Compress-Encrypt Video Steganography

Swetha V
PG Student
Department of Computer Science & Engineering
Jawaharlal College of Engineering and Technology

Liji L Dominic
Assistant Professor
Department of Computer Science & Engineering
Jawaharlal College of Engineering and Technology

Ambikadevi Amma T
Professor
Department of Computer Science & Engineering
Jawaharlal College of Engineering and Technology

Abstract

Preserving the secrecy of confidential information has its own importance right from the early era of computing. Steganography is the art of hiding the messages inside other harmless medium in such a way that an enemy cannot even sense the presence of the secret message. But if the presence of information behind media is revealed or detected, the objective of steganography is violated and security is compromised. This paper proposes and analyzes a methodology for video steganography that processes the text before hiding it behind a cover video. In pre-processing process, the text is first compressed and then modified using a key. The proposed method combines the idea of video steganography, cryptography and compression techniques which provide enhanced security. The RC4 is used for encrypting the compressed text and 4LSB method is used to conceal the processed text inside the cover media, while the purpose of compression is to enhance security and increase embedding capacity.

Keywords: 4LSB, Compress, Encrypt, LZW, RC4, Video Steganography

I. INTRODUCTION

Communication is the lifeblood of any organization and is one of the most important needs of human beings. Internet is changing the way people communicate and think. Internet communication has become an integral part of the Infrastructure of today’s world. Most networks are fundamentally insecure. If the principals on the network are trusted without good cause, the network is inherently insecure even if one has the best infrastructure. Many applications transfer security sensitive data such as credentials used for authentication, or bank transaction details or data such as credit card numbers across networks. To preserve data integrity and confidentiality, the channel between communication end points must be secured. For end-to-end protection, wide spectrum of attacks requires a wide spectrum of technologies. Encryption provides an obvious approach to information security. However, encryption clearly marks a message as containing interesting information, and will attract the attacker and in turn the encrypted message becomes subject to attack.

Data security basically aims at protecting data from unauthorized users or hackers and providing high security. In the recent era, the area of data security has gained more attention due to the massive increase in internet users for the purpose of data transfer. In order to enhance the security features, many techniques such as digital watermarking, cryptography and steganography are developed. Cryptography is the process of scrambling the original text into a seemingly unreadable format for others. Steganography is the process of hiding the one information into other media like text, image or audio file, in such a way that a third person cannot even sense the presence of the hidden message. While cryptography is a method to conceal information by encrypting it to cipher texts using an unknown key and transmitting it to the intended receiver, steganography provides further security by hiding the cipher text into another cover medium. Watermarking is the practice of imperceptibly altering work to embed a secret message. The process of inserting information into a digital signal is called Digital watermarking. The main aim of digital watermarking is to protect the integrity and authenticity of digital media. Digital watermarking directly embeds a watermark containing owner identification into the host signal in such a way that the hacker can’t remove the watermark without reducing the quality of the signal or an image.

To hide secret information in some other source of information without leaving any apparent evidence of data alteration steganographic techniques can be used. All of the traditional steganographic techniques have limited information-hiding capacity approximately 10% or less. This is because the principle of those techniques was either to replace all the least significant bits of a multivalued image with the secret information or to replace a special part of the frequency components of the vessel image. Data containing both the cover signal and the embedded information is known as stego data. Occasionally, especially when referring to image Steganography, the cover image can be called as Vessel or Container. Steganographic technologies are a very important part of the future of Internet security and privacy on open systems such as the Internet.

Security, capacity and robustness are three main aspects of steganography. All these factors are inversely proportional to each other creating steganographic dilemma. The basic difference between steganography and cryptography is that cryptography just scrambles the message to be conveyed but not hide the secret data of communication whereas steganography is used to hide the
cryptographic scramble secret information for secured communication. Cryptography when combined with steganography adds two fold securities in data communication. In this method AVI video is chosen as cover media for hiding the secret message. The aim is to hide a text behind a video file. The LSB are found to be good for hiding more secret information data. In both investigative and security manner, this secured steganography method deals with the idea of cryptography, video steganography and the use of compression techniques. As video is a bunch of images or frames, initially the video file is converted into series of frames, then search for a change of frame in which embedding of the data is done. The information added to the frames is called digital steganography. Individual pixels in an image are taken and the insertion and retrieval depends on their threshold value. The reverse is applied for retrieval.

