A Comprehensive Review of Audio Steganalysis Methods

Hamzeh Ghasemzadeh¹*, Mohammad H. Kayvanrad²

Abstract: Recently, merging signal processing techniques with information security services has found a lot of attention. Steganography and steganalysis are among those trends. Like their counterparts in cryptology, steganography and steganalysis are in a constant battle. Steganography methods try to hide the presence of covert messages in innocuous-looking data, whereas steganalysis methods try to reveal existence of such messages and to break steganography methods. The stream nature of audio signals, their popularity, and their wide spread usage make them very suitable media for steganography. This has led to a very rich literature on both steganography and steganalysis of audio signals. This paper intends to conduct a comprehensive review of audio steganalysis methods aggregated over near fifteen years. Furthermore, we implement some of the most recent audio steganalysis methods and conduct a comparative analysis on their performances. Finally, the paper provides some possible directions for future researches on audio steganalysis.

1. Introduction

Digital computers have revolutionized every aspect of our lives. Information security is one of the branches of science that has benefited entirely from this invention. In particular, merging signal processing techniques with information security services has found a lot of attentions. Some of these new trends include multimedia encryption systems and their cryptanalysis [1], multimedia secret sharing, steganography, steganalysis, and watermarking. Each of these techniques serves a different purpose. For example, multimedia encryption addresses confidentiality and watermarking serves the purpose of copyright protection. Among these techniques, steganography is exceptionally interesting. Arguably, the main purposes of steganography are privacy and preventing traffic analysis. In other words, while encryption prevents unauthorized access to the data, it cannot conceal pattern of communications. Therefore, an adversary who is watching the channel can obtain very valuable information including time of communications, frequency of communications, size of messages, identity of senders and recipients, and much more.

Steganography is best described in terms of subliminal channels. Subliminal channels were first introduced under the prisoner's problem [2]. Two accomplices in a crime are apprehended and imprisoned in two different cells. The warden, who wants to gather some information, allows them

to communicate as long as he can read their messages. Apparently, the culprits want to talk about escape plan, but they cannot use encrypted messages. Therefore, they use innocuous looking communications and hide their messages inside them.

Steganalysis is the countermeasure of the warden against such hidden messages. Because steganography changes content of cover signal, usually they leave a trail and thus can be detected. Therefore, steganalysis is a decision problem and if it is solved for a specific method with a high probability, that method is considered broken [3]. If embedding method and statistical model of cover is known, optimal detector can be constructed. Absence of this information has led to steganalysis systems based on feature extraction and machine learning techniques [4]. Figure 1 presents typical block diagram of such systems.

Steganalysis is broadly divided into targeted and universal methods. In the targeted paradigm, the warden knows the embedding algorithm. On the other hand, universal methods do not have such assumptions [3]. We can also divide universal methods into blind and semi-blind systems. Blind systems are constructed only from cover signals. On the other hand, semi-blind systems use both classes for determining the decision boundaries. Additionally, two different models of passive and active wardens exist. A passive warden listens to communications without any interfering and his goal is only detection of the hidden

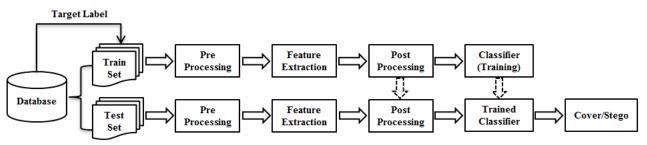


Figure 1. Block diagram of typical steganalysis systems

¹ Department of Communicative Sciences and Disorders, Michigan State University, MI, USA

³ Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran

^{*}ghasemza@msu.edu

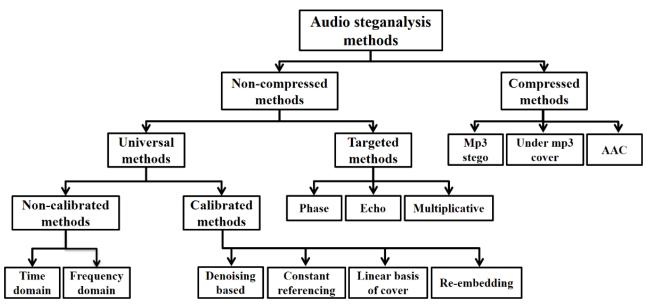


Figure 2. Classification of audio steganalysis methods

messages. Whereas, an active warden tries to prevent communications of hidden messages [3]. Other terms that need to be distinguished are active and passive steganalysis. Passive steganalysis only determines the existence of hidden messages whereas active steganalysis goes one step further and provides the warden with extra information such as the length of hidden message and/or its location.

Current literature lacks a comprehensive review of audio steganalysis methods. This motivated us to review near fifteen years of work on audio steganalysis. This survey is organized as follows: The next two sections are devoted to detection of non-compressed audio signals. To this end, universal and targeted steganalysis methods are respectively reviewed in sections 2 and 3. Sections 4 and 5 are devoted to steganalysis of compressed audio signals. In section 6 comparative analyses of some audio steganalysis methods are presented. Section 7 describes some possible directions for future research on audio steganalysis and finally the paper is concluded in section 8. Figure 2 summarizes the classification that is used in this paper.

2. Universal non-compressed methods

2.1. Non-calibrated features

In these methods, features are extracted directly from the signal and they can be categorized according to their domain of features. Although in some cases such distinction is very hard, yet this methodology is both very common and informative. Figure 3 depicts a simple schematic of feature extraction in non-calibrated methods.

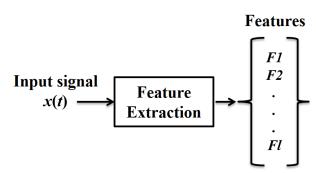


Figure 3. Non-calibrated feature extraction

2.1.1. Time domain features

2.1.1.1. IDS and Steganalysis

Preventing information piracy and its unauthorized disclosure could be an important application of steganalysis. This fact motivated Dittmann to incorporate steganalysis into intrusion detection systems (IDS) [5]. Such applications need real-time systems; therefore, simple features were used with simple thresholding for making the decision. Their investigation showed that ratio of ones and zeros in least significance bit (LSB) of covers and stegos were different.

2.1.1.2. Linear prediction residue

This technique was based on the assumption that audio signals are stationary and their samples are highly correlated over short periods of time. Embedding imposes faint changes on the correlation between neighboring samples. Therefore, if the warden known the correlation between cover samples, he can detect the abnormalities induced by the embedding process. This idea was implemented in [6]. In this work, linear prediction code (LPC) was used to extract the correlation between neighbouring samples. Furthermore, to extract correlations of different frequency bands, LPC was applied on the wavelet coefficients of all sub-bands. Finally, two different sets of features were fused together. The first set

provided statistical details of signal in the frequency domain. The second set modeled irregularities in the correlation between samples of the signal. These sets were calculated as higher order statistics (HOS) of wavelet coefficients and LPC prediction errors, respectively. The method was cross-validated for different embedding capacities of the Steghide [7]. Considering the trade-off between missed detection and false alarm, the best results were achieved when the system was trained with signals embedded at 20% of maximum capacity.

2.1.1.3. Statistical model of histogram

Another possibility is determining distribution of wavelet coefficients. Fu et al. proposed to extract different statistics from wavelet coefficients and their frequency domain counterparts [8]. This method was further improved by applying principle component analysis (PCA) for reducing dimensionality of feature space. The method was tested on different wavelet domain steganography algorithms. The results demonstrated that applying PCA reduces dimensionality of the feature space considerably with a negligible degradation in the performance.

2.1.1.4. Chaotic-based Steganalysis

Most of the aforementioned works were based on linearity and stationarity assumptions about audio signals. However, more recent studies have found evidences of some phenomena, which linear model cannot describe. Researchers have been able to model these phenomena with chaos theory. Based on such models, noise of steganography would change chaotic structure of signal. Therefore, by extracting features from the chaotic structure of audio signals one can distinguish between covers and stegos. This path was followed in [9]. After investigating different chaotic-based measurements, it was concluded that false neighbour fraction and Lyapunov spectrum were more discriminative for steganalysis. The steganalysis system was tested on different settings and was compared with other methods. The work concluded that performances of all methods were comparable for detection of watermarking, but chaotic-based features detected steganography methods better.

2.1.1.5. Features based on Markov process

Markov processes are the simplest generalization of independent processes and they have been very popular for steganalysis. These processes have the interesting property that the past has no influence on the future samples, as long as the present value is determined. If such a system has only finite number of states, it is called a Markov chain. A Markov chain is completely determined by its transition matrix defined as:

$$P_{xy} = p(X_{n+1} = y | X_n = x)$$
 (1)

In [10], Markov transitions of the second order derivatives of audio was used for steganalysis. The transition range of the Markov chain was limited to [-4, 4] to keep the feature dimension low. Furthermore, this leads to extracting features from smooth regions of signals rather than regions with dramatic changes. Additionally, this work argued that besides embedding strength, complexity of audio signals also affects performance of steganalysis. This work proposed the

following metric for fast measurement of audio signal complexity:

complexity:
$$C_{x(t)} = \frac{N \sum_{i=0}^{N-2} x''(t)}{(N-2) \sum_{i=0}^{N-1} x(t)}$$
(2)
where $x''(t)$ and N denote the second order derivative on

where, x''(t) and N denote the second order derivative and length of the signal. Based on this metric, audio signals were divided into three categories of low, middle, and high complexities and each category was tested separately. The simulations demonstrated that these features had good performance even for audio signals with high levels of complexities. This method was later refined with transition range of [-6, 6] and a bigger database for evaluation [11].

2.1.1.6. Autoregressive time delay neural network

Determining the appropriate feature extraction is one of the biggest challenges in steganalysis. Considering the difficulty of modeling temporal characteristics of audio signals, this decision becomes even harder. A special type of network known as autoregressive time delay neural network (AR-TDNN) was proposed in [12] to address this problem. AR-TDNN has the advantage that feature extraction does not need to be specified explicitly. In other words, the network implements both feature extraction and classification parts of the system. Implementation of these networks has two parts. In the TDNN part, samples of the input are fed into the network and it combines feedback of the output with delayed samples of the input for extracting useful patterns. On the other hand, the autoregressive part of system recognizes the sequence of the previously learned patterns. The efficacy of the system was tested on both LSB and discrete wavelet transform (DWT) embedding algorithms.

