Significance of Steganography on Data Security

Venkatraman.S[*], Ajith Abraham[+], Marcin Paprzycki[+]
[*]*Dept. of Computer Science & Engineering, University of Madras, INDIA*
[+]*Dept. of Computer Science, Oklahoma State University, USA*
svraman83@hotmail.com ,{ aa, marcin }@cs.okstate.edu

Abstract

With the ever increasing amount and variety of data to be stored and transmitted in various mediums, the specification of security which has to be established at various levels of medium access and the accompanying issues of authentication and authorization has become a critical factor. Various steganographic, watermarking and data-embedding algorithms have usually manipulated the actual data in order to either hide any coveted information or to provide some level of access control over the medium. The mediums are usually images, video, audio etc., wherein specific portions or the overall space is usually 'corrupted' with 'significant' data. This paper is an attempt to bring out the significance of the steganographic techniques that are employed in information processing algorithms for data security. It deals with the problem of data security, focusing mainly on images, and tries to state the various properties and characteristics that the steganographic algorithms should possess. The paper also highlights the technique of masking used in the conventional steganographic LSB algorithms and in its variants.

1. Introduction

The growing use of the Internet has led to a continuous increase in the amount of data that is being exchanged and storage in various digital media. This has led to some unexpected cases involving both benevolent and malevolent usage of digital data. Security and authentication techniques like digital watermarks; steganographic methods and other data embedding algorithms have contributed much to enhance the various security features and to preserve the intellectual property. In this respect, steganographic techniques have been the most successful in supporting hiding of critical information in ways that prevent the detection of hidden messages [3]. While cryptography scrambles the message so that it cannot be understood, steganography hides the data so that it cannot be observed. Different types of steganographic techniques employ color composition, luminance, unusual sorting of color palettes, exaggerated noise, relationship between color indices etc. The framework for steganography can be given in terms of the prisoners' problem [2]. The main objectives of the security or steganographic algorithms should be such as to provide confidentiality, data integrity and authentication [1]. Applications for such a data-hiding scheme include in-band captioning, covert communication, image tamper proofing, authentication, embedded control, and revision tracking [16]. As data security is proving to be one of the foremost concerns of any system administrator, let it be a LAN or across the Internet, any distribution system must provide [1]

- Secure content distribution
- Secure Access Key Distribution
- Authentication of Source and sink consumer devices
- Renewability of content protection system

The rest of the paper is arranged as follows: Section 2 deals with the basic requirements and characteristics of the data embedding algorithms; Section 3 concerns the basic techniques used in steganography. Section 4 briefs on the measures used in data embedding algorithms. Section 5 summarizes and concludes the paper.

2. Requirements

Most steganographic techniques proceed in such a way that the data which has to be hidden inside an image or any other medium like audio, video etc, is broken down into smaller pieces and they are inserted into appropriate locations in the medium in order to hide them. The aim is to make them unperceivable and to leave no doubts in minds of the hackers who 'step into' media-files to uncover 'useful' information from them. To achieve this goal the critical data has to be hidden in such a way that there is no major difference between the original image and the 'corrupted' image. Only the authorized person knows about the presence

Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04)
0-7695-2108-8/04 $ 20.00 © 2004 IEEE

of data. The algorithms can make use of the various properties of the image to embed the data without causing easily detectable changes in them. Such methods include: noise insertions, manipulation of image properties like luminance, chrominance, etc. Many steganographic techniques cause changes in pixel relations through unusual sorting of color palettes, exaggerated noise or difference in relationships between color in color indexes.

Data embedding or water marking algorithms necessarily have to guarantee that

- Presence of embedded data is not visible;
- Ordinary users of the document are not affected by the watermark that is the normal user does not see any ambiguity in the clarity of the document;
- The watermark can be made visible by the creator (and possibly the authorized recipients) when needed; this implies that only the creator has the mechanism to break the data embedded inside the document.
- The watermark is difficult for the other eavesdropper to comprehend and to extract them from the channels

2.1 Perceptual Transparency

One of the most important considerations while designing any algorithm that is used for data hiding is that it should perform its operation without raising any suspicion of the eavesdropper. Most steganographic techniques or data embedding techniques implicitly employ limitation of the Human Auditory System (HAS) or Human Visual System (HVS) to embed data. Some advanced perceptual models can also be used to determine the best way to embed data in order to conceal its identity [8].

The noise or any modulation induced by the originator should not change the characteristics of the cover image and should not produce any kind of distortions. The perceptual transparency signifies this technique that should not be sacrificed. The technique fails if the embedding algorithm arouses curiosity or suspicion in the minds of the attacker. Also in some cases like copyright protection using watermarks for protecting intellectual property, it is necessary that the integrity of the original work may be maintained so that they can be extracted out from the medium when the situation warrants [11]. Applications that don't are not too critical on the technique used or the perceptual transparency might increase the information content by increasing the amount of noise or causing geometrical changes in the cover.