II. METHODOLOGY

Because of the limitation of the existing system, it is necessary to develop a new system that provides better security and embedding capacity. Preserving the secrecy of confidential has its own utmost importance right from the early era of computing. Steganography is a technique that hides the existence of secret message by masking it behind a cover media. The main purpose of steganography is violated if the presence of secret information behind media is detected. The proposed technique for steganography is named as Compress-Encrypt-Steganography. This method pre-processes the text before hiding it behind a cover image. In the proposed method AVI video is used as the cover medium. Each frame will be an image. In the sender side, the proposed method involves 3 stages namely frame conversion, preprocessing and embedding. In frame conversion phase as video is a bunch of frames, for easy processing the video is converted into component frames. In pre-processing process, the text is first compressed using LZW, which is a dictionary method of compression. The compression is followed by encryption. RC4; a stream cipher method is used for encrypting the compressed text. The next stage is embedding, where data is concealed behind cover image using 4LSB method. Whether data is hidden or not is represented in the first frame of the video. Using 4LSB method, the key for the encryption is hidden in the \(\lfloor n/3 \rfloor\)th frame and the preprocessed text is hidden in the \(\lfloor n/2 \rfloor\)th frame. After embedding process the frames are recombined to form the stego video. The stego video is then passed through the communication channel. In the receiver side, the proposed method involves 3 stages namely frame conversion, undo stage and extraction. On receiving the stego video, initially the stego video is converted into frames in the frame conversion stage. The frames are then forwarded to extraction stage. In the extraction phase, using reverse 4LSB algorithm the key is extracted from the \(\lfloor n/3 \rfloor\)th frame and the preprocessed text from the \(\lfloor n/2 \rfloor\)th frame. The key and the preprocessed text is the moved to undo stage. In undo stage the preprocessed text is first decrypted using the key and decompressed. The output of the undo stage will be the secret message. The sender and receiver side algorithm for the system is as follows:

Transmitting Side Algorithm:
1) Select cover video and split it into frames.
2) Select cover frame from cover video.
3) Enter the secret message.
4) Encrypt and compress the secret message.
5) Hide the key inside the \(\lfloor n/3 \rfloor\)th cover frame using 4LSB algorithm.
6) Hide pre-processed secret message inside the \(\lfloor n/2 \rfloor\)th cover frame using 4LSB algorithm to get stego frame.
7) Replace original frames chosen for hiding secret message by those stego frames.
8) Recombine frames to form Stego video and transfer it using any communication channel.

Receiver Side Algorithm:
1) Get the stego video and split it into number of frames.
2) Select stego frame along with adjacent frames.
3) Extract key and preprocessed data from corresponding stego frame by the use of reverse 4LSB algorithm.
4) Perform decryption and decompression on the extracted data.
5) Display secret message.
6) Exit.

The important algorithms or key steps included in the proposed system are explained below:

A. Compression using LZW Method:

Compression is the art or science of representing information in a more compact form. There are many well-known methods for data compression. Based on different types of data, each method use different ideas and produce different output, they are all based on the same principle i.e.; removing redundancy from the original data in the source file. Here the text message is compressed using LZW (Lempel–Ziv–Welch) compression algorithm. Lempel–Ziv–Welch (LZW) is a universal lossless dictionary based data compression algorithm. Dictionary-based algorithms encode variable-length strings of symbols as single tokens instead of encoding single symbols as variable-length bit strings. In LZW variable-length strings of symbols or characters that commonly occur together are represented using fixed-length code words. While receiving the data, the LZW encoder and decoder dynamically build up the same dictionary. As scanning progress LZW places longer and longer repeated entries into a
dictionary. If the element has already been placed in the dictionary, then the code for an element is emitted, rather than the string itself.

**Algorithm 1. LZW Compression Algorithm**

```plaintext
BEGIN
s = next input character;
While not EOF
{
c = next input character;
if (s+c) exists in the dictionary
{
s = s+c;
}
else
{
output code for s
add string (s+c) to the dictionary with a new code;
s = c;
}
Output the code for s;
END
```