2.1.2. Frequency domain features

These methods primarily use frequency domain features for distinguishing between covers and stegos.

2.1.2.1. Steganalysis based on MFCC

Mel-frequency cepstrum coefficients (MFCC) are one of the most well-known features in speech processing applications. Cepstrums are frequency components of the logarithm of the magnitude of spectral power of the signal and it can be interpreted as the rate of power changes in different frequency regions. MFCC is a modified version of cepstrum and it reflects some of characteristics of the human auditory system (HAS).

Based on potency of MFCC for speech recognition applications, it was adopted for audio steganalysis in [13]. This work focused on steganalysis of VoIP channels and it used three sets of features. The first set was statistical characteristics of the signal, which included entropy, LSB ratio, flipping rate of LSB, and some other time domain statistical moments. For the second set, 29 MFCCs were calculated. The third set was calculated based on the hypothesis that removing speech-relevant portions of the signal would be beneficial for steganalysis. Therefore, the signal in the range of 200 to 6819.59 hertz was filtered. Then MFCCs of the filtered signal were calculated. Efficacy of this method was tested on five steganography and four watermarking methods. Furthermore, different combinations of feature sets were investigated. The tests confirmed that removing speech-relevant portions was beneficial for steganalysis.

2.1.2.2. MFCC of derivative of audio signals

Another work argued that derivative of the signal is more informative for steganalysis [14]. This work showed that taking the second order derivative of signal improves discriminative properties of its high-frequency regions. An experiment was conducted to justify this idea. To that end, the whole spectrum was divided into 80 different regions. After measuring discriminative properties of each region it was concluded that high frequency regions were more informative. To investigate the potency of the method, the system was contrasted with ordinary MFCC and wavelet-based MFCCs. Results showed that wavelet-based MFCC was better than ordinary MFCC and the derivative-based MFCC was the best.

2.1.2.3. Frequency domain feature fusion

To implement a more powerful system different features were combined in [15]. First, short-term characteristics of signal were extracted from 12 MFCCs. The second set consisted different moments of spectral characteristics of the second order derivative of the signal. Finally, features based on audio quality metric [16] and LPC residue [6] were added to them and a vector of 52 features was produced. Feature selection was conducted based on their F-scores and the system was tested in both targeted and universal scenarios.

2.1.2.4. Reversed-psychoacoustic model of human hearing

Ghasemzadeh et al. argued that employing features based on models of human auditory system (HAS) is counter intuitive and it would lead to discarding vital information for steganalysis [17]. According to this work, a steganography method is insecure if its stego signals are distinguishable from their covers. Therefore, in its most basic form, the human perception system should be oblivious to steganography induced noise. Based on this, it was argued that if features are extracted based on models of HAS they should lose their significance. To address this, an artificial auditory system was designed that could virtually hear effect of steganography and therefore could discriminate stegos from covers. This new model was called reversed Mel and it had the maximum deviation from HAS. That is, it had finer

resolution in high frequencies and coarser resolution in low frequencies. Equation (3) shows the proposed R-Mel scale.

$$R - Mel = 1127 \times ln(1 + \frac{F_s/2 - f}{700})$$
(3)

where, f and F_S denote the given frequency in hertz and the sampling rate of the signal, respectively. Figure 4.A presents a comparison between Mel and R-Mel scales. In addition, calculation of cepstrum coefficients relies on a set of triangular windows. Figure 4.B compares windows constructed based on Mel and R-Mel scales.

This work was further improved in [18]. First, efficacy of maximum deviation from HAS was justified. To that end, spectrums of covers and their steganography noise were divided into L sub-bands and then signal to noise ratio of sub-band B_i was defined as:

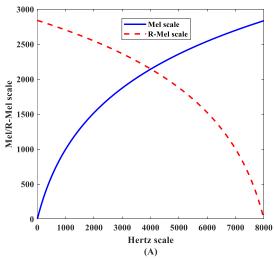
$$SNR_{i} = 10log_{10} \left(\frac{\int\limits_{Bi} \left| C\left(e^{jw}\right) \right|^{2}}{\int\limits_{Bi} \left| N\left(e^{jw}\right) \right|^{2}} \right), 1 \le i \le L$$

$$(4)$$

where, C(w) and N(w) denote the spectrum of cover and its steganography noise. Comparing plot of SNR_i for different steganography methods with frequency resolution of both Mel and R-Mel scales showed that R-Mel scale was more suitable for steganalysis purposes. The work also used HOS for providing better discriminative properties. Finally, genetic algorithm (GA) [19] was invoked to find the optimum sub-set of features. This method was tested on a wide range of data hiding algorithms including both LSB and non-LSB methods. Furthermore, results of both targeted and universal scenarios were investigated. The results showed that HOS and feature selection based on GA improve results of steganalysis considerably.

2.2. Calibrated features

Finding features that depends less on the contents of signal and more on the presence of hidden message is one of the most challenging problems in steganalysis. Such features can reflect reliably the presence of hidden messages. Apparently, if the cover is known such features can be extracted very easily. But, in a realistic scenario this



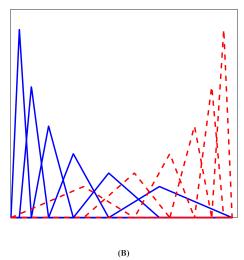


Figure 4. Comparison between R-Mel and Mel (A) R-Mel scale vs Mel scale (B) Filter banks constructed based on R-Mel and Mel

assumption does not hold. Therefore, researchers have proposed estimation as a practical solution to this dilemma.

2.2.1. Self-generated estimation of cover

This idea was proposed in [16] and it is based on the following formulation of steganography:

$$s(t) = c(t) + n(t) \tag{5}$$

where, s(t), c(t), and n(t) denote the stego, cover, and steganography noise, respectively. It is noteworthy that equation (5) holds for both cover-independent (like LSB embedding) and cover dependent (like echo) noises. If n(t) is found effectively, virtually a good estimation of the cover signal would be available. This is the main rationale behind steganalysis methods of this category. Basically these methods estimate the cover by applying a noise removal procedure on the signal. It is expected that different amount of noise is extracted from stegos and covers. These methods have employed different metrics to quantify such discrepancies. Let s(t) and $\tilde{c}(t)$ denote original signal and its estimated cover, Figure 5 shows block diagram of these methods.

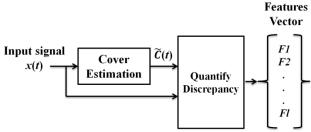


Figure 5. Self-calibrated feature extraction

2.2.1.1. Steganalysis based on Audio Quality Metrics (AQM)

This idea was proposed by Ozer el al. [16] and the wavelet-thresholding was employed for noise removal and estimation of the cover. If discrepancies between estimated cover and stego signals are measured accurately, stegos and covers can be distinguished. Authors noticed that, in speech coding applications, researchers have been trying to address a similar problem for measuring artifacts of speech coders for a long time. Results of those endeavours is a set of measurements known as audio quality metrics (AQM), which are categorized into three groups of time, frequency, and perceptual domains. AQMs measure discrepancies between short frames of the reference signal and its modified version and the final metrics are calculated as the average value of each metric over all of frames.

The same methodology was adapted for steganalysis where signals under inspection and their denoised version were considered as reference and modified signals, respectively. Then their discrepancies were evaluated through AQM metrics. Finally, these metrics were used as steganalysis features. In [16] different combinations of feature selections and machine learning methods were tested.

2.2.1.2. Steganalysis based on Hausdorff Distance

Liu et al. argued that AQM have been designed specifically for objective assessment of the quality of audio signals and not steganography impairments [20]. They noted

that, this argument is especially true about perceptual metrics. Therefore, it is very likely that these features have limited capability to capture the effects of embedding. To alleviate such problems, some researchers have used Hausdorff distance for measuring the effect of message embedding [20, 21]. Hausdorff distance is fundamentally a max-min measure which has found numerous applications in template matching and content-based retrieval problems. These works also used multi-resolution decomposition capability of the wavelet transform for magnifying discontinuities of the signal after embedding.

To extract features, estimated cover and the signal under scrutiny were segmented and decomposed into different sub-bands. Then, Hausdorff distance between each pair of sub-bands was calculated. Final features were HOS of Hausdorff distances over all frames. The system was tested on Steghide [7] and the results were compared with other steganalysis methods. Furthermore, discriminative ability of different sub-bands was investigated. It was shown that lower levels of wavelet decomposition produce more potent features.

Geetha et al. used the same idea for feature extraction, but investigated efficacy of six decision tree classifiers on different steganography and watermarking methods [21].

2.2.1.3. Modelling the noise based on GMM and GGD

Another solution for distinguishing between stego and cover is directly comparing distribution of their wavelet coefficients. Work of [22] used Gaussian mixture model (GMM) and generalized Gaussian distribution (GGD) for this purpose. To that end, two different methods were proposed. The first method modeled distribution of coefficients of all sub-bands of wavelet with GMM and used them for steganalysis. In the second method, GGD and GMM were employed for capturing distributions of coefficients of signal and its de-noised version. Then, deviations between the two distributions were calculated with four different distance measures. Unfortunately, this work did not provide any result on the performance of the system.

The same idea was investigated more properly in [23]. The authors showed that spread spectrum embedding causes the histogram of these coefficients to become flatter around zero. GMM and GGD were employed to capture this flatness. The results showed that GMM captures artifacts of spread spectrum hiding more accurately. Potency of individual features were measures on two different spread spectrum methods but the system was not tested on combination of more than one features. Results showed that the best performance is achieved when three Gaussian kernels are mixed.

