Cryptographic Semantic System for Provision of the Performance Mode of Perfect Strong

Yevgen Samojlik [1] [0000-0002-0646-5162], Viktor Gnatyuk [1] [0000-0002-4916-7149], Volodymyr Klimchuk [1] [0000-0001-8378-002X], Oleksii Shylo [1] [0000-0002-0887-2677], Oleksandr Tkachenko [2] [0000-0002-4410-7956] and Yuliia Burmak [3] [0000-0002-5410-6260]

[1] National Aviation University, Kyiv, Ukraine
[2] State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, Kyiv, Ukraine
[3] Kyiv College of Communication, Kyiv, Ukraine
nezakoo@gmail.com

Abstract. Perfectly stable existing cryptosystems with a theoretically proven ideal theoretic-informational stability have a limited scope of use primarily due to the need of complying with the condition of non-exceed during encryption of the so-called keys and unicity distance. Natural languages are characterized by relatively small values of the unicity distance. It means that one often has to change the keys of ciphers. Unicity distance can be significantly increased by using the artificial language of display of the applied area with an alphabet of large dimension. However, in this case, the problem of passing information spread is weakened, but not eliminated fully. A perfectly stable crypt-semantic method of protecting text messages, which is deprived of the above-mentioned disadvantage, is synthesized. An algorithm for the implementation of this method of protection is provided. This method provides an ideal (formally proved) strong of crypt-semantic ciphers, concerning the protection of the meaning of text information that does not depend on either the ratio between the total length of the texts and the length of the cipher keys, nor the statistical properties of the pseudorandom number generator (PRNG), which is included in the scheme of this method implementation.

Keywords: text information protection, crypt-semantic security method, perfectly strong cryptosystem, unicity distance, cryptosystem synthesis.

1 Introduction

In practice, methods for symmetric cryptography, the classification and characteristics of which are widely covered in scientific publications (see, in particular, [1]) are often used to ensure the confidentiality of information circulating between a limited range of authorized agents of the key (password) information holders. However, in many cases these methods do not provide an absolute guarantee of protection (in other words, they do not provide one hundred percent guarantees at the formal level regarding the impossibility of their disclosure). It is always possible to perform direct search for possible options for the password, although for this, in most cases, it is necessary to perform an unreasonably large amount of computational work and, finally, to "break" the protection, that is, to make a substantiated conclusion regarding the content of the source information. The exceptions are the methods underlying the perfectly strong cryptosystems. The perfection of strong is understood as an opponent’s interception of any cryptograms in any quantities that are encrypted by these methods, which does not reduce uncertainty in the possible choice of open texts [2,3]. The Mauborgne/Vernam scheme [4] is an example of a perfectly strong cryptosystem with a theoretically proven ideal of theoretical and informational strong, which ensures the impossibility of unequivocal restoration of open text messages, even when a crypto-analyst has samples of encrypted messages of arbitrarily large total length, and the crypto-analyst has unlimited time and unlimited computing resources to decrypt intercepted cryptograms [1]. The Mauborgne/Vernam scheme provides an absolute that is, theoretically unlimited, level of encrypted texts strong, but provided under the condition that the total length of these texts does not exceed the length of the cipher key. If the perfection of the cryptosystem strong is theoretically proved, any cryptographic analysis of its design, that is, the analysis of its theoretical and informational model loses its meaning. In this case, only the vulnerability of the decisions made on the variants of technical implementation of this model can be discussed.

However, the Mauborgne/Vernam scheme its own disadvantages, which considerably narrow the areas of its use. First of all, it is the need to adhere to the basic condition for the perfect strong of the cryptosystem - non-exceeding of the so-called keys and unicity distance in the encryption [1,2]. It means that it is necessary to ensure that the current total length of encrypted text messages in the encryption process does not exceed the length of the cipher key. Otherwise, there is a theoretical possibility of disclosing a cipher key, since during decoding there is the possibility of receiving only one (single) meaningful message that is identical to the transmitted one [1,2]. If the unicity distance does not exceed, then as a result of decryption by the method of a direct search of variants of implementation of the cipher key two and more meaningful messages are received, which leads the crypto-analyst to the deprivation of the possibility of unambiguous installation of the true key cipher. Numerous studies of many authors point the relatively small values of the unicity distance when encrypting messages, composed of symbols of the alphabet of any of the natural languages [1,3]. It leads to the need of frequent changes in key information, which is a problem for many applications. In addition, it is necessary to ensure randomness and equal probability of choosing the key implementation options [1], which also creates some difficulties during the practical implementation of this scheme.

