ANALYSIS OF ASPECTS OF MESSAGES HIDING IN TEXT ENVIRONMENTS

ANALIZA ZAGADNIEŃ UKRYCIA WIADOMOŚCI W ŚRODOWISKU TEKSTOWYM

Olesya Afanasyeva

Pedagogical University in Krakow, Technics Institute

e-mail: olesya@afanasyev.kiev.ua

Abstract: In the work are researched problems, which arise during hiding of messages in text environments, being transmitted by electronic communication channels and the Internet. The analysis of selection of places in text environment (TE), which can be replaced by word from the message is performed. Selection and replacement of words in the text environment is implemented basing on semantic analysis of text fragment, consisting of the inserted word, and its environment in TE. For implementation of such analysis is used concept of semantic parameters of words coordination and semantic value of separate word. Are used well-known methods of determination of values of these parameters. This allows moving from quality level to quantitative level analysis of text fragments semantics during their modification by word substitution. Invisibility of embedded messages is ensured by providing preset values of the semantic cooperation parameter deviations.

Keywords: steganography, steganosystem, message concealment, text environment, message interpretation, semantics, semantic vocabularies.

Streszczenie: Niniejsza praca dotyczy badań nad zagadnieniami, które powstają podczas ukrywania wiadomości w środowisku tekstowym, przesyłanym przez elektroniczne kanały komunikacyjne oraz internet. Przeprowadzona jest analiza wyboru lokalizacji w środowisku tekstowym (TE), w którym oryginalne słowo jest zastępowane słowem ukrywanej wiadomości. Wybór i zastąpienie słów w środowisku tekstowym jest dokonywane w oparciu o analizę semantyczną fragmentu tekstu składającego się ze wstawionego słowa oraz środowiska tekstowego (TE). Dla przeprowadzenia takiej analizy wykorzystano koncepcję parametrów semantycznych grupy słów (wyrażenia) oraz wartości semantycznej pojedynczego słowa. Zastosowano dobrze znane metody określania wartości tych parametrów. To pozwala na przejście od analizy semantyki fragmentów tekstu na poziomie jakościowym do analizy na poziomie ilościowym w przypadku modyfikacji tych fragmentów poprzez zamianę pojedynczych słów. Niewidzialność (niemożliwość wykrycia) wbudowanych komunikatów jest zapewniona przez wstępne określenie wartości odchylenia dla parametru semantycznej zgodności analizowanego wyrażenia.

Słowa kluczowe: steganografia, steganosystem, ukrycia wiadomości, tekstowe środowisko, interpretacja wiadomości, semantyka, słowniki semantyczne.

1. Introduction

Steganographic methods of message concealment in different types of digital environments are quite widespread and vital. This is based on the main characteristic of steganography, which is concealment of the message transfer itself when using steganographic methods. This goal can be reached if, during a message insertion into a certain environment, changes caused by insertion of the corresponding data in the environment cannot be noticed. Using digital networks and other tools for digital communication favors the development of steganographic methods of message concealment. A quite widespread type of digital environments, used in information transfer networks, is a text environment (TE). Specificity of TE is that alphabet symbols and words are used in this environment, and their usage in steganography is not fully enough investigated, disregarding control characters.

Known methods of steganographic message concealment in TE were developed by Wayner [1]. These methods are based on using the distribution probability of different alphabet symbols in a certain text. Symbols or phrases of a concealed message are chosen on the basis of these probabilities. To implement this method, a class of imitating functions is introduced. This class allows to use systems of formal grammatics, which allows to perform message concealment in the output structures and so on.

One of the tasks arising in the research scope of using TE for steganography is the task of choosing the parameters which characterize a degree of invisibility of a message inserted in the environment, a degree of message recognizability when an anomaly is detected in the TE and so on [2].

This paper contains analysis of a method of inserting separate message words into the TE which is passed on using semantic parameters. A definition of additional parameters is formulated, among which is a concept of a technical parameter of message invisibility, a concept of TE interpretation story is introduced, of load of certain story fragments and several other concepts necessary for deeper study of message concealment problems in TE.

This paper also contains research of a task of modifying TE if its size is not enough for a message insertion. Such expansion is proposed to implement on the basis of data of text descriptions of a message interpretation and the TE itself.

