BPCS-Steganographyとディジタル絵封筒システム
BPCS-Steganography and Digital Picture Envelope System

野崎 剛一*, 河口 英二†
Koichi NOZAKI* and Eiji KAWAGUCHI†

長崎大学*
Nagasaki University

九州工業大学†
Kyushu Institute of Technology

BPCS-Steganographyは, 24ビットのフルカラー画像BMPファイルをダミー画像データとするステガノグラフィ技術で, 画像データのビットプレーン上に情報の埋め込む. このシステムの最大の特徴は, 情報の埋め込み容量が大容量であることである. 我々は, このステガノグラフィ技術を新たなデジタル絵封筒システムとして電子メールシステムへ応用した. 本稿では, ステガノグラフィ技術を使ったインターネット秘匿メールシステムのモデルに関して記述している. このシステムは, 大学のような限られた機関のメンバーを利用者として想定しているが, それ以外の機関でも利用でき, 極めて安価で容易に電子メールの機密保護を可能とするものである.
キーワード : ステガノグラフィ, 画像の複雑さ, ビットプレーン, 情報秘匿, 電子メール

BPCS-Steganography is a new steganographic technique. It uses a true color image file (24 bits BMP file) as the information hiding dummy image. Embedding is made on the bit-planes of the image. The most important feature of this steganography is that its embedding capacity is very large. We will give a new concept to this steganography a Digital Picture Envelope. This paper describes a model of an anonymous covert mailing system for Internet communication using steganography. We develop a real-life secure mailing system in which a sender can send a secret message even to a non-acquainted person in an anonymous way. The users of this system are assumed to be members of a closed organization. But it is not primarily limited only to such users. It is a quite easy-to-use system with very cheap cost.
Keywords : Steganography, Image Perception, Image Complexity, Information Hiding, Image Bit-Plane, anonymous mail

1. Introduction

In this Internet age, there are many people who want communicate with a distant partner anonymously and covertly.

We may be able to realize this hope by using steganography. Modern steganography has a relatively short history because people did not pay much attention to this skill until the Internet security became a social concern. Most people did not know what steganography was because they did not have any ways to know the meaning. Even today ordinary dictionaries do not contain the word "steganography." Books on steganography are still very few [1], [2].

In the present paper we will show the BPCS-Steganography and our basic model of an anonymous and covert e-mailing system. Finally, we show our future schedule to make it a real life system.

*総合情報処理センター
〒852-8521 長崎市文教町1番14号
Information Science Center, Nagasaki University
〒852-8521 1-14 Bunkyo-machi, Nagasaki
E-mail : nozaki@net.nagasaki-u.ac.jp
†工学部電気工学科
〒804-8550 福岡県北九州市戸畑区仙水町1-1
Kyushu Institute of Technology
〒804-8550 1-1 Sensui-cho, Tobata-ku, Kitakyushu, JAPAN
E-mail : kawaguch@know.comp.kyutech.ac.jp

2. Steganography

In recent years, several steganographic programs are posted on internet home pages. Most of them use image data for the container (or, carrier) of the secret information. Some of them use the least significant bits of the image data to hide secrets. Other program embeds the secret information in a specific band of the spatial frequency component of the carrier. Some other program makes use of the sampling error in image digitization. However, almost those steganographies are insufficient in terms of information hiding capacity. They can embed only 5-15 % of the carrier image. Therefore, current steganography is more oriented to water marking of computer data than to secret human-to-human communication applications.

We have invented a new technique to hide secret information in a color image. This is not based on a programming technique, but is based on the property of human vision system. Its information hiding capacity is as large as 50% of the original image data. This could open a new step for a steganography toward a secure internet communication age.

Digital images are categorized in either binary (black-and-white) or multi-valued pictures despite their actual color. We can decompose an n-bit image into a set of n binary images by bit-slicing operations[3][4]. Therefore, binary image analysis is essential to all digital image processings. Bit-slicing is not necessarily the best in the Pure-Binary Coding system (PBC), but in some case the Canonical Gray Coding system (CGC) is much better [5].

