2009 International Conference on Computer Engineering and Applications
IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore

Knots of Substitution Techniques of Audio Steganography

Mazdak Zamani [1], Azizah A. Manaf [2], and Rabiah B. Ahmad [3]

University of Technology Malaysia

Abstract. This paper sets the main focus on the evaluation of transparency, robustness and capacity of the embedding function of Substitution Techniques of digital audio steganography algorithms. Lots of steganography techniques have been described in this literature. Here, beside the evaluation of embedding parameters for the existing techniques, two problems -weaknesses- of Substitution Techniques are investigated which if they could be solved, the large capacity - strength- of Substitution Techniques would be practical.

Keywords: audio steganography, substitution techniques, LSB.

1. Introduction

Steganography is the study of techniques for hiding the existence of a secondary message in the presence of a primary message. The primary message is referred to as the carrier signal or carrier message; the secondary message is referred to as the payload signal or payload message. Steganography itself offers mechanisms for providing confidentiality and deniability; it should be noted that both requirements can also be satisfied solely through cryptographic means [9].

Steganography and watermarking describe methods to embed information transparently into a carrier signal. Steganography is a method that establishes a covered information channel in point-to-point connections, whereas watermarking does not necessarily hide the fact of secret transmission of information from third persons. Besides preservation of the carrier signal quality, watermarking generally has the additional requirement of robustness against manipulations intended to remove the embedded information from the marked carrier object. This makes watermarking appropriate for applications where the knowledge of a hidden message leads to a potential danger of manipulation. However, even knowledge of an existing hidden message should not be sufficient for the removal of the message without knowledge of additional parameters such as secret keys [2].

Steganographic algorithms can be characterized by a number of defining properties [8]. Three of them, which are most important for audio Steganographic algorithms, are introduced below.

- Transparency evaluates the audible distortion due to signal modifications like message embedding or attacking. In order to meet fidelity constraint of the embedded information, the perceptual distortion introduced due to embedding should be below the masking threshold estimated based on the HAS/HVS and the host media [4].

[1] PhD Student, Faculty of Computer Science and Information System, University of Technology Malaysia, 54100 Kuala Lumpur, Malaysia (zmazdak2@siswa.utm.my)

[2] Full Professor, College of Science and Technology, University of Technology Malaysia, 54100 Kuala Lumpur, Malaysia (azizah07@citycampus.utm.my)

[3] Lecturer, Centre for Advanced Software Engineering, University of Technology Malaysia, 54100 Kuala Lumpur, Malaysia, (rabiah@citycampus.utm.my).

- Capacity of an information hiding scheme refers to the amount of information that a data hiding scheme can successfully embed without introducing perceptual distortion in the marked media [4].
- Robustness measures the ability of embedded data or watermark to withstand against intentional and unintentional attacks. Unintentional attacks generally include common data manipulations such as lossy compression, digital-to-analog conversion, re-sampling, re-quantization, etc. whereas intentional attacks cover a broad range of media degradations which include addition white and colored noise, rescaling, rotation (for image and video steganography schemes), resizing, cropping, random chopping, and filtering attacks [4].

2. Comparison and evaluation of existing techniques

Image steganography has been widely studied by researchers. There are a variety of methods used in which information can be hidden in images. Some of them are described here given by Lee et al. [1], Chan et al. [13], Chang et al. [5], and Hsu et al. [12].

In image steganography almost all data hiding techniques try to alter insignificant information in the cover image. Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image. For instance, a simple scheme proposed by Lee et al. [3], is to place the embedding data at the least significant bit (LSB) of each pixel in the cover image [10]. The altered image is called stego-image. Altering LSB doesn't change the quality of image to human perception but this scheme is sensitive a variety of image processing attacks like compression, cropping etc.

When using a 24-bit image, one can store 3 bits in each pixel by changing a bit of each of the red, green and blue color components, since they are each represented by a byte. An 800 × 600 pixel image, can thus store a total amount of 1,440,000 bits or 180,000 bytes of embedded data For example a grid for 3 pixels of a 24- bit image can be as follows:

(01010101 01011100 11011000)

(10110110 11111100 00110100)

(11011110 101100101 01101011)

When the number 300, which binary representation is 101101100 is embedded into the least significant bits of this part of the image, the resulting grid is as follows:

(00101101 00011100 11011101)
(10100111 11000100 00001101)
(11010011 01011100 01100010)

Here the number 300 was embedded into the first 8 bytes of the grid, only the 5 bits needed to be changed according to the embedded message. On average, only half of the bits in an image will need to be modified to hide a secret message using the maximum cover size.

