EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 3
Issue: 6
570-574
Year: 2016

VIDEO STEGANOGRAPHY USING DNA SEQUENCE

Suman Chakraborty and *Prof. Samir Kumar Bandyopadhyay

Department of Computer Science and Engineering University of Calcutta, India.

***Corresponding Author: Prof. Samir Kumar Bandyopadhyay**

Department of Computer Science and Engineering University of Calcutta, India.

Article Received on 21/04/2016 **Article Revised on 12/05/2016** **Article Accepted on 31/05/2016**

ABSTRACT

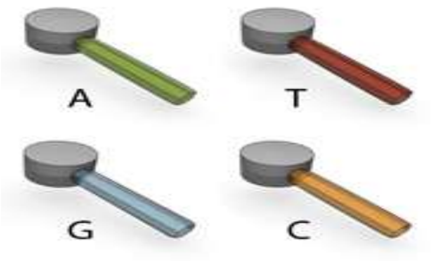
Cost of an information directly proportionate with transmission cost. Transmission cost depends upon size of data. In case of video, required more data to transmit than non video information, that's why video transmission costly. In this paper proposed a method which can use small size of video as cover media to hide secret information by compressing secret information. Compression done by DNA sequence image compression procedure. Due to use of comparatively small size of video, makes this procedure economically viable and faster secret transmission.

KEYWORDS: HLSB, DNA sequence, Nucleotide, Compression, Steganography.

2. INTRODUCTION

Steganography is one of the popular and tested procedure for secret data transmission .Beauty of Steganography that unauthorized user can't realize the present of secret information. Root of this technique comes from Greek culture. In Greek, Steganography means cover writing. In video Steganography secret information embedded with in a video file and size of secret information very less compare with cover file size. This difference of size makes this technique more secure.

All of the information needed to build and maintain an organism — whether it's a human, a dog, or a bacterial cell — is contained in its DNA. DNA molecules are composed of four nucleotides, and these nucleotides are linked together much like the words in a sentence. Together, all of the DNA "sentences" within a cell contain the instructions for building the proteins and other molecules that the cell needs to carry out its daily work.



Determining the order of the nucleotides within a gene is known as **DNA sequencing**. The earliest DNA sequencing methods were time consuming, but a major breakthrough came in 1975 with the development of the process called **Sanger sequencing**. Sanger sequencing is named after English biochemist Frederick Sanger, and it is sometimes also referred to as **chain-termination sequencing** or **dideoxy sequencing**. Some 25 years after its creation, the Sanger method was used to sequence the human genome, and, with the addition of many technological improvements and modifications, it remains an important method in laboratories across the world today.

The Sanger method relies upon a variation of the replication process in order to determine the sequence of nucleotides in a segment of DNA. Before Sanger sequencing can begin, however, researchers must first make many copies of, or **amplify**, the DNA segment they wish to sequence. This is done either by cloning the DNA or by triggering the polymerase chain reaction (PCR). Once the DNA has been amplified, it is heated so that the two strands separate, and a synthetic primer is added to the mixture. The primer's sequence is complementary to the first piece of target DNA, which means that the primer and the DNA target bind with each other. At this point, the target sequence is exposed to a solution that contains DNA polymerase and all of the nucleotides required for synthesis of the complementary DNA strand — along with one special ingredient.

As described above, the next major step in the Sanger process is to expose the target sequence to DNA polymerase and significant amounts of all four nucleotides. In their unbound form, nucleotides have three phosphate groups and are formally called **deoxynucleotide triphosphates**, **dNTPs** (where the "N" is a placeholder for A, T, G, or C). During the construction of a new DNA strand, a molecule called a hydroxyl group (which contains an oxygen atom and a hydrogen atom) attaches to the sugar of the last dNTP in the strand and chemically binds to the phosphate group on the next dNTP. This binding causes the DNA chain to grow. In Sanger sequencing, however, a special type of "dummy" nucleotide is included with the regular dNTPs that surround the growing DNA strand. These special nucleotides are known as **dideoxynucleotide triphosphates**, or **ddNTPs** (Figure 2), and they lack the crucial hydroxyl group that is attached to the sugar of dNTPs. Therefore, whenever a ddNTP is added to a growing DNA strand, it is unable to chemically bind with the next nucleotide in the chain, and the DNA strand stops growing.[1]