In the article [5] the method of increasing the unicity distance by the synthesis of the artificial display of the applied region with an alphabet of large dimension is proposed. It is proved that the performance indicators of perfectly strong cryptosystems, which use the mechanism of enlarging the language alphabet of the display of text messages, have significantly higher values compared with the indicators of other methods to provide a regime of perfect secrecy. In particular, in article [5] the effectiveness of the method of textual information protection with the enlargement of the language alphabet of displaying this information is comparable to the efficiency of the Mauborgne / Vernam scheme. As a criterion of efficiency, the win index in the key length is taken (under other equal conditions):

$$\boldsymbol{Z} = \frac{\boldsymbol{H_1}(\mathbf{K})}{\mathbf{H}(\mathbf{K})}, \tag{1}$$

where H(K) is the encryption operator of the entropy of the security system as the size of the encryption key space, which depends on the number of cipher keys possible for use; K – the number of possible encryption keys used in the protection system with the enlarged alphabet (H(K)= log(K));

H1(K) is the entropy of the cipher key in the case of use of the Mauborgne/Vernam scheme for the protection of information in the mode of perfect secrecy (the length of the key in this case should be equal to the length of the message [1]). The following expression is used to determine the entropy of such a key [5]:

$$H_1(\square) = log_2(B_1^{\square 1}), \tag{2}$$

where n_1 is the encryption operator is the length of the message, which is written in the language of human communication (Ukrainian, Russian, English etc.), which has an alphabet B1.

While the entropy of the key used to protect information by using the enlarged alphabet, the display language of this information is calculated as [5]:

$$\square(\square) = log_2 \prod_{\square=1}^{\square}[B - (s - 1)], \tag{3}$$

B is the alphabet of the table form, that is, the number of possible combinations of tabular lines of the semantic dictionary, which is artificially created as a result of statistical and semantic domain analysis, n – the number of rows in the tabular form, and s – the number of columns in the tabular form.

In study [5] it is shown that the win Z for the effectiveness of a cryptosystem with an enlarged alphabet, compared with the Mauborgne/Vernam scheme, may be 50 or more. Nevertheless, the problem of the distribution of passinformation, has been weakened, but not eliminated in this case.

In this paper, a perfectly strong crypt-semantic method of protecting text messages, which is deprived of the above-mentioned defects, is proposed and substantiated. A formal justification is provided to point this method provision of an absolute (formally proved) strong of crypt-semantic ciphers to protect the meaning of information, which does not depend either on the ratio between the total length of the texts and the length of the cipher keys, nor on the statistical properties of the pseudorandom number generator (PRNG) that is included in the scheme of implementation of this method [7-12]. Crypt-semantic security system allows you encrypt as many volumes of text messages as you want in perfect strong regardless of the values of the unicity distance.

2 Rationale of the Possibility to Create Cryptosystems, Which Provide the Mode of Perfect Strong Without the Necessary Changing of the Key Cipher

The process of encrypting text messages, provided in written or spoken form, will be presented in the following form:

$$\{\mathbf{S}\} \xrightarrow{\mathbf{P}_{\mathbf{SF}}} \{\mathbf{F}\} \xrightarrow{\mathbf{P}_{\text{Ш}}(\mathbf{x})} \{\mathbf{F}_{\text{Ш}}\}, \qquad (4)$$

where **{S}** is the encryption operator is semantic image-space, elements of which are used during the formation of outgoing text messages;**{F}** – space of reflections by means of the chosen language of semantic images, taken from space **{S}**; **$\{F_{\text{Ш}}\}$** – space of encrypted reflections of semantic images from space **{S}**; **P_{SF}** – the operator of the transformation of semantic images into their reflections within the framework of the chosen language, **$P_{\text{Ш}}(x)$** – encryption operator, which depends on the choice of the cipher key x.

The process of decoding of spoken messages will be presented in the following form:

$$\{\mathbf{F}_{\text{Ш}}\} \xrightarrow{\mathbf{P}_{\text{рш}}\,(\mathbf{x})} \{\mathbf{F}\} \xrightarrow{\mathbf{P}_{\mathbf{FS}}} \{\mathbf{S}\}, \qquad (5)$$

where **$P_{\text{рш}}$ (x)** is the encryption operator is decryption operator, which depends on the choice of the cipher key x, **P_{FS}** – the operator of the transformation of reflection of semantic images into output semantic images. We use crypt-semantic approach to ensure perfectly strong information security to reproduce operators **P_{SF}** and **P_{FS}**, while operators **$P_{\text{Ш}}(x)$** та **$P_{\text{рш}}(x)$** will be implemented by any of the known methods of symmetric cryptography.