Since there are 256 possible intensities of each primary color, changing the LSB of a pixel results in small changes in the intensity of the colors. The human eye cannot perceive these changes - thus the message is successfully hidden. With a well-chosen image, one can even hide the message in the LSB without noticing the difference.

Redundant embedding of the message content is proposed in different publications. For example, the message can be embedded periodically as proposed by Honsinger [6] or Kutter [11]. This periodic embedding results in the characteristic peaks in the autocorrelation function (ACF). These peaks reflect the applied geometrical transforms. However, an attacker can also calculate the ACF with the aim of predicting the message. Further publications based on redundant embedding use cyclic properties of the message pattern [3] or use redundant embedding in video sequences [1,9]. A steganography method based on the autocorrelation function, which especially addresses local nonlinear geometrical distortions, was proposed by Voloshynovskiy et al. in [5].

Invariant transform domains are applicable in increasing robustness. Some transforms are inherently not affected by specific geometrical transforms. For example, replacing the DCT transform with an invariant transform like log-polar mapping (LPM), which is also called the Fourier-Mellin transform, as described by

Ruanaidh and Pun [7], has some theoretical advantages. After applying the DFT, which is invariant to translation, every amplitude in the DFT at position (u, v) is projected in a new coordinate space (ρ, θ) via the projection:

$u = e^{\rho} \cos(\theta)$, $v = e^{\rho} \cos(\theta)$

In this new coordinate system, rotation and scaling are converted into translation. By calculating the amplitude of the DFT of the LPM, the resulting domain is invariant against rotation, scaling, and translation (RST). Because of practical problems, the authors suggested embedding the message into the translation-invariant DFT domain and adding a (second) template message into the RST-invariant LPM domain. Another approach that uses the properties of the LPM domain was proposed by Lin et al. [2].

Template insertion is another technique for increasing the robustness of steganography techniques. In the case of image steganography, a template is inserted in the image. This template is used to recover the original image format and does not carry any steganography content. One of those methods was proposed by Pereira and Pun [8]. The template consists of points that are randomly arranged in the DFT domain. Their radii vary between two limiting frequencies and are chosen (magnitude and phase) via a secret key. Peaks are generated by increasing the coefficients at the calculated positions. The message detection process consists of two steps. First, the template is detected. This information is used to calculate a linear transform. Second, the information about the linear transform is used to retrieve the embedded message. As with redundant embedding, an attacker can also use information about the template to attack the embedded message, it is described by Herrigel et al. [10].

Further approaches have considered a number of properties for embedding, for example, geometry recovery by using the original stego image as proposed by [12]. These methods require the original image instead of using a template. The original image is used to identify the geometrical distortions and to undo them. The main disadvantage is the fact that blind or oblivious detection is not possible with these methods after a geometrical attack. Using regions of interest (ROI) for steganography as proposed by Su et al. [13] is currently difficult to achieve without human interaction because semantically meaningful regions have to be identified. However, content-based steganography based on robust segmentation, as presented in the next section, is a generalized variant of steganography of ROIs.

While newer methods also have to face the previously described problem of geometrical distortions, they attempt to use semantic information in the image—the content of the image—for synchronization. Thus, they are classified as content-based steganography algorithms. In "Towards Second Generation steganography Schemes," Kutter et al. [12] outlined a scheme that is based on significant features concerning perception. These features should be invariant to noise, covariant to geometrical transforms, and independent of cropping. For feature extraction, the image is decomposed using the Mexican-Hat wavelet as proposed by Manjunath et al. [14]. The detected features are used for an image segmentation using Voronoi diagrams (i.e., partitioning of a given space). The resulting segments are used for embedding a message with an existing steganography scheme. The detected feature is used as a reference origin for the steganography process.

For the detection of the message, the same features have to be extracted and the image has to be segmented. The authors reported that the feature location may move by 1 or 2 pixels, which has to be compensated for by a limited search. A similar scheme which is well described is presented by Bas et al. [1].