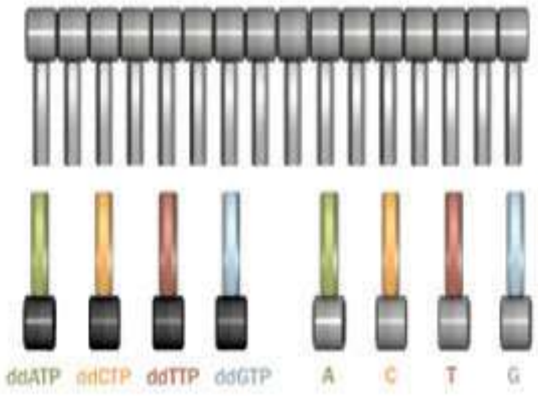


Figure 2: The four ddNTPs

In video Steganography a video file as a cover media can contain secret information around 5-10% of it's size. It's very low rate of embedding. Also increase the cost of secret transmission .In this paper we overcome this problem by compress the size of secret image around 99% using DNA sequence compression technique.

The paper is organized as follows. Introduction of the proposed technique in section **2**, Related work of the proposed technique is explained in section **3**. The methodology of the proposed technique is described in section **4**. Section **5** concludes the paper.

3. Related Works

3.1 Image Compression using DNA sequence

3.1.1 Methods For DNA Sequence Compression

A massage representing a DNA sequence, with the combination of a, c, t, g is equivalent to DNA sequence of the original data. Compressing the DNA sequence by- 1) 2-bits encoding method, 2) Exact matching method, 3) Approximate matching method, 4) For the approximate matching method. These compression techniques would be produce equivalent digital form of the DNA sequence and one of procedure will produce minimum number of bits.[2-3]

We consider three standard edit operations in our approximate matching algorithm. These are:

1) *Replace*. This operation is expressed as (*R*, *p*, *char*) which means replacing the character at position *p* by character *char*.
2) *Insert*. This operation is expressed as (*I*, *p*, *char*), meaning inserting character *char* at *p*.
3) *Delete*. This operation is written as (*D*, *p*), meaning deleting the character at position *p*.

Let *C* denote "copy," then the following are two ways to convert the string "gaccttca" to "gaccgtca" via different edit.

C C C C R C C C
g a c c g t c a
g a c c t t c a
or
C C C C I C D C C
g a c c g t c a
g a c c t t c a

The first involves one replacement operation.The second involves one insertion and one deletion. It can be easily seen that there are infinitely many edit sequences to transform one string to another A list of edit operations that transform a string *v* to another string *u* is called an *Edit Transcription* of the two strings.[2]This will be represented by an edit operation sequence λ(*u*,*v*) that orderly lists the edit operations. For example, the edit operation sequence of the first edit transcription in the above example is λ(*gaccgtca*, *gaccttca*) = {(*R*,4, *g*)} ; and for the second edit transcription, λ(*gaccgtca*, *gaccttca*) = {(*I*,4,*g*),(*D*,6)}.If we know the string *u* and an edit operation sequence λ(*u*,*v*) from *v* to *u*, then the string *u* can be constructed correctly using λ.There are many ways to encode one string given another. Using the above example, we describe four ways to encode "gaccgtca" using string "gaccttca" supposing that the string "gaccttca" is located earlier in the sequence.

1) 2-bits encoding method. In this case,we can simply use 2 bits to encode each character; i.e., 00 for *a*, 01 for *c*, 10 for *g*, 11 for *t*. Thus "10 00 01 01 10 11 01 00" encodes "gaccgtca." It needs 16 bits in total.