2.2.2.Constant referencing (CIAQM)

Previous methods used original signal for estimating the cover; therefore, they are self-referencing. Avcibas showed that self-referencing (e.g., denoising) leads to dependence of features on the content of signals [24]. To alleviate this problem, he proposed method of constant referencing. That is, two fixed reference signals were selected, one of which was a cover and the other one was its stego version. Let $r, r+\varepsilon$, and M denote the reference cover, its stego version, and the metric that measures their discrepancies,

respectively. For every incoming signal (s), the proposed features were defined as:

$$f(s,r,r+\varepsilon) = M(s,r) - M(s,r+\varepsilon) \tag{6}$$

It was shown that features of equation (6) were independent from the cover signal. The steganalysis system was constructed with different AQMs as the discrepancy metrics (*M*) and linear regression for classification. The efficacy of this method was tested on six different embedding algorithms. Simulations showed that constant referencing outperforms self-referencing methods. Also, results showed that this improvement is more noticeable for steganography methods.

2.2.3. Estimation of cover space model

While previous works used discrepancies between a reference and the incoming signal for making the decision, another approach is possible if the high dimensional model of covers space is known. In this fashion the warden checks every signal x(t) with this model, if it fits the model, it is a cover and a stego otherwise. Unfortunately, a perfect model does not exist for empirical cover sources [4]. But this idea could be exploited for estimating a model for empirical cover signals. This approach was pursued by Johnson et al [25]. First short time Fourier transform (STFT) was invoked to capture both time and frequency regularities of audio signals. Then PCA was invoked to extract a set of linear bases that were localized in both time and frequency domains. These linear bases form a vector quantizer and the residual signal from this quantization can indicate accuracy of this sub-space for modeling the audio signal. Therefore, features were calculated as HOS of quantization error. Simulations were conducted on LSB embedding methods and results showed that this method achieves reasonable accuracy in LSB embedding when at least 4-bits are used for embedding.

2.2.4. Re-embedding calibration

Previous calibration methods are common in the sense that characteristics of cover signals were estimated. Another possibility is estimation of stego characteristics. This path was pursued in [26]. The paper addressed both targeted and universal cases. In the targeted paradigm the embedding algorithm is known; therefore, signals were embedded with a random message and the same embedding algorithm. Then, difference between features extracted from original signal and its embedded version were used for steganalysis. This technique was later extended into universal paradigm, where embedding algorithm is not known. For this purpose, notion of bit-plane sensitivity was defined as the amount of noise that is introduced in each bit-plane after random embedding. Sensitivity of different bit-planes of a wide range of data hiding algorithms were investigated. Analysis showed that 1-LSB was the most sensitive bit-plane; therefore, LSB reembedding was employed as a universal calibration method. Additionally, potency of cepstrum features were compared with energy of filter banks and it was shown that energy of filter banks had far better performance. Finally, due to positive effect of feature normalization and superior performance of GA for feature selection [27], features were normalized and GA was invoked for feature selection.

2.3. Summary

In this section a summary of the investigated methods is presented. First, we take a brief look on the relation between the reviewed papers. Figure 6 presents how different works are related to each other.

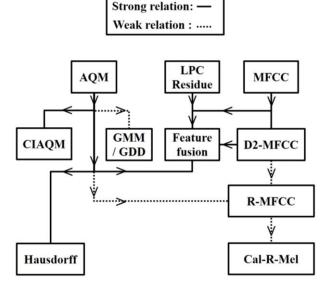


Figure 6. Relationship between different papers

| AQM: | [16] | Fusion: | [15] |
|--------|----------|------------|----------|
| LPC: | [6] | D2-MFCC: | [14] |
| MFCC: | [13] | R-MFCC: | [17, 18] |
| CIAQM: | [24] | Hausdorff: | [20, 21] |
| GMM: | [22, 23] | Cal-R-Mel | [26] |

One of the major differences between the reviewed papers was the embedding methods that they had investigated. Those methods are presented in three different tables. Table 1 summarizes the watermarking methods, the symbol that is used for referring to each of them, and their embedding domain. These methods include: 2A2W [28], COX [29], DSSS [30], Echo [30], FHSS, LSB watermarking [28], spread spectrum audio watermarking (SSAW) [31], WAWW [28], and multi carrier spread spectrum (MCSS) [32].

Table 1. Specifications of watermarking methods

| Symbol | Name | Domain | Ref. |
|-----------|------|---------|------|
| W1 | 2A2W | wavelet | [28] |
| W2 | COX | Freq. | [29] |
| W3 | DSSS | Time | [30] |
| W4 | Echo | Time | [30] |
| W5 | FHSS | Freq. | - |
| W6 | LSB | 1-LSB | [28] |
| W7 | SSAW | Freq. | [31] |
| W8 | VAWW | wavelet | [28] |
| W9 | MCSS | Freq. | [32] |

Table 2 summarizes the LSB steganography methods and the symbol that is used for referring to each of them. These methods include: Heutling [28], Hide4pgp [33], Invisible Secret [34], LSB Matching [35], Steganos [36], Steghide [7], Stools [37].

Table 2. Specifications of LSB-based steganography methods

| Symbol | Name | Domain | Ref. |
|--------|------------------|---------|------|
| SL1 | Heutling | 1-LSB | [28] |
| SL2 | Hide4pgp | 1- 4LSB | [33] |
| SL3 | Invisible Secret | 1-LSB | [34] |
| SL4 | LSB Matching | 1-LSB | [35] |
| SL5 | Steganos | LSB | [36] |
| SL6 | Steghide | 1-LSB | [7] |
| SL7 | Stools | LSB | [37] |

Information of non-LSB methods are presented in Table 3. These methods include: Addition Method (AM) [8], Amplitude Modulation (AMod) [38], DWT-FFT [39], DWT fusion [40], mp3Stego [41], Publimark [28], Quantization Index Method (QIM) [8], Stego wave [42], Stochastic Modulation (StMod) [43], WaSpStego [13], Wavelet LSB [8], Integer wavelet (I-Wavelet) [44], and DSSS in the frequency domain (DSSS + DCT) [45].

Table 3. Specifications of non-LSB-based steganography methods

| Symbol | Name | Domain | Ref. |
|--------|-------------|-----------------|------|
| SN1 | AM | wavelet | [8] |
| SN2 | AMod | Freq. | [38] |
| SN3 | DWT-FFT | Wavelet + Freq. | [39] |
| SN4 | DWT fusion | wavelet | [40] |
| SN5 | Mp3stego | Side Info | [41] |
| SN6 | Publimark | no information | [28] |
| SN7 | QIM | wavelet | [8] |
| SN8 | Stego wave | no information | [42] |
| SN9 | StMod | Time | [43] |
| SN10 | WaSpStego | wavelet | [13] |
| SN11 | Wavelet LSB | wavelet | [8] |
| SN12 | I-Wavelet | DWT | [44] |
| SN13 | DSSS + DCT | Freq. | [45] |

Table 4 presents which embedding algorithms were investigated by each steganalysis method.

Different criteria of true positive rate (TPR), true negative rate (TNR) and accuracy (Ac.) may be defined to report efficacy of classification. Other important differences between the reviewed methods were domains of feature extraction, number of features, type of classifier, number of clean files in their databases, size of training set, and average of their performance. These parameters are summarized in Table .

3. Targeted non-compressed methods

Despite having wide dynamic range, HAS exhibits low differential sensitivity. Additionally, HAS is insensitive to the absolute value of phase and can only perceive relative phases [30]. Based on these characteristics different embedding methods have been proposed. This section presents steganalysis systems that have been designed specifically for their detection.

3.1. Phase coding methods

steganalysis of phase coding system was investigated in [46]. This work observed that phase embedding preserves the relative phase of each block, but the phase between consecutive blocks is changed. To capture signatures of phase embedding, signal was segmented and unwrapped phase of all segments were calculated. The steganalysis features were statistical moments of the absolute difference between phases of consecutive segments. The system was tested for different length of blocks and sub-blocks.

3.2. Echo embedding methods

Echo hiding methods have a bank of kernels and depending on the bits of message, one of them is convolved with each segment of cover [47]. Therefore, kernels play a major role on the characteristics of echo hiding methods. For example, it is possible to design kernels that achieve better robustness. Positive-negative (PN) and forward-backward (FB) are examples of kernels that have been designed for such

Table 5. Summary of universal steganalysis papers. Notations of the table are as follows -: not reported parameters, ★: non-applicable parameters, M: music files and S: speech files

| Method | Reference | Feature | Feature | Classifier | Clean | Train | Universa | l Results | Targeted | Results |
|------------------|-----------|----------|---------|------------|-------|-------|----------|-----------|----------|----------|
| | | Domain | No. | | No. | Size | TPR | TNR | TPR | TNR |
| IDS | [5] | Bits | 3 | Threshold | 30 | - | - | - | | |
| LPC | [6] | Wavelet | 40 | SVM | 500 | 50% | - | - | 98.93 | 95.76 |
| Wavelet + PCA | [8] | Wavelet | 36 | NN | 400 | 62.5% | - | - | 96.22 | 95.11 |
| Chaotic | [9] | chaotic | 22 | SVM | 2554 | 50% | - | - | M:69.14 | M:58.56 |
| | | | | | | | | | S: 91.37 | S: 89.89 |
| Markov | [10] | Time | 81 | SVM | 12000 | 50% | - | - | Ac. = | = 92.2 |
| Markov | [11] | Time | 169 | SVM | 19380 | 70% | - | - | Ac. = | = 97.3 |
| AR-TDNN | [12] | Time | | NN | 150 | | - | - | 68.48 | 78.29 |
| MFCC | [13] | Cepstrum | 36 | SVM | 389 | 80% | - | - | 66.04 | - |
| D2-MFCC | [14] | Freq. | 29 | SVM | 12000 | 50% | - | - | Ac. = | = 85.9 |
| Feature Fusion | [15] | Freq. | 12 | SVM | 600 | 67% | Ac. | = 91 | Ac. | = 90 |
| AQM | [16] | Time + | 19 | SVM | 664 | 50% | 81.8 | 79.7 | 93.5 | 92.75 |
| | | Freq. | | | | | | | | |
| RMFCC | [17] | Freq. | 29 | SVM | 4169 | 70% | | | 97.29 | 94.71 |
| RMFCC + HOS+GA | [18] | Freq. | 21 | SVM | 4169 | 70% | 94.4 | 99.1 | 99.1 | 99.0 |
| Hausdorff | [20] | Wavelet | 25 | SVM | 994 | 90% | - | - | 95 | 88 |
| Hausdorff + tree | [21] | Wavelet | 25 | J48 Tree | 200 | 75% | - | - | 88.64 | 72.89 |
| DWT+GMM | [23] | Wavelet | 1 | SVM | 1000 | 50% | | | 93.44 | 91.22 |
| CIAQM | [24] | Time + | 19 | linear | 100 | 50% | - | - | 95 | 95 |
| | | Freq. | | regression | | | | | | |
| STFT+PCA | [25] | STFT | 4 | SVM | 1800 | 80% | - | - | 56.95 | 98.1% |
| Cal-R-Mel | [26] | Freq. | 15 | SVM | 4169 | 90% | M:98.7 | M:99.8 | M:99.5 | M:99.7 |
| | | | | | | | S:87.8 | S: 96.7 | S:94.9 | S: 93.9 |