We emphasize that in expressions (4) and (5) operators **$P_{\text{Ш}}(x)$** and **$P_{\text{рш}}$ (x)** do not depend on the semantic characteristic of space elements **{S}**, which means that encryption is carried out without consideration of the meaning of text messages (which is why the methods of classical cryptography are universal in relation to the meaning of the information).

We also emphasize that in this case the operator of the transformation of the reflection form of any semantic image by means of any natural language in its semantic content of **P_{FS}** has determined (by grammar rules) character.

The axiom, concerning the uniqueness of the inverse transformator from "intercepted sample of cipher graph" to "the true meaning of the sample of the original open message", lies on the basis of the known methods of cryptographic analysis of encrypted text messages. That is, if

$$\{\mathbf{F}_{\text{Ш}}\} \xrightarrow{\mathbf{P}_{\text{ш}\mathbf{S}}\,(\mathbf{x})} \{\mathbf{F}\}, \qquad (6)$$

where **$P_{\text{ш}S}(x)$** is the encryption operator is the operator of the inverse transformation of encrypted text messages directly into the content of outgoing open messages, then **$P_{\text{ш}S}(x)$** has determined character.

Therefore, whatever theory or hypothesis a cryptographic analyst would have put as the truth of the cipher key variant identified, he always has the opportunity to use an effective tool to verify the correctness of the results of his cryptanalytic research. Namely, if the key variant, determined by him, provides the transformation of a meaningless sample of the cipherogram in the text, having a specific content, then the results of cryptanalysis in accordance with the above axiom on the uniqueness of the transformation are considered successful. Consequently, any known cipher (with the exception of the Mauborgne/Vernam scheme, which, under certain conditions, does not ensure the uniqueness of the transformation), from a theoretical point of view, may be compromised.

The main purpose of the study, presented in this publication, is to create a method of information protection with the emphasis on violations of privacy. The method provides a view of the absolute confidentiality of protected messages from a formal point, provided when the total amount of encrypted information in a formally unlimited extent exceeds the length of the passwords used. (Implementation of such a method will, on the one hand, give one hundred percent confidence in ensuring the confidentiality of messages, and on the other hand, will allow get rid of the problem of distributing cipher keys, since in the case reviewed in this paper, the need to keep the distance of unity is eliminated with the optimal choice of length of the cipher key).

To achieve this goal, we will use the crypt analogy approach to ensure confidentiality, which involves the dependence of operators $\mathbf{P_{SF}}$ and $\mathbf{P_{FS}}$ from the semantic characteristics of elements of the space of semantic images **{S}**, that is, operators of encryption / decryption of the crypt semantic security system $\mathbf{P_{Ш}(x, S)}$ and $\mathbf{P_{рш}(x,S)}$ should be dependent both on the choice of the cipher key, and on the choice of an element from the space of semantic images. That is, the process of encryption will be carried out in the following form:

$$\{S\}\xrightarrow{\mathbf{P_{ш}(x, S)}}\{\mathbf{F_{Ш}}\}, \tag{7}$$

where $\mathbf{P_{ш}(x, S)}$ is the encryption operator is the encryption operator of the crypt semantic system, which is dependent on both the selection of the key of the cipher x, and from the choice of an element from the space of semantic images **{S}**.

In this case, the encryption will be done taking into account the meaning of the text messages, and the above axiom for the uniqueness of the transformation "intercepted encryption sample" will not be true to its true meaning according to expression (6).

Let’s determine the mental activity of man as a psychophysiological process of creating (generating) virtual semantic flows, representing a finite-dimensional determined (not random) sequences of discrete semantic images (**SO**), which possess the property of meaningfulness. In the process of mental activity, the subject can create both statistically and semantically unrelated streams. However, due to the lexicographic effects, inherent in any language of reflection of semantic images [6], a separate defined semantic flow is possible and appropriate to be regarded as discrete non-random sequence **SO**, which reflects the direction of human thoughts. At the same time, we will assume that semantic image **SO** is a discrete element of the semantic flow, perceived by the subject as a virtual logically consistent semantic construct that has signs of semantic completeness (that is, self-sufficiency content that does not require mandatory additional explanations). Since a specific semantic flow (**SP**) is determined to have an identified sequence of **SO**, then **SP** achieves the property of meaningfulness as well [12-15].