2) Exact matching method.Wecan use (repeat position, repeat length) to represent an exact repeat. This way, for example,if we use 3 bits to encode an integer, 2 bits to encode a character, and use 1 bit to indicate if the next part is a pair (indicating an exact repeat) or a plain character, then the string "gaccgtca" can be encoded as {(0,4), *g*,(5,3)}, relative to "gaccttca." Thus, a 17-bit binary string "0 000 100 1 10 0 101 011" is required to encode the {(0,4), *g*,(5,3)}.

3) Approximate matching method. In this case, the string "gaccgtca" can be encoded as {(0,8),(*R*,4, *g*)}, or "0 000 111 100 100 10" in binary, with *R* encoded by 00, *I* encoded by 01, and *D* encoded by 11, and 0/1 indicating whether the next item is a doubleton or triple. A total of 15 bits is needed. 4) For the approximate matching method, if we use the edit operation sequence, then the

string "gaccgtca" can be encoded as {(0,8),(*I*,4, *g*),(*D*,6)}, or "0 000 111 1 01 100 10 1 10 110," in total 21 bits.

3.1.2 Method-I for Image Compression

Step1. Take DNA sequence of an animal whose image will compress.
Step2. This DNA sequence assign as a DNA sequence of image.
Step3. Four nucleotides in DNA sequence (A,C,G,T),2 bit is used to represent 4 nucleotides.

Like 00 for A,01 for C,10 for G and 11 for T.

Step4. Represent DNA sequence in digital form D.
Step5. D is the compress binary form of that image.[4]

3.1.3 Method II for Image Compression

Step-1. Take DNA sequence of an animal whose image will be compressed
Step-2. This DNA sequence assign as a DNA sequence of image
Step-3. Compress DNA sequence using -1) 2-bits encoding method, 2) Exact matching
method, 3) Approximate matching method, 4) For the approximate matching method, one of the method will produce minimum no. of bits.
Strp-4. Compress DNA sequence is the compress digital[4]

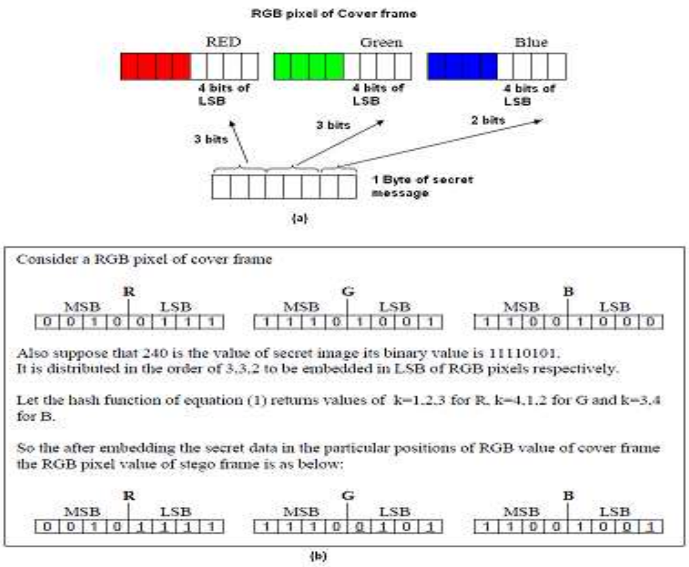
3.2 Hash based Least Significant Bit (HLSB) technique for Video Steganography

A video stream (AVI) consists of collection of frames and the secret data is embedded in these frames as payload. The information of the cover video (AVI) such as number of frames (n), frame speed (fp/sec), frame height (H) and width (W) are extracted from the header. The cover video is then broken down into frames. Now the proposed LSB based technique has been applied to conceal the data in the carrier frames. The size of the message does not matter in video steganography as the message can be embedded in multiple frames.