2. Insertion of message words into the text environment

An invisibility degree of a message inserted into *TE* is a quite important parameter. This degree is especially important for *TE*, because the message itself is also a text. In the paper [2,3] a set of parameters is given which as a whole characterizes the invisibility degree. The mentioned parameters are closely related to elements of a system of text information perception by the user. Besides parameters given in [4,5], let us introduce a technical parameter of invisibility.

Definition 1. A technical parameter of invisibility π is going to be defined as ratio of the word number in the message text V_i to the word number in TE , where a message is being planned to be inserted.

$$\pi = [\sum_{j=1}^m V_i(s_j)] / [\sum_{j=1}^n TE_i(s_j)]; n > m.$$

This parameter is changed in $[0,1]$ range. A value $\pi = 1$ corresponds to a situation when $V_i = TE$, a situation when the message size is equal to the size of TE . We will not take this case into account. A value $\pi = 0$ corresponds to a situation when there is no message in TE , but we will not consider this case either. The π parameter is called a technical invisibility parameter because it defines the choice of the TE size relative to V_i . To estimate the π parameter beforehand and, respectively, estimate the necessary size of TE , we can assume that $\pi = \alpha \cdot 10^{-3}$, where α takes values from 1 to 9. This parameter characterizes a certain degree of invisibility because it can be stated that, the more is a size of a text where a certain message V_i of a fixed size is placed, dispersed along all the text, the more difficult it is to detect in this text.

In order to insert V_i into any TE , the latter has to have a certain semantic interpretation which will be described by the following relation:

$$J(TS) = J(\psi_1, \dots, \psi_n) = J\{\psi_1(j(\varphi_{11}), \dots, j(\varphi_{1g})), \dots, \psi_n(j(\varphi_{n1}), \dots, j(\varphi_{nq}))\},$$

where J and j indicate a semantic interpretation of a sentence ψ_j from TE , φ_{ij} is a phrase of a sentence i . Obviously, semantic interpretations V_i and TE have to agree within a certain degree, or $J(V_i) \propto J(TE)$. Otherwise it will be easy to detect words in TE which are related to V_i . This brings up a task of determining the connection between a degree of V_i visibility in the TE and a degree of agreement between $J(V_i)$ and $J(TE)$. A solution of this task is based on using concepts of semantic vocabularies S_C [6,7]. To solve it, a vocabulary $S_C(TE)$, limited by elements with a text TE , is generated on the basis of TE . It is clear that $S_C(TE)$ can be extended so that $S_C(TE) \subset S_C^*$. This means there is an interpretation area $W_i^*(TE)$, that can include an interpretation area limited by the given TE . If we choose one of the possible ways to determine semantic meaningfulness σ^Z for elements from $S_C(TE)$, it becomes possible to determine a measure of semantic consistency σ^u between the elements from TE and the elements from $S_C(TE)$ and $S_C^*(TE)$. Instead of σ^u parameter, an opposite parameter of semantic contradiction σ^S can be used. If a $S_C(TE)$ is built on the basis of $W(TE)$, there is a possibility to determine acceptable values of σ^S parameter for a given TE . Because $S_C(V_i)$ is also defined for V_i , we can establish acceptable range of $\sigma^S(V_i)$ values for V_i as well. If $S_C(V_i) \subset S_C(TE)$, then V_i can be built in the TE on the basis of using words from TE .

In this case the following tasks arise:

- a task of replacing the word numbers from the TE , which were replaced by words from V_i , by a algorithm of recognizing words from V_i in the TE environment,
- a task of choosing places in the TE to place words from V_i .

A word sequence from V_i in the TE can be natural, or match with the word sequence in V_i , or can be defined by a certain algorithm.

3. Extending the text environments

Because for every $S_C(TE)$ there is an extension $S_C^*(TE)$, words can be used for V_i from $S_C(W_i^*) = S_C^*(TE) - S_C(TE)$. In this case, when building a message V_i in the TE , a replacement of words from the TE with words from $S_C(W_i^*)$ is performed. Obviously, such replacement should not cause visible anomalies in the TE .

Based on the size of TE , which are necessary to use to insert V_i into them, it becomes obvious that a process of detecting V_i in the TE has to be performed with certain tools of TE analysis. This analysis is implemented on the basis of detecting such words x_i in the TE , for which σ^S with their surroundings has a certain meaning. Let us review the following statement.