Assuming that the main formal characteristics of both **SO** and **SP** is the level of abstraction **(aggregation, generalization)** of their representation, then the flow of semantic images (**SO**) in the general case can be represented as follows:

$$\boldsymbol{SP^{(i)} = SO_1{}^{(i-1)},\ SO_2{}^{(i-1)}, \dots, SO_k{}^{(i-1)}, \dots, SO_N{}^{(i-1)}}, \tag{8}$$

where $SP^{(i)}$ is the encryption operator is semantic flow of *i*-level of abstraction i; *i* is index of abstraction level; $SO_k{}^{(i-1)}$ – *κ*-element of $SP^{(i)}$, that has more detailed abstraction level of representing semantic measure; *κ* – number of *SO* in sequence of semantic images, which show *SP*; *N* is length of sequence *SO*, which comply *SP*.

If the comparative semantic images are perceived by the subject as identical in terms of meaning, but he/she has doubts about the truth of such perception, it is expedient to determine the semantic relation as plausible in this case. That is, if the symbol ^ is chosen as a sign of the likelihood relation, then any two semantic units of any $TZ_a{}^{(i)}{}_k$ and $TZ_b{}^{(i)}{}_k$, which are of the same level of abstraction, are in relation to semantic plausibility

$$\boldsymbol{TZ_a{}^{(i)}{}_k \ \wedge\ TZ_b{}^{(i)}{}_k}\ , \tag{9}$$

If the probability that $TZ_a{}^{(i)}{}_k$ is identical in essence of the content to $TZ_b{}^{(i)}{}_k$ is less than 1.

It is important to emphasize that the relation of plausibility is determined between the semantic units of any, but the same level of abstraction.

In this paper, to obtain the dependence of the encryption operator from {**S**}, it is proposed to replace the true semantic image of the outgoing message (that is, one of the elements of space {**S**}) to another element of the semantic thesaurus of this applied area, the semantic image of which is in relation to semantic plausibility with the true meaning of the outgoing message, that is

$$\boldsymbol{SO_k{}^{(1)} \wedge \{SO_1{}^{(1)} \wedge\ SO_2{}^{(1)} \wedge\ \dots\dots \wedge SO_M{}^{(1)}\}}, \tag{10}$$

where $SO_k{}^{(1)}$ is the encryption operator – is true, for example, *κ*-semantic image of the *i*-level of abstraction in the semantic flow $SP^{(i+1)}$ of (i+1)-abstraction level, that is the sequence of semantic images of the i-abstraction level, which represent $SP^{(i+1)}$ (see (8));

$\{SO_1{}^{(1)} \wedge\ SO_2{}^{(1)} \wedge\ \dots\ \dots\ \wedge SO_M{}^{(1)}\}$ – a set of *M,* plausible for the true semantic images of the original message, which are part of the semantic thesaurus;

$\kappa \neq \{1, 2, \dots, M\}$.

Note 1. More detailed descriptions of semantic thesaurus will be provided in subsequent publications.

Under these conditions, crypto-analytic work is meaningless, since he will lose the opportunity to distinguish the true meaning of encrypted messages from many other false, but plausible semantic images.

3 Synthesis of the crypt-semantic method to protect the meaning of information

Assuming that the functional space of the application area of a certain i-th information system Πi is specified, the target use of which is related to the transmission, processing and storage of text messages in the form

$$\mathbf{8\Pi_i(\Phi_i,Z_i)} \text{ is } \{\mathbf{\Phi_{i,},Z_i}\} = \{\Phi_{i,1}, \Phi_{i,2}, \dots\ \Phi_{i,N};\ Z_{i,1}, Z_{i,2}, \dots Z_{i,M}\}, \qquad (11)$$

where $\mathbf{\Phi_i}$ is the encryption operator – the functional structure of the i-th information system, consisting of elements of the set Φ_i admissible for the performance of functions $\Phi_{i,к}$, where к=1, 2, …. , N;

$\mathbf{Z_i}$ - the functional structure of the i-th information system, consisting of elements of the set Z_i bounding conditions Z_i, where k = 1, 2, …,M, where the execution of the functions of the i-th information system is admissible;

$\Phi_{i,k}$ - k-functional element from a given set of allowable functions for the i-th information system Φ_i, where k = 1, 2, …, N;

$Z_{i,k}$ - k-th bounding condition from a given set of boundary conditions Z_i, where k = 1, 2, …, M;

N, M - number of items in accordance with sets Φ_i and Z_i.

Expression (11) specifies the functionality of the applied space $\mathbf{\Phi_i}$ of i-th information system and the space of $\mathbf{Z_i}$ conditions, in which this information system can be used for its intended purpose.

