A Secure Audio Steganography Approach

Mazdak Zamani[1], Azizah Bt Abdul Manaf[1], Rabiah Bt Ahmad[1], Farhang Jaryani[1],
Hamed Taherdoost[1], Akram M. Zeki[2]
Universiti Teknologi Malaysia, Malaysia[1]
International Islamic University Malaysia, Malaysia[2]
zmazdak2@siswa.utm.my
azizah07, rabiah {@citycampus.utm.my}
fjaryani, akramzeki{@yahoo.com}
hamed.taherdoost@gmail.com

Abstract

A wide range of steganography techniques has been described in this paper. 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. Furthermore, a novel, principled approach to resolve the problems is presented. Using the proposed genetic algorithm, message bits are embedded into multiple, vague and higher LSB layers, resulting in increased robustness.

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 [6].

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. 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 [13].

- **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 [9].

- **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 [13].

2. Comparison and evaluation of existing techniques

Some steganographic experts e.g. [7] introduced a method based on statistical analysis of Pairs of Values (PoVs) that are exchanged during message embedding. Pairs of Values that differ in the LSB

only, for example, could form these PoVs. This method provides very reliable results when we know the message placement (such as sequential). However, we can only detect randomly scattered messages with this method when the message length becomes comparable with the number of samples in the audio.

Existing cryptographic and steganographic mediums suffer from a myriad of attacks. Johnson [9] has studied such attacks on image steganography, whereas Pal et al. [1] has studied similar attacks in the context of audio steganography. Even though cryptography and steganography are exposed to so many probable attacks, very few people have given a thought to find alternate ways to transmit information. The goal of steganalysis is to defeat steganography methods by identifying the presence of hidden information.

Fisk et al. [5] point out the weaknesses of TCP/IP protocol suite and discuss how those weaknesses could be used as covert channels for secret communication, whereas Bao et al. [8] focus on using communication accessories like email headers etc for secret communication. Avcibas et al. [11] proposed a steganalysis technique based on image quality metrics. Fard et al. [3] proposed a novel (Genetic Algorithm) GA evolutionary process to make a secure steganographic encoding on JPEG images. Martín et al. [9] experimentally investigated if stego-images, bearing a secret message were statistically "natural". Koval et al. [12] discussed the problem of performance improvement of non-blind statistical steganalysis of additive steganography in real images. Luo et al. [4] presented a secure LSB steganography system against sample pair analysis, such as RS, SPA and DIH method by adopting chaotic technique and dynamic compensation skill.

Embedding a message by this method is not limited to the DCT domain. However, the DCT domain has been extensively studied because this is the transform used in Joint Picture Expert Group (JPEG) compression, where extensive studies on perceptibility were performed. Further advantages of using the DCT domain include the fact that frequency decomposition in frequency bands is efficient, DCT transform is widely used in image and video compression schemes, and the DCT coefficients affected by compression are well known. A considerable number of image steganography techniques share this architecture. Yet they differ chiefly in the signal design, the embedding, and the retrieval of the message content [10].

Amplitude modulation of the DFT coefficients is applied by many steganography techniques [5]. One advantage of the DFT transform is the resulting shift (translation) invariance. Another one is the ease of considering the human perception by weighting frequencies. The properties of the DFT have been studied extensively in image processing literature.

One of the results obtained there is the fact that the phase information is more important for the image content than the magnitude [8]. Consequently, Ruanaidh et al. [7] propose embedding the message in the phase information of the image, which is comparable to phase modulation in communication theory, in contrast to the previously described amplitude modulation. For a blind retrieval of the message, an optimal statistical detector is proposed by Ruanaidh et al. Various methods for steganography digital images in the wavelet domain have been proposed. Among other reasons, the development of new compression schemes led to new steganography techniques. Barni et al. proposed a steganography method based on the wavelet decomposition [14]. The wavelet decomposition decomposes the input image in high and lowpass components with different orientations [9].

The patchwork technique is also applied to image steganography, as, for example, proposed by Pitas and Kaskalis [11]. They split the image into two subsets, and in one subset the pixel values are increased, whereas in the other subset pixel values are decreased. Further patchwork techniques are block based, like the techniques proposed by Langelaar et al. [6] or Bruyndonckx et al. [5].

The similarity between this method and the correlation-based method is shown by Cox et al. [11]: This algorithm can be formulated as a correlation by defining a pattern with the same dimensions as the coefficient matrix. The pattern values are determined by the influence of the corresponding coefficients: It is zero for coefficients not considered in the evaluation. The pattern values for the pair coefficients are either 1 or −1. Thus, the sign of the correlation directly depends on the relation of the pair coefficients.

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 [14]. 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. [4].

Further approaches have considered a number of properties for embedding, for example, geometry recovery by using the original stego image as proposed by [7]. 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.