Table 4. Details of steganography methods investigated by each steganalysis paper

| | | | | | | | | | | ganaly | | | | | | | | | |
|------|-----|-----|-----|-----|------|------|------|----------|------|--------|----------|------|------|------|------|------|----------|------|------|
| W1 | [5] | [6] | [8] | [9] | [10] | [11] | [12] | [13] | [14] | [15] | [16] | [17] | [18] | [19] | [21] | [23] | [24] | [25] | [26] |
| W2 | | | | ✓ | | | | | | | ✓ | ✓ | ✓ | | ✓ | | | | ✓ |
| W3 | | | | ✓ | | | | √ | | ✓ | ✓ | | | | | ✓ | ✓ | | |
| W4 | | | | ✓ | | | | | | ✓ | ✓ | | | | ✓ | | ✓ | | |
| W5 | | | | ✓ | | | | | | | ✓ | | | | | ✓ | ✓ | | |
| W6 | | | | | | | | ✓ | | ✓ | | | | | | | | | |
| W7 | | | | | | ✓ | | | | | | ✓ | ✓ | | | | | | ✓ |
| W8 | | | | | | | | ✓ | | | | | | | | | | | |
| W9 | | | | | | | | | | | | | | | | | | | ✓ |
| SL1 | | | | | | | | ✓ | | | | | | | | | | | |
| SL2 | | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ |
| SL3 | | | | | ✓ | ✓ | | | ✓ | | | | | | | | | | |
| SL4 | | | | | ✓ | ✓ | | | ✓ | | | | | | | | | | |
| SL5 | | | | ✓ | | | | | | | ✓ | | | | | | ✓ | | |
| SL6 | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| SL7 | | ✓ | | | | | | | | | ✓ | | | | ✓ | | | | |
| SN1 | | | ✓ | | | | | | | | | | | | | | | | |
| SN2 | | | | | | | | | | | | | | | ✓ | | | | |
| SN3 | | | | | | | ✓ | | | | | | | | | | | | |
| SN4 | | | | | | | | | | | | | | | ✓ | | | | |
| SN5 | ✓ | | | ✓ | | | | | | | | | | | | | | | |
| SN6 | | | | | | | | ✓ | | | | | | | | | | | |
| SN7 | | | ✓ | | | | | | | ✓ | | | | | | | | | |
| SN8 | | ✓ | | | | | | | | | | | | | | | | | |
| SN9 | | | | ✓ | | | | | | | | | | | | | | | |
| SN10 | | | | | | | | ✓ | | | | | | | | | | | |
| SN11 | | | ✓ | | | | | | | | | | | | | | | | |
| SN12 | | | | | | | | | | | | | ✓ | | | | | | ✓ |
| SN13 | | | | | | | | | | | | | ✓ | | | | | | ✓ |

purposes. The mathematical representation of famous kernels are as follows:

$$h_{Basic}[n] = \delta[n] + \alpha \delta[n - d_i] \tag{7}$$

$$h_{PN}[n] = \delta[n] + \alpha \delta[n - d_i] - \alpha \delta[n - d_i']$$
 (8)

$$h_{PN}[n] = \delta[n] + \alpha \delta[n - d_i] - \alpha \delta[n - d'_i]$$

$$h_{FB}[n] = \delta[n] + \alpha \delta[n - d_i] + \alpha \delta[n + d_i]$$
(8)
If binary encoding of message is used, then $i \in \{0,1\}$.

Targeted steganalysis of the basic kernel was first investigated in [48]. It was shown that for short windows, stegos have a peak in their power cepstrums. Therefore, the power cepstrum of the signal was calculated on short windows and different moments of local maxima of all windowed-cepstrums were used for steganalysis. The method investigated effect of different parameters of the system.

Steganalysis of echo hiding with PN kernels was discussed in [49]. They showed that the power cepstrum and the complex cepstrum of the stego signal exhibit peaks in delay positions of the kernels. To detect these artifacts, the audio signal was segmented into 20ms chunks. Skewness of both the power cepstrum and the absolute value of the complex cepstrum over all frames were calculated. The steganalysis feature was defined as kurtosis of the calculated values of skewness. Simulation results showed that features based on the absolute value of complex cepstrum were more discriminative than power cepstrum-based features.

Xie et al. took one step further in the steganalysis of echo systems and presented an active attack [50]. The proposed system was based on studying the behaviour of cepstrum of a sliding window. Based on this work, using a sliding window W_s smaller than the segment length, four situations will exist: 1) when W_s is inside a zero embedded segment, 2) when it is inside a one embedded segment, 3) when it is crossing from a one (zero) embedded into another one (zero) embedded segment, and 4) when it is crossing the border of two segments with different hidden bits. The work showed that if the peak position of each sliding cepstrum is recorded, it exhibits high density around the value of delays. Furthermore, cepstrum peak location aggregation rate (CPLAR) was introduced as the ratio of the number of times that the cepstrum peak location was at one of those positions to the total number of windowed cepstrums. CPLAR was used for deciding about the presence/absence of the hidden message. The method was also able to find the length of segmentation in the embedding system. Extensive simulations were presented on different kinds of echo kernel.

The results showed that if PN kernels are used, detection and estimation of its parameters are harder.

Recently, a more secure echo method was proposed [47]. This work proposed that segment length and parameters of echo kernels be varied in a pseudo random fashion. True simulation it was shown that this strategy makes CPLAR feature obsolete. Also, they showed this technique improves performance of the system in the presence of active warden.

3.3. Multiplicative embedding methods

A multiplicative steganography scheme may be expressed as:

$$s(t) = c(t)(1 + m(t)) \tag{10}$$

According to equation (5) the steganography induced noise is equal to:

$$n(t) = c(t).m(t) \tag{11}$$

Apparently, this noise is cover dependent and multiplicative. Steganalysis of such embedding methods was presented in [51]. It was claimed that the conventional steganalysis methods cannot detect multiplicative systems properly. To address this problem, the logarithm of absolute values of audio samples were used. In this fashion multiplicative noise was transformed into additive noise. Then, wavelet transform was applied on the new signal and statistical moments of signal and its sub-band coefficients were extracted for steganalysis. These features were mixed with the statistics of the linear prediction residue of each sub-band [6]. The experimental results showed this technique improves detection of multiplicative steganography systems.

3.4. Summary

Targeted steganalysis papers are summarized in table 6.

4. Mp3 Steganalysis

The mpeg-1 layer III (mp3) can provide high compression rate and good quality. These characteristics have turned mp3 into one of the most popular formats of audio signals. Consequently, it provides a suitable medium for steganography.

Over the past decade different mp3 steganography methods have been developed. They include Mp3stego [41], under mp3 cover (Ump3c) [52], Mp3stegz [53], Huffman table swapping [54], quantization step parity [55], Linbit [56], Bv-Stego [57], and other methods [58-60]. Current literature has only investigated steganalysis of Mp3stego and Ump3c. This section reviews their findings.

4.1. Steganalysis of Mp3stego

Mp3 encoding procedure consists of two nested loops. In this fashion, the "inner loop" does the actual quantization

of the data and determines the suitable step of quantizer to meet with the available bit budget. On the other hand, the "outer loop" controls distortion of the encoding process and keeps it beyond the level of human perception. Mp3stego changes the termination condition of inner loop until the embedding rule is satisfied. Therefore, hiding of the message happens while the audio signal is being encoded into MP3 bit stream.

Using notation of mp3 standard, the "part2_3_length" determines the number of bits that is used for encoding data of each frame. Mp3stego embeds bits of message as the parity of the "part2_3_length" variable by controlling when the "inner loop" is terminated [41]. Because Mp3stego hides the information during the compression process, many parts of the mp3 are changed. Thus, steganalysis of Mp3stego can be accomplished in different ways. They include variances of block sizes, numbers of different block lengths, MDCT coefficients, statistics of the bit reservoir, and statistics of quantization step.

4.1.1. Histogram of block length

Westfeld investigated Mp3stego algorithm and concluded that modified frames are one step size smaller than their normal counterparts [61]. In other words, if the maximum length of frames were fixed, stego files would had lower bit rates. Of course this is not the case with mp3 standard. Actually, the algorithm adjusts length of blocks to achieve the target average bit rate (e.g. 128k). Therefore, if bit rate of one block is decreased, next blocks will use those extra bits. This work showed that while the mean of block size in covers and stegos is the same their variance is not.

4.1.2. Number of different block lengths

Investigating histograms of stegos and covers reveals that their block length is different. Specifically, Mp3stego leads to more different block lengths to be used. Based on this idea an ultra-light-weight steganalysis system was proposed in [5]. This system used length of first block (F) as an estimation of the expected value of block length. Then the number of different block lengths in the file was calculated (C). After investigating the ratio of F/C for covers and stegos, they proposed the empirical value of 4.6 as the appropriate threshold. In this fashion, the ratio of F/C was larger than 4.6 for covers.