It is necessary to provide protection of the meaning of transmitted processed and stored information by means of the i-th information system, from violations of confidentiality at the level of providing absolute guarantees of the impossibility of breaking the security system, from both theoretical (formal) and practical points of view. In this case, the amount of information to be protected, (for any length of the cipher key) should not be limited to the length of this key, and the length of the key should be chosen depending on the permissible value of the probability of making non-false solutions in the process of the semantic flow decoding.

Formal way to complete the task

Crypt-semantic approach should be used to complete the task. In order to implement operators of encryption/decryption $\mathbf{P_{SF}}$ та $\mathbf{P_{FS}}$ in expressions (4) and (5) to the methods of semantic cryptography, it is important to show the space of given application area of the i-th information system $\mathbf{\Pi_i}$ on the space of semantic images $\mathbf{S_i}$, considering semantic relations between them, which are determined by the given space of restrictive conditions $\mathbf{Z_i}$, that is

$$\mathbf{\Pi_i(\Phi_i, Z_i)} \xrightarrow{\mathbf{P_{\Phi S}(Z_i)}} \{\mathbf{S_i}, \mathbf{Z_i}\}, \qquad (12)$$

where $\mathbf{P_{\Phi S}(Zi)}$ is the encryption operator – operator of $\mathbf{\Pi_i}$ reflection on $\mathbf{S_i}$, considering $\mathbf{Z_i}$.

In fact, {$\mathbf{S_i}$, $\mathbf{Z_i}$} specifies the structure of the semantic thesaurus of a given application area, the formal synthesis of which must be carried out within the framework of the corresponding linguistic corps [6]. In the tasks of crypt semantics, the bases of linguistic corpuses are semantic dictionaries (another name - thesaurus of knowledge), which contains a whole set of semantic elements (semantic images) that are in a certain way interconnected. It represents a complete linguistic representation of the space of those applied areas, for which the corresponding semantic dictionaries were created. The synthesized structure of the thesaurus, which is supposed to be used during the construction of a security system, involves the dependence of operators P_{SF} and P_{FS} on the semantic characteristics of the space elements of semantic images{S}.

The crypt-semantic method of protection of text messages' meaning (by creating a symmetric cryptosystem) is proposed. In the system, encryption/decryption operators $\mathbf{P_{Ш}(x, S)}$ та $\mathbf{P_{рш}(x, S)}$ are dependent not only on the choice of cipher key from the space **x**, but also on the choice of an element from the space of semantic images $\mathbf{S_i}$.

According to this method, each semantic image of $\mathbf{S_i}$ is placed in a certain way by a certain set of other semantic images of the same $\mathbf{S_i}$, which are in this semantic way with respect to semantic plausibility in the structure of the thesaurus (see expression (9)). This structure of the thesaurus enables replacement of encrypted messages with other plausible messages that do not reflect the true meaning of outgoing messages under certain conditions during encryption. While decrypting, if the password is known, making back replacements of plausible encrypted messages is the true sense. It is clear that for an attacker the password is unknown. Consequently, there are no possibilities for making back replacements, as shown by further analysis of this method of protection.

The proposed method of protection involves the creation of a security system in two stages: first, a semantic dictionary (thesaurus) $\mathbf{T_Z}$ of application area $\mathbf{\Pi_i}$. Software and hardware for the implementation of encryption/decryption operators $\mathbf{P_{Ш}(x, S)}$ and $\mathbf{P_{рш}(x, S)}$ are developed (see expressions (4) - (5)), where the created thesaurus is used.

Operators of encryption/decryption of the crypt-semantic security system $\mathbf{P_{Ш}(x, S)}$ and $\mathbf{P_{рш}(x, S)}$, following this method, are dependent both from the choice of the cipher key from **x**, and from the choice of the element from the space of semantic images **S**. The encryption process involves generation of the outgoing message (i.e., one of the elements of the space **{F}**), and also of the representative sample - the set of encrypted mappings of false but plausible semantic images (that is, in a certain way a certain number of elements from space $\{\mathbf{F_{Ш}}\}$, the source semantic images of which are related to the likelihood ratio with the encrypted display of the semantic image of the true sample of a text message), occurring along with the encrypted display of the true semantic image. The work of any crypto-analyst, as a rule, loses its meaning under such conditions, since he/she is deprived of the opportunity to distinguish the true meaning of encrypted messages from false, but plausible semantic images.