Statement 1. If there is a certain function $F[x_i, f(\sigma_1^S, \dots, \sigma_k^S)]$, that describes relations between $\sigma^S(x_i, x_j)$ in the range of $\psi_i(\varphi_j, x_k)$, where φ_j are phrases of a sentence ψ_i , x_k is a built-in word from V_i , then a value of $F(x_i, \varphi_i)$ exists, which is acceptable in the TE and has the corresponding interpretation of σ^S parameter.

According to the statement condition let us assume that a $\psi_i(\varphi_1, \dots, \varphi_k)$ exists in TE for which the formula $F[x_i, f(\sigma_1^S, \dots, \sigma_k^S)]$ has an interpretation corresponding to the acceptable value of a σ^S parameter, or $\sigma^S \leq \delta\sigma^S$, where $\delta\sigma^S$ is a boundary value of semantic contradiction. If $x_i \in V_i$ and belongs to $S_C(TE)$, then $x_i \in \varphi_i \in \psi_i$ does not cause a semantic anomaly because TE is generated on the basis of grammatics $\Gamma(\gamma_1, \dots, \gamma_k)$, used to generate TE , and also thanks to the fact that $W(TE)$ and $S_C(TE)$ are used to generate TE . If $[(x_i \in \varphi_i) \& (x_i \notin S_C(TE))]$, then $(x_i \in S_C(W_i^*))$. This means that $(x_i \in S_C(W_i^*))$ is semantically acceptable. If a $\sigma^S[(x_i \& x_{i+1}) \vee (x_i \& x_{i-1})] \leq \delta\sigma^S$ takes place, the statement condition is true. Let us assume that a $[(x_i, x_k) \in \varphi_i] \& [\sigma^S(x_i, x_k) > \delta\sigma^S]$ takes place, then in $S_C(W_i^*)$ we choose such $x_g \propto x_i$ that $\sigma^S(x_i \& x_g) \leq \delta\sigma^S$. If the given replacement $x_i \rightarrow x_g$ cannot be performed so that $\sigma^S(x_i, x_g) \leq \delta\sigma^S$, consecutive replacements of $x_j \in \psi_i$ with $x_r \in S_C(W_i^*)$ in ψ_i are performed so that the condition $F[x_g, f(\sigma_1^S, \dots, \sigma_k^S)]$ was true for all couples $x_h \& x_k$. The mentioned condition is a generalization of a local condition $\sigma^S(x_i, x_g) \leq \delta\sigma^S$ for the whole ψ_i . Because $(x_g, \dots, x_h) \subset S_C(W_i^*)$, where W_i^* is a semantically acceptable extension of W_i , it is possible to generate $\psi_i^* \propto \psi_i(TE)$, and the corresponding $x_i \in V_i$ is semantically acceptable.

4. Choosing the placement of a message word in the text environment

Besides choosing a word from V_i , to replace it with a word in TE , or to choose a word from TE as a word from V_i , a task arises of choosing places in TE for inserting words from V_i [8,9]. The words x_i from TE , that, according to the statement 1, can be used to replace the first word from V_i , can exist in many places of TE . Thus, we need to define the signs which could indicate from all the possible areas of TE the one where performing the changes would meet the necessary level of semantic contradiction as much as possible. To solve this task, let us introduce the following conditions and definitions.

Condition 1. Each TE must have its own story.

Definition 2. A story H exists in TE in the case when each word x_i of each phrase φ_i and each sentence ψ_i have interpretational descriptions in the subject area of W_i , represented in the corresponding semantic vocabularies S_C .

Definition 3. A story is a function of semantic meaningfulness $H(TE) = \mathfrak{I}(x_i, \varphi_i, \psi_i, \eta)$, which characterize the components mentioned in the definition 2, ordered in most cases by natural numeration η in TE .

Because TE inherently has serial structure, the story can be considered a linear function corresponding to the numeration of elements.

Condition 2. To place a single message V_i , it is reasonable, and in certain cases necessary, to use a uniform TE , where uniformity is determined by presence of a common story.