4.1.3. Statistics of MDCT coefficients

Mp3stego increases quantization step of encoder. Apparently, this technique decreases the absolute value of MDCT coefficients. If statistical distributions of the MDCT coefficients are known, this extra distortion could be detected. This idea was pursued by Qiao et al. [62]. They extracted different statistical metrics from signal to capture any

Table 6. Summary of targetd steganalysis papers. Notations of the table are as follows -: not reported parameters, ★: non-applicable parameters

| Ref. | Target | Domain | Feature | Classifier | Cover | Train | Perfor | Performance | | |
|------|----------------|---------|---------|------------|-------|-------|--------|-------------|--|--|
| | method | | No. | | No | size | TPR | TNR | | |
| [46] | Phase | Freq. | 5 | SVM | 800 | 25% | 98.2 | 95 | | |
| [48] | Echo | Freq. | 8 | SVM | 1200 | 50% | 82.17 | 87.83 | | |
| [48] | Echo | Freq. | 1 | SVM | 300 | 17% | 94 | 89.2 | | |
| [50] | Echo | Freq. | 1 | Threshold | 200 | × | - | - | | |
| [51] | Multiplicative | wavelet | 40 | SVM | 450 | 44% | 94.8 | 96.4 | | |

anomaly in its MDCT coefficients. First, GGD was used to model distribution of the MDCT coefficients in each frame. Second set of features were calculated from different subbands of the second order derivatives of the MDCT coefficients. The third and fourth sets of features were Markov transition probabilities of inter-frame and intra-frame MDCT coefficients. In the simulation, different combinations of features were considered and the best result was achieved when all features were used. This work was improved in [63], which added feature selection to the method and used a bigger database for evaluation.

Jin et al. used the same concept and extracted joint probability of adjacent MDCT coefficients in the same channel or the neighbor channels [64]. The work showed that:

- 1) Features along the minor diagonal direction were more discriminative.
- 2) Features in the center of the transitions matrix were more powerful.

Based on these observations a method for reducing the number of features was proposed. This work was extended in [65] and features were extracted from difference of absolute values of MDCT in the inter and intra frames of mp3.

4.1.4. Calibrated steganalysis

Audio signal has a wide dynamic range, implying that some of its portions have faster transitions and are more complex. Apparently, reconstruction of these portions need more bits. To solve this problem, mp3 standard benefits from a short buffer mechanism called *bit reservoirs*. This technique keeps the average bit rate of the frames constant but allows individual frames to have different bit rates. In this fashion, frames with low complexities are encoded with fewer bits and the extra bits are saved for frames that are more complex.

Signals embedded with Mp3stego typically have larger quantization steps. Therefore, their encoding requires fewer bits and those extra bits are stored in *bit reservoirs*. Consequently, statistics of the bit reservoir of stego would be different. This idea was presented in [66]. Furthermore, recompression calibration was used to remove effect of audio contents from the extracted features. The system was tested on mp3 with different bit rates.

4.1.5. Statistics of quantization step

Yan et al. observed that average value of the quantization steps between stegos and covers are the same but the difference between quantization steps of adjacent granules is increased [67]. According to the mp3 standard,

both granules of the same frame use the same psychoacoustic model. Furthermore, it is logical to assume that consecutive granules have similar characteristics. This paper argued that embedding proves reduces those similarities. Assuming that the current granule is selected for embedding and its parity does not agree with the embedding bit, the algorithm increases the quantization step and adds the extra bits to the bit reservoir. If the next granule uses those extra bits, its quantization step is decreased. Therefore, the difference between quantization steps of two consecutive granules would increase.

4.2. Steganalysis of Ump3c

Ump3c is another steganography tool for mp3 that hides information in the LSB of *global gain* of each granule. Therefore, unlike Mp3stego, it works directly on mp3 files. It is noteworthy that Ump3c and Mp3stego have the same capacity.

Active steganalysis of Ump3c was conducted by Jin et al. [68]. Their method is a modified version of the regular-singular (RS) image steganalysis. In the RS steganalysis, an invertible flipping function is defined. Then effect of applying the flipping function on the noise of signal is evaluated. Based on this criterion three groups of regular, singular and neither are defined for the case of increase, decrease, or no change in the noise of signal. The method also uses a mask for determining which samples should be flipped. Number of regular groups should be larger than number of singular groups for covers.

For steganalysis of Ump3c, global gains of mp3 were extracted. After dividing them into groups of M samples, noise of each group was measures as:

$$f(G) = \sum_{i=1}^{M-1} |gg_k[i] - gg_k[i+1]|$$
 (12)

where, k is the index of group. The flipping function was defined as:

$$F_1: 0 \leftrightarrow 1, 2 \leftrightarrow 3, \dots, 254 \leftrightarrow 255 \tag{13}$$

$$F_{-1}: -1 \leftrightarrow 0, 1 \leftrightarrow 2, \dots, 255 \leftrightarrow 256$$
 (14)

The number of regular and singular groups were determined for both flipping functions. These statistics were used for calculating the amount of hidden data. The system was tested with different embedding rates and different masks.

4.3. Summary

Summary of reviewed papers are presented in Table

Table 7. Summary of mp3 steganalysis papers. Notations of the table are as follows SI: side information, BPB: bit per bit, R: maximum capacity ratio, -: not reported parameters, ★: non-applicable parameters

| Ref. | Target Method | Domain | Feature No. | Classifier | Train size | Database Spec. | | Perfor | mance |
|------|----------------------|--------|--------------------------------------|------------|------------|----------------|--------------|---------|-------|
| | | | | | | Clean No. | Min Capacity | TNR | TPR |
| [5] | Mp3stego | SI | - | Threshold | × | - | 12.6% R | 84.4 | 79.3 |
| [62] | Mp3stego | MDCT | 214 | SVM | 75% | 1000 | 16% R | Ac. = 9 | 4.1 |
| [63] | Mp3stego | MDCT | <200 | SVM | 60% | 5000 | 40% R | Ac. = 9 | 1.35 |
| [64] | Mp3stego | MDCT | 64 kb:41 128 kb:115 192 kb:212 | SVM | - | 3000 | 10% R | 86.48 | 94.37 |
| [66] | Mp3stego | SI | 1 | SVM | 50% | 1200 | 0.001% BPB | 75 | 73.38 |
| [67] | Mp3stego | SI | 1 | Threshold | × | 1456 | 10% R | 80 | 96.72 |
| [68] | Ump3c | SI | × | × | × | 200 | 3.3% R | × | × |

5. Steganography and steganalysis of AAC

Advanced audio coding (AAC) is another popular compressed audio format which is used by many audio/video streaming services and websites. AAC is the successor of mp3 and both of them use perceptual coding and entropy coding for achieving high compression rate while maintaining high quality of signal. There are lots of similarities between the two, but ACC can achieve the same level of quality for 70% of bit rate of mp3 [69].

5.1. AAC steganography

AAC bit stream has different components and some of them have been exploited for steganographic purposes. They include Huffman table information, MDCT coefficients, and quantization parameters.

Possibility of changing Huffman coding section for embedding was pursued in [70]. Another possible place for data hiding is the sign bit of code words [71]. This work used an interesting technique for minimizing the distortion at the expense of reducing the capacity. If XOR of two consecutive sign bits did not agree with bit of message, sign bit of the smaller coefficient was changed [71]. LSB embedding in MDCT coefficients is another promising trend which was proposed in [72]. For this purpose, embedding was done during the encoding process. More specifically, embedding algorithm enforces usage of a smaller scale factor and exploit the extra bit for carrying the message.

Some other embedding methods are direct adaptation of mp3 algorithms to AAC format. Huffman tables of AAC can encode values of MDCT between 0 and 15. Frames with larger MDCT use an especial symbol known as escape sequence. LSB of escape sequence was used for hiding information in [73]. This method is counterpart of Linbit embedding in mp3 [56]. Bitrate steganography on AAC, hides message as the parity of the number of bits used for each frame [74]. Conducting a comparison with mp3stego shows that AAC bitrate steganography is the adaptation of mp3Stego to AAC format.

5.2. AAC steganalysis

5.2.1. Calibrated Markov transition probability

Work of [75] showed that data hiding algorithm of [70] changes correlation between adjacent scale factor bands. Based on this observation, Markov transition probability between indexes of Huffman codebook of consecutive scale factor bands were used. The method only investigated correlation between tables 1 to 10, so steganalysis system was constructed with 100 features. Finally, potency of steganalysis features were improved by re-compression calibration. Simulation results showed that calibration improves the results by about 10%.

5.2.2.Difference of inter and intra frame probabilities

Transitions in audio signals are very smooth, so adjacent samples are highly correlated in both time and frequency domains. This characteristic was exploited for steganalysis of Huffman table sign method [71]. This approach was proposed in [76] where statistical characteristics of inter-frame and intra-frame MDCT coefficients were used for steganalysis. The method

constructed a rich model consisting of 1296 features and used ensemble classifier. To construct the rich model, frames of AAC were divided into two groups of short and long based on their block types. Then, Markov transition probability and accumulative neighbouring joint density of first and second order derivative of inter and intra frames of each group were used as the final features. Experiment results, showed detection rate of 85.34% for embedding capacity of 50%.

6. Evaluation of Audio Steganalysis methods

6.1. Non-compressed methods

Reviewing previous works on audio steganalysis shows some shortcomings:

- 1- Most image steganalysis methods have been tested on BOSS or BOWS databases. Therefore, their reported results are comparable. On the other hand, in audio steganalysis such standard databases are not present and each work has used a different database. Evaluating performance of audio steganalysis methods on the same database makes a fair comparison between them easier, and alleviates this problem for future referencing.
- 2- Different audio steganography techniques have been proposed, most of which are non-LSB methods. Referring to table 4, it is evident that there is not a balance between the amount of work on LSB and non-LSB methods. Therefore, non-LSB methods have not been investigated properly.
- 3- In practical situations the warden does not have any prior knowledge about the embedding algorithm. Thus, universal steganalysis resembles more closely with practical situations. According to table 5, most of the previous works have only been investigated in targeted scenario. Consequently, universal steganalysis has not been investigated properly.
- 4- Previous works have shown that complexity of signal plays an important role on the performance of steganalysis system [11]. But, many of existing works have not investigated this property.