Instead of creating a triangulation of the image data, Dittmann et al. [13] proposed a scheme based on self-spanning patterns (SSP). These SSPs are also based on image feature points. The initial pattern, which is represented by a polygon, is spanned over four feature points. Information carrying patterns are spanned around the previous pattern, resulting in a set of polygons with a given traverse direction.

An estimation of images parameters is proposed by Alghoniemy and Tewfik [8] which is also based on the wavelet decomposition. Previous proposed methods like [6] suggested using image moments which are invariant against geometrical transforms [5]. However, their main disadvantage is the missing robustness against cropping, which is addressed by the method presented in [9]. The scaling parameter is estimated using the edges standard deviation ratio (ESDR) and the rotation angle is estimated using the average edges angles difference (AEAD). These estimations are based on the wavelet maxima which are extracted from the low-frequency components of the wavelet decomposition. The ESDR and the AEAD show increased

robustness against cropping. However, they are not completely robust against general affine transforms. Therefore, Alghoniemy and Tewfik propose to combine this method with exhaustive search strategies.

Local messages are proposed by Tang and Hang [13]. Their scheme uses the same feature extraction method as proposed by Kutter et al. [10]. These extracted features build the centers of nonoverlapping image disks. The messages are embedded and extracted in each image disk. Before embedding and detection, the image disks are normalized using the normalization method proposed by Alghoniemy and Tewfik [15]. The message is then embedded in the DFT domain.

Segmentation or region-based image steganography algorithms are proposed by Nikolaidis and Pitas [7] and by Celik et al. [11]. In contrast to the methods in which a region of interest has to be selected manually, these methods use image segmentation methods to group the pixels of an image according some statistics.

The method proposed by Celik et al. [10] is based on a color clustering using a k-means clustering method. The cluster centers are identified and a Delauny triangulation of these cluster centers results in image regions that message bits are embedded. Thus, this steganography technique is related to the triangulation method used in [14]. However, different features are used for the triangulation. The method proposed by Nikolaidis and Pitas [6,14] is based on the iterated conditional modes (ICM) for clustering. The resulting regions are merged and the largest regions of the final results are used for embedding the message. Before steganography, the regions are approximated by ellipsoids. The bounding rectangle of each ellipsoid is used for the embedding and detection of the message.

A simple method, which is called data hiding in block parity (DHBP), is proposed by Fu and Au [7], which is similar to the steganography method of binary images proposed by Wu et al. [11,15]. The block-sum parity (even or odd number of pixels) encodes the information. The method proposed by Baharav and Shaked [13] uses different dither matrices (instead of one). The message determines the dither matrices used. The dither matrix influences the pixel distribution in the output image. The scheme is shown in Figure 4.8. For decoding the distance of the message, the dithered image (region) is used to determine the embedded value.

Binary text documents are considered by Mei et al. [5,14]. The interesting aspect of this approach is that different patterns in text documents (similar to strokes) are identified and modified. The modifications change the number of pixels, but the main properties remain the same. Thus, a maximum quality is achieved.

3. Conclusion

The steganographic algorithms were primarily developed for digital images and video sequences; interest and research in audio steganography started slightly later. In the past few years, several algorithms for the embedding and extraction of message in audio sequences have been presented. All of the developed algorithms take advantage of the perceptual properties of the human auditory system (HAS) in order to add a message into a host signal in a perceptually transparent manner. Embedding additional information into audio sequences is a more tedious task than that of images, due to dynamic supremacy of the HAS over human visual system. On the other hand, many attacks that are malicious against image steganography algorithms (e.g. geometrical distortions, spatial scaling, etc.) cannot be implemented against audio steganography schemes. Consequently, embedding information into audio looks more secure due to existing less steganalysis techniques for attacking to audio. Furthermore, Natural sensitivity and difficulty of working on audio caused there are not algorithms and techniques as mush as exist for image. Therefore, regarding nowadays audio files are available anywhere, working on audio and improvement in related techniques is needed.

The theory of substitution technique is that simply replacing either a bit or a few bits in each sample will not be noticeable to the human eye or ear depending on the type of file. This method has high embedding capacity (41,000 bps) but it is the least robust. It exploits the absolute threshold of hearing but is susceptible to attacks. The obvious advantage of the substitution technique, the reason for choosing this technique, is a very high capacity for hiding a message; the use of only one LSB of the host audio sample gives a capacity of 44.1 kbps. Obviously, the capacity of substitution techniques is not comparable with the capacity of other

more robust techniques like spread spectrum technique that is highly robust but has a negligible embedding capacity (4 bps).