3. Why Still Substitution Techniques of Audio Steganography

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 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 seems more secure due to 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) [2].

4. The Remained Problems of Substitution Techniques of Audio Steganography

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 can 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, this paper briefly addresses following problems of substitution techniques of audio steganography:

a. Having low robustness against attacks which try to reveal the hidden message

b. Having low robustness against distortions with high average power

4.1. First Problem

One type of robustness that is very critical for security is withstanding against the attacks which try to reveal or extract the hidden message. This paper is to improve this type of robustness. With an intelligent algorithm we hope to reach a more robust substitution technique, as such, extracting the hidden message become inaccessible to adversary.

Certain way to withstand against these attacks is making more difficult discovering which bits are modified. Thus, the algorithm may not change some sample due to their situations. This selecting will improve the security of the method and robustness of the technique, because if somebody tries to discover the embedded message, he has to apply a specific algorithm to read some bits of samples. But if modified samples are secret, nobody can discover the

message. It is remarkable that if we achieve float target bits, it will be novel.

As we know in samples LSBs are more suspicious, thus embedding in the bits other than LSBs could be helpful to increase the robustness. Furthermore, discovering which samples are modified should be uncharted. To reach to the level of ambiguity, the algorithm will not use a predefined procedure to modify the samples but will decide, according to the environment, in this case the host file; as such it will modify indistinct samples of audio files, depending on their values and bits status. Thus, some of the samples which algorithm determines they are suitable for modifying will modify and other samples may not change. This ambiguity in selecting samples will thus increase security and robustness of the proposed algorithm.

4.2. Second Problem

A significant improvement in robustness against unintentional attacks -for example signal processing manipulation- will be obtained if an embedded message is able to resist distortions with high average power. To achieve this robustness the message could embed in deeper layers. But, selecting the layer and bits for hosting is critical because the random selection of the samples used for embedding introduces low power additive white Gaussian noise (AWGN). It is well known from psychoacoustics literature [1] that the human auditory system (HAS) is highly sensitive to the AWGN. This fact limits the number of bits that can be imperceptibly modified during message embedding [3]. Embedding the message bits in deeper layers absolutely causes bigger error and it will decrease the quality of transparency. Thus, the algorithm which embeds the message bits in deeper layers should modify other bits intelligently to decrease the amount of this error and reserve the transparency.

Predictably, substitution techniques try to modify the bits of samples in accordance with a directive that is defined in algorithm. The target bits are definite, and the amount of resultant noise is not controlled. Of course, there are some better techniques that try to adjust the amount of resultant noise in substitution techniques. These improved algorithms alter other bits else than target bit in sample to decrease the amount of resultant noise. A key idea of the improved algorithm is message bit embedding that causes minimal embedding distortion of the host audio. It is clear that, if only one of 16 bits in a sample is fixed and equal to the message bit, the other bits can be flipped in order to minimize the embedding error. For example, if the original sample value was $0\ldots01000_2 = 8_{10}$, and the message bit was zero is to be embedded into 4th LSB layer, instead of value $0\ldots00000_2 = 0_{10}$ that the standard algorithm would produce, the proposed algorithm produces a sample that has value $0\ldots00111_2 = 7_{10}$, which is far closer to the original one. However, the extraction algorithm remains the same; it simply retrieves the message bit by reading the bit value from the predefined layer in the stego audio sample. In the areas where the original and message bit do not match, the standard coding method produces a constant error with 8-Quantization Steps (QS) amplitude [1].

The improved method introduces a smaller error during message embedding. If the 4th LSB layer is used, the absolute error value ranges from 1 to 4 QS, while the standard method in the same conditions causes a fixed absolute error of 8 QS.

What would be improved is a level of intelligence in those substitution algorithms which try to adjust the sample bits after modifying the target bits. The basic idea of the proposed algorithm is embedding that cause minimal embedding distortion of the host audio. What is clear as much as intelligence the alteration algorithms have, the amount of resultant noise could be improved. Because the total noise will be less, when we are able to alter and adjust more samples.

5. The Solution

Accordingly, there are two following solutions for mentioned problems:

1. *The solution for first problem*: Making more difficult discovering which bites are embedded by modifying the bits else than LSBs in samples, and selecting the samples to modify privately-not all samples.

2. *The solution for second problem*: Embedding the message bits in deeper layers and other bits alteration to decrease the amount of the error.

To integrate these two solutions, "embedding the message bits in deeper layers" that is a part of second solution also can satisfy "modifying the bits else than LSBs in samples" of second solution. In addition, when we try to satisfy "other bits alteration to decrease the amount of the error" of second solution, if we ignore the samples which are not adjustable, also "selecting not all samples" of first solution will be satisfied.

Thus, intelligent algorithm will try to embed the message bits in the deeper layers of samples and alter other bits to decrease the error and if alteration is not possible for any samples it will ignore them.