2.2 Information Capacity

The amount of information that can be embedded into a medium without modifying the medium also characterizes the robustness of the technique. Steganographic capacity is the size of information that can be hidden relative to the size of the cover image. The hidden information and the cover image should withstand any kind of transformations, such as rotation, blurring, denoising, adding noise, sharpening, scaling and other linear and non-linear filtering techniques.

2.3 Tamper Proof

Tamper proofing is used to indicate that the host signal has been modified from its authored state. Modification to the embedded data indicates that the host signal has been changed in some way. Even though the medium is not restricted in steganography, but mechanisms should be provided to detect the possible 'corruption' of the medium. This property assumes significance in watermarking and copyright protection schemes, where the copyright has to be effective even after modifying.

One of the main goals of data embedding or watermarking algorithms is to ensure that the embedded data remains uncorrupted and also recoverable; its goal is not to restrict or regulate access to the host signal. A class of processes is always used in conjunction instead of a single process to achieve all possible goals. No single method is capable of achieving the desired properties of an undetectable data-embedding scheme without sacrificing some amount of bandwidth. There is a tradeoff between the amount of embedded data and the degree of immunity to host signal modification. As discussed in [6], it is not possible to achieve the twin goals of an embedded data rate and a high resistance to modification, by constraining the degree of host signal degradation. However bandwidth can be traded for robustness by exploiting redundancy. In [7],[12],[13],[14],[15] some of the most important steganographic tools that are in use are discussed.

3. Techniques

Given the proliferation of digital images, and given the high degree of redundancy present in a digital representation of an image (despite compression), there has been an increased interest in using digital images as cover-objects for the purpose of steganography. The

simplest of such techniques essentially embed the message in a subset of the LSB (least significant bit) plane of the image, possibly after encryption. In this section the focus is on LSB embedding in digital images.

When dealing with steganography in images it is important to choose an image carrier size and palette carefully since manipulation is more evident in small or well-known images. Based on the same premise, palettes with drastic changes in color are also unsuitable. It is recommended to use grey-scaled palettes, since there is no drastic change between shades. It has to be noted, that one of the more important weaknesses of the LSB is that it is vulnerable to lossy compression i.e. transforming an image to JPEG. However, as long as that compression is lossless, the medium maintains its state and there are no transformations in its behavior.

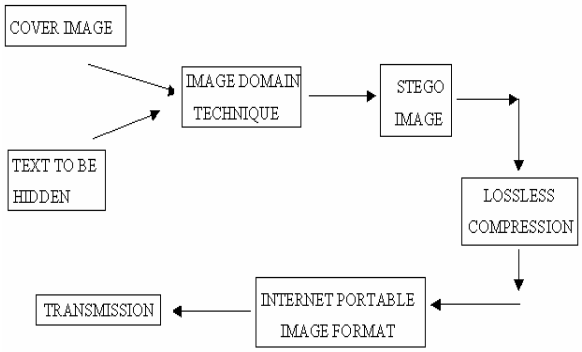

Fig 1: Basic flowchart of Steganographic text embedding.

The techniques for hiding the text behind digital images are broadly classified into two categories: (1) Image Domain Techniques - are entirely dependent upon the image's format (i.e. the way the pixels are arranged inside an image representation). Since pixels are represented by bits, bit manipulation is performed to 'invisibly' modify the color value of certain pixels. As a result, to the human eye the new image looks like the exact replica of the original image. Image domain techniques are generally applied to lossless formats. (2) Transform or Frequency Domain Techniques - are independent on image formats and thus can be applied to lossy formats as well. They involve algorithms and tools that manipulate the image by applying transforms such as DCTs and Wavelet Transformations. They hide messages in more significant areas of the cover image and may manipulate image properties such as their luminance. Hence in these techniques we observe a trade-off between the amount of data to be hidden and the robustness of the image.

3.1 LSB Coding

Least Significant Bit coding is one of the simplest methods for inserting data into digital signals in noise free environments. Probability of changing an LSB in one pixel is not going to affect the probability of changing the LSB of the adjacent or any other pixel in the image.

To a computer, an image is an array of numbers that represent light at various points (pixels). These pixels make up the image's Raster Data. For instance, an image of size 640*480 pixels and 256 colors (8 bit/pixel) contains up to approximately 300 KB of data. The message to be hidden should be compressed before being embedded so that a larger amount of information can be hidden. To hide the image in the LSB's of each byte of a 24-bit image, we can store 3 bits in each pixel. A 1024*768 image has the potential to hide a total of 2,359,296 bits (294,912 bytes) of information. To the human eye, the resulting stego-image will look identical o the cover image.