A story function $H(TE) = \mathfrak{I}(x_i, \varphi_i, \psi_i, \eta)$ in TE structure will be called a story line (Hl). For convenience, a story line Hl , corresponding to TE structure, goes from the beginning to the end of the text. Function Hl will possess different values in definition interval, depending on the numeration value $Hl(\eta)$. We will call these values a story load $Hl(\eta)$ in the selected interval. Let us assume that one step of $Hl(\eta)$ will be defined within the interval of a single phrase. The suitability of this limitation is based on using semantic parameters regarding the text elements.

Definition 4. A semantic load $\sigma^n(fr_i)$ of a fragment Hl is defined by a semantic meaningfulness of this fragment, $\sigma^Z(fr_i)$.

A semantic meaningfulness $\sigma^Z(fr_i)$ is defined by the number of cases when fr_i is used in the selected TE , which can be described as follows:

$$\sigma^Z(fr_i) = \sum_{i=1}^n sgr(fr_i),$$

where sgr is an identification function of a fragment fr_i , described by the following relation:

$$[(fr_i @ fr_j) \rightarrow (sgr(fr_i) = 1)] \& [(fr_i \neq fr_j) \rightarrow (sgr(fr_i) = 0)],$$

where fr_j is a current fragment from TE , fr_i is a fragment for which $\sigma^Z(fr_i)$ is being defined. Operation $fr_i @ fr_j$ means a definition of identity degree of meaningfulness between fr_i and fr_j . This operation is quite important, so we will examine it in more detail. It means comparing the key words or phrases, depending on the size of fr_i , from the fragments fr_i and fr_j and defining their semantic correlations. In most cases the TE structure is serial, so we will consider the structure of fr_i and fr_j serial as well. Let us introduce a rule of defining structural difference between fr_i and fr_j . For a more convenient description let us assume that $fr_i = \varphi_i$, $fr_j = \varphi_j$. In this case we will examine structural difference between φ_i and φ_j if $(x_{ik} \in \varphi_i) = (x_{jm} \in \varphi_j)$.

Rule 1. If relations between $x_{ik} * x_{ik+1}$ and $x_{jm} * x_{jm+1}$ are interpreted as a conjunction or a disjunction, there is no structural difference in φ_i and φ_j when $x_{ik} = x_{jm}$. If this relation is interpreted as an implication in one of the phrases and as a conjunction or a disjunction in the other one, there is a structural difference between φ_i and φ_j .

This leads us to a definition of σ^Z for a certain uniform TE for all $fr_i \in TE$. In this case, to determine a semantic load σ^N , we will limit ourselves with a step of change of value $\sigma^N(H)$, which is defined by a single phrase. Let us write the function $H(TE) = \mathfrak{I}(x_i, \varphi_i, \psi_i, \eta)$ as $H(TE) = \mathfrak{I}(\varphi_i, \eta)$. Then the function $\sigma^N(H)$ can be represented as:

$$\sigma^N(H) = \mathfrak{I}(\sigma^N(\varphi_i), \eta).$$

Because $\sigma^N(\varphi_i)$ is defined by a certain number depending on a numeration index of φ_i in TE , we can write that $\sigma^N(H) = \mathfrak{I}(\sigma^N(\varphi_i))$. Let us assume, based on known data, that to insert $x_i \in V_i$ into $\varphi_i \in TE$, such φ_i are chosen, for which $\sigma^N(\varphi_i) \approx \min[\sigma^N(H)]$.

As stated before, S_C is a description of text interpretations of words x_i and phrases φ_i used in TE . Based on using S_C , let us introduce the concept of key words or phrases which will be defined by their semantic parameters $\sigma^{KL}(x_i)$, or $\sigma^{KL}(\varphi_i)$.

Definition 5. Key words x_i or phrases φ_i will be such x_i or φ_i , for which the sizes of description of text interpretations are greater than interpretation descriptions of other x_j or φ_j , and the difference value is formally described as:

$$\{\forall i \exists j (i, j \in \eta) [sg(j(\varphi_i)) - j(sg(\varphi_j))] > \beta\} \rightarrow \sigma^{KL}(\varphi_i).$$

In a certain TS a situation can occur when key elements are distinguished within it in the form of x_i or φ_j . Obviously, a sentence ψ_i , used by $\sigma^{KL}(\varphi_i)$, should not be used for placement of $x_i \in V_i$ regardless of $\sigma^Z(\varphi_i)$ value, if φ_i is a key phrase.