3. BPCS-Steganography

3.1 The Bit-Plane Complexity Segmentation Steganography

In 1997 the authors invented a new steganographic method named "BPCS-Steganography." (BPCS : Bit-Plane Complexity Segmentation) The most important feature of this steganography is that it has a very large data hiding capacity [6], [7]. It normally embeds 50% or more of a container image file with information without increasing its size. We made an experimental program (for Windows) and located it on a Web site for free downloading [8]. We also have several introductory Web pages to BPCS-Steganography and its applications [9].

Steganography can be applied to variety of information systems. Some key is used in these systems when it embeds/extracts secret data. One natural application is a secret mailing system [10] that uses a symmetric key. Another application pays attention to the nature of steganography whereby the external data (e.g., visible image data) and the internal data (any hidden information) cannot be separated by any means. We will term this nature as an "inseparability" of the two forms of data.

3.2 The complexity of binary images

There is no standard definition of image complexity. Kawaguchi discussed this problem in connection with the image thresholding problem, and proposed three types of complexity measures[11][12][13].

In the present paper we adopted a black-and-white border image complexity.

3.2.1 The definition of image complexity

The length of black-and-white border in a binary image is a good measure for an image complexity. If the border is long, the image is complex, otherwise it is simple. The total length of black-and-white border corresponds to the summation of the number of color-changes along the rows and columns in an image. For example, a single black pixel surrounded by white background pixels has the boarder length of 4.

We will define the image complexity α by the following.

$$\alpha = \frac{k}{\text{The max. possible B - W changes in the image}}$$

Where, k is the total length of black-and-white

border in the image. So, the value ranges over:

$$0 \leq \alpha \leq 1$$

α is calculated over the whole image area. It gives us the global complexity of a binary image. In addition, we can also use α for a local image (e.g., 8×8 pixel size area) complexity.

Our perception to see a binary image is that the informative regions are simple, while the noise-like regions are complex. We will use α as our complexity measure in this paper.

3.3 Conjugation of binary image

Let P be a $2^N \times 2^N$ size black-and-white image with black area as the foreground and white area as the background. W and B denote all white and all black patterns, respectively. We introduce two checkerboard patterns Wc and Bc, where Wc has a white pixel at the upper-left position, and Bc is its complement, i.e., the upper-left pixel is black (See Fig. 1). We regard black and white pixels have logical value of "1" and "0", respectively.

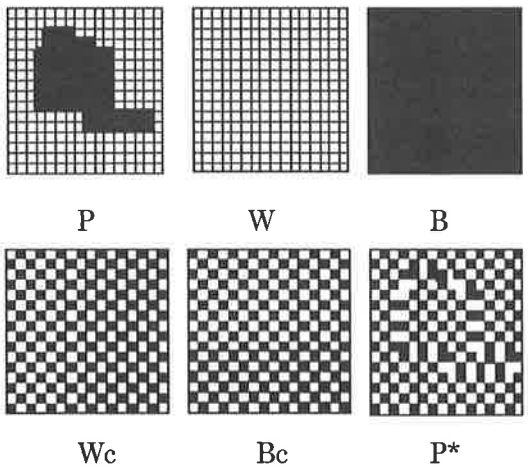

Fig. 1 Illustration of each binary pattern

P is interpreted in a way that. Pixels in the foreground area have B pattern, while pixels in the background area have W pattern.

Now we define P* as the conjugate of P which satisfies:

1) The foreground area shape is the same as P.
2) The foreground area has the Bc pattern.
3) The background area has the Wc pattern.

Correspondence between P and P* is one-to-one. The following properties hold true for such conjugation operation which can be easily proved.

"$\oplus$" designates the exclusive OR operation.

A) $P^* = P \oplus Wc$ (3.1)
B) $(P^*)^* = P$ (3.2)
C) $P^* \neq P$ (3.3)

The most important property about conjugation is the following.

D) Let $\alpha(P)$ be the complexity of a given image P, then we have,

$$\alpha(P^*) = 1 - \alpha(P). \quad (3.4)$$

It is evident that the combination of each local conjugation (e.g., 8×8 area) makes an overall conjugation (e.g., 512×512 area).

(3.4) signifies that every binary image pattern P has its counterpart P*. The complexity value of P* is always symmetrical against P regarding $\alpha = 0.5$.

3.4 The file embedding algorithm

3.4.1 The Bit-Plane Complexity Segmentation

Steganography is our new steganographic technique which has a large information hiding capacity[14].

