UDK 004.056(043.2)

STEGANALYSIS OF GRAPHIC CONTAINER

A. S. Shmatok, PhD, ***A. B. Petrenko*** PhD, ***A. B. Yelizarov,*** PhD,
V. A. Tytov, E. A. Borysenko

National Aviation University

Sh_al_st@mail.ru

In this paper, the authors justified and tasked by to development algorithm of the software will perform a passive attack on the container, as amended by LSB him hidden information. The aim of the task was to automate the process of passive steganalysis graphics container with the help of statistical pattern recognition theory.

Keywords: information security, steganography, steganalysis, statistical pattern recognition theory, testing the assumption of normality, algorithm.

Обґрунтовано та поставлено завдання щодо розроблення алгоритму програмного продукту, який здійснюватиме пасивну атаку на контейнер, із внесеною в нього методом НЗБ прихованої інформації. Метою було автоматизувати процес пасивного стегананалізу графічного контейнеру за допомогою статистичної теорії розпізнавання образів.

Ключові слова: захист інформації, стеганографія, стеганоаналіз, статистична теорія розпізнавання образів, перевірка нормальності вибірки, алгоритм.

Foreword

Even with optimal conditions for making attack the task of extracting the hidden message from the container may be very difficult. We can definitely confirm about the existence of hidden information only after its separation in an explicit form. Sometimes the goal of stegoanalysis is not the steganography recovery algorithm in general, but the search, for example, of certain steganography key, which is used for selection of container bits in steganography transform.

The main goal of stegoanalysis is the simulation of steganography systems and its research to give qualitative and quantitative assessment of reliability of using of steganography conversion and also construction of methods of detecting of hidden information in a container, its minor revision or destruction.

One of the main tasks of stegoanalysis is the investigation of possible traces of using of steganography facilities and development of methods that allow eliciting facts of their using. Usage exact steganography transformation requires from steganography analyzer of individual attention to his research. Research of massages, which are hidden by out of many existing steganography methods, is quite time-consuming process [1].

Analysis of research and publishing

Analyzed sources emphasize the principal directions of steganography analyzer activity and help to set a goal of algorithm elaboration and to classify developed algorithm [1; 2].

Implementation of the algorithm itself due to statistical pattern recognition theory is described in the works [3; 4].

Problem statement

Steganography system considered to be broken if the trespasser, at least, succeeded to prove the existence of hidden messages in intercepted container. It is expected that the trespasser is able to perform all types of attacks and has unlimited computational capabilities. If he cannot confirm the hypothesis that secret message is hidden in the container, the steganography system is deemed to be stable.

Generally, there are several stages of steganography system hacking:

- Detection of fact of hidden information presence.
- Extraction of hidden message.
- Transformation (Modifying) of hidden information.
- Prohibition on implementing of any information transfer, including the hidden information.

The first two steps refer to passive attacks on steganography system and the others to the active (or malicious) attacks.

Insofar as we are interested not so much in the content of the container as the mere fact of hidden information presence, it was decided the solution about the attack on the ground of known mathematical model of the container or its part [2].

Rationale is that the number of multimedia traffic, which is handled in modern automated systems, is growing literally day by day, the steganography analyzer has a problem of process automation of stegoanalysis images as the most common container for information hiding.

Process automation of stegoanalysis allows more detailed study only suspect objects; it saves resources of automated systems.

Statement of basic material

As we know, when hidden information is injected in image, its statistical performances are suffered. If we take the hidden information in the container to be the signal and bit image space, where we added information by LSB method, to be the information channel, then we can estimate that without output signal we will get noise. Hereof we can estimate that LSB image without added information will make noise.

In the signal acquisition classical theory against the noises, signals and noises are described usually by fully known to characteristics distribution law. However, in practice we can fulfill these conditions seldom, especially in the part of priori knowledge of characteristics. Actually, if the general form of distribution law of signals and noise is possible to settle in many cases from common physical or technical considerations, then, for example, such characteristics as amplitude or intensity of the signal and noise, which are depended on the nature of the modulation, radio conditions, the availability of external sources, prevent the instability of reception equipment elements and many other factors, which are usually a priori unknown. In this case, the direct usage of the results of the identification classical theory is difficult, so it is necessary to summarize it. It is universal in occurrence the case of prior uncertainty, which can be complete based on the statistical pattern recognition theory. As a result, we arrive at the adaptive signal determination against the noise, which rejects the teach procedure, which in concerned parametric case is come down to signal and noise parameter estimation (and in general non-parametric case — to the estimation of its distribution laws), and discovery procedure (namely decision making), wherein instead of signal and noise parameters itself it is used their estimates obtained during the teaching [3].

As an algorithm of discovery of hidden information in graphical container will be used parametric adaptive signal acquisition against noise.