Pseudo Code for LSB Insertion Algorithm

```
{
buffer: = buffer containing the pixel info(320*200);
n: = number of characters in the file to be encoded;
for I: = 1 to n
  begin
   char: = getnextChar();
   bit_in_char: = char AND ox01;
   pixel: = getNextPixelFromBuffer;
   If (bit_in_char == 0)    //access pixels sequentially
      pixel: = pixel AND oxfe;        //inserting 0 in the LSB
  else
      pixel: = pixel OR ox01;         //inserting 1 in the LSB
  putPixelBackIntoBuffer;
  char: = char>>1;                    //shift right 'char' by 1 bit
  end //buffer contains the hidden message(new pixel info)
end for
 }
```

One of the disadvantages of the LSB Coding methods is that the binary sequences require exact preservation of the signal for the successful extraction of the hidden message. Hence they should be used in contexts that do not require more sophisticated approaches. Noisy Transmission, filtering, cropping, color space conversion or resampling could destroy the hidden message. Also they are susceptible to lossy compression that will cause their original information to be lost.

3.2 Random Pixel Manipulation Technique

In the LSB technique, the information is hidden in sequential fashion. Hence the risk of information being uncovered is relatively high as such approach is susceptible to all 'sequential scanning' based techniques. The Random Pixel Manipulation Technique attempts at overcoming this problem, where pixels are chosen in a random fashion instead of a sequential one.

In this technique, a stego-key is chosen. A stego-key is nothing but a string, which can be effectively manipulated to obtain a random number sequence. The stego-key provides a seed value, which is an integer that helps us to generate a repeated sequence of unique pseudorandom numbers ranging from 0 to N; where N is the number of pixels available. This sequence is then used to 'scramble' the hidden data. At the receiving end the stego-key is used to uncover the data (it plays the role of a password). It provides the same seed value and consequently the same sequence of unique random numbers as generated in the sender's side. Thus the embedded data that is distributed randomly throughout the image is recovered bit by bit, packed and regrouped to fully regenerate the hidden original data. Thus random pixel manipulation technique can be utilized to add additional level of trust to the robust implementation of the LSB based steganography. The sequence of events is flowcharted in Figure 3.

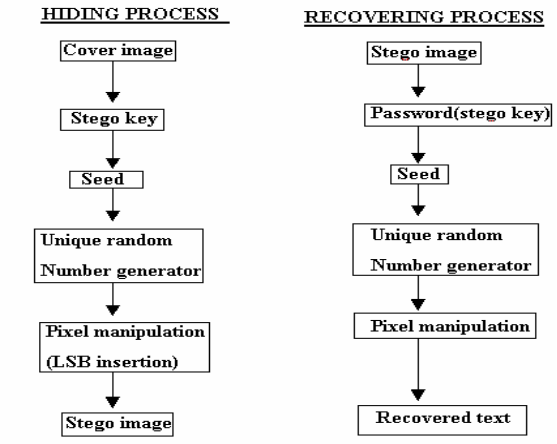

Fig 3: Flowchart of Random Pixel Manipulation Technique

3.3 Masking

The masking properties of the human visual system allow perceptually significant embedding to be unnoticed by an observer under normal viewing conditions [11]. "Masking" refers to the phenomenon where a signal can be imperceptible to an observer in the presence of another signal (referred to as the masker.) Masking systems perform analysis of the image and use the information about the capabilities of the "observer" to determine appropriate regions to place the message data. Masking systems can also use the analysis to vary the strength (amplitude) of the embedded data based upon local image characteristics to maximize robustness. These systems can embed in either the spatial or a transform domain. Based on the local document characteristics the robustness of the masking system can be increased.

4. Measures

Security, embedding distortion and embedding rate can be used as schemes to evaluate the performance of the data hiding schemes.

4.1. Entropy

A steganographic system is perfectly secure when the statistics of the cover data and the stego data are identical, which means that the relative entropy between the cover data and the stego-data is zero. Entropy considers the information to be modeled as a probabilistic process that can be measured in a manner that agrees with intuition [10].The information theory approach to steganography holds the systems' capacity to be modeled as the ability to transfer information. More information regarding information theory and its application to steganography can be found at [10].

