoding and decoding of information will be different for the forward and reverse steganography transformations. However, we have provided this situation, taking $K_m \neq K_m^*$; $K_m \in \mathcal{K}$, $K_m^* \in \mathcal{K}^*$; $m = 1, 2, ..., z$, where, in general, K_m and K_m^* are respectively some fixed keys of direct and reverse transformations. Strictly speaking,

for such systems the expressions (11.11) and (11.12) should be changed as follows:

$$F_w : \mathcal{M} \overset{\left(K_w^{a,c}\right)_e}{\longrightarrow} \mathcal{M}_c \nearrow \mathcal{M}_c \times \mathcal{C} \overset{K_w}{\longrightarrow} \mathcal{S}, \tag{11.14}$$

and

$$F_w^* : \mathcal{S} \overset{K_w^*}{\longrightarrow} \mathcal{M}_c \times \mathcal{C} \searrow \mathcal{M}_c \overset{\left(K_w^{a,c}\right)_d^*}{\longrightarrow} \mathcal{M}, \tag{11.15}$$

because the encryption key is not equal to the decryption key: $\left(K_w^{a,c}\right)_e \neq \left(K_w^{a,c}\right)_d^*$, where $K_w^{a,c} = \left(\left(K_w^{a,c}\right)_e, \left(K_w^{a,c}\right)_d^*\right)$ are the operations on the selected container and message.

Next we ought to discuss the question concerning the order of using keys from sets $\mathcal{K}$ and $\mathcal{K}_a$, acting on elements in the process of embedding and retrieval information in/from the container. As follows from (11.11) and (11.14), and also from the physical characteristics of considered processes, an additional key $K_w^{a,c}$ (and K_w^a) from the set of keys $\mathcal{K}_{a,c}$ (or $\mathcal{K}_a$ resp.) during the deposition of information is used before the main process of embedding, and the reverse operation should be done after the message is retrieved from the container. In the case of a composite additional key (e.g. encryption and noiseless coding of the message M_i), consisting in general of v subkeys taken from the corresponding subsets of the set $\mathcal{K}_a$: $\mathcal{K}_a = \{\mathbf{K}_1^a, \mathbf{K}_2^a, ..., \mathbf{K}_v^a\}$; $\mathbf{K}_1^a = \{K_{11}^a, K_{12}^a, ..., K_{1n}^a\}$, ..., $\mathbf{K}_v^a = \{K_{v1}^a, K_{v2}^a, ..., K_{vm}^a\}$, the order of using such subkeys during the extracting of information will be reversed (or mirrored) in comparison with the deposition process.

We now turn to the keys which form the set $\mathcal{K}$. We said already that in the simplest case, the key used for embedding of information may include a method or a deposition technique. For steganographic systems of information transmission one of the possible types of containers (text, graphics, audio sounds, image, etc.) can be selected, and for each of them one of the possible algorithms or deposition methods may be implemented (for example, in the case of a text document container the Line-shift coding method, Word-shift coding or otherwise). Recall that only the algorithm of embedding is valuable in the task of protecting intellectual property rights within a particular electronic document. Thus, one subset of keys (say $\mathcal{K}_1$) of the set of keys $\mathcal{K}$ is appropriate for containers of type $\mathcal{C}_1$ and algorithms (methods) for deposition of information messages $M_i \in \mathcal{M}(\mathcal{K}_1)$: $\mathcal{K}_1 = \{\mathcal{K}_1^c, \mathcal{K}_1^a\}$; $\mathcal{K}_1^c = \{K_1^c, K_2^c, ..., K_n^c\}$, $\mathcal{K}_1^a = \{K_1^a, K_2^a, ..., K_h^a\}$. Another subset of keys (say $\mathcal{K}_2$) of the set $\mathcal{K}$ forms tags discussed above, which determine in general the choice of the container elements for depositing the relevant elements of the message. It is evident that these keys may also depend on the type of the used container and the used algorithm of embedding. Therefore, for simplicity, we assume that $\mathcal{K}_2 = \{K_1^2, K_2^2, ..., K_d^2\}$.