In this section we try to fill these gaps. To that end, we used database of [18] which contained 4169 excerpts. Then, we embedded each cover with different methods and at different embedding capacities with random messages. LSB methods of Hide4pgp [33] and Steghide [7], non-LSB method of integer wavelet (I-wavelet) [44], and watermarking method of COX [29] were used for this purpose. In this manner our database had a total number of 54197 excerpts. We implemented some of the most recent and well-cited audio steganalysis methods and evaluated them in the both targeted and semi-blind universal paradigms. We used 10fold cross validation with support vector machine (SVM) for this purpose. Furthermore, if a certain method had preprocessing (ex. feature normalization) or post-processing (feature selection) they were also implemented. Finally, complexity of all files were measured with the metric of equation (2) and then they were divided in three distinct sets. These regions were defined as follow:

low complexity: $C \le 0.06$ (15)

medium complexity: $0.06 \le C \le 0.12$ (16)

 $high\ complexity: 0.12 \le C$ (17)

First, we have evaluated each method in the targeted paradigm. Table 8 presents results of this analysis. The best results are shown in bold face letters.

Receiver operating characteristics (ROC) of different feature sets for detection of Steghide at capacity of 0.25 BPS for low and medium complexities are shown in figure 7.

Table 8. Accuracy of different feature sets in targeted scenario. Capacity is expressed in terms of bit-per-symbol (BPS) and accuracies lower than random guess (50%) are marked by ×. Finally, the number of features in each method is mentioned after their name in the ().

| | Capacity/ | | D2-MFCC (29) | R-MFCC (29) | R-MFCC+ | Markov (169) | Cal-R-Mel (15) |
|-----------|-----------------|------------|--------------|-------------|--------------------|--------------|----------------|
| Method | Param. | Complexity | [14] | [17] | HOS+GA (7) [18] | [11] | [26] |
| | | low | 94.6 | 89.6 | 99.7 | 99.9 | 100 |
| | C = 4 | medium | 95.4 | 87 | 99.7 | 100 | 100 |
| | | high | 91.6 | 85.7 | 99.7 | 100 | 100 |
| Hido4nan | | low | 93.5 | 87.6 | 99.5 | 99.2 | 99.9 |
| Hide4pgp | C = 2 | medium | 86.6 | 86.8 | 99.1 | 99.9 | 99.9 |
| | | high | 67.2 | 71.9 | 98.5 | 99.9 | 100 |
| | | low | 77.2 | 88.5 | 99.1 | 99.1 | 100 |
| | C = 1 | medium | 60.6 | 77.2 | 97.5 | 99.5 | 99.7 |
| | | high | × | × | 95.5 | 99.7 | 99.9 |
| | | low | 60.3 | 83.3 | 98.7 | 78 | 99.7 |
| | C = 0.5 | medium | × | 57.9 | 95.9 | 92.1 | 99.4 |
| | | high | × | × | 93.9 | 97.7 | 99.9 |
| Steghide | | low | × | 65.3 | 97.5 | 63.5 | 99.3 |
| Stegmae | C = 0.25 | medium | × | × | 93.6 | 83.1 | 99.4 |
| | C = 0.23 | high | × | × | 91.2 | 94.2 | 99.9 |
| | | low | × | × | 94.9 | 54.1 | 98.8 |
| | C = 0.12 | medium | × | × | 89.6 | 68.1 | 99.1 |
| | | high | × | × | 86.8 | 86.9 | 99.2 |
| | | low | 94.4 | 88.7 | 99.7 | 99.7 | 100 |
| | C = 2 | medium | 94.8 | 86.8 | 99.5 | 100 | 100 |
| | | high | 88.1 | 84.7 | 99.7 | 100 | 100 |
| | | low | 88.6 | 88.4 | 99.4 | 99.3 | 99,9 |
| | C = 1 | medium | 77.4 | 86.7 | 98.6 | 99.8 | 99.8 |
| | | high | 55.6 | 59.2 | 97.7 | 99.8 | 99,9 |
| | | low | 59.9 | 81.4 | 98 | 97.4 | 99.7 |
| | C = 0.5 | medium | × | 50.6 | 95 | 99 | 99.4 |
| I-wavelet | | high | × | × | 92.9 | 99.5 | 99.8 |
| | | low | × | 58.7 | 96.7 | 85.3 | 99.1 |
| | C = 0.25 | medium | × | × | 90.4 | 95.8 | 99 |
| | | high | × | × | 86.8 | 99 | 99.7 |
| | | low | × | × | 91.3 | 64 | 98.6 |
| | C = 0.12 | medium | × | × | 82.5 | 89.5 | 98.4 |
| | | high | × | × | 73.8 | 96.9 | 98.2 |
| | | low | 95.4 | 95.9 | 99.9 | 100 | 99 |
| COX | $\alpha = 0.01$ | medium | 96 | 94 | 99.4 | 100 | 98.7 |
| | | high | 92.8 | 92.6 | 99.6 | 100 | 99.5 |

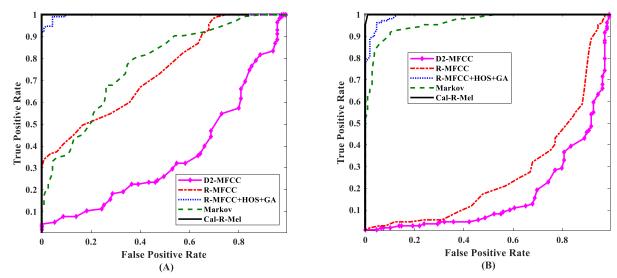


Figure 7. ROC of different feature sets for detection of Steghide at capacity of 0.25 BPS in the targeted paradigm (A) low complexity (B) medium complexity

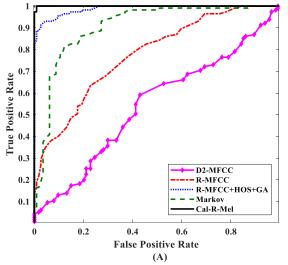


Figure 8. ROC of different feature sets in the universal paradigm

Investigating results of table 8 and figure 7 shows that Cal-R-Mel has the best performance. Also, for most feature sets we see a negative correlation between complexity and performance. That is, signals with higher complexities are harder to detect. But at the same time, Markov feature does not follow this pattern. It is quite possible that using a better complexity measure more meaningful results are achieved.

Performance of different methods in the universal paradigm is compared. These results are shown in table 9 with the best results shown in the bold face letters.

Table 9. Comparison of different methods in the universal paradigm

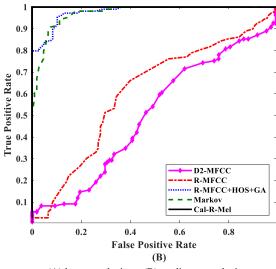
| Method | Ref. | Feature | Complexity | | Results | |
|--------|------|---------|------------|------|---------|------|
| | | No | | TPR | TNR | Ac. |
| | | | low | 60.8 | 54.3 | 57.5 |
| D2- | [14] | 29 | medium | 47.1 | 45.1 | 46.1 |
| MFCC | | | high | 34.2 | 42.5 | 38.4 |
| | | | low | 71.6 | 70.7 | 71.2 |
| R-MFCC | [17] | 29 | medium | 63.9 | 51.2 | 57.5 |
| | | | high | 45.7 | 33.3 | 39.5 |
| R- | | | low | 93.6 | 98.3 | 96 |
| MFCC+ | [18] | 7 | medium | 89.8 | 93.7 | 91.7 |
| HOS+GA | | | high | 85.3 | 93.8 | 89.6 |
| | | | low | 68.5 | 96 | 82.2 |
| Markov | [11] | 169 | medium | 75.8 | 98.7 | 87.3 |
| | | | high | 90.5 | 98.9 | 94.7 |
| | | | low | 99.9 | 97.9 | 98.9 |
| Cal-R- | [26] | 15 | medium | 99.5 | 98.2 | 98.8 |
| Mel | | | high | 99.4 | 99.3 | 99.4 |

ROC of different feature sets in the universal paradigm and for low and medium complexities are shown in figure 8.

Comparing results of table 9 and figure 8 shows that Cal-R-Mel outperforms other methods by a large margin.

6.2. Mp3stego methods

Three different methods for steganalysis of Mp3stego were implemented and they were tested for different categories of signal complexity. Results of these simulations are presented in table 10.



(A) low complexity (B) medium complexity

ROC of different feature sets for detection of Mp3stego at capacity of 12.5% for low and medium complexities are shown in figure 9.

Based on results of table 10 and figure 9 we can conclude that method of differential quantization, outperforms other steganalysis methods.

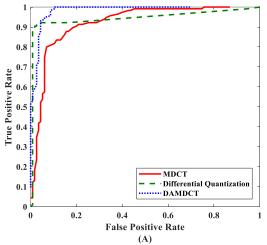
7. Future Works

This section discusses some directions that may be worth further investigation.

- 1- Steganalysis is implementation of a passive warden. On the other hand, active warden can measure robustness of steganography methods. This approach has not been addressed adequately in audio steganography. Investigating, robustness of existing methods to active warden and findings more robust methods can address this shortcoming.
- 2- Most of the reviewed papers have used the detection of watermarking systems (table 4) as an estimation of performance of steganalysis system on robust steganography methods. Comparing objectives of steganography and watermarking methods shows that undetectability is not the main concern of watermarking. Therefore, reliable detection of watermarking systems cannot be interpreted as reliable detection of robust steganography methods. Detection of robust steganography methods should to be analyzed separately.
- 3- Referring to table 4, it is evident that there is not a balance between the amount of works on LSB and non-LSB methods. Although, we tried to address this in section 6, yet non-LSB methods need more investigations.
- 4- According to table 5, most of the previous works have only investigated targeted steganalysis. We tried to address this by evaluating some methods under semi-blind scenario in section 6. Future works may focus on blind universal audio steganalysis.
- 5- Most of successful image steganalysis methods are based on calibration. A more thorough investigation of cover estimation techniques in audio and

| Table 10. Accuracy of different feature sets. Capacity is expressed in percentage of maximum embedding capacity. The number of |
|--|
| features in each method is written in the (). |

| | | M | DCT (4 | 8) | Differen | tial Quanti | zation (1) | DAMDCT (34) | | |
|----------|------------|------|--------|------|----------|-------------|------------|-------------|------|------|
| Capacity | Complexity | | [63] | | | [67] | [65] | | | |
| | | TPR | TNR | Ac. | TPR | TNR | Ac. | TPR | TNR | Ac. |
| | low | 89.1 | 94.7 | 91.9 | 98.6 | 94.8 | 96.7 | 89.9 | 97 | 93.4 |
| 100 | medium | 88.4 | 96.4 | 92.4 | 98.8 | 95.9 | 97.3 | 88.3 | 95.9 | 92.1 |
| | high | 88.1 | 94.7 | 91.4 | 98.5 | 94.7 | 96.6 | 87.1 | 94.1 | 90.6 |
| | low | 83.3 | 91.1 | 87.2 | 97.8 | 93.4 | 95.6 | 89.5 | 96.4 | 92.9 |
| 50 | medium | 81 | 92.8 | 86.9 | 97.9 | 94 | 96 | 87.1 | 96 | 91.5 |
| | high | 79.2 | 91.4 | 85.3 | 96.9 | 92 | 94.4 | 86.8 | 94.4 | 90.6 |
| | low | 78.3 | 88.2 | 83.2 | 96.7 | 91.7 | 94.2 | 88.2 | 96.3 | 92.2 |
| 25 | medium | 75 | 90.1 | 82.5 | 98.2 | 92.5 | 95.4 | 87.8 | 96 | 91.9 |
| | high | 74 | 89 | 81.5 | 95.5 | 88.6 | 92 | 86.1 | 94.1 | 90.1 |
| | low | 78.7 | 88.1 | 83.4 | 97 | 91.6 | 94.3 | 88.9 | 96.5 | 92.7 |
| 12.5 | medium | 75.3 | 90.2 | 82.7 | 98.3 | 92.4 | 95.3 | 88.9 | 95.7 | 92.3 |
| | high | 73.1 | 89.4 | 81.3 | 96.5 | 88.5 | 92.5 | 87.7 | 94.1 | 90.9 |



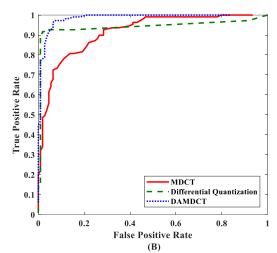


Figure 9. ROC of different feature sets for detection of mp3stego at capacity of 12.5%

(A) low complexity (B) medium complexity

proposing an efficient method would lead to better steganalysis systems.

- 6- While there is a rich literature on steganalysis, cover space has not been considered properly. Specifically, the impact of cover contents (speech/music or genre of music), sampling frequency, quantization depth and etc. on steganalysis systems should be investigated.
- 7- The idea of signal complexity and its impact on steganalysis was proposed in [10]. Our analysis in section 6 showed that sometimes existing metric does not behave as expected. Conducting a formal experiment signal complexity and proposing a better measure for steganalysis applications would be fruitful for both steganography and steganalysis applications.
- 8- Most of current steganalysis methods process a frame of the signal and then use moments of their results for steganalysis. A framed based method is another approach that has not been considered yet. Such method may make a decision about every frame and then use an appropriate rule for making the final decision. Also, such methods would be very beneficial for steganography and its results can be

- used for better understanding the cover space and implementing adaptive steganography methods.
- 9- Previous works have shown the effect of audio contents on the result of steganalysis [9]. Therefore, implementing a proper clustering method before steganalysis and investigating its effect seems to be fruitful.
- 10- Current literature has only considered steganalysis of mono signals. Investigating the correlation between different channels of stereo signals may improve performance of steganalysis.
- 11- Referring to table 5, most methods have used SVM for classification. On the other hand, recent image steganalysis methods suggest that better performances may be achieved if rich models with ensemble classifiers are employed. Furthermore, recent trends in pattern recognition tasks have turned considerably in favor of deep learning techniques. But, such approaches have not been investigated for image or audio steganalysis.
- 12- The main focus of mp3 methods has been on Mp3stego. Consequently, most of mp3 methods have not been investigated at all. Steganalysis of these methods should be addressed.

- 13- Different encoders for mp3 are available and they exhibit quite different characteristics [77]. Studying the effect of these differences on steganalysis systems seems another direction that needs investigation.
- 14- Due to wide spread usage of AAC in many audio/video streaming services it could be one of the best covers for steganography. Unfortunately, neither steganography nor steganalysis of AAC has found adequate attentions.

8. Conclusion

This work conducted comprehensive review of audio steganalysis literature and classified them into different categories. Reviewing the audio steganalysis literature showed that their main contributions had been their feature extractions. Therefore, we specifically paid more attention to this part. Considering the type of cover signal, the systems were broadly divided into non-compressed (wave) and compressed (mp3 and AAC) methods. Furthermore, considering the absence/presence of prior knowledge about the embedding algorithm, the systems were divided into two sub-groups of universal and targeted methods. For a better comparison between different works, each subsection of the paper was concluded with a summary of the relevant papers. Also, to conduct a fair comparison between different methods, some of them were implemented and were tested on the same database, on both LSB and non LSB steganography methods, and in both targeted and universal paradigms. In the end, some future directions for audio steganalysis were discussed.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments and valuable contributions.

References

- [1] H. Ghasemzadeh, M. T. Khass, and H. Mehrara, "Cipher Text Only Attack on Speech Time Scrambling Systems Using Correction of Audio Spectrogram," *The ISC International Journal of Information Security*, vol. 9, 2017.
- [2] G. J. Simmons, "The prisoners' problem and the subliminal channel," in *Advances in Cryptology*, 1984, pp. 51-67.
- [3] R. Böhme, Advanced statistical steganalysis: Springer Science & Business Media, 2010.
- [4] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, *et al.*, "Moving steganography and steganalysis from the laboratory into the real world," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*, 2013, pp. 45-58.
- [5] J. Dittmann and D. Hesse, "Network based intrusion detection to detect steganographic communication channels: on the example of audio data," in *Multimedia Signal Processing*, 2004 IEEE 6th Workshop on, 2004, pp. 343-346.
- [6] X.-M. Ru, H.-J. Zhang, and X. Huang, "Steganalysis of audio: Attacking the steghide," in *Machine Learning and Cybernetics*, 2005. *Proceedings of*

- 2005 International Conference on, 2005, pp. 3937-3942.
- [7] S. Hetzl and P. Mutzel, "A graph—theoretic approach to steganography," in *Communications and Multimedia Security*, 2005, pp. 119-128.
- [8] J.-w. Fu, Y.-c. Qi, and J.-S. Yuan, "Wavelet domain audio steganalysis based on statistical moments and PCA," in *Wavelet Analysis and Pattern Recognition*, 2007. *ICWAPR'07*. *International Conference on*, 2007, pp. 1619-1623.
- [9] O. H. Koçal, E. Yürüklü, and I. Avcibas, "Chaotic-Type Features for Speech Steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 3, pp. 651-661, 2008.
- [10] Q. Liu, A. H. Sung, and M. Qiao, "Novel stream mining for audio steganalysis," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 95-104.
- [11] Q. Liu, A. H. Sung, and M. Qiao, "Derivative-based audio steganalysis," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 7, p. 18, 2011.
- [12] S. Rekik, S.-A. Selouani, D. Guerchi, and H. Hamam, "An Autoregressive Time Delay Neural Network for speech steganalysis," in *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on,* 2012, pp. 54-58.
- [13] C. Kraetzer and J. Dittmann, "Mel-cepstrum-based steganalysis for VoIP steganography," in *Electronic Imaging* 2007, 2007, pp. 650505-650505-12.
- [14] Q. Liu, A. H. Sung, and M. Qiao, "Temporal derivative-based spectrum and mel-cepstrum audio steganalysis," *Information Forensics and Security, IEEE Transactions on*, vol. 4, pp. 359-368, 2009.
- [15] Y. Wei, L. Guo, Y. Wang, and C. Wang, "A blind audio steganalysis based on feature fusion," *Journal of Electronics (China)*, vol. 28, pp. 265-276, 2011.
- [16] H. Özer, B. Sankur, N. Memon, and İ. Avcıbaş, "Detection of audio covert channels using statistical footprints of hidden messages," *Digital Signal Processing*, vol. 16, pp. 389-401, 2006.
- [17] H. Ghasemzadeh and M. K. Arjmandi, "Reversed-Mel cepstrum based audio steganalysis," in Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on, 2014, pp. 679-684.
- [18] H. Ghasemzadeh, M. Tajik Khas, and M. Khalil Arjmandi, "Audio Steganalysis Based on Reversed Psychoacoustic Model of Human Hearing," *Digital Signal Processing*, vol. 51, pp. 133-141, 2016.
- [19] H. Ghasemzadeh, "A metaheuristic approach for solving jigsaw puzzles," in *Intelligent Systems* (ICIS), 2014 Iranian Conference on, 2014, pp. 1-6.
- [20] Y. Liu, K. Chiang, C. Corbett, R. Archibald, B. Mukherjee, and D. Ghosal, "A novel audio steganalysis based on high-order statistics of a distortion measure with hausdorff distance," in *Information Security*, ed: Springer, 2008, pp. 487-501.
- [21] S. Geetha, N. Ishwarya, and N. Kamaraj, "Audio steganalysis with Hausdorff distance higher order statistics using a rule based decision tree paradigm,"