The terms “adaptation”, “adaptive algorithms” are widely interpreted. As an adaptive algorithm we will agree to understand such decision algorithm in the construction of which it is used the pertaining for overcoming the prior uncertainty. The leaning objective is the formation on the basis of the observed realization of examined assessment process (fetch) of unknown distribution functions (under the nonparametric prior uncertainty) or estimation of unknown distribution parameters of the (under the priori parametric uncertainty). Then these estimates are used instead of unknown probability characteristics by the synthesis of decision algorithm.

Thus, adaptive detection algorithms and differentiation of signals on noise background are found substitution ally in sufficient statistic of likelihood ratio resulting from studying of unknown parameters estimation or unknown likelihood function.

It can be supervised learning by classified teaching selection for which it is priori known for what of the hypotheses belongs each of its elements, or it can be no supervised leaning by unclassified observed sample for which statistics are formed, which are used for decision making [4].

Let us assume, that the random array of image S (T) $(N = 20)$ is fed to the input of the adaptive detector (Fig. 1), in a normally distributed manner with unknown average and intensity against the additive normal noise ξ which is irrespective of it with unknown average and intensity.

Fig. 1. Array of images S

Under the statistical pattern recognition theory specified problem of adaptive detection is a special case of recognition of one-dimensional normal collection with unknown and different mean (1) and variance (2) and is in deciding on belonging of experimental sample collection (x1, x2, ..., xn) of the input process to one of two classes s2, which is characterized by the presence of cumulative process, formed by the signal and noise with unknown mean AE = AC + (average variance) and S1 and it is characterized by single noise with unknown mean and variance.

Adaptive detector learning procedure consist of estimation of unknown parameters: mean values of the total variance and variance of noise (Fig. 2), under the character and in terms of which are used

maximum likelihood estimates obtained by classified training set of class S1 and S2 class.

Further, due to the wavelet transform algorithm, we remove noise of image, if the image does not contain hidden information; statistical performance hardly will change, but if the hidden information is contained in a container, then its statistical performance will approach to a clean container.

Further we compare the statistical performance of the array of images that was forthcoming, with the performances of an array of cleared images (Fig. 3, 4).

For example, we compare the mass of images with no present embedded hidden information. (Fig. 5).

$$a = \frac{1}{m+1} \cdot \sum_{i=0}^{m} msumkz_i. \quad (1)$$

$$\sigma = \frac{1}{m} \cdot \sum_{i=0}^{m} \left(msumkz_i - a \right)^2. \quad (2)$$

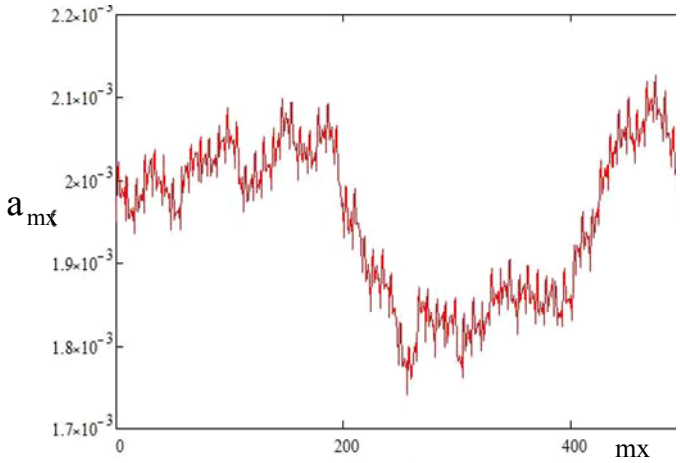
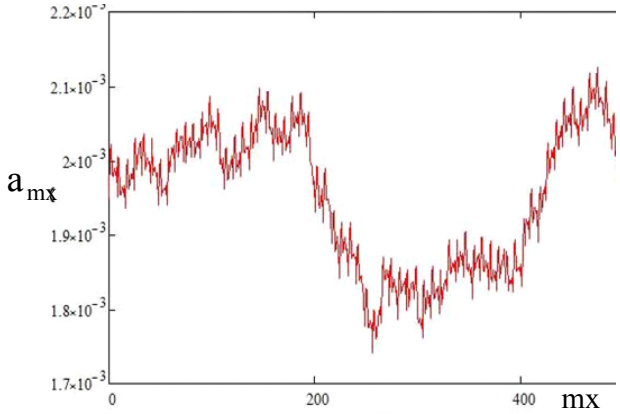
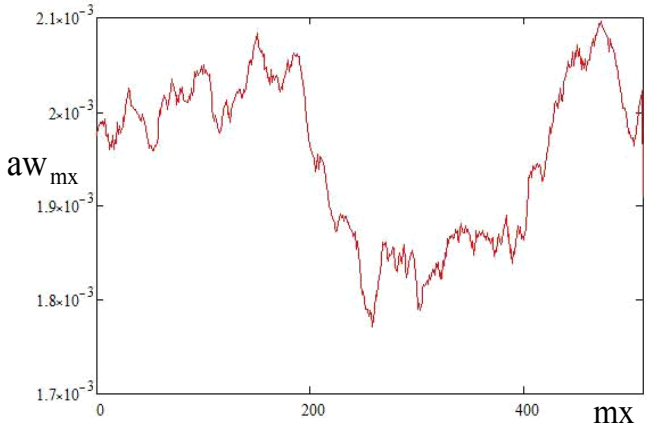


