DOI: 10.5604/01.3001.0013.0910

STEGANOGRAPHY – THEORY AND PRACTICE

Vladimir Barannik, Bogdan Gorodetsky, Natalia Barannik
University of Ivan Kozhedub Kharkiv National University of Air Force, Faculty/Department combat use and operation of ACS

Abstract. *Analysis of modern interstate conflicts, trends in the development of forms of warfare. It is shown that confrontation is characterized by various forms, is hidden in nature and is carried out mainly in the political, economic, informational and other spheres. It is proved that a significant part of hybrid wars are information operations used for the destructive impact on society, commercial activities, politics and economics using information and communication space and technologies. The article expresses the need to create a theoretical basis for combating cyber attacks in special telecommunication systems as an integral part of the national security of the state. The development of methods for hiding information as well as providing information during video streaming and images in networks is underway. The basic calculations are given at the initial stages of information hiding and methods for ensuring the latent transfer of data in telecommunication systems.*

Keywords: cyber attacks, telecommunications equipment, cyber security, information hiding

STEGANOGRAFIA – TEORIA I PRAKTYKA

Streszczenie. *W artykule dokonano analizy współczesnych konfliktów międzypaństwowych oraz kierunków rozwoju form walki zbrojnej. Pokazano, że konfrontacja charakteryzuje się różnymi formami, jest ukryta w naturze i prowadzona jest głównie w sferze politycznej, gospodarczej, informacyjnej i innych. Udowodniono, że znaczna część wojen hybrydowych to operacje informacyjne wykorzystywane do destrukcyjnego oddziaływania na społeczeństwo, biznes, politykę i ekonomię, wykorzystujące informacyjno-komunikacyjną przestrzeń i technologię. W artykule wyrażono potrzebę stworzenia teoretycznej podstawy zwalczania cyberataków w specjalnych systemach telekomunikacyjnych jako integralnej części bezpieczeństwa narodowego państwa. Opracowano metodę ukrytej informacji, a także dostarczania informacji w przesyłaniu strumieni wideo i obrazów w sieciach. Podstawowe obliczenia są podane na początkowych etapach ukrywania informacji i metod zapewnienia bezpieczeństwa cybernetycznego w systemach telekomunikacyjnych.*

Słowa kluczowe: cyberataki, sprzęt telekomunikacyjny, bezpieczeństwo cybernetyczn

Introduction

Computer viruses are by far the most common cause of information loss in the world. The term was coined in 1983 by Fred Cohen. The exact definition of a computer virus no one can still give, because it is impossible to distinguish any differences or purely inherent only to them signs. Since all these features are present in other programs. But there is still one characteristic of many viruses is the ability to reproduce. Computer viruses can even adapt to the environment and move.

Currently, there is no system, server or computer that gives 100% guarantee that there is no virus. That day thanks to cybernetic and virus attacks a lot of information is stolen, and sell it to DarckNet. With each passing day, virus attacks are becoming more complex, more hidden, and they are becoming more difficult to find. According to the conclusion of large companies that will produce anti-virus programs, in 2018, cyber attacks will become even more. In spite of the tendency of viruses, and regardless of their structure, program code, can see the virus has begun to add a neural network. This is due to the fact that virus programs have become more flexible, and change their code depending on the situation, without human intervention. Also analyzing the traffic thanks to the program AirShark, I saw the virus programs have become multi-channel, that is not how the virus used to communicate with the server on one channel that immediately exposed him. New viruses use other programs and open ports only for themselves, thus enabling the virus to spread and steal information without any interference. In this paper, I want to demonstrate methods of masking virus programs, antivirus programs, as well as methods of countering virus attacks.