- Expert Systems with Applications, vol. 37, pp. 7469-7482, 2010.
- [22] Z. Kexin, "Audio steganalysis of spread spectrum hiding based on statistical moment," in *Signal Processing Systems (ICSPS)*, 2010 2nd International Conference on, 2010, pp. V3-381-V3-384.
- [23] W. Zeng, R. Hu, and H. Ai, "Audio steganalysis of spread spectrum information hiding based on statistical moment and distance metric," *Multimedia Tools and Applications*, vol. 55, pp. 525-556, 2011.
- [24] I. Avcıbas, "Audio steganalysis with contentindependent distortion measures," *IEEE Signal Processing Letters*, vol. 13, 2006.
- [25] M. K. Johnson, S. Lyu, and H. Farid, "Steganalysis of recorded speech," in *Electronic Imaging 2005*, 2005, pp. 664-672.
- [26] H. Ghasemzadeh and M. K. Arjmandi, "Universal Audio Steganalysis Based on Calibration and Reversed Frequency Resolution of Human Auditory System," *IET Signal Processing*, 2017.
- [27] H. Ghasemzadeh, M. T. Khass, M. K. Arjmandi, and M. Pooyan, "Detection of vocal disorders based on phase space parameters and Lyapunov spectrum," *Biomedical Signal Processing and Control*, vol. 22, pp. 135-145, 2015.
- [28] A. Lang and J. Dittmann, "Profiles for evaluation and their usage in audio wet," in *IS&T/SPIE's 18th Annual Symposium, Electronic Imaging*, 2006.
- [29] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *Image Processing, IEEE Transactions on*, vol. 6, pp. 1673-1687, 1997.
- [30] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM systems journal*, vol. 35, pp. 313-336, 1996.
- [31] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *Signal Processing, IEEE Transactions on*, vol. 51, pp. 1020-1033, 2003.
- [32] M. Li, M. K. Kulhandjian, D. A. Pados, S. N. Batalama, and M. J. Medley, "Extracting spread-spectrum hidden data from digital media," *IEEE transactions on information forensics and security*, vol. 8, pp. 1201-1210, 2013.
- [33] H. Repp, "Hide4PGP," *Available: www. heinz-repp. onlinehome. de/Hide4PGP. htm,* 1996.
- [34] I. secret. *Invisible secret*. Available: http://www.invisiblesecrets.com/
- [35] T. Sharp, "An implementation of key-based digital signal steganography," in *Information hiding*, 2001, pp. 13-26.
- [36] Steganos. Steganos. Available: http://www.steganos.com
- [37] Stools. Stools. Available: http://info.umuc.edu/its/online_lab/ifsm459/s-tools4/
- [38] K. Gopalan, S. J. Wenndt, S. F. Adams, and D. M. Haddad, "Audio steganography by amplitude or phase modification," in *Electronic Imaging 2003*, 2003, pp. 67-76.
- [39] S. Rekik, D. Guerchi, S.-A. Selouani, and H. Hamam, "Speech steganography using wavelet and

- Fourier transforms," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, pp. 1-14, 2012.
- [40] M. F. Tolba, M.-S. Ghonemy, I.-H. Taha, and A. S. Khalifa, "High capacity image steganography using wavelet-based fusion," in *Computers and Communications*, 2004. Proceedings. ISCC 2004. Ninth International Symposium on, 2004, pp. 430-435.
- [41] F. A. Petitcolas, "mp3stego," ed, 1998.
- [42] S. wave. *Stego wave*. Available: http://www.jjtc.com/Steganography/tools.html
- [43] J. Fridrich and M. Goljan, "Digital image steganography using stochastic modulation," in *Electronic Imaging* 2003, 2003, pp. 191-202.
- [44] S. Shirali-Shahreza and M. Manzuri-Shalmani, "High capacity error free wavelet Domain Speech Steganography," in *Acoustics, Speech and Signal Processing*, 2008. ICASSP 2008. IEEE International Conference on, 2008, pp. 1729-1732.
- [45] R. M. Nugraha, "Implementation of Direct Sequence Spread Spectrum steganography on audio data," in *Electrical Engineering and Informatics* (ICEEI), 2011 International Conference on, 2011, pp. 1-6.
- [46] W. Zeng, H. Ai, and R. Hu, "A novel steganalysis algorithm of phase coding in audio signal," in Advanced Language Processing and Web Information Technology, 2007. ALPIT 2007. Sixth International Conference on, 2007, pp. 261-264.
- [47] H. Ghasemzadeh and M. H. Keyvanrad, "Toward a Robust and Secure Echo Steganography Method Based on Parameters Hopping," in *Signal Processing and Intelligent Systems*, 2015.
- [48] W. Zeng, H. Ai, and R. Hu, "An algorithm of echo steganalysis based on power cepstrum and pattern classification," in *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on, 2008*, pp. 1344-1348.
- [49] Y. Wang, H. Wen, Z. Jian, and Z. Wu, "Steganalysis on positive and negative echo hiding based on skewness and kurtosis," in *Industrial Electronics and Applications (ICIEA)*, 2014 IEEE 9th Conference on, 2014, pp. 1235-1238.
- [50] C. Xie, Y. Cheng, and Y. Chen, "An active steganalysis approach for echo hiding based on Sliding Windowed Cepstrum," *Signal Processing*, vol. 91, pp. 877-889, 2011.
- [51] Y.-C. Qi, L. Ye, and C. Liu, "Wavelet domain audio steganalysis for multiplicative embedding model," in *Wavelet Analysis and Pattern Recognition*, 2009. *ICWAPR* 2009. *International Conference on*, 2009, pp. 429-432.
- [52] C. Platt, "UnderMP3Cover," ed, 2004.
- [53] Z. Achmad, "MP3Stegz," ed, 2008.
- [54] D. Yan and R. Wang, "Huffman table swapping-based steganograpy for MP3 audio," *Multimedia Tools and Applications*, vol. 52, pp. 291-305, 2011.
- [55] Y. Diqun, W. Rangding, and Z. Liguang, "Quantization step parity-based steganography for MP3 audio," *Fundamenta Informaticae*, vol. 97, pp. 1-14, 2009.

- [56] D.-H. Kim, S.-J. Yang, and J.-H. Chung, "Additive data insertion into MP3 bitstream using linbits characteristics," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, 2004, pp. iv-181-iv-184 vol. 4.
- [57] R. J. Menon, "Mp3 steganography and steganalysis," University of Rhode Island, 2008.
- [58] L. Gang, A. N. Akansu, and M. Ramkumar, "MP3 resistant oblivious steganography," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 2001, pp. 1365-1368.
- [59] M. H. Shirali-Shahreza and S. Shirali-Shahreza, "Real-time and MPEG-1 layer III compression resistant steganography in speech," *IET information security*, vol. 4, pp. 1-7, 2010.
- [60] M. Zaturenskiy, "MP3 files as a steganography medium," in *Proceedings of the 2nd annual conference on Research in information technology*, 2013, pp. 23-28.
- [61] A. Westfeld, "Detecting low embedding rates," in *Information Hiding*, 2003, pp. 324-339.
- [62] M. Qiao, A. H. Sung, and Q. Liu, "Feature mining and intelligent computing for MP3 steganalysis," in *Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS'09. International Joint Conference on*, 2009, pp. 627-630.
- [63] M. Qiao, A. H. Sung, and Q. Liu, "MP3 audio steganalysis," *Information Sciences*, vol. 231, pp. 123-134, 2013.
- [64] C. Jin, R. Wang, D. Yan, P. Ma, and K. Yang, "A novel detection scheme for MP3Stego with low payload," in *Signal and Information Processing* (ChinaSIP), 2014 IEEE China Summit & International Conference on, 2014, pp. 602-606.
- [65] C. Jin, R. Wang, and D. Yan, "Steganalysis of MP3Stego with low embedding-rate using Markov feature," *Multimedia Tools and Applications*, vol. 76, pp. 6143-6158, 2017.
- [66] D. Yan and R. Wang, "Detection of MP3Stego exploiting recompression calibration-based feature," *Multimedia Tools and Applications*, vol. 72, pp. 865-878, 2014.
- [67] D. Yan, R. Wang, X. Yu, and J. Zhu, "Steganalysis for MP3Stego using differential statistics of quantization step," *Digital Signal Processing*, vol. 23, pp. 1181-1185, 2013.
- [68] C. Jin, R. Wang, D. Yan, and X. Yu, "Steganalysis of UnderMP3Cover," *Journal of Computational Information Systems*, vol. 8, pp. 10459-10468, 2012.
- [69] K. Brandenburg, "MP3 and AAC explained," in Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding, 1999.
- [70] J. Zhu, R.-D. Wang, J. Li, and D.-Q. Yan, "A huffman coding section-based steganography for AAC audio," *Information Technology Journal*, vol. 10, pp. 1983-1988, 2011.
- [71] J. Zhu, R. Wang, and D. Yan, "The sign bits of huffman codeword-based steganography for aac audio," in *Multimedia Technology (ICMT)*, 2010 International Conference on, 2010, pp. 1-4.

- [72] S. Xu, P. Zhang, P. Wang, and H. Yang, "Performance analysis of data hiding in MPEG-4 AAC audio," *Tsinghua Science & Technology*, vol. 14, pp. 55-61, 2009.
- [73] Y. Wang, L. Guo, Y. Wei, and C. Wang, "A steganography method for AAC audio based on escape sequences," in *Multimedia Information Networking and Security (MINES)*, 2010 International Conference on, 2010, pp. 841-845.
- [74] Y. Wei, L. Guo, and Y. Wang, "Controlling bitrate steganography on AAC audio," in *Image and Signal Processing (CISP)*, 2010 3rd International Congress on, 2010, pp. 4373-4375.
- [75] Y. Ren, Q. Xiong, and L. Wang, "Steganalysis of AAC using calibrated Markov model of adjacent codebook," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, 2016, pp. 2139-2143.*
- [76] Y. Ren, Q. Xiong, and L. Wang, "A Steganalysis Scheme for AAC Audio Based on MDCT Difference Between Intra and Inter Frame," in *International Workshop on Digital Watermarking*, 2017, pp. 217-231.
- [77] R. Böhme and A. Westfeld, "Statistical characterisation of MP3 encoders for steganalysis," in *Proceedings of the 2004 workshop on Multimedia and security*, 2004, pp. 25-34.