It is clear that the keys of sets $\mathcal{K}_1^c$ and $\mathcal{K}_1^a$ may interest us as independent parameters only in the analysis of the system stability to cracking, because the choice of the type of container is not accidental, but of determined character in solving in particular the problems of copyright protection. Therefore, we will only use the combined keys related to the set $\mathcal{K}$ in the formal description of processes under consideration. Considering the foregoing expressions (11.11) and (11.12), in a general case they take the form:

$$F_w : \mathcal{M} \overset{\left(K_{1w}^a\right)}{\longrightarrow} \mathcal{M}_1 \overset{\left(K_{2w}^a\right)}{\longrightarrow} \mathcal{M}_2 \; ... \; \mathcal{M}_{v-1} \overset{\left(K_{vw}^a\right)}{\longrightarrow} \mathcal{M}_v \;\nearrow\; \mathcal{M}_v \times \mathcal{C} \overset{K_w}{\longrightarrow} \mathcal{S}, \tag{11.16}$$

and

$$F_w^* : \mathcal{S} \overset{K_w^*}{\longrightarrow} \mathcal{M}_v \times \mathcal{C} \;\searrow\; \mathcal{M}_v \overset{\left(K_{vw}^a\right)_d^*}{\longrightarrow} \mathcal{M}_{v-1} \; ... \; \mathcal{M}_2 \overset{\left(K_{2w}^a\right)_d^*}{\longrightarrow} \mathcal{M}_1 \overset{\left(K_{1w}^a\right)_d^*}{\longrightarrow} \mathcal{M}, \tag{11.17}$$

where $\mathcal{K}_w = \left\{K_w^1, K_w^2\right\}$; $K_w^1 \in \mathcal{K}_1$, $K_w^2 \in \mathcal{K}_2$; $\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2\}$; $\mathcal{K}_a = \{\mathbf{K}_1^a, \mathbf{K}_2^a, ..., \mathbf{K}_v^a\}$, and $K_{1w}^a \in \mathbf{K}_1^a$, $K_{2w}^a \in \mathbf{K}_2^a$, ..., $K_{vw}^a \in \mathbf{K}_v^a$; $(\mathcal{K}_a)^* = \{(\mathbf{K}_1^a)^*, (\mathbf{K}_2^a)^*, ..., (\mathbf{K}_v^a)^*\}$, and $(K_{1w}^a)^* \in (\mathbf{K}_1^a)^*$, $(K_{2w}^a)^* \in (\mathbf{K}_2^a)^*$, ..., $(K_{vw}^a)^* \in (\mathbf{K}_v^a)^*$.

We can rewrite (11.16) – (11.17) in a general form corresponding to (11.7) – (11.8), but taking into account the order of operations:

$$F : \mathcal{M} \times \mathcal{K}^a \times \mathcal{K} \times \mathcal{C} \to \mathcal{S}, \tag{11.18}$$

and

$$F^* : \mathcal{S} \times \mathcal{K} \times (\mathcal{K}^a)^* \to \mathcal{M} \times \mathcal{C}. \tag{11.19}$$

According to (11.18) the selected key from one set of the additional keys ($\mathcal{K}_a$) is used for the preliminary single or multiple (v-fold) conversion of the deposited message, and then the selected keys of the other set of keys ($\mathcal{K}$) are used directly when implementing the deposition operation of the message $M_i \in \mathcal{M}$ into the container $C_j \in \mathcal{C}$.

The expressions (11.18) and (11.19) allow us to perform the operation of synthesis of the transmitting part of steganography system which by definition (11.9) is called the two-key steganography system. As is seen in Fig. 11.5, this part of the system consists of the following blocks:

1) *the source of the message flow* $\mathcal{M}$, which generates a specific message, say M_i;

2) *the source of flow of empty containers* (e.g. document-containers) $\mathcal{C}$, which furnishes a specific container $C_j \in \mathcal{C}$ for the embedding there the message M_i;

3) *pre-conversion block* (encryption, encoding, etc.) of the message M_i;
4) *a source of additional keys*, i.e. set $\mathcal{K}_a$ containing v various sets of keys, $\mathcal{K}_a = \{\mathbf{K}_1^a, ..., \mathbf{K}_v^a\}$; $\mathbf{K}_1^a = \{K_{11}^a, ..., K_{1n}^a\}$, ..., $\mathbf{K}_v^a = \{K_{v1}^a, ..., K_{vm}^a\}$;
5) *a source of keys* $\mathcal{K}$, and
6) *block of deposition* of the message M_i in the container C_j, forming steganomessage S_q based on the key $K_w \in \mathcal{K}$, $w = 1, 2, ..., l$; $S_q \in \mathcal{S} = \{S_1, S_2, ..., S_r\}$, $q = 1, 2, ..., r$.

