An Automatically Changing Feature Method based on Chaotic Encryption

1Wang Li, 2Gang Luo and 3Lingyun Xiang
1College of Information Science and Engineering, Hunan University, Changsha, 410082, China
2College of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, 410114, China

Abstract: In practical applications, in order to extract data from the stego, some data hiding encryption methods need to identify themselves. When performing data hiding, they embed some specific logo for self-identification. However, it is unavoidable to bring themselves the risk of exposure. Suppose each hidden method has a corresponding logo S and the attacker has a logo set Φ which consists of some hidden methods’ logos. Once he find the logo S which matches a logo in Φ, he can easily recognize the very method. To solve this problem, we propose a method based on synchronization to hide the specific sign or logo. First, the sender generates a key using a public variable parameter which is always changing from time to time. Then, we can calculate the hidden data’s location in the cover from the key. According to the locations, we can embed the logo into the cover. As the public parameter is changing from time to time, each transmission of hidden data has a unique location sequence. When the stego reaches its receiver, according to the public parameter, the receiver could generate the key and get the hidden data’s location to extract the secret data correctly. Experimental results verify that the data hiding method performs well and hardly has impacts on the cover’s quality and has little impacts on the robustness, imperceptibility and capacity of the original stego-cover. Besides, it is able to recover the key with linear time complexity when the critical information which is used to generate the key is missing.

Key words: Data hiding, automatically changing, chaotic encryption, synchronization

INTRODUCTION

Data hiding techniques are mainly about researching on hiding data into cover. Avoiding being noticed by adversaries is their main purpose. Until now, it consists of Steganography, Digital Watermark, Visual Cryptography and etc. With data hiding techniques, secret communications become possible. Data hiding techniques are being applied to military, business, finance areas, such as conveying military intelligence, the privacy data of e-business, secret contracts between business partners and so on. Besides, data hiding techniques are utilized for identification, copyright protection, integrity authentication and so on.

Nowadays, main researches on data hiding techniques can be roughly divided into three areas: Spatial domain, transform domain and file format.

For spatial domain methods, LSB and LSB matching are the most commonly used algorithms. Researches on data hiding in spatial domain focus on capacity improvement, imperceptibility advancement and resistance against steganalysis (Zhang, 2011; Onoorni et al., 2011; Geetha et al., 2011; Zeki et al., 2011; Zanganeh and Ibrahim, 2011; Rajagopalan et al., 2014).

For transform domain methods, though data hiding methods are more complicated, they often achieve better robustness than spatial domain ones. The QIM and adaptive embedding algorithms (Wang et al., 2002; Podilchuk and Zeng, 1998) are typical of these methods. There are also some researches (Cancellaro et al., 2011; Liao et al., 2011; Khan et al., 2011; Bouabid et al., 2012) on techniques about reversible watermarking of encrypted media. Some (Qiao and Nahrstedt, 1998; Shi and Bhargava, 1998; Wu and Kuo, 2005; Qiu and Wang, 2012) are on encryption techniques for the compressed image and some (Chen et al., 2003; Abdulaset F et al., 2010) focus on robustness and imperceptibility.

Style of file (such as XML, HTM, PDF, etc.) can be utilized in data hiding as well (Bo et al., 2009; Shahrizaz, 2006; Cantrell and Dampier, 2004; Liu and Tsai, 2007; Liu et al., 2008), according to the character of the file’s style. Such techniques are based on specific style of files which are different from each other.

In some cases, when extracting hiding information, we always needs a logo to mark the specific steganography algorithm. This brings great unsafety to the hiding system. For example, the software ffmpeg (Schroeder, 2005) embeds a fixed sequence of

Corresponding Author: Gang Luo, College of Information Science and Engineering, Hunan University, Changsha, 410082, China 2404
“0110011001100110011001101101111010100011000101” into the carrier without any hiding processing. As long as an attacker calculates this specific sequence, he or she can confirm that the carrier he or she catches is encrypted by Fienocode. Unfortunately, there are few researches on how to resolve such problems. Our purpose is to resolve this kind of problems. We put forward a method that can make the logo’s appearance different from time to time. Meanwhile, it makes sure that the receiver can extract the logo with 100% accuracy and recognizes the steganography tool correctly. And when the key is missing, it can recover the key with limited time cost in linear time complexity.