1. The use of methods of hiding information in real life

In our lives, we are often faced with hidden information and we often use it, even from childhood, we begin to invent with friends special privetstvie, kakieto movement that denote some words or sentence. In our time, young people use the abbreviation or designation in correspondence, they do it to parents or postoronii people could not read their information. Examples, word of two letters – "ku", it means greeting or another example "SPS" – many thanks, the third example of the "XS" – who the hell knows. Examples are very many, I'm not even surprised if our young people already have a wordlist to conceal information from prying people. As we have seen that we are using the concealment of information very often and dismantled classic examples and definitions we can give. Concealment of information is the transformation of known information into a form that is known only to the person to whom we want to convey this information. In this section, a more in-depth look at the use of hiding information, namely steganography (which means hiding digital information in images and video streams) and viruses (the use of various methods of hiding information that would not be detected by anti-virus programs).

1.1. Classification of methods of masking virus attacks

Masking virus attacks is becoming almost the most urgent problem at the moment, every day more and more viruses are found, constantly creating new algorithms to find the virus. As we know that viruses go two steps ahead from antivirus programs. People who create viruses are people who are big enough programmers, they know how to read program codes very well [7]. Next, we continue to talk about the masking of virus attacks, it is now possible to classify the masking of viruses as shown in figure 1.

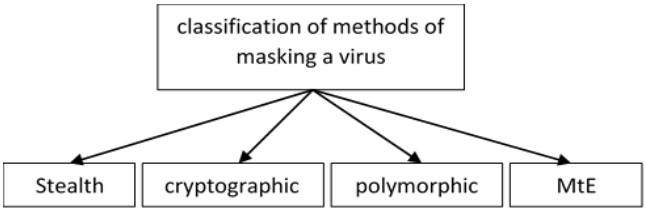


Fig. 1. Classification of methods of masking a virus

STEALTH – viruses

Virus authors replace some components of the operating system, such as interrupts, drivers so that the virus program becomes invisible to other programs. Such viruses are called viruses-the stealth, or stealth-viruses (Stealth, invisible).

Stealth-viruses are always resident. The resident module intercepts calls of the operating system to the affected files and sectors of disks and "substitutes" initial objects instead of them.

So Stealth-viruses are hiding from an experienced user and many anti-virus tools that perform early virus scanning by changing the length of the checksum files and the content of the boot sectors.

It should be noted that Stealth-the virus is invisible only when its resident module is in the RAM of the computer.

Encryption viruses

In order to make detection more difficult, some viruses encrypt their code. The concept of a computer virus is closely related to another concept – signature. A signature is a piece of code found in all copies of the virus and only in them. The signature uniquely identifies the presence or absence of the virus.

It is obvious that to search for the body of viruses on the disks by anti-virus tools is impossible (not enough memory!). Keep complete virus program codes. Therefore, the developers of anti-virus tools do the following: to find viruses, they save their signatures. This search for virus codes is called signature search.

The encryption technique itself is as follows: every time a virus infects a new program, it encrypts its own code using a new key. The encryption key depends on the target file (for example, its name or length). As a result, two instances of the same virus can differ significantly from each other, and have different lengths! This makes it difficult to detect the virus using signature search. After all, the encrypted code no longer has the same signature.

Encryption viruses, when receiving control, first decrypt their code using the decryption procedure, and then perform all other actions. Encryption viruses are sometimes called "Ghost" viruses.

Polymorphic virus

All further improvements in virus algorithms are already dictated by the survival of viruses when working under various anti-virus tools.

Encryption viruses hide the signature of their code. But the encrypted code must be decrypted before execution, hence a decryption procedure is required, which itself cannot be encrypted as it is executed before the main virus code. The decryptor contains a specific code and is large enough to serve as a signature. This is used by anti-virus programs that use the decryption procedure code as a signature. The authors of the viruses on this approach said polymorphic viruses. These viruses use not only different keys for encryption, but also different encryption procedures (respectively, decryption). Two instances of such a virus have no matching code sequence!

Viruses, through the use of various decryptors, can completely change their code, called polymorphic viruses (polymorphic).

These viruses are supplemented by decryption generators. Such a generator creates for each new copy of the virus its own decoder, different from all the others, but performs the same function. It's complicated. Such tasks are already related to programming automation.

The first polymorphic virus was written in the United States by mark Washburn (Mark Washburn) as an experimental. It was called "V2Px", or "V-1260", and although it was not distributed as a virus, it served as an example for virus authors.