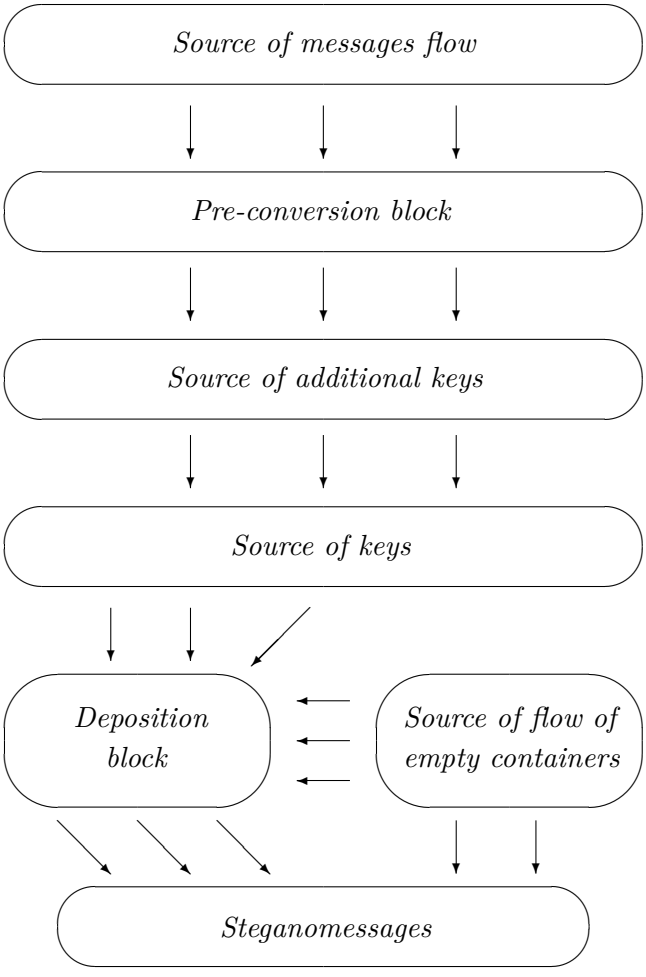

Figure 11.6: Scheme of functioning of the transmitter part of a steganosystem.

The receiver part of the system performs encryption of information using known keys. It is synthesized on the basis of expressions (11.11) and (11.13) in an analogous way. In fact this part of the system works in a reverse order in comparison with the transmitter part, therefore we omit the details.

Structural diagram of the main stages of the multi-key steganography system concerning its transmitter and receiver part is shown in Fig. 11.7.

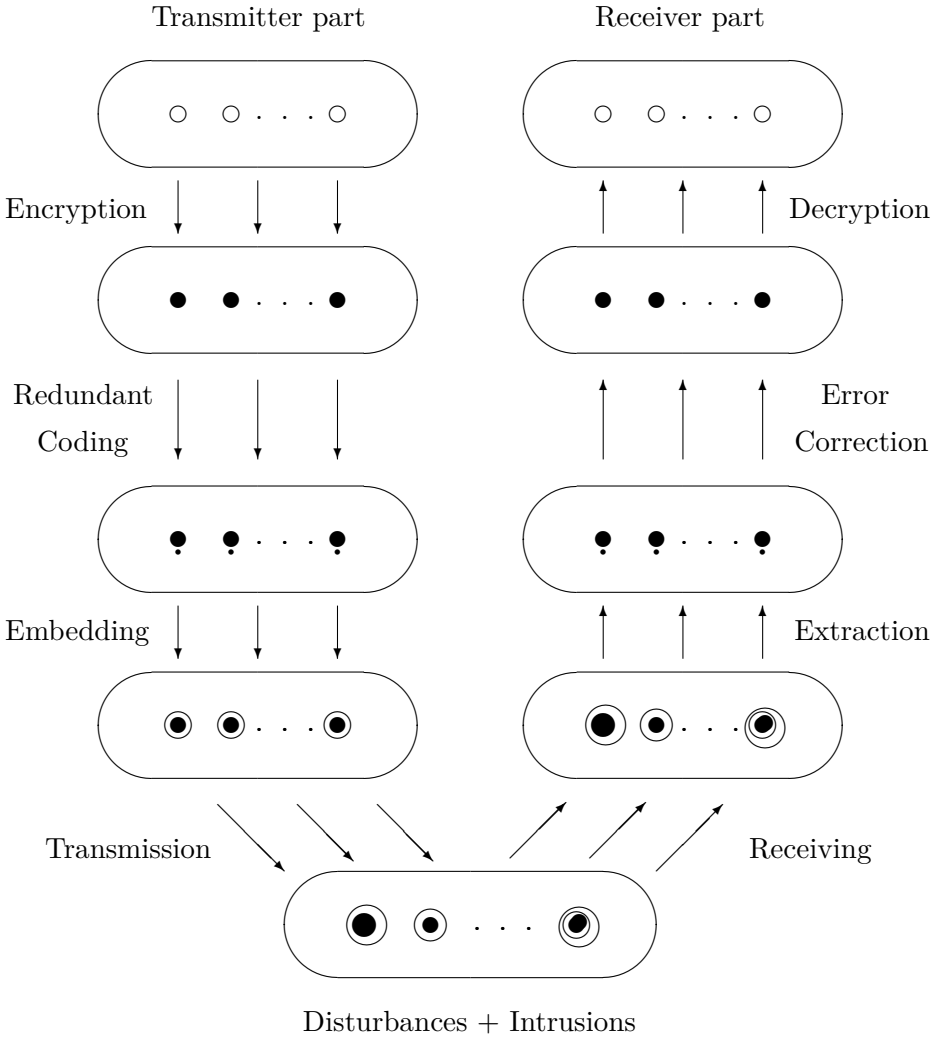

Figure 11.7: Graphic representation of the synthesized cryptographic system.