Fig. 2. Determination of the statistical performance of the array of image where mean values (1) and variance are calculated (2)



a



b

Fig. 3. Comparative graphs of mean values of clear (*a*) and cleared array of images (*b*)

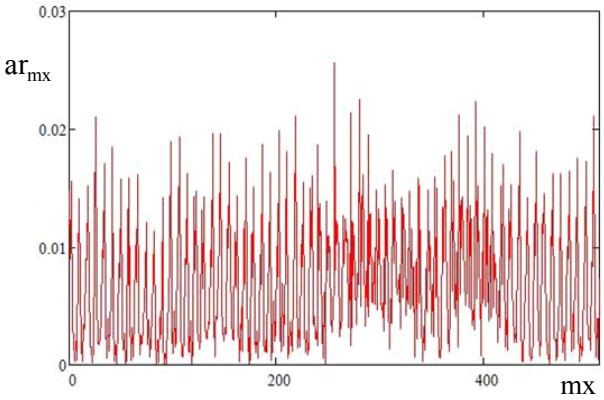


Fig. 4. Compare forthcoming array with the array of cleared images

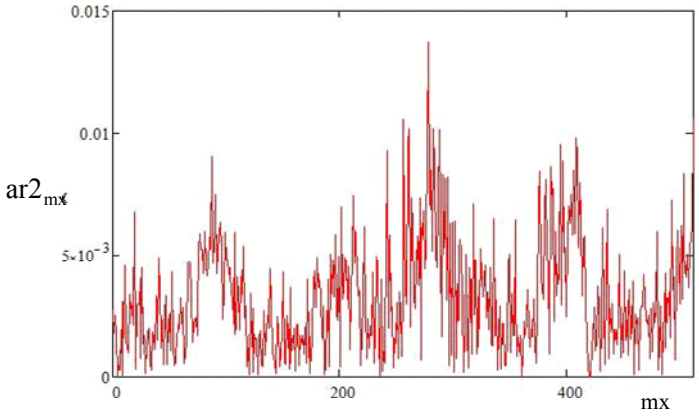


Fig. 5. Compare the array of images without embedded information with the array of cleared images

Due to the data received we define the mean values threshold averages for class determination to which the image belongs: S1 — suspected existence of hidden information, S2 — clear container. Thus the threshold value C = 4. If C > 4, the image is classified as S1, C < 4 and if it refers to the class S2.

Conclusions

The authors discovered a software algorithm, making possible to detect the presence of hidden information in a graphical representation, and to automate actual process. Algorithm of detection of hidden information in a graphical container uses both passive and active stegoanalysis methods that allow detecting and deleting of hidden information in the image. Therefore the usage of such algorithm is reasonable for creation of software tool that will detect the fact of existence of hidden information.

REFERENCES

1. *Ivanov V.* To show all that is hidden: stegoanalysis – steganography detection, principles and methods [electronic resource] / V. Ivanov // Steganography & travelling. — 2012. http://nestego.-blogspot.com/2012/05/blog-post_21.html.

2. *Ivanov V.* Types of attack on steganography system:[electronic resource] / V. Ivanov // Steganography & travelling. — 2012. http:// nestego.-blogspot.-com/2012/05/blog-post_18.html..

3. *Fomin Y. A.* Statistical theory of pattern recognition / Y. A. Fomin, G. P. Tarlovskiy. — M. : Radio and communication, 1986. — 264 c.

4. *Levin B. R.* Theoretical foundations of statisti-cal radio engineering / B. R. Levin. — M. : Sov. radio, 1974–1976. — B. 1–3. – book 1 — 552 p., book 2 — 392 p., book 3 — 288 p.

ЛІТЕРАТУРА

1. *Иванов В.* Показать все что скрыто: Стегоанализ — выявление стеганографии, принципы и методы: [Электронный ресурс] / В. Иванов // Стеганография & путешествия, 2012. — Режим доступа: http://nestego.blogspot.com/2012/05/blog-post_21. html.

2. *Иванов В.* Виды атак на стеганографическую систему: [Электронный ресурс] / В. Иванов // Стеганография & путешествия, 2012. http://nestego. blogspot.com /2012/05/blog-post_18.html.

3. *Фомин Я. А.* Статистическая теория распознавания образов / Я. А. Фомин, Г. Р. Тарловский. — М. : Радио и связь, 1986. — 264 с.

4. *Левин Б. Р.* Теоретические основы статистической радиотехники / Б. Р. Левин. — М. : Сов. радио, 1974–1976, кн. 1–3; кн. 1 — 552 с., кн. 2 — 392 с., кн. 3 — 288 с.

Стаття надійшла до редакції 05.12.2013