At the moment, we know a huge number of polymorphic viruses. It would seem that polymorphic viruses implement complex algorithms, so only highly qualified specialists can write them.

Mutant engine (MtE) viruses

But, unfortunately, at present, not only highly qualified specialists can create polymorphic viruses.

In 1991. in Bulgaria, a well-known author of viruses that call themselves Dark Avenger (black Avenger), an algorithm for creating polymorphic viruses was developed. This is a very complex algorithm that generates decoders that are completely different from each other. Their size ranges from 0 to 512 bytes, and almost all processor commands can be found in the body. This algorithm the author called Mutation Engine (machine mutations), abbreviated he is called MtE or DAME (Dark Angel MuTation Engine).

Viruses attached to them mollem MtE for the generation of interpreters, called MtE-viruses. These are semi-automatic viruses.

Even by the number of automated developments for the creation of polymorphic viruses, the wide spread of polymorphic viruses becomes evident. Not all of these types of viruses are hidden in video streams or images. At the moment, the most popular viruses have mutation angina (It), they can be controlled in real time and adaptively adjust to the situation.

1.2. Steganography

First, consider the definition of what is Steganography. Steganography-a method of transmission or storage of information, taking into account the secrecy of the fact of such transfer (storage). This term was introduced in 1499 the Abbot of the Benedictine monastery of St. Martin in Spygame Johannes Trithemius in his treatise "Steganography" (lat. Steganographia), encrypted under a magic book.

Before we consider steganography, we will consider cryptography. Cryptography is the science of methods of ensuring confidentiality (impossibility of reading information to outsiders), data integrity (impossibility of imperceptible change of information), authentication (authentication of authorship or other properties of the object), as well as the impossibility of refusal of authorship. Initially, cryptography studied methods of information encryption-reversible conversion of open (source) text based on a secret algorithm or key into encrypted text (ciphertext). Traditional cryptography forms a section of symmetric cryptosystems in which encryption and decryption is performed using the same secret key. In addition to this section, modern cryptography includes asymmetric cryptosystems, electronic digital signature systems (EDS), hash functions, key management, obtaining hidden information, quantum cryptography.

Cryptography does not protect against fraud, bribery or blackmail of legitimate subscribers, theft of keys and other threats to information arising in secure data transmission systems.

Cryptography is one of the oldest Sciences and has a history of several thousand years.

Unlike cryptography, which hides the contents of a secret message, steganography hides the fact of its existence. Typically, the message will look like something else, such as an image, article, shopping list, letter or Sudoku. Steganography is usually used in conjunction with cryptography methods, thus complementing it.

The advantage of steganography over pure cryptography is that messages do not attract attention. Messages, the fact of which encryption is not hidden, cause suspicion and can be incriminating in those countries in which cryptography is prohibited [1]. Thus, cryptography protects the content of the message, and steganography protects the fact of the presence of any hidden messages.

In the late 1990s, there were several areas of steganography:

- Classic,
- Computer,
- Digital,
- Network.

The most relevant areas of steganography are digital and network. Digital steganography-the direction of classical steganography, based on the concealment or introduction of additional information into digital objects, causing some distortion of these objects. But, as a rule, these objects are multimedia objects (images, video, audio, textures of 3D objects) and distortion, which are below the threshold of sensitivity

of the average person, does not lead to noticeable changes in these objects. In addition, in digitized objects, initially having an analog nature, there is always quantization noise; further, when playing these objects, there is additional analog noise and nonlinear distortion of the equipment, all this contributes to greater invisibility of hidden information.

1.3. Network steganography

Recently, methods have become popular when hidden information is transmitted through computer networks using the features of data transmission protocols. Such methods are called "network steganography". This term was first introduced by Syperski Krzysztof (Pol. Krzysztof Szczypiorski) in 2003. Typical network steganography methods include changing the properties of one of the network protocols. In addition, the relationship between two or more different protocols can be used to more reliably hide the transmission of a secret message. Network steganography covers a wide range of methods, in particular:

- WLAN-steganography is based on methods that are used to transmit steganograms in wireless networks (Wireless Local Area Networks). A practical example of WLAN steganography is the HICCUPS system (Hidden Communication System for Corrupted Networks).
- LACK-steganography-hiding messages during conversations using IP-telephony. For example: using packets that are delayed or intentionally corrupted and ignored by the receiver (this method is called the LACK – Lost Audio Packets Steganography) or hiding information in header fields that are not used.