In some cases, information about the number of L elements in a subset of semantic images $\{F^k_{Ш}\}$ may be useful for crypto-analytics. These subsets are related by the ratio of plausibility to the encrypted reflection of the semantic image of the true sample text message S_k. This information is considered known and, if L is small relative to 1, then the probability of making an error-free decision on S_k in the decryption process may be sufficiently large to consider the decryption result to be useful to the intruders. Therefore, from the point of view of legal users of the security system, depending on the specific conditions of the application area $\mathbf{\Pi_i}$, it is expedient to increase the value of L to a certain permissible level of L_0, excess of which makes the work of crypt-analytics meaningless. Under these conditions, in the general case (if not carrying out a probabilistic analysis of the stream of output semantic images) the probability of making an error-free decision in the decryption process regarding the truth of S_k will be determined from the expression

$$p(S_k) \approx 1/ L_0. \tag{13}$$

Therefore, the minimum possible value of the key length of the cipher n_{min} is expedient to choose from the ratio below:

$$n_{min} \geq | \log_2 L_0|. \tag{14}$$

That is, the length of the encryption key for crypt-symantec encryption should be chosen depending on the permissible value of the probability of making false solutions in the process of decoding the semantic flow.

Algorithm for the implementation of the crypt-semantic method of text information security

Let the following notations will be accepted:

S – is a subject that synthesizes a sample of an output text message D in the process of solving application tasks in a given application area;

D – is a sample of the open (unencrypted) text message, semantic content of which requires security (S:D, where «:» - sign of visual or vocal perception);

S_F – is the operator of the processing of the original sample according to the grammar rules of the language (natural or artificial), adopted to display text messages of a given scope of application, using the means of a specially developed linguistic corpus;

F_D –is a sample of the correct open text message, created according to grammar rules and additional information about the structure of this message (adopted system of text units, localization of text units in a discrete sequence of these units, results of marking (markup) of outgoing message according to linguistic characteristics, etc.);

S_C – is the operator of analysis of a sample of a corrected text message, developed for matching elements of the semantic thesaurus of the application area and determining the parameters of the localization of this message in the structure of the thesaurus;

C_D – is a sample of an open text message that directly reflects the meaning of the original text message and corresponds to accepted grammatical rules and semantic constraints;

S_Z – is an encryption operator that converts the output sample of an open text message, such as an encrypted text message, that has a plausible, but most likely other meaning (which provides ambiguity in the perception of the meaning of the encrypted message);

Z_D – is a sample of an encrypted text message (synthesized from the semantic units of the used thesaurus), the semantic content of which is most likely different from the true meaning of an open text message;

S^0_Z – is decryption operator, which provides the transformation of ambiguous meaning of an encrypted text message, such as a text message that uniquely reflects the true meaning of an open text message;

$Z^0{}_D$ – is a sample of a decoded text message that completely matches the C_D, ie. $Z^0{}_D \leftrightarrow C_D$, where the symbol ↔ means the convergence of the form and meaning of the messages $Z^0{}_D$ and $C_{D.}$

Then the algorithm for the implementation of the cryptanalysis method should reproduce the sequence of operations for the processing of text messages, shown in Fig. 1 in the form of a diagram of forms representation of the processed sample of a text message and the operators of the transformation of these forms by the means of protection system.

Confidentiality of text messages according to this method is provided as follows. A sample of an unencrypted output text message D, by means of a specially designed linguistic corpus, is processes according to the rules S_F of language grammar (natural or artificial), adopted to display text messages of a given application area. As a result, the sample of an open text message F_D is formulated in accordance with the rules of grammar and the additional information about the structure of this message (the adopted system of linguistic units, the localization of linguistic units in the discrete sequence of these units, the results of marking (markup) of the output message according to the linguistic characteristics, etc.). Next, the analysis of the S_C sample F_D for matching elements of the semantic thesaurus, is provided. It is specially designed to display the meaning of text messages and semantic links between them in the form of corresponding semantic relationships for a given application area. As a result, a sample of the output open message C_D is received. It reflects the true meaning of the original text message and corresponds to accepted grammatical rules and semantic constraints without any distortion.

Next step is the execution of encryption with the use of any known symmetric cipher S_Z, which is based on the use of a pseudorandom sequence (number) generator (PRNG). Namely, PRNG is set to the original state according to the password. Then, by way of gamma (performance of *XOR*- add by module 2) the elements of the message C_D are replaced with the plausible elements from the thesaurus. (The choice of methods for replacing the true meaning of messages with plausible depends on the synthesized structure of the thesaurus, in particular on the number of the calculated levels of abstraction of its semantic units.)