METHODOLOGY

Problem statements and analysis: To hide the specific logo, we need to encrypt it first. However, a proficient can get the logo with Chosen-plaintext attack when the logo is encrypted by the same key all the time. To avoid this, we need a key that can automatically change itself. A changing mechanism is what we need. Using an automatically changing key, the key of the sender’s must synchronize with the receiver’s. When sender’s key doesn’t match the receiver’s, we need a mechanism to restore the key’s synchronization. Considering all the requirements listed above, we take system time as the changing factor of the key. The system time has 4 key important natures: Linearity, uniqueness, finiteness in a finite interval and dynamics. The uniqueness and dynamics nature of time are necessary to the key’s automatically changing. And in case of being damaged or tampered, we need an efficient way to recover the key. The finiteness and linearity nature of the time can be used to recover the key. We also need a method to hide the encrypted logo. Comparing to hiding a logo in a fixed position, it is safer to hide the logo in random positions. The chaotic system performs well in producing pseudo-random sequences. Among chaotic mappings, the skew tent mapping has excellent pseudo-randomness and it obeys uniform distribution. With the skew tent mapping, we calculate the embedding locations so that it has the ability to resist statistical analysis.

Merits of the method:

- Choosing different keys from time to time
- Being able to figure out the missing or damaged key in linear time complexity
- Resistance to statistics analysis

Description of the method: First, P stands for plaintext, VI denotes the initial vector of 128 bits. T is the time that generated by the computer system and its precision is second. In this study, we choose the format of time as "YYYYMMDDHHMSS". Specifically, we encrypt the time T with MD5 algorithm (128 bits) and get a sequence of hex numbers: $T_{n}=MD5(T)$, $T_{n}=t_{n-16}, t_{n-15}, \ldots, t_{n-1}$.

S is the logo that need to be encrypted and its length is 1 byte (each character is 1 byte and 1 byte = 8 bits). VI and S are divided into the following groups:

$$VI = v_1, v_2, \ldots, v_{15}$$

$$S = s_1, s_2, \ldots, s_{16}$$

(Each vi or si is 8 bits).

According to $T_{n}$ and VI, the system generates the secret key K in the following way:

$$K = VI \oplus T_{n}$$

$$k_i = v_i \oplus t_{n-i} \oplus [0, 15]$$

The logo S is encrypted to M in the following way:

$$m_i = k \mod 6 \oplus s_i = b_{n,3}\cdot b_{n,2}\cdot b_{n,1} \oplus b_{n,0} \oplus s_i, i \in \{0, 1, 2, \ldots, 15\}$$

$$M = m_1, m_2, \ldots, m_{15} = b_{n,3}\cdot b_{n,2}\cdot b_{n,1} \oplus b_{n,0} \oplus s_i \in \{0, 1\}$$

The M is what we need to hide into the carrier. Its length is the same with S, the length of M is also 1 byte. And $T^0$ is the jth created time and the corresponding secret key is $K^0$ (j is an integer). $T^0$ is different from time to time and the $K^0$ is created by $T^0 \mod VI$. When i $\neq j$, $K^0 \neq K^0$.

Logistic map is often used to design random encryption system. Logistic mapping not only has low computing complexity but also has good pseudo-random performance. However, the skew tent mapping’s distribution is more uniform than logistic mapping’s (Xiang et al., 2008) and it has stronger pseudo-randomness as well (Masuda and Aihara, 2001). Therefore, the chaotic sequence generated by the skew tent mapping is better for cryptographic purpose:

- Logistic mapping:

$$x_{i+1} = 3.57x_{i}(1-x_{i}), u \in [0, 4], x \in (0, 1) \quad (1)$$

- Skew tent mapping:

$$f(x) = \begin{cases} \frac{x}{p}, & x \in [0, p) \\ \frac{1-x}{1-p}, & x \in [p, 1] \end{cases} \quad (2)$$
We set \( r \) (\( r \) is 8 bits) as follows:

\[
 r = (k_0 \oplus k_1 \oplus k_2 \cdots \oplus k_{15}) \mod 16
\]

To get the hidden data’s location by skew tent map, \( x_0 \), \( p \), \( N \) is initiated as follow:

\[
V_L = k_{max} \oplus k_{max}, \mod 16
V_R = k_{max}, \mod 16
U_L = k_1 \oplus k_2 \oplus \cdots \oplus k_7
U_R = k_7 \oplus \cdots \oplus k_1
x_0 = (V_L \oplus V_R)^{2^4}, P = (U_L \oplus U_R)^{2^4}, N = k_0 \oplus k_1 \oplus k_2 \cdots \oplus k_{15}
\]

We embed the encrypted logo \( M \) into a cover grayscale \( I \). \( M \) is 1 byte and the length of the cover \( I \) is \( L \) bytes. Suppose there is a logo set \( \Phi = \{S_0, S_1, S_2, \ldots, S_{15}\} \) and the \( L_{br} = \max \{\text{length}(S_i) + \epsilon, \epsilon < N \) and \( \epsilon \) is agreed on by both the sender and receiver. The cover \( I \) is divided into \( 8L_{br} \) blocks. There are \( g = \lfloor L/8L_{br} \rfloor \) pixels in each block and the block set is denoted as \( \text{Block} = \{B_0, B_1, B_2, B_{30}\} \). The ith bit of \( M(b_{in}) \) is embedded into the block \( B_i \). To embed the bit \( b_{in} \) into the block \( B_i \), we set \( d_i = g \times x_0 \) as the embedding location of \( B_i \) and embed \( b_{in} \) into the \( d_i \)th pixel of \( B_i (0 < i < 8L_{br}-1) \). Then we convert pixel’s value from decimal to binary, in the form of \( b_i, b_i, b_i, b_i, b_i, b_i, b_i, b_i \) (0-255). To hide the bit \( b_{in} \), if \( b_i = b_{in} \), we do nothing to \( b_i \), otherwise we replace \( b_i \) with \( b_{in} \).

