Derivative-Based Audio Steganalysis

QINGZHONG LIU, Sam Houston State University

ANDREW H. SUNG, New Mexico Tech

MENGYU QIAO, South Dakota School of Mines and Technology

This article presents a second-order derivative-based audio steganalysis. First, Mel-cepstrum coefficients and Markov transition features from the second-order derivative of the audio signal are extracted; a support vector machine is then applied to the features for discovering the existence of hidden data in digital audio streams. Also, the relation between audio signal complexity and steganography detection accuracy, which is an issue relevant to audio steganalysis performance evaluation but so far has not been explored, is analyzed experimentally. Results demonstrate that, in comparison with a recently proposed signal stream-based Mel-cepstrum method, the second-order derivative-based audio steganalysis method gains a considerable advantage under all categories of signal complexity—especially for audio streams with high signal complexity, which are generally the most challenging for steganalysis—and thereby significantly improves the state of the art in audio steganalysis.

Categories and Subject Descriptors: I.5.4 [Pattern Recognition]: Applications—*Signal processing*; K.6.m [Management of Computing Information]: Miscellaneous—*Insurance; security*

General Terms: Algorithms, Design, Security

Additional Key Words and Phrases: Audio, steganography, steganalysis, derivative, Mel-cepstrum, Markov, signal complexity, SVM

ACM Reference Format:

Liu, Q., Sung, A. H., and Qiao, M. 2011. Derivative-based audio steganalysis. *ACM Trans. Multimedia Comput. Commun. Appl.* 7, 3, Article 18 (August 2011), 19 pages.

DOI = 10.1145/2000486.2000492 <http://doi.acm.org/10.1145/2000486.2000492>

1. INTRODUCTION

Steganography is the creation of a media embedded with secret content in such a way that no one apart from the sender and the intended recipients know the existence of the secret. Digital steganography provides an easy means for covert communications on the Internet by hiding data in digital cover files such as images, audios, and videos; it has the advantage that the steganograms or digital media carrying secret content, unlike ciphertexts or cryptograms, do not reveal themselves as containing secrets. Thus, steganography has created a threat for national security and law enforcement due to the variety of unlawful purposes for which it can conceivably be used.

This research was supported by the Institute for Complex Additive Systems Analysis, a research division of New Mexico Tech.

Authors' addresses: Q. Liu, Department of Computer Science, Sam Houston State University, Huntsville, TX 77341; email: qx1005@shsu.edu; A. H. Sung, Department of Computer Science and Institute for Complex Additive Systems Analysis, New Mexico Tech, Socorro, NM 87801; email: sung@cs.nmt.edu; M. Qiao, Department of Mathematics