As we have mentioned that replacing the complex regions in each bit-plane of a color image with random binary patterns is invisible to human eye. We can use this property for information hiding (embedding) strategy. Our practical method is as follows.

In our method we call a carrier image a "dummy" image. It is a color image in BMP file format, which hides (or, embeds) the secret information (files in any format). We segment each secret file into a series of blocks having 8 bytes data each. They are regarded as a 8×8 image patterns. We call such blocks the secret blocks. The file embedding algorithm into a dummy image takes the following steps.

1) Transform the dummy image from PBC to

CGC system.

2) Segment each bit-plane of the dummy image into informative and noise-like regions by using a threshold value (α_0). A typical value is α_0=0.3.
3) Segment the secret file into the series of secret blocks.
4) If a block (S) is less complex than the threshold (α_0), then conjugate it to make a more complex block (S*). It becomes more complex than α_0
5) Embed each secret block into the noise-like regions of the bit-planes (or, replace all the noise-like regions with series of secret blocks). If the block is conjugated, then record it in a "conjugation map."
6) After the secret file embedding, embed the conjugation map, too.
7) Convert the embedded dummy image from CGC to PBC.

The Decoding algorithm (i.e., the extracting operation of the secret information from an embedded dummy image) is just the reverse procedure of the embedding steps.

The novelty in BPCS-Steganography is itemized as follows.

A) Segmentation of each bit-plane of a color image into "Informative" and "Noise-like" regions.
B) Introduction of the B-W boarder based complexity measure (α) for region segmentation
C) Introduction of the Conjugation operation to convert simple secret blocks to complex blocks.
D) Using CGC image plane instead of PBC plane

4. Problems of an encrypted mailing system

There are two kinds of cryptography systems: symmetric and asymmetric.

Symmetric cryptography systems use the same key to encrypt and decrypt a message, while asymmetric cryptography systems use one key (the public key) to encrypt a message and a different key (the private key) to decrypt it.

In a symmetric system a message sender and receiver use a same encoding/decoding key. In this system, however, the sender and the receiver must negotiate on what key they are going to use before they start communication. Such a negotiation must be absolutely secret. They usually use some second channel (e.g., fax or phone). However, the second channels may not be very secure. There is another problem in this situation that if the sender is not acquainted with the receiver, it is difficult to start the key-negotiation in secret. Further more, the more secure the key system is, the more inconvenient the system usage is.

An asymmetric system uses a public key and a private key system. The public key is open to the public, and it is used for message encoding when a sender is sending a message to the key owner. However, if the public key is counterfeited, this system does not work at all. So, the public key must be authenticated. Therefore, this system needs a special authentication bureau, which is an organization that all the people in the world can trust in. In reality, it can only exist if it commercially pays. Therefore, this system is expensive and time consuming for users.

5. A Digital Picture Envelope

We have started to develop a secure and easy-to-use e-mailing system according to the BPCS-Steganography method. We do not intend to develop a new "message reader-and-sender" or "message composer", but we are developing three system components that constitute an Anonymous Covert Mailing System (ACMS). A message sender inserts (actually, embeds) a secret message in an envelope using steganography and sends it as an e-mail attachment. The receiver receives the attached envelope and opens it to receive the message. An "envelope" in this system is actually an image file that is a container, vessel, cover, or dummy data in the terminology of

steganography. This system can solve all the problems mentioned in the previous section.

The following items are the conditions we have set forth in designing the system.

(1) The name of the message sender can be anonymous.
(2) The message is hidden in the envelope and only the designated receiver can open it.
(3) Sender can send a secret message even to an unacquainted person.
(4) It is easy to use for both sender and receiver.
(5) The system is inexpensive in installing and using.

We expect this ACMS is used in a closed organization (such as in a company), but there are no restrictions for any group (or the general public) to use it.

5.1 Components of the system

ACMS is a steganography application system (or program). It makes use of the inseparability of the external and internal data. The system can be implemented differently according to different programmers or different specifications. Different ACMS' are incompatible in operation with others. In other words, an organization using ACMS must use one single ACMS. However, each member in the organization can use it in a customized way. The customization is made when the user installs it on his/her own computer. In the following description, M_i denotes a member i, and M_j denotes a member j.