This technique takes eight bits of secret data at a time and conceal them in LSB of RGB (Red, Green and Blue) pixel value of the carrier frames in 3, 3, 2 order respectively. Such that out of eight (08) bits of message six (06) bits are inserted in R and G pixel and remaining two (02) bits are inserted in B pixel. The detailed technique has been depicted in Figure 1. This distribution pattern is taken because the chromatic influence of blue to the human eye is more that red and green pixel. Thus the quality of the video is not sacrificed but we could increase the payload. Also this small variation in colours in the video image would be very difficult for the human eye to detect .The embedding positions of the eight bits out of the four (4) available bits of LSB is obtained using a hash function of the form,

$$k = p\%n \text{ --------------------------(1)}$$

where, k is LSB bit position within the pixel, p represents the position of each hidden image pixel and n is number of bits of LSB which is 4 for the present case. The bits are distributed randomly during fabrication which increases the robustness of the technique compared to other LSB based techniques.[5,6] After concealing data in multiple frames of the carrier video, frames are then grouped together to form a stego video, which is now an embedded video to be, used as normal sequence of streaming. The intended user follows the reverse steps to decode the secret data. During decoding the setgo video is again broken into frames after reading the header information. Using the same hash function which is known to the intended user, the data of the secret message is regenerated. The extracted stream of the secret information is used to authenticate the video.[7]



Consider a RGB pixel of cover frame

Also suppose that 240 is the value of secret image its binary value is 11110101.
It is distributed in the order of 3,3,2 to be embedded in LSB of RGB pixels respectively.

Let the hash function of equation (1) returns values of k=1,2,3 for R, k=4,1,2 for G and k=3,4 for B.

So the after embedding the secret data in the particular positions of RGB value of cover frame the RGB pixel value of stego frame is as below:

Figure 1: Proposed hash based LSB embedding technique (a) shows secret data embedded in 4 bits of LSB in 3,3,2 order in corresponding RGB pixels of carrier frame and (b) example of embedding of bits using.

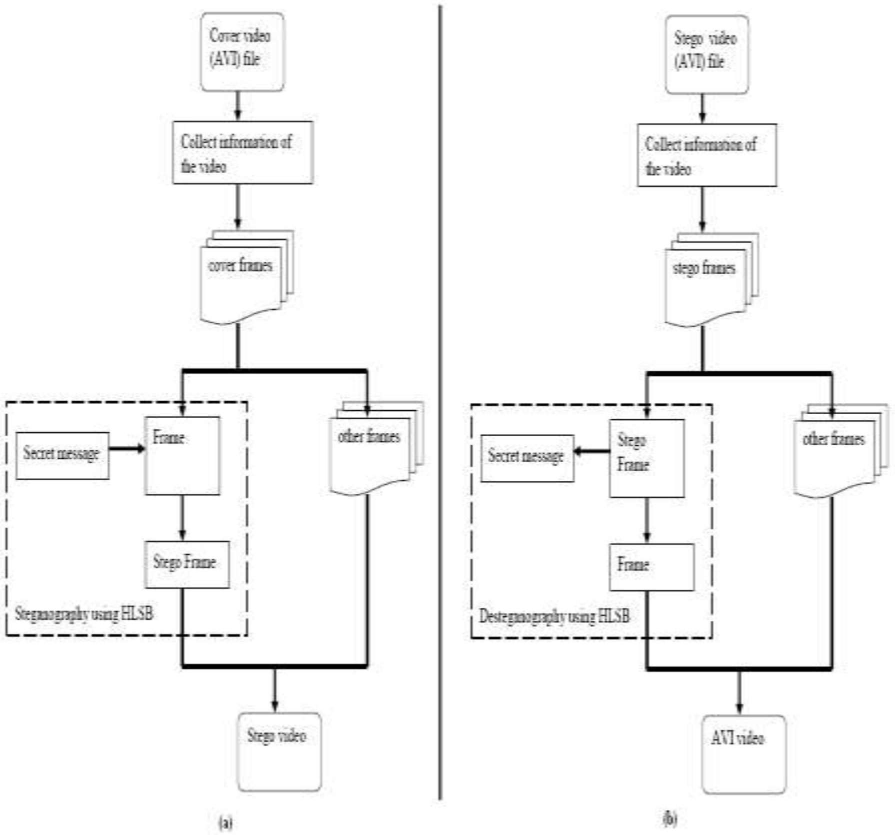


Figure 2: Block diagram of HLSB Video Steganography technique (a) Encoding and (b) Decoding[7]