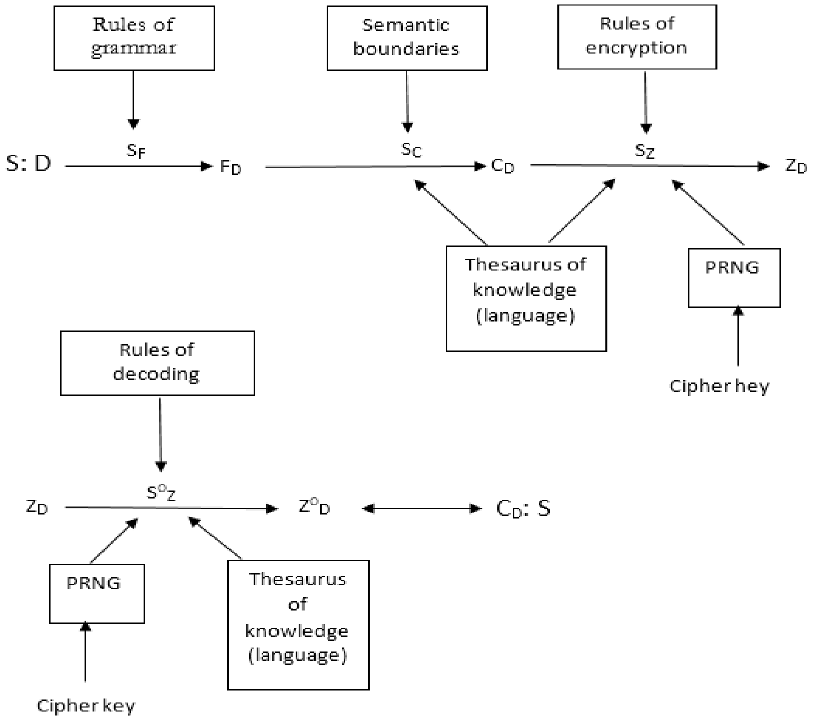

Fig. 1. Algorithm for implementing crypt-semantic security of text messages.

As a result, a sample of an encrypted text message ZD is received. It has a plausible, but, probably, other meaning (that provides ambiguity in the perception of the meaning of encrypted messages). To decode message ZD, operator of reverse transform S0Z is used. Namely, with the help of a known password, the PRNG is set at the same initial state that was encrypted, and then the operation XOR is used repeatedly, so the elements of the ambiguous meaning of the encrypted message ZD , by choosing from thesaurus, are replaced with elements of the message Z0D, which reflects the true meaning of the original text message of CD and corresponds to the accepted grammar rules and semantic constraints of the given application area.

Briefly, the goals of the method can be defined as inclusion of the lexicographic system in the symmetric system of cryptographic information protection in such a way that in the process of encryption the semantic ambiguity of encrypted message samples is provided. In this case, the means of the lexicographic system used must be capable of carrying out grammatical and semantic analysis of messages within the given application area.

4 Discussion and Conclusions

The possibility of creating perfectly stable systems of semantic content maintenance of text messages, the levels of stability of which do not depend on the length of the cipher key, is substantiated. It has been shown that in order to ensure absolute confidentiality of messages provided under conditions of the exceeded total amount of encrypted information, which makes the length of the passwords bigger, it is important for the encryption / decryption operators of the security system be dependent not only on the selection of the cipher key, but also on the choice of an element from the space of semantic images of the application. The above described condition can be fulfilled if a crypt-semantic approach is used to ensure confidentiality. To obtain the dependence of the operator encryption on the elements of the space of semantic images, it is proposed to replace true semantic image of the outgoing message (that is, one of the elements of this space) with the element of the semantic thesaurus of this application area, the semantic image of which is in relation to semantic plausibility with the true semantic means of the outgoing message. Under these conditions, the work of crypto-analytics is meaningless, since they will be deprived of the opportunity to distinguish the true meaning of encrypted messages from many other false, but plausible semantic images.

The synthesis of the crypt semantic method of protecting the meaning of text messages, which is capable of providing a mode of perfect strong, is performed. According to this method, the space of a given application area of the use of the i-th information system is reflected on the space of semantic images of this system due to the semantic relations between them, determined by the given space of restrictive conditions. That is, the structure of the semantic thesaurus of a given application area is synthesized. Every semantic image from the space of semantic images is brought into conformity with the set of other semantic images from the same space that are in this semantic way in relation to semantic plausibility in thesaurus' structure. This structure of the thesaurus enables replacement of messages during encryption with other plausible messages that do not reflect the true meaning of outgoing messages. And while decrypting, if the password is known, making back replacements of plausible false messages in the true sense can be performed.

The crypt-semantic method involves the creation of a security system in two stages: first, the creation of a semantic dictionary (thesaurus) of the applied area, and then the development of software and hardware implementation of the encryption / decryption operators, which uses the created thesaurus.