An ACMS consists of the following three components.

(1) Envelope Producer (EP)
(2) Message Inserter (MI)
(3) Envelope Opener (EO)

We express M_i's ACMS as $ACMS_i$ (i.e., customized ACMS by M_i). Hence, it is described as $ACMS_i = (EP_i, MI_i, EO_i)$.

EP_i is a component that produces M_i's envelope (E_i). E_i is the envelope (actually, an image file) which is used by all other members in the organization when they send a secret message to M_i. E_{0i} is produced from an original image (E_0). M_i can select it according to his preference. E_i has both the name and e-mail address of M_i on the envelope surface (actually, the name and address are "printed" on image E_i), as illustrated in Fig. 2.

Fig. 2 An example of an envelope

It will be placed at an open site in the organization so that anyone can get it freely and use it any time. Or someone may ask M_i to send it directly to him/her.

MI_i is the component to insert (i.e., embed according to the steganographic scheme) M_i's message into another member's (e.g., M_j)'s envelope (E_j) when M_i is sending a secret message ($Mess_i$) to M_j. One important function of MI_i is that it detects a key (Key_j) that has been hidden in the envelope (E_j), and uses it when inserting a message ($Mess_i$) in E_j.

EO_i is a component that opens (extracts) E_i's "message inserted" envelope $E_i(Mess_j)$ which M_i received from someone as an e-mail attachment. The sender (M_j) of the secret message ($Mess_j$) is not known until M_i opens the envelope by using EO_i.

5.2 Customization of an ACMS

Customization of an ACMS for member M_i takes place in the following way. M_i first decides a key (Key_i) when he installs the ACMS onto his computer. Then he types in his name ($Name_i$) and e-mail address ($EAdrs_i$). Key_i is secretly hidden (according to a steganographic method or some other method) in his envelope (E_i). This Key_i is eventually

transferred to a message sender's MI_j in an invisible way. $Name_i$ and $EAdrs_i$ are printed out on the envelope surface when M_i produces E_i by using EP_i (cf. Fig. 2). Key_i is also set to EO_i at the time of installation. $Name_i$ and $EAdrs_i$ are also inserted (actually, embedded) automatically by MI_i any time M_i inserts his message ($Mess_i$) in another member's envelope (E_j). The embedded $Name_i$ and $EAdrs_i$ are extracted by a message receiver (M_j) by EO_j. Fig. 3 illustrates the scheme of this AC-Mailing system.

5.3 How it works

When some member (M_j) wants to send a secret message ($Mess_j$) to another member (M_i), whether they are acquainted or not, M_j gets (e.g., downloads) M_i's envelope (E_i), and uses it to insert his message ($Mess_j$ by using MI_j. When M_j tries to insert a message, M_i's key (Key_i) is transferred to MI_j automatically in an invisible manner, and then is actually used. M_i can send $E_i(Mess_j)$ directly, or ask someone else to send it to M_i as an e-mail attachment.

M_j can be anonymous because no sender's information is seen on $E_i(Mess_j)$. $Mess_j$ is hidden, and only M_i can see it by opening the envelope. It is not a problem for M_j and M_i to be acquainted or not because M_j can get anyone's envelope from an open site. ACMS is a very easy-to-use system because users are not bothered by any key handling, as the key is always operated automatically. As ACMS doesn't need any authorization bureau, this system can be established with very low cost. All these features overcome the drawbacks of an encrypted mailing system .

6. Anti reverse-engineering strategy

The ACMS is secure only if the key is not stolen (i.e., not disclosed by anyone). The location of the hidden key in the envelope is kept secret by the system developer, but some people may be interested in reverse-engineering the execution programs (EP and MI). General techniques to make a program difficult to be reverse-engineered include the following.