4.2. Mean Squared Error & SNR

The (weighted) mean squared error between the cover image and the stego-image (embedding distortion) can be used as one of the measures to assess the relative perceptibility of the embedded text. Imperceptibility takes advantage of human psycho visual redundancy, which is very difficult to quantify. Mean square error (MSE) and Peak Signal to Noise Ratio (PSNR) can also be used as metrics to measure the degree of imperceptibility:

$$MSE= [\ \Sigma^{M}_{i=1}\ \Sigma^{N}_{j=1}\ (\ f_{ij} - g_{ij})^2\ ]MN$$
$$PSNR = 10\ \log_{10}(L^2/MSE),$$

where M and N are the number of rows and number of columns respectively of the cover image, f_{ij} is the pixel value from the cover image , g_{ij} is the pixel value from the stego-image, and L is the peak signal value of the cover image (for 8-bit images, L=255). Signal to noise ratio quantifies the imperceptibility, by regarding the message as the signal and the message as the noise.

Thus, the higher the SNR, the more perceptible is the message.

$$SNR = \sigma^2_S / \sigma^2_N$$

4.3. Correlation

Correlation is one of the best known methods that evaluate the degree of closeness between two functions. This measure can be used to determine the extent to which the original image and the stego-image are close to each other, even after embedding data. Localization, that is detection of the presence of the hidden data relies on the use of cross correlation function R_{XY} of two images X and Y, defined as[8],

$$R_{XY}(\alpha,\beta) = \Sigma_i \Sigma_j X(i,y)\, Y(i-\alpha, j-\beta)$$

4.4. Ensuring Integrity-using Checksums

In order to ensure the integrity of data and the cover medium, mechanisms should be employed that either detect that the medium has been altered or is able to withstand such changes and corrects them to the original state. Checksums could be used to alert the user of possible contamination or tampering. For monochrome images the application of checksums is going to straightforward with the checksums being calculated for the appropriate number of bits required to represent each of the pixels. For color images, the checksum scheme can be extended three times to the three-color planes. The checksum could also be calculated in a new coordinate system, for e.g., hue-saturation-intensity plane instead of RGB plane, and the resulting checksum could be embedded in the original coordinate plane.

5 Conclusion

Given the high degree of redundancy present in a digital representation of multimedia content, there has been an increased interest in using it for the purpose of steganography. The paper suggested how a variation of the LSB insertion algorithm can be used for achieving better security and also improved covertness. Analyzing data in which information has been hidden is called *steganalysis*, and results of steganalysis can be used to change or improve embedding techniques. No technique of information hiding can ensure perfect secrecy; however, by combining steganography with other techniques, such as cryptography, a higher chance of success can be achieved. One should think of steganography, not as a replacement to cryptography but as a vital supplement to it. Even though the cousins in the spy craft family - steganography and cryptography - have their relative merits and demerits, when combined suitably can provide excellent security mechanisms that are much in need at present.

6. References

[1] Ahmet M. Skicioglu, 'Protecting Intellectual Property In Digital Multimedia', IEEE Computer, 2003.

[2] G. J Simmons, Prisoner's Problem and the Subliminal Channel (The), CRYPTO83 – Advances in Cryptology, pp. 51-67, 1984.

[3] N. Johnson and S. Jajodia, "Exploring steganography: seeing the unseen," *IEEE Computer*, pp. 26-34, February 1998.

[4] Johnson, N. Steganography. http:// www.jjtc.com/ stegdoc/stegdoc.html

[5] Bernd Girod, Joachim J. Eggers and R. B¨auml,A Communications Approach to Image Steganography ,Proceedings of SPIE Vol. 4675, Security and Watermarking of Multimedia Contents IV, San Jose, Ca., 2002.

[6] W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding, IBM Systems Journal, Vol. 35, Nos 3&4, 1996

[7] *Steganography tools*, www.cotse.com/tools/ stega.htm

[8] M. Swanson, M. Kobayashi, and A. Tewfik, "Multimedia data embedding and watermarking technologies," *Proceedings of the IEEE*, Vol. 86, No. 6, pp. 1064-1087, 1998.

[9] http://www.ece.purdue.edu/~ace

[10] Rafael C.Gonzalez, Richard E.Woods,Digital Image Processing, Pearson Education , 2003

[11] R. Wolfgang, C. Podilchuk and E. Delp, "Perceptual watermarks for images and video," to appear in the *Proceedings of the IEEE*, 1999.

[12] http://www.crosswinds.net/shetzl/ steghide/ index.html

[13] http://www.cl.cam.ac.uk/fapp2/ steganography/ mp3stego/

[14] http://www.spammimic.com/

[15] H. Berghel, L. O'Gorman, Protecting ownership rights through digital Watermarking, IEEE Computer Mag., pp 101-103, 1996.

[16] Lisa M. Marvel, Charles G.Boncelet Jr. and Charles T.Retter, "Spread Spectrum Image Steganography", IEEE Trans on Image Processing, Vol. 8, No. 8, 1999.