**Algorithm 2. LZW Decompression Algorithm**

**B. Encryption using RC4 Algorithm:**

Encryption is the process scrambling the original information in order to conceal its meaning which will prevent any unauthorized access. Hence, the goal of encryption is to ensure secrecy. The symmetric key algorithm, RC4 is a stream cipher designed in 1987 by Ron Rivest for RSA Security. Both encryption and decryption uses same algorithm. In RC4, data stream is simply XORed with the generated key sequence. The key stream that is completely independent of the plaintext used uses a variable length key from 1 to 256 bit to initialize a 256-bit state table. The pseudo-random bits and pseudo-random stream which is XORed with the plaintext to give the cipher text are generated subsequently using the state table. The algorithm can be broken into two phase: initialization, and stream generation. In the initialization stage using the key, K as a seed the 256-bit state table, S is populated. When the state table is once setup, as data is encrypted, it continues to be modified in a regular pattern. The initialization process can be summarized by the pseudo-code:

```plaintext
j = 0;
for i = 0 to 255:
S[i] = i;
for i = 0 to 255:
j = (j + S[i] + K[i]) mod 256;
swap S[i] and S[j];
```

Once the initialization process is completed, the algorithm enters into the stream generation phase. The process may be summarized as shown by the pseudo code below:

```plaintext
i = j = 0;
```
for \((k = 0\) to \(N-1\))
{
    i = (i + 1) \mod 256;
    j = (j + S[i]) \mod 256;
    \text{swap } S[i] \text{ and } S[j];
    pr = S[ (S[i] + S[j]) \mod 256 ]
    \text{output } M[k] \text{ XOR } pr
}
Where \(M[0..N-1]\) is the input message consisting of \(N\) bits.

**C. Data Hiding using 4LSB Algorithm:**

The idea of the LSB algorithm is to insert the bits of the hidden message into the least significant bits of pixels. LSB (Least Significant Bit) substitution is the process of adjusting the least significant bit pixels of the carrier image. It is a simple approach for embedding message into the image. The Least Significant Bit insertion varies according to number of bits in an image. Video is a sequence of images displayed at faster rates taking the advantage of human vision system. An extremely simple steganographic method is to hide the information at pixel level. Each frame or image is made up of no. of individual pixels. Each of these pixels in an image is made up of a string of bits the 4 least significant bit of 8-bit true color image is used to hold 4-bit of our secret message image by simply overwriting the data that was already there. By experimentation, it has been proved that the impact of changing the 4 least significant bits is almost imperceptible. In hiding process, the last 4 bits of image or frame pixel is replaced with 4 bits of our secret data. For this secret data which is also sequence of bytes are broken down into set of 4 bits. To hide each character of secret message we need two pixels. So the number of characters that we can hide in \((mxm)\) image is given by the following equation.

\[
\text{Total size of one frame} \div 8 = \text{-----------------------------} \quad (1)
\]

Suppose size of a single frame is 160KB, then for 1LSB, maximum data that can be hidden is \(1\times20KB = 20KB\). For 2LSB it is \(2\times20KB = 40KB\). For 3LSB it is \(3\times20=60KB\). For 4LSB it is \(4\times20KB =80KB\). If steganographic process go beyond 4LSB, i.e. for 5LSB it is \(5\times20KB=100 \text{ KB}\), means that size of the data can be hide is more than 50%, hence it is look like visible watermarking. For implementing steganography proposed method is using 4LSB algorithm. Any data change in least significant bit does not change the value of data significantly. For example in 8 bit data highest value is 1111 1111. A change in LSB bit from 1 to 0 will only change the value from 255 to 254 which mean it changes 1/256 times so this change is negligible and unrecognizable to human eye. In similar way as number of bits goes on increasing change in 4LSB do not change pattern significantly.

**III. Result and Discussion**

The proposed system that combines the idea of steganography, cryptography and compression technique provides three times secure system compare to pure steganographic system. Cascading multiple algorithms will increase time complexity but in turn the system provides enhanced security and embedding capacity.

**IV. Conclusion**

Steganography is a technique that screens the existence of information by masking it behind some other seemingly innocuous media. But if the existence of information behind media is detected, the main purpose of steganography is violated. In order to improve the security of steganography, the proposed technique processes the information before masking it behind cover image. Compression involved reduces the size of information and modifying information using key changes its structure. Thus mere extraction of the information from the video does not provide it in its original form. Encryption algorithm improves the lowest matching performance between the image and the secret information by changing the statistical characteristics of the secret information to enhance the anti-detection of the image steganography. By combining the concepts of steganography, cryptography and compression technique time complexity of the system increases but in turn triple security is gained. The use of the video based Steganography method is more eligible than other multimedia files, because it provides high embedding capacity. Video is bunch of many frames and data is hidden inside specified frame which enhances security as normal observer cannot detect inside certain frames played at high rate by video.

**References**


http://www.forevid.org/

http://www.steganosaur.us/