A subject area of interpretation $W_i(TE)$, formed on the basis of TE , can be extended by semantically consisted fragments. Such extension can be considered reasonable if it can be performed considering the subject area V_i , which can favor

the increase of invisibility level of V_i in the TE . Extension of TE is performed only in those cases when the selected TE is not widely known, for example, a text fragment of a well-known published work. This is one of the signs for choosing a certain type of texts, that are the most convenient for placing V_i in their environment.

An extension of TE is initiated in the case when during the TE analysis, performed by a steganosystem (SS), it becomes clear that a size of TE is not sufficient for a placement of V_i with a given invisibility degree μ . The extension process is performed within the scope of SS , which uses the rules of sentence generation of text extension ΔTE . These rules are formed on the basis of grammatical rules of the language which TE is created in [10,11]. Such rules have to meet the following conditions.

Condition 3. A sentence generated for ΔTE must be semantically redundant.

Condition 4. A sentence ψ_i for ΔTE is formed on the basis of an extended semantic vocabulary $S_C^* = S_C \cup \Delta S_C$, which describes the corresponding extension of a subject area of ΔW_i interpretation.

Condition 5. Extension of S_C^* and the corresponding ΔW_i is generated on the basis of using components from V_i , which could be x_i or φ_j from V_i .

Condition 6. Extension ΔTE is performed on the basis of analysis of V_i interpretation description, or $J(V_i)$, and is performed dispersedly along all TE .

Extension of TE is performed on the basis of analysis of an interpretation $J(V_i)$, which is represented as $S_C(V_i)$. Because V_i is formed so that semantic load V_i would be maximal, which is based on the requirement of V_i size to be minimal in order to reach the higher level of μ . This circumstance leads to complications while choosing the insertion place, because in this case a single word from V_i during its insertion requires an environment where semantic meaningfulness of its components is also high. This is necessary to reach the given consistency level of $x_i \in V_i$ with $x_j \in TE$. Because of this, while inserting V_i into TE , it is not always reasonable to try to generate V_i of the shortest possible size. It is obvious that an optimal V_i is a message which is the most similar to TE by its semantic parameters. This means that it is reasonable to adapt V_i to a TE which would be semantically uniform with semantics of V_i . In this case a task arises of optimization of semantic load value V_i relatively to the semantic load of TE and the given value of V_i invisibility in the TE environment. Let us examine the following definition.

Definition 6. A semantic load λ of a fragment fr_i of a text τ_i of an environment TE is defined by a ratio of average meaningfulness of the components of the fr_i fragment to the average meaningfulness of the components of the whole TE or its part that participates in the definition of a semantic load, which can be described by the following relation:

$$\lambda(fr_i) = \frac{\sum_{i=1}^b \sigma_i^Z(x_i \in fr_i)}{\sum_{j=1}^m \sigma_j^Z[(x_j \notin fr_j) \& (x_j \in TE)]}; m < n .$$

It can be seen from this relation that $\lambda(fr_i) = 1$ in case when $fr_i = TE$ and in all other cases it takes values within the range $(0,1)$ except «0» value, because $\lambda(fr_i)=0$ when $fr_i = 0$. In most cases V_i is formed in such a way that $\lambda(V_i)$ is quite high while n value is relatively not so high because $\lambda(V_i)$ is different from $\lambda(fr_i \in TE)$. Because a separate text is examined, we will call $\lambda(V_i)$ a personal semantic load and will write it as $\lambda^P(V_i)$.

Definition 7. A personal semantic load $\lambda^P(V_i)$ is defined by an average value of semantic meaningfulness of all V_i components, which is formally described by the following relation:

$$\lambda^P(V_i)=[\sum_{i=1}^m \sigma_i^Z \in V_i)]/m.$$

A message is mainly generated in such a way that V_i has the maximal $\lambda^P(V_i)$. To represent this factor with the most precision, defining the value $\sigma_i^Z(x_i \in V_i)$ by the size of $j(x_i)$ in S_C is not sufficient. Thus, to generate $j(x_i)$ from V_i for all or a single component a_{ij} from $x_i = \langle a_{i1}, \dots, a_{ik} \rangle$, a parameter of personal meaningfulness $a_{ij}(\zeta)$ is introduced, which is assigned by the corresponding a_{ij} when calculating $\sigma^Z(x_i)$.