The principle of operation of the LACK is as follows. The transmitter (Alice) selects one of the voice stream packets, and its payload is replaced by bits of a secret message – a steganogram that is embedded in one of the packets. Then the selected packet is deliberately delayed. Each time an excessively delayed packet reaches a recipient unfamiliar with the steganographic procedure, it is discarded. However, if the receiver (bean) knows about the hidden link, it retrieves the hidden information instead of deleting the received RTP packets.

We also denote the classification of steganography algorithms. All algorithms of hidden information embedding can be divided into several subgroups:

- Working with the digital signal itself. For example, the LSB method.
- The "Sealing-in" of hidden information. In this case, there is an overlay of the hidden image (sound, sometimes text) on top of the original. Often used for embedding digital watermarks (CVZ).
- Use of file format features. This can include writing information to metadata or to various other unused reserved fields in the file.

2. Information hiding method

In this section, we will consider an information hiding algorithm, using steganography, which was described in Section 1, and we will also consider information encryption. This method of information hiding allows you to transmit hidden messages or commands using images or video streams. It can also be used as a hidden digital key. This method allows you to transfer information with almost no data loss and skirting various firewalls and anti-virus programs. To disguise information, I use several types of data encryption and digital steganography techniques to embed sensitive information into raster images of mlm video streams [7].

2.1. Data encryption methods

In this section, we will look at two types of data encryption that I used, MD5 and AES, the first I will use the AES cryptographic encryption method.

Mathematical foundations of the cipher. A finite Galois field is used to describe the algorithm GF(2^8) are polynomials of the form

$$b_7x^7 + b_6x^6 + b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0 \quad (1)$$

Degree less than 8, and the coefficients $b_7, b_6 \cdots b_0\{0,1\}$. Operations in the field are performed modulo m (x). In total, the field GF(2^8) reads $2^8 = 256$ polynomials.

The second method I use is MD5 cryptographic encryption, because it is more compressed.

For calculations, 4 variables of 32 bits size are initialized and the initial values are set by hexadecimal numbers (byte order little-endian, first low byte):

A = 01 23 45 67; // 67452301h
B = 89 AB CD EF; // EFCDAB89h
C = FE DC BA 98; // 98BADCFEh
D = 76 54 32 10. // 10325476h

These variables will store the results of the intermediate calculations. The initial state of ABCD is called the initializing vector.

Let's define more functions and constants that we need for calculations.

It will require 4 features for four rounds. We introduce functions from three parameters-words, the result will also be the word formula (1) was taken from [1]:

1 round : FunF (X , Y , Z) = (X ∧ Y) ∨ (¬ X ∧ Z) (1.1)
2 round: FunG (X , Y , Z) = (X ∧ Z) ∨ (¬ Z ∧ Y) (1.2)
3 round: FunH (X , Y , Z) = X ⊕ Y ⊕ Z (1.3)
4 round: FunI (X , Y , Z) = Y ⊕ (¬ Z ∨ X), (1.4)

where ⊕ , ∧ , ∨ , ¬ bitwise logical operations XOR, AND, OR, NOT.