It is clear that the main part of this scenario is bit alteration that it should be done by intelligent algorithms which use either genetic algorithms or a symbolic AI system.

6. Genetic Algorithm Approach

As Figure 1 shows, there are four main steps in this algorithm that are explained below.

Alteration: At the first step, message bits substitute with the target bits of samples. Target bits are those bits which place at the layer that we want to alter. This is done by a simple substitution that does not need adjustability of result be measured.

Modification: In fact this step is the most important and essential part of algorithm. All results and achievements that we expect are depending on this step. Efficient and intelligent algorithms are useful here. In this stage algorithm tries to decrease the amount of error and improve the transparency. For doing this stage, two different algorithms will be used.

One of them that is more simple likes to ordinary techniques, but in aspect of perspicacity will be more efficient to modify the bits of samples better. Since transparency is simply the difference between original sample and modified sample, with a more intelligent algorithm, I will try to modify and adjust more bits and samples than some previous algorithms. If we can decrease the difference of them, transparency will be improved. There are two example of adjusting for expected intelligent algorithm below.

Sample bits are: 001**0**1111 = 47
Target layer is 5, and message bit is 1
Without adjusting: 001**1**1111 = 63 (difference is 16)
After adjusting: 001**1**0000 = 48 (difference will be 1 for 1 bit embedding)

Sample bits are: 001**00**111 = 39
Target layers are 4&5, and message bits are 11
Without adjusting: 001**11**111 = 63 (difference is 24)
After adjusting: 000**11**111 = 31 (difference will be 8 for 2 bits embedding)

Another one is a Genetic Algorithm which the sample is like a chromosome and each bit of sample is like a gene. First generation or first parents consist of original sample and altered sampled. Fitness may be determined by a function which calculates the error. It is clear, the most transparent sample pattern should be measured fittest. It must be considered that in crossover and mutation the place of target bit should not be changed.

Verification: In fact this stage is quality controller. What the algorithm could do has been done, and now the outcome must be verified. If the difference between original sample and new sample is acceptable and reasonable, the new sample will be accepted; otherwise it will be rejected and original sample will be used in reconstructing the new audio file instead of that.

Reconstruction: The last step is new audio file (stego file) creation. This is done sample by sample. There are two states at the input of this step. Either modified sample is input or the original sample that is the same with host audio file. It is why we can claim the algorithm does not alter all samples or predictable samples. That means whether which sample will be used and modified is depending on the status of samples (Environment) and the decision of intelligent algorithm.
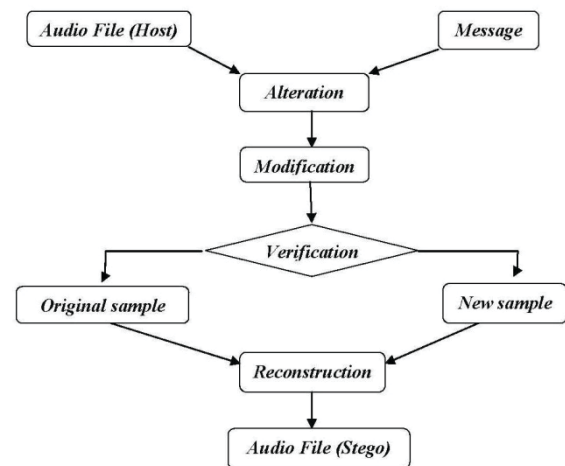

Figure 1: Approach Diagram

7. Conclusions

A new approach is proposed to resolve two problems of substitution technique of audio steganography. First problem is having low robustness against attacks which try to reveal the hidden message and second one is having low robustness against distortions with high average power. An intelligent algorithm will try to embed the message bits in the deeper layers of samples and alter other bits to decrease the error and if alteration is not possible for any samples it will ignore them. Using the proposed genetic algorithm, message bits could be embedded into multiple, vague and deeper layers to achieve higher capacity and robustness.

8. Acknowledgements

9. 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] 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.

[5] Cvejic N. and Seppänen T. "Increasing the capacity of LSBbased audio steganography", *Proc. 5th IEEE International Workshop on Multimedia Signal Processing*, St. Thomas, VI, December 2002, pp. 336-338.

[6] 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.

[7] 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.

[8] 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.

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

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

[11] Martín Alvaro, Sapiro Guillermo and Seroussi Gadiel, "Is Image Steganography Natural?" *IEEE Transactions On Image Processing*, Vol. 14, No. 12, December, 2005.

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

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

[14] Voloshynovskiy, S. V., et al., "StegoWall: Blind Statistical Detection of Hidden Data,", *Proceedings of Electronic Imaging 2002, Security and Watermarking of Multimedia Contents IV*, San Jose, CA, January 2002, pp. 57–68.