Like all multimedia data hiding techniques, audio steganography has to satisfy three basic requirements. They are perceptual transparency, capacity of hidden data and robustness. Noticeably, the main problem of audio substitution steganography algorithm is considerably low robustness.

There are two types of attacks to steganography and therefore there are two type of robustness. One type of attacks tries to reveal the hidden message and another type tries to destroy the hidden message. Substitution techniques are vulnerable against both types of attacks. The adversary who tries to reveal the hidden message must understand which bits are modified. Since substitution techniques usually modify the bits of lower layers in the samples -LSBs, it is easy to reveal the hidden message if the low transparency causes suspicious. Also, these attacks could be categorized in another way: Intentional attacks and unintentional attacks. Unintentional attacks like transition distortions could destroy the hidden message if is embedded in the bits of lower layers in the samples -LSBs. As a result, in this paper the authors briefly address following problems of substitution techniques of audio steganography:

- Having low robustness against attacks which try to reveal the hidden message
- Having low robustness against distortions with high average power

4. Acknowledgements

This work is part of a project supported by the Ministry of Science, Technology and Innovation of Malaysia whose title is “Development of Digital Audio Information Hiding Systems For High-Embedding-Capacity Applications” (01-01-06-SF0524).

5. References

[1] Alghoniemy, M., and A. H. Tewfik, “Geometric Distortion Correction in Image Watermarking,” *Proceedings of Electronic Imaging 2000, Security and* Watermarking of Multimedia Contents II, Vol. 3971, San Jose, CA, January 2000.

[2] Avcibas I., Memon N. and Sankur B. “Steganalysis using image quality metrics”. *IEEE Transactions on Image Processing*, vol. 12, pp. 221–229, Feb. 2003.

[3] Bas, P., J.-M. Chassery, and B. Macq, “Geometrically Invariant Watermarking Using Feature Points,” *IEEE Transactions on Image Processing*, Vol. 11, No. 9, September 2002, pp. 1014–1028.

[4] Bender, W., et al., "Techniques for data hiding", *IBM Systems Journal*, Vol. 35, Nos 3&4, pp. 313-36, 1996.

[5] Chan Y. K. and Chang C. C. “Concealing a Secret Image Using the Breadth First Traversal Linear Quad tree Structure”. *IEEE Proceedings of Third International Symposium on Cooperative Database Systems for Advanced Applications*, pp. 194-199, 2001

[6] Chen, B., and G. W. Wornell, “Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding,” *IEEE Transactions on Information Theory*, Vol. 47, No. 4, May 2001.

[7] Eggers, J. J., and B. Girod, “Blind Watermarking Applied to Image Authentication,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, May 2001.

[8] Fridrich, Jessica and others. “Steganalysis of LSB Encoding in Color Images.” *Proceedings of the IEEE International Conference on Multimedia and Expo*. 1279–1282. New York: IEEE Press, 2000.

[9] Fu, M. S., and O. C. Au, “Data Hiding Watermarking for Halftone Images,” *IEEE Transactions on Image Processing*, Vol. 11, No. 4, 2002, pp. 477–484.

[10] Kelton W. D. and Law A. *Simulation Modeling and Analysis*. Mcgraw-Hill Science, USA, 2006.

[11] Lee, Y. K. and Chen L. H. “High Capacity Image Steganographic Model”. *IEEE Proceedings Vision, Image and Signal Processing*, pp. 288-294, 2000.

[12] Lin, C.-Y., et al., “Rotation, Scale, and Translation Resilient Watermarking for Images,” *IEEE Transactions on Image Processing*, Vol. 10, No. 5, May 2001.

[13] Nikolaidis, A., and I. Pitas, “Region-Based Image Watermarking,” *IEEE Transactions on Image Processing*, Vol. 10, No. 11, November 2001, pp. 1726–1740.

[14] Pal S.K., Saxena P. K. and Mutto S.K. “The Future of Audio Steganography”. *Pacific Rim Workshop on Digital Steganography*, Japan, 2002.

[15] Westfeld, A., “F5-A Steganographic Algorithm,”, *Information Hiding: 4th International Workshop*, Vol. 2137 of Lecture Notes in Computer Science, April 2001, Pittsburgh: Springer-Verlag, pp. 289–302.