Define a table of constants

T [1 ... 64] is a 64-element data table constructed as follows formula (2) was taken from [2]:

$$T[n] = \text{int}(2^{32} \cdot n) \quad (2)$$

Each 512-bit block goes through 4 stages of calculation in 16 rounds. To do this, the block is represented as an array X of 16 words of 32 bits. All rounds are of the same type and have the form: [abcd k s i], defined as, formula (3) was taken from [10]:

$$a = b + ((a + Fun(b , c , d) + X [k] + T [i]) \lll s) \quad (3)$$

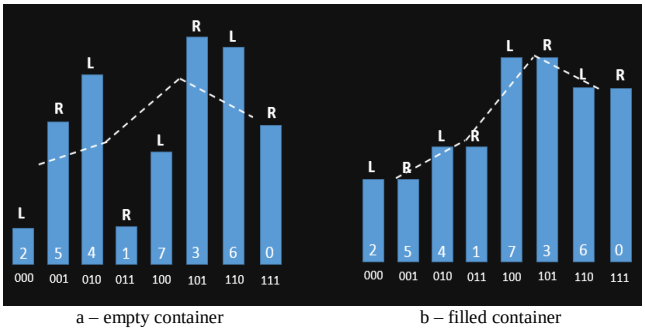
where k is the number of a 32-bit word from the current 512-bit message block, and ... ⋘ s is the cyclic shift to the left by s bit of the resulting 32 — bit argument. The number s is given separately for each round.

2.2. A method of masking data

Further we used steganography in which statistical methods of analysis: histogram method

This method, proposed in 2000 [1], Andres Westfield and Andreas Pfitzmann, also known as "Chi-square"-method. Let's try to explain its essence.

The entire raster is analyzed, for each color is the number of points of this color in the raster (for simplicity, we are talking about an image that has one color plane). The method assumes that the number of points of two adjacent colors ("adjacent" colors – colors that differ only in the least significant bit) differs significantly for a normal image (an empty container) (Fig. 2). And the number of pixels of such colors is approximately the same for the filled container (Fig. 2).



a – empty container b – filled container

Fig. 2. A method of masking data

Visually it is possible to imagine the algorithm in such a way, it is quite simple to understand.

Strictly speaking, the algorithm consists in sequential execution of the following steps:

- Theoretically, the expected frequency of pixels of color i after the introduction of the message is calculated as follows formula (4) was taken from [6]:

$$n_i^* = \frac{1}{2}|\{colour | sortedIndexOf(colour) \in \{2i, 2i+1\}\}| \quad (4)$$

- The measured frequency of occurrence of a symbol of a particular color is defined as formula (5) was taken from [6]:

$$n_i = |\{colour | sortedIndexOf(colour) = 2i\}| \quad (5)$$

- Chi-square criterion for the number of degrees of freedom k-1 is calculated as follows formula (6) was taken from [9]:

$$x_{k-1}^2 = \sum_{i=1}^{k} \frac{(n_i - n_i^*)^2}{n_i^*} \quad (6)$$

Of course, we have tested the applicability of this method [11] to detect filled Starokonstantinov and the results shows in Fig. 3.

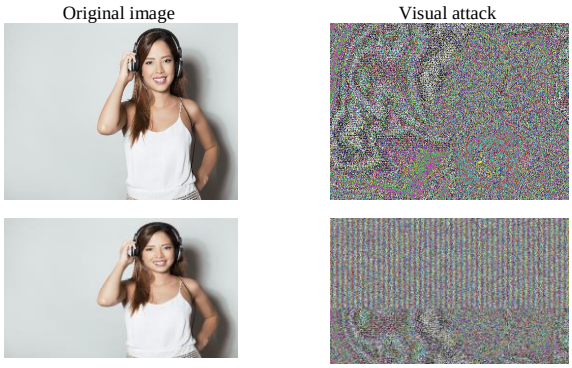


Fig. 3. Results on the test

The Chi-square distribution thresholds for p=0.95 and p=0.99 are 101.9705929 and 92.88655838, respectively. Thus, for zones where the calculated Chi-square value is less than the threshold one, we can accept the initial hypothesis "the distribution of frequencies of neighboring colors is the same, therefore, it is a filled "stegocontainer"".

Indeed, if you look at the images for a visual attack, it's easy to notice that these areas contain an embedded message. Thus, for embedded messages with high entropy, the method works.

To combat this threat, you can use image compression, or rather use what changes the program code [13], for this we use compression algorithm Deflate. The data algorithm is used to compress PNG files.