The state of the system is determined by the type of the transmission channel and effects on the channel caused by external influences (indicated on this scheme as disturbances and intrusions). This structure can be considered as universal.

In solving problems concerning the protection of intellectual property rights of text documents, by the channel we understand any information and communication environment where the protected text document is since its establishment until the commencement of proof of authorship. Thus, the main difference between the developed here mathematical model from the models known earlier is the division of the set of keys used for coding information into two types. This allows to estimate more accurately the resistance of systems to cracking using, for example, the probabilistic approach for all the parameters of the model.

This means that the selection of keys for the process of embedding of an information message can be carried out not in a deterministic manner, but randomly independent from at least two sets of parameters corresponding to two various types of keys ($\mathcal{K}$ and $\mathcal{K}_a$).

In our opinion the greater number of the coding keys, in particular addition of several kinds of cryptographic keys, as well as random selection of keys, without doubt will enhance the security of the sent information and resistance of the system against attacks of unauthorized persons. Moreover, random choice of cryptographic keys should further decrease the risk of decryption of hidden messages by improper persons.

Bibliography

[1] Brassil J., Low S. H., Maxemchuk N. F., O'Gorman L. *Electronic marking and identification techniques to discourage document copying,* IEEE Journal on Sel. Areas in Commun. **13**, 8 (1995), 1495–1504.

[2] Gribunin V. G., Okov I. N., Turincev I. V., *Digital Steganography*, Solon-Press, Moscow 2002 (in Russian).

[3] Konahovich G. F., Puzyrenko A. U., *Computer Steganography*, MK-Press, Kiev 2006 (in Russian).

[4] Kuznetsov A., Smirnov A., Meleshko E., *The mathematical model and flow diagram of the steganography system*, Technika v Silskogospodarskomu Virobnictvi, Galuzevie Mašinobuduvannia, Avtomatizacja **1**, 25 (2012), 273–281.

[5] Low S. H., Maxemchuk N. F., Lapone A. M., *Document identification for copyright protection using centroid detection*, IEEE Trans. on Commun. **46**, 3 (1998), 372–383.

[6] Shannon C. E., *A Mathematical theory of communication*, Bell Syst. Techn. J. **27**, (1948), 379–423.

[7] Shannon C. E., *Communication theory of secrecy systems*, Bell Syst. Techn. J. **28**, (1949), 656–715.

[8] Shutko N. P., *Text steganography as an effective instrument of protection of the copyright on electronic document*, Proceedings of the International Conference on New Electronic Technologies and Their Industrial Implementation, Zakopane 2013, p. 147.

[9] Shutko N. P., Romanenko D. M., Urbanovich P. P. *Mathematical model of the text steganography on the base of modifying the spatial and color settings of text characters*, Proc. of BSTU; Phys.-Math.-Sciences and Informatics, Mińsk 2015, **6**, 152–157.

[10] Urbanovich N., *Development, analysis of efficiency and performance in an electronic textbook methods of text steganography*, Printing Future Days: 4th International Scientific Conference on Printing and Media Technology, Chemnitz 2011, 189–193.

[11] Urbanovich N., Plaskovitsky V., *The use of steganography techniques for protection of intellectual property rights*, Przegląd elektrotechniczny **8**, (2012), 342–344.

[12] Urbanovich P., Urbanovich N., Chourikov K., Rimorev A., *Niektóre aspekty zastosowania metod steganograficznych do przechowywania powiadomień tekstowych*, Przegląd elektrotechniczny **7**, (2010), 95–97.

Pavel Urbanovich, born in 1955. He graduated from the Belarusian State University of Informatics and Radioelectronics (Mińsk); Ph. D. degree (1983) in radioelectronics, Doctor of Science (1993) in computer science, Professor of informatics (1995). In 2000 P. Urbanovich joined KUL. He is the author of more than 350 scientific publications and textbooks. Research interests: methods and means of the reliability improving of the information systems; information protection in computer systems; databases; cryptography and steganography; computer models and software tools to analyze and evaluate environmental problems.

Nadzeya Shutko, born in 1987. She graduated from the Belarusian State Technological University (Mińsk), in 2016 completed postgraduate studies; works as a teacher at the Faculty of Information Technology of BSTU. Research interests: steganography; copyright protection in the Internet.