Methods of TE extension while inserting V_i provide for solving the following tasks:

- 1) Generation of text fragments fr_i which are meant to extend TE in the corresponding place of TE .
- 2) Choosing a place in the TE environment for its extension.
- 3) Defining the size ΔTE which is a set of single fragments fr_i , or $\Delta TE = \sum_{i=1}^m fr_i$.

To generate fr_i fragments, a generator of text fragments (GTF) is used. This generator operates on the basis of using $S_C(V_i)$ and $S_C(TE)$ and using the rules of output of text fragments fr_i . To perform the corresponding output, we need criteria, based on which particular steps of an output can be performed and a situation which corresponds to the completion of an output is determined.

An output system consists of the following elements:

- rules of output or transformation of text fragments, which we will write as $\Gamma\{\gamma_1, \dots, \gamma_m\}$ and which are related to the grammatical rules of the language used to build TE ,
- criteria of choosing regular text fragments from $J(V_i)$, $J(TE)$ and from intermediate fr_i generated during the output of a fragment fr_i , which will be written as $K(tm_i, tm_j)$,
- criteria of choosing a regular transformation rule which is supposed to be used during the current output step, a system of these criteria will be written as $K(\gamma_i)$.

Formally an output system as a whole can be described by the following relation:

$$\mathcal{L} = \{J(V_i), J(TE), K(tm_i, tm_j), \Gamma(\gamma_1, \dots, \gamma_m)\}.$$

After using $\mathcal{L}$ we obtain a fragment tm_i which we need to place in the TE environment. A process of $fr_i(tm_i)$ output can be described by this relation: $fr_i = \mathfrak{H}[J(V_i), J(TE)]$, because $S_C(V_i)$ and $S_C(TE)$ are the basic input data for performing transformations. An obtained fragment fr_i can be a sentence ψ_i which is a semantically complete form. This means that, to choose a place in the TE for $fr_i = \psi_i$, basic semantic parameters can be used that characterize ψ_i in general. One of these parameters is semantic contradiction σ^S . Obviously, to determine them, we need to use semantic meaningfulness $\sigma^Z(tm_i)$ and semantic meaningfulness of ψ_{i-1} and ψ_{i+1} sentences which create surroundings for $\psi_i = tm_i$. To determine a placement of ψ_i in the TE we have to transfer the numeration of all sentences used in the TE . Let us assume that this numeration is $\eta(\psi_i)$. Because extension of ψ_i should not cause semantic anomalies in the TE , such insertion place is chosen in the TE where ψ_i is not going to change the average value of semantic consistency $\sigma^U(TE)$. To successfully solve this task, we need to find out the initial conditions, specified as $J(V_i)$, $J(TE)$, that make it possible to determine fr_i , for which there are two compatible sentences in the TE , between which a sentence $\psi_i = fr_i$ can be placed without causing a semantic anomaly, or $[\sigma^S(\psi_{i-1}, \psi_i) > \vartheta\sigma^S] \vee [\sigma^S(\psi_i, \psi_{i+1}) > \vartheta\sigma^S] = \alpha(\sigma^S)$, where $\alpha(\sigma^S)$ is a semantic anomaly, which is occurrence of an unacceptable value of semantic contradiction.

The next task which requires our attention is defining the size of extension of a text environment ΔTS . The size ΔTS will be measured in a number of ψ_i sentences obtained as a result of their output from $J(V_i)$, $J(TE)$. We can limit ourselves with the size of ΔTE , that does not exceed $V_i(\psi_i)$, measured in a number of sentences which V_i consists of. This does not mean ΔTE can be greater than $V_i(\psi_i)$. For convenience, sizes of descriptions will be written as $R(V_i)$, $R(\Delta TE)$, $R(TE)$, respectively. This does not mean $R(\Delta TE) \leq R(V_i)$. Obviously, $R(\Delta TE)$ can be greater than $R(V_i)$. This could be required when ensuring a given level of invisibility of a message V_i inserted into TE . Thus, the value of necessary extension is based on a requirement to ensure a given safety level of a message V_i . In this case, safety of a message transfer is defined by a set of parameters listed in [1,12].