Algorithm of Encoding[7]

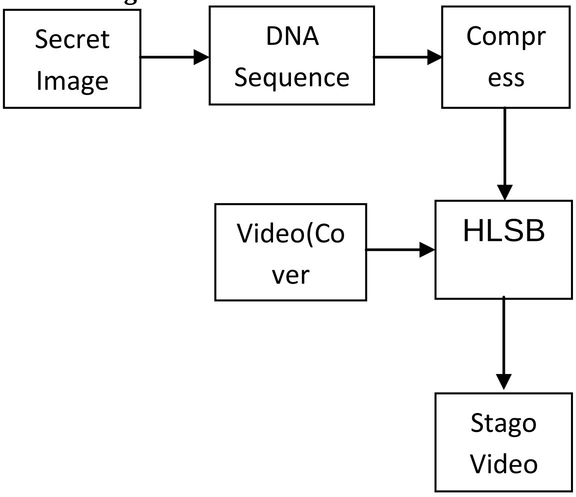
Step 1: Input cover video file or stream.
Step 2: Read required information of the cover video.
Step 3: Break the video into frames.
Step 4: Find 4 LSB bits of each RGB pixels of the cover frame.
Step 5: Obtain the position for embedding the secret data using hash function given in equation 1.
Step 6: Embed the eight bits of the secret image into 4 bits of LSB of RGB pixels of the cover frame in the order of 3,3,2 respectively using the position obtained from step 5.
Step 7: Regenerate video frames.

Algorithm of Decoding[7]

Step 1: Input stego video file or stream.
Step 2: Read required information from the stego video.
Step 3: Break the video into frames.
Step 4: Find 4 LSB bits of each RGB pixels of the stego frame.
Step 5: Obtain the position of embedded bits of the secret data using hash function given in equation 1.
Step 6: Retrieve the bits using these positions in the order of 3, 3, 2 respectively.
Step 7: Reconstruct the secret information.

4. Proposed Methodology

4.1 Flow Diagram



4.2 Encoding Algorithm

Step-1: Compress the Secret Image by DNA sequence image compression method (Method –I or Method-II).
Step-2: Embed the secret information in Cover Video by HLSB procedure.
Step-3: Transmit Stago video.

5. CONCLUSION

In the proposed method we can use small size of video as cover media to hide secret information by compressing secret information.

REFERENCES

1. http://www.nature.com/scitable/topicpage/the-order-of-nucleotides-in-a-gene-6525806
2. M. Crochemore, W. Rytter, Jewels of Stringology, World Scientific, 2002.
3. European Bioinformatics Institute, <http://www.ebi.ac.uk/>.
4. Prof. Samir Kumar Bandyopadhyay and S Chakraborty, “Image Compression using DNA sequence” , International Journal of Computer Science & Engineering Technology, December 2011; 1(11).
5. Mritha Ramalingam, Stego Machine Video Steganography using Modified LSB Algorithm, in World Academy of Science, Engineering and Technology, 2011; 74: 502-505.
6. Juan Jose Roque and Jesus Maria Minguet, SLSB: Improving the Steganographic Algorithm LSB, in the 7th International Workshop on Security in Information Systems (WOSIS 2009), Milan, Italy, pp.1-11, 2009.
7. Kousik Dasgupta, J.K. Mandal and Paramartha Dutta, “Hash Based Least Significant Bit Technique For Video Steganography (HLSB), “International Journal of Security, Privacy and Trust Management (IJSPTM), April 2012; 1(2).