It has been shown that in some cases, the length of the cipher key for crypt-semantic encryption should be chosen depending on the permissible value of the probability of making non-error solutions in the process of decoding the semantic flow. An appropriate expression is provided for choosing the minimum possible length of the cipher key.

An algorithm for the implementation of the crypt-semantic security method the meaning of text messages is synthesized.

References

1. Sushko S.O., Kuznetsov G.V., Fomichova L. YA., Korablyev A.V.: Encyclopedia of cryptanalysis. Dnipropetrovsk: National Munity University, (2010) (in Ukrainian)
2. Shannon C.E. A Mathematical Theory of Communication //Bell System Technical Journal, 1948. - Vol. 27, no. 4. - Pp. 379-423, 623-656.
3. Shannon C.E. Predication and Entropy in Printed English // Bell System Technical Journal, 1951. Vol. 30, no. 1. Pp. 50-64.

4. Van Tilborg H.C.A. X.K.A. Encyclopedia of cryptography and security. – New York: Springer, 2005. – 684 p.
5. Synthesis of quite proof cryptosystem with increased unicity distance for cloud computing / Klimchuk, V., Samoylik, E., Gnatyuk, V., Prysiazhnyy, D., Buryachok, V. – CEUR Workshop Proceedings, 2018, pp. 596-607
6. Cabinet-type linguistics / Shyrokov V.A., Bugakov O.V., T.O. Gryaznukhina T.O. – K.: Dovira, 2005. – 471 p.
7. S. Gnatyuk, A. Okhrimenko, M. Kovtun, T. Gancarczyk, V. Karpinskyi, Method of Algorithm Building for Modular Reducing by Irreducible Polynomial, Proceedings of the 16th International Conference on Control, Automation and Systems, Oct. 16-19, Gyeongju, Korea, 2016, pp. 1476-1479.
8. S. Gnatyuk, T. Zhmurko, P. Falat, Efficiency Increasing Method for Quantum Secure Direct Communication Protocols, Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2015), Warsaw, Poland, September 24-26, Vol. 1, 2015, pp. 468-472.
9. Hu Z., Gnatyuk S., Kovtun M., Seilova N. Method of searching birationally equivalent Edwards curves over binary fields, Advances in Intelligent Systems and Computing, Vol. 754, pp. 309-319, 2019.
10. S. Gnatyuk, V. Kinzeryavyy, M. Iavich, D. Prysiazhnyi, Kh. Yubuzova, High-Performance Reliable Block Encryption Algorithms Secured against Linear and Differential Cryptanalytic Attacks, CEUR Workshop Proceedings, Vol. 2104, pp. 657-668, 2018.
11. Tynymbayev S., Gnatyuk S.A., Aitkhozhayeva Y.Z., Berdibayev R.S., Namazbayev T.A. Modular reduction based on the divider by blocking negative remainders, News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences, №2 (434), pp. 238-248, 2019.
12. Kalimoldayev M., Tynymbayev S., Gnatyuk S., Ibraimov M., Magzom M. The device for multiplying polynomials modulo an irreducible polynomial, News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences, №2 (434), pp. 199-205, 2019.
13. Iavich M., Gagnidze A., Iashvili G., Gnatyuk S., Vialkova V. Lattice based Merkle, CEUR Workshop Proceedings, Vol. 2470, pp. 13-16, 2019.
14. Fedushko S., Benova E. Semantic analysis for information and communication threats detection of online service users. Procedia Computer Science, Volume 160, 2019, Pages 254-259. https://doi.org/10.1016/j.procs.2019.09.465
15. Zh. Hu, S. Gnatyuk, T. Okhrimenko, V. Kinzeryavyy, M. Iavich, Kh. Yubuzova, High-Speed Privacy Amplification Method for Deterministic Quantum Cryptography Protocols Using Pairs of Entangled Qutrits, CEUR Workshop Proceedings, Vol. 2393, pp. 810-821, 2019.
16. Korobiichuk I., Syerov Y., Fedushko S. (2020) The Method of Semantic Structuring of Virtual Community ContentMechatronics 2019: Recent Advances Towards Industry 4.0. MECHATRONICS 2019. Advances in Intelligent Systems and Computing, vol 1044. Springer, Cham. pp 11-18. https://doi.org/10.1007/978-3-030-29993-4_2
17. Gnatyuk S., Kinzeryavyy V., Kyrychenko K., Yubuzova Kh., Aleksander M., Odarchenko R. Secure Hash Function Constructing for Future Communication Systems and Networks, Advances in Intelligent Systems and Computing, Vol. 902, pp. 561-569, 2020.