5. Summary

This paper contains research of the tasks which arise during message concealment in text environments, transferred via electronic communication channels and the Internet network. An analysis of place choosing in TE , where a text word can be replaced with a message word, has been performed. Word choosing and replacing in TE is performed on the basis of a semantic analysis of a text fragment consisting of a word being inserted and the surrounding words. To perform this analysis,

a concept of semantic parameters of consistency between words and semantic meaningfulness of a single word is used. Known methods of determining the values of these parameters are used in the paper, which allows to go from qualitative level to quantitative level of semantic analysis of text fragments when modifying them by replacing words. Invisibility of such message insertion is ensured by establishing the given values of deviation of semantic consistency parameter.

For a case when semantics of a *TE* text is substantially different from the message semantics, it is proposed to perform a *TE* extension in such a way that the extended text would be more suitable for concealment of an inserted message.

An analysis of tasks is performed, that occur during the implementation of the corresponding extension. Input data for solving the extension tasks are interpolate message descriptions and interpretative descriptions of *TE*. The process of such extension is based on using a system of text fragment output. This output system uses grammatical rules of sentence building in the language used for describing *TE* and the message. These rules are extended by the conditions which take into account the values of the corresponding semantic parameters and the logical structure of a sentence, which is a text fragment supposed to be output.

Problems of message concealment in *TE* are quite vital because there are widespread services of text transferring in the Internet network. The approaches, developed in this paper, to creating methods of message concealment in *TE* are new solutions of the problems of text steganography.

6. References

- [1] Afanasyeva O.Y., Korostil Y.M. Hiding data in text environments and their parameters. Scientific Journals Maritime University of Szczecin, 38(110), 2014 pp. 10-14.
- [2] D. Salomon. Compression of data, image and sound. – Moscow: Technosfera, 2004. – C. 368. – ISBN 5-94836-027-X.
- [3] D. Vatolin, A. Ratushnyak, M. Smirnov, V. Yukin. Methods of data compression. Structure of archivers, image and video compression. – Dialog-MIFI, 2002. – C. 384. – ISBN 5-86404-170-X.
- [4] Arnold M. Schmucher M., Wolthusen S. Techniques and Applications of Digital Watermarking and Content Protection. London: Artech, 2003. 274 p.
- [5] Cox I.J., Miller M.L., Bloom J.A. Digital Watermarking. London: Morgan Kaufmann Publishers. 2002. 534 p.
- [6] Afanaseva O.Y. Methods of semantic transformations in steganosystems. Modeling and information technologies. A collection of studies IPME NASU, Kyiv, B 56, 2010, pp. 188-196.
- [7] D.C. Wu, W.H. Tsai, A steganographic method for images by pixel-value differencing, Pattern Recognition Letters 24 (9–10) (2003) 1613–1626.

- [8] Afanaseva O.Y., Durnyak B.V. Basic characteristics of digital steganographic systems. Modeling and information technologies. A collection of studies IPME NASU, Kyiv, B 51, 2009, pp. 17-25.
- [9] Durnyak B.V., Korostil O.Y. Designing of advertisement in the Internet network on the basis of its semantic analysis. UPA, Lviv, 2014. 136 p.
- [10] Afanaseva O.Y. Research of semantic features of displaying graphic image. Modeling and information technologies. A collection of studies IPME NASU, Kyiv, B 40, 2007, pp. 81-92.
- [11] K. Solanki, A. Sarkar, B.S. Manjunath, YASS: yet another stegano-graphic scheme that resists blind steganalysis, in: Proceedings of the Ninth International Workshop on Information Hiding, Saint Malo, France, 11–13 June 2007, Lecture Notes in Computer Science, vol. 4567, pp. 16–31.
- [12] J. Cheng, A.C. Kot, Steganalysis of halftone image using inverse halftoning, Signal Processing 89 (6) (2009) 1000–1010.

Dr. hab. technical sciences Olesia Afanasieva, prof. Pedagogical University in Krakow. Specializes in the following scientific research fields: steganography hiding of messages in digital environments of various types, safety of complex technical objects, information systems security methods, experimental data analysis with cognitive graphics means.