References

[1] Ablamejko S., Lagunovskij D.: Obrabotka izobrazhenij: tehnologija, metody, primenenie. Amalfeja, Minsk 2000.
[2] Barannik V. V., Alimpiev A. M., Bekirov A., Barannik D. V., Barannik N. D.: Detections of sustainable areas for steganographic embedding. Proc. East-West Design & Test Symposium (EWDTS) 2017, 555–558 [DOI: 10.1109/EWDTS.2017.8110028].
[3] Barannik V., Barannik D., Bekirov A., Lekakh A.: Steganographic method based on the modification of regions of the image with different saturation. Proc of. 14th International Conference: Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering 2018, 542–545 [DOI: 10.1109/TCSET.2018.8336260].
[4] Christophe E., Lager D., Mailhes C.: Quality criteria benchmark for hiperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing 43(9)/2005, 2103–2114.
[5] Computer viruses, Moscow, 2011.
[6] Gonzalez R., Woods R.: Tsyfrova obrobka zobrazhen. Tekhnosfera, Kiev 2018.
[7] Gribunin V., Okov I., Turincev I.: Tsifrovaya steganografiya. Solon-Press, Moscow 2018.
[8] Grundmann M., Kwatra V., Han M., Essa I.: Efficient hierarchical graph based video segmentation. IEEE CVPR, 2010.
[9] Konakhovich G. M, Puzyrenko A.: Computernaya steganographiya. Teoriya i praktika. To Press, Kiev 2016.
[10] Melnik A. M.: Informatsiyni systemy ta merezhi. Bulletin "Lviv Polytechnic" 673, 2010, 365–374.
[11] Miano J.: Compressed image file formats: JPEG, PNG, GIF, XBM, BMP. Addison-Wesley Professional, New York 1999.
[12] Miano J.: Formats and image compression algorithms in action. Triumph, Kiev 2013.
[13] Pratt W., Chen W., Welch L.: Slant transform image coding. Proc. Computer Processing in Communications. Polytechnic Press, New York 1969.
[14] Savyn L.V. Setetsentrychnaya y setevaya voyna. Vvedenye v kontseptsyyu. Evrazyyskoe dvyzhenye, Moscov 2011.
[15] Sindeev M., Konushin A., Rother C.: Alpha-flow for video matting. Technical Report, 2012.
[16] Stankiewicz O., Wegner K., Karwowski D., Stankowski J., Klimaszewski K., Grajek T.: Encoding mode selection in HEVC with the use of noise reduction. Proc. International Conference on Systems, Signals and Image Processing (IWSSIP), Poznan, 2017, 1–6.
[17] Wallace G.: Overview of the JPEG (ISO/CCITT) – Still image compression: image processing algorithms and techniques. Processing of the SPIE 1244, 1990, 220–233.
[18] Wallace G.: The JPEG Still Picture Compression Standard. Communication in ACM 34(4)/1991, 31–34.
[19] Wang S., Zhang X., Liu X., Zhang J., Ma S., Gao W.: Utility-Driven Adaptive Preprocessing for Screen Content Video Compression. IEEE Transactions on Multimedia 19(3)/2017, 660–667.

Bogdan Gorodetsky
e-mail: homich1997g@gmail.com

Cadet of Ivan Kozhedub Kharkiv National University of Air Force, Department combat use and operation of ACS. Participant of the Ukrainian scientific and practical conference. Scientific activities: information security, information and psychological influence.

ORCID ID: 0000-0003-2183-3557

Natalia Barannik
e-mail: Barannik_V_V@ukr.net

National university of civil defence of Ukraine. Active participant
of scientific conferences, computer science olympiads.

ORCID ID: 0000-0001-6420-1838

Ph.D. Vladimir Barannik
e-mail: vvbar.off@gmail.com

Chief of the Department of combat command and control of the ACS of Ivan Kozhedub Kharkiv National University of Air Force, Doctor of Technical Sciences, professor.

ORCID ID: 0000-0002-2848-4524

otrzymano/received: 01.12.2018 *przyjęto do druku/accepted: 28.02.2019*