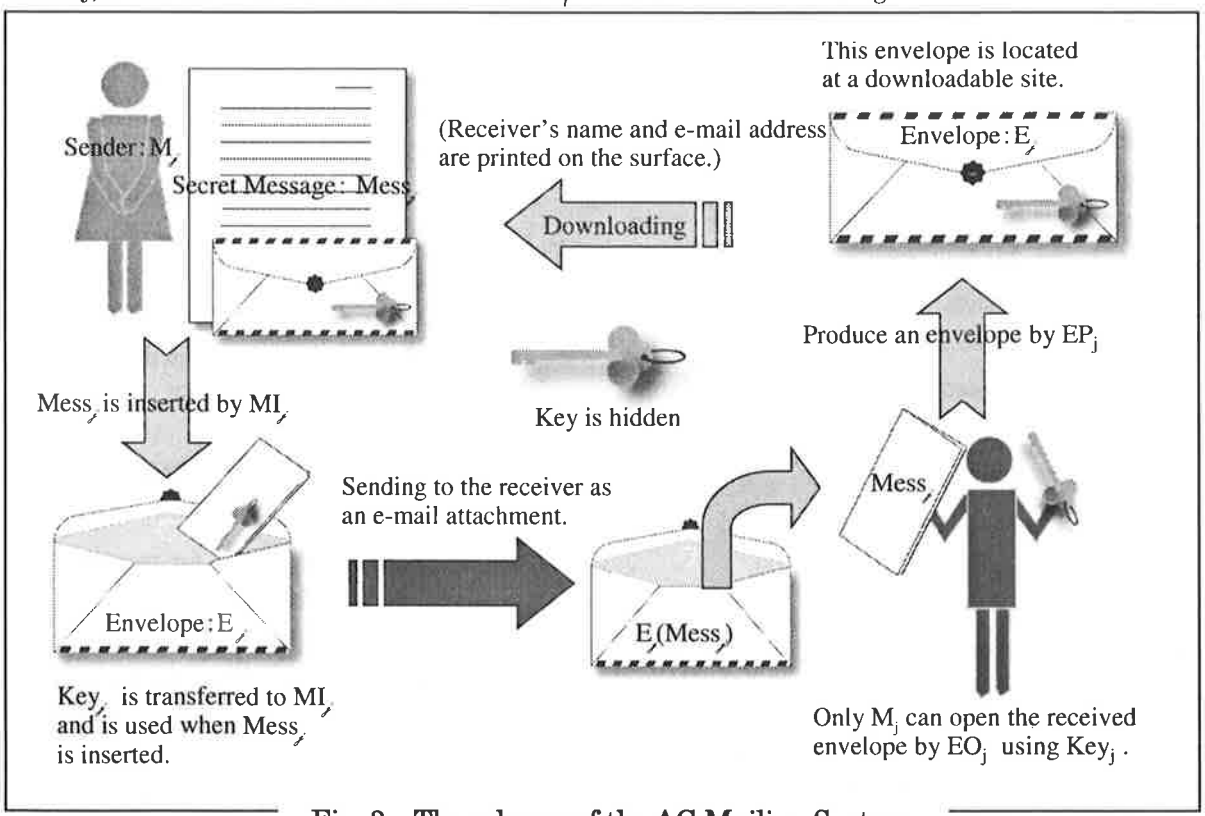

Fig. 3 The scheme of the AC-Mailing System.

(1) To make the structure of the program a very tangled one, or a "spaghetti program."
(2) To make it "manually untraceable" by way of inserting very complicated subprograms.
(3) To set a lot of branches in the program-flow according to non-algorithmic conditions.

One practical method for (3) is to use "time intervals" between two (or more) instructions along a program-flow. In a normal running mode of the computer, the flow branches into a correct direction, but in a reverse-engineering mode the computer speed may be shifted down for tracing and the flow strays into incorrect directions. This confuses the reverse-engineers completely. It is difficult to make a theoretical analysis of how secure the system is, but it is practically safe if the programs are made very carefully.

7. E-mail Security and Internet Privacy

We use the Transmission Control Protocol/Internet Protocol (TCP/IP) over the Internet communication. TCP/IP allows information to pass through intermediate computers makes it possible for a third party to interfere with communications. So, there may be some eavesdropping, tampering or impersonation on our Internet communications. Normally, most users of the Internet do not monitor or interfere with the network traffic that continuously passes through their machines. Although e-mail is one of the most popular uses of the Internet, it is less secure and in many ways more dangerous than a postcard. An e-mail message travels through many computers and at each computer many people can access it. So if we want to protect our confidential information from these attackers, we must use encryption programs.