Figure 1 shows the embedding process and the embedding algorithm is as follows:

---

**Input:** \( T \) is system time, \( VI \) is initial vector, \( 8L \) is the bit length of cover 1 and it is divided into \( 8L_{br} \) blocks, \( M \) is the encrypted form of \( S \), its length is \( 8L \) bits.

**Output:** \( d_0, d_1, d_2, d_{8L_{br}-1} \)

**Algorithm:**

1. \( K = \text{MD5}(T \oplus VI) \)
2. \( M = \{0, 1, 2, \ldots, 14\} \)
3. \( r = (k_0 \oplus k_1 \oplus k_2 \oplus \cdots k_{15}) \mod 16 \)
   - If (\( r = 8 \)) change \( T \) and go to step 1.
   - \( V_L = k_{max}, \mod 16 \)
   - \( V_R = k_{max}, \mod 16 \)
   - \( U_L = k_1 \oplus k_2 \oplus \cdots \oplus k_7 \)
   - \( U_R = k_7 \oplus \cdots \oplus k_1 \)
   - \( x_0 = (V_L \oplus V_R)^{2^4}, P = (U_L \oplus U_R)^{2^4}, N = k_0 \oplus k_1 \oplus k_2 \cdots \oplus k_{15} \)

4. For \( i = 0 \) to \( 8L_{br}-1 \)
   - \( x_{i+1} = \text{Temp}(x_0, p, N) \)
5. \( g = \text{floor}(L/8L_{br}) \)
6. For \( j = 0 \) to \( 8L_{br} \)
   - \( d_j = \text{floor}(g \times x_0) \)
7. For \( i = 0 \) to \( 8L_{br}-1 \)
   - \( b_i = \text{Temp}(x_0, p, N) \)

**Embedding the ith bit of \( n_{in} \) of \( M \) into the jth pixel of \( I \)**

The sender embeds \( M \) into \( I \) and the original \( I \) becomes stego-carrier \( I_m \). The receiver already has \( VI \) and receives \( T \) from the sender. During the extraction phase, he generates \( K \) with the \( T \) and \( VI \). The receiver gets the sub-location sequence \( d_0, d_1, d_2, \ldots, d_{8L_{br}-1} \) to extract the message \( M_{stg} \) which contains \( M \). With the \( K \), he decrypts \( M_{stg} \) and gets the message \( S \) which contains logo \( S \).

Figure 2 shows the extracting process and the extracting algorithm is as follows:

---

**Receive time**

**Fig. 1:** Embedding process

**Fig. 2:** Extracting process
Input: T is received from the sender, VI is initial vector, 8L is the bit length of stego-carrier I_a
Output: S
Algorithm:
Step 1: K = MD5(T) + VI
Step 2: r = (k0|k1|...|k16) mod 16
        If r = 8 change T and go to step 1.
        V_a = k_0|v0|k_1|v1|...|k_16|v_16
        V_b = k_0|v0|k_1|v1|...|k_16|v_16
        V_c = k_0|v0|k_1|v1|...|k_16|v_16
        x_0 = (V_a XOR V_b)^2
        P = (U|a|U_b)^2
        N = k_0|k_1|...|k_16
Step 3: For i = 0 to 8L_a - 1
        x_a = Tent(x_i, p, N)
Step 4: g = floor (L/8L_a)
Step 5: For j = 0 to 8L_p - 1
        d_j = floor (g < x_i)
Step 6: For i = 0 to 8L_p - 1
        log_i = 2log_j + d_j
Step 7: Extract M_a from I_a. (M_a is 8L_a bits)
Step 8: S = Decrypt(M_a)
        if (S_c=S) Extracting Success.
        else recover T

To extract the hidden logo S, the extracting algorithm needs to extract 8L_a bits from I_a and get the message M_a which is I_a bytes. It decrypts M_a to the form of S_c = S|S_p, S_c = s_0|s_1|...|s_{16}. We compare S_c with each logo in the set \Phi. If we find that there is a logo S_c\in\Phi and S_c = S, then we get the correct logo.

Analysis of the algorithm: Both the sender and receiver need to have the same logo set \Phi. The algorithm embeds an uncertain logo that is from the logo set \Phi into the cover I. Because the receiver doesn't know which logo the sender selects, to define a fixed length I_a is necessary. In this study, I_a meets the following requirements: I_a = \max \{ |I| = \text{length}(S_c), S_c \in \Phi \} Adding a new logo to the set \Phi needs to reset I_a. There is a difference between embedding and extracting: The T is generated by the system at the sender side while the receiver receives T instead of generating it. And the usage of skew tent mapping makes sure that the logo bits are uniformly hidden in the cover. With a dynamic T, the hidden locations are different from time to time.

In addition, when T (system time) is damaged or tampered by some attacker, we still have a chance to obtain it. Suppose we lost T, and we have T_0 which is got from the last transmission. We received the M at the real time T_{real}. (T_c(T_{real}, T_{sends}) Here is a integer). We can try each T from either T_0 to T_{real} or T_{real} to T_0. It is certain that we can deduct the correct T which is equal to T_c. And if the searching direction is from T_0 to T_{real} we need to set the time cost to \Delta t_{real}. And if the time interval is from T_{real} to T_0, we need to set the time cost as \Delta t_{real}. In addition, we denote the time cost to locate T_c as \Delta t and \Delta t_{max} (\Delta t_{sends}, \Delta t_{real}). And the searching of T_c is in a linear way, so the time complexity to find T_c is O(\Delta t).

Experiments: Let T_c = 2014030122212, VI = 7154A05BCC 3E552D3E5E21B716A31EF and the images size be 512x512, 512x256 and 256x256. Each image size consists of 456 images. The images we use for experiments are all from http://sipi.usc.edu/database/. Let secret logo be m. Set m = “Hunan University is a good university”. We choose LSB embedding method for the experiments, because it is easy to operate and causes little distortion to the original images.

Figure 3 shows the impacts on 1000 locations where T_i is changing from T_i = 2014030122212 to T_i = 2014030122213. y is the degree of location’s difference between T_c and T_i. d_{i,j} is the ith location generated by the T_c. d_{i,j} is the ith location generated by T_i. y_i is the degree of difference between d_{i,j} and d_{i,j}. y_i = |d_{i,j} - d_{i,j}|/8I. 8I is the bit length of S, i.e. 0, 1, 2..., 8I-1. The average degree of difference from 8I locations is:

\[
\bar{y} = \frac{\sum_{i=1}^{8I} y_i}{8I}
\]

In this experiment, we calculate the \bar{y} = 33.99%.

Fig. 3: Key sensitivity to initial value
DISCUSSION

In this study, we analyzed the feasibility of the algorithm. In the process of embedding, we encrypt logo with a dynamic key. Then embed the encrypted logo into cover in a random way. In this way, the attacker has little chance to find the logo. When in extracting process, we need to extract an encrypted message that contains the encrypted logo. We decrypt the encrypted message and get the message which contains logo. In this way, we can use a set of data hiding algorithms which is based on logo, instead of only one. By conducting experiments on testing how sensitive the key to the chaotic mapping’s parameter, we find that the hiding location sequence’s (d_1, d_2, d_3, …) changing rate is close to 100% when T changes 1 bit. When the time T_e for generating key is damaged, the algorithm can recover the damaged T_e in a linear computation complexity with limited cost. By embedding logos into more than 1300 images of different sizes, measuring by both the MSE and PSNR shows that the algorithm has little impact on the robustness, capacity and invisibility of original one. However, though the proposed algorithm can resolve the risk but there is still an efficient problem. The system time T for K’s generation is not hidden during the message’s transmission, it could be tampered easily. Though there is a way for T’s recovery, it degrades the efficiency of the algorithm to some extent.

CONCLUSION

The data hiding software which relies on some fixed logo to identify themselves can perform data hiding and data extracting efficiently. However, the fixed logo may bring great risk to the data hiding algorithm, because it can be easily caught by an attacker especially when the attacker has knowledge of the algorithm. To solve this problem, this study propose a method to enable the logo to change itself automatically by adding time variable. In theory, each change of the logo is different from each other. And the method hides the logo bit by bit with a random sequence which is generated by the skew tent mapping. The random sequence is used to calculate the bit’s hiding location. In this way, the attacker can hardly extract the logo even though he has the knowledge of the data hiding algorithm. Furthermore, when the time is damaged, there is a remedy for the receiver to generate the key in limit time cost. Though the proposed method resolves the problems brought by the fixed logo, there is still room for improvements. The future study will focus on improving the robustness of the algorithm. Besides, we need to find an efficient way to protect the system time which is exposed during the transmission.
ACKNOWLEDGMENTS

This study was supported by National Natural Science Foundation of China (No. 61202439, 61103215) and the Fundamental Research Funds for the Central Universities of China.

REFERENCES