Today there are many encryption programs currently available. Some are secure and others are weak. Most e-mail encryption products fall into two main standards or protocols. The two de facto ad hoc standards are S/MIME and PGP. But almost e-mail uses are not concerning about e-mail security and think it cumbersome and complex. So, there are few people who use these encryption programs.

8. Conclusion and Future schedule

Today, it is difficult and inconvenient to make information sent via e-mail private, or to ensure its integrity and the authenticity of the sender. Because of its speed, convenience and ease of use, e-mail has become a very important tool for almost faculties, stuffs and students in universities. This new digital envelope technology is so secure that we can send any important confidential information via e-mail and it is a quite new and easy-to-use. Our project is still at a prototype- implementation stage. The basic system design has already been finalized. We will test and investigate the actual usage of the system within some groups in a college or university. We may also test this system for "inter-organization" use by introducing a "user-group identifying capability" in the system. The final goal of the project is to develop a "world-wide use" version as soon as possible.

References

[1] Stefan Katzenbeisser and Fabien A.P. Petitcolas (eds) : "Information hiding techniques for teganography and digital watermarking", Artech House (2000).

[2] Neil F. Johnson, Zoran Duric and Sushil Jajodia : "Information Hiding", Kluwer Academic Publishers (2001).

[3] Hall, Ernest L. : "Computer Image Processing and Recognition", Academic Press, New York (1979).

[4] Jain, Anil K. : "Fundamentals of Digital Image Processing", Prentice Hall, Englewood Cliffs, NJ (1989).

[5] Kawaguchi, E., Endo, T. and Matsunaga, J. : "Depth First picture expression viewed from digital picture processing", IEEE Trans. on PAMI, vol.5, no.4, pp.373-384 (1988).

[6] M. Niimi, H. Noda and E. Kawaguchi : "An image embedding in image by a complexity based region segmentation method", Proceedings of International Conf. on Image Processing'97, Vol.3, pp.74-77, Santa Barbara, Oct. (1997).

[7] E.Kawaguchi and R.O. Eason : "Principle and applications of BPCS Steganography", Proceedings of SPIE: Multimedia Systems and Applications, Vol.3528, pp.464-463 (1998).

[8] URL
http://www.know.comp.kyutech.ac.jp/BPCSe/Dpenv-e/DPENVe-pro_down.html

[9] URL
http://www.know.comp.kyutech.ac.jp/BPCSe/

[10] E. Kawaguchi, et al : "A concept of digital picture envelope for Internet communication" in Information modeling and knowledge bases X, IOS Press, pp.343-349 (1999).

[11] Kawaguchi, E. and Taniguchi, R. : "Complexity of binary pictures and image thresholding - An application of DF-Expression to the thresholding problem", Proceedings of 8th ICPR, vol.2, pp.1221-1225 (1986).

[12] Kawaguchi, E. and Taniguchi, R. : "The DF-Expression as an image thresholding strategy", IEEE Trans. on SMC, vol.19, no.5, pp.1321-1328 (1989).

[13] Kamata, S, Eason, R. O., and Kawaguchi, E. : "Depth-First Coding for multi-valued pictures using bit-plane decomposition", IEEE Trans. on Comm., vo.43, no.5, pp.1961-1969 (1995).

[14] Koichi Nozaki, et al : "A Large Capacity Steganography Using Color BMP Images", Proc. ACCV'98, pp.112-119 (Jan. 1998).

(2002年 10月22日 原稿受付)
(2003年 3月14日 採録決定)

Authors' Profile

Koichi Nozaki

He received the B.E. electrical engineering in 1975 from Kyushu University, Japan. He worked as a research associate and is currently a lecturer of the Information Science Center, Nagasaki University. He has contributed in computer science education and management of the large computer network system of Nagasaki University. His current research interests include steganography, image processing, and computer network system.

Eiji Kawaguchi

He received a Doctor of Engineering Degree in 1972 from Department of Electronics Engineering at Kyushu University, Japan. He started his academic career at universities from 1969 in engineering. Currently he is a professor of Electrical, Electronics and Computer Engineering Department at Kyushu Institute of Technology. His research interests include speech recognition, pattern understanding, image processing, knowledge engineering, information hiding, natural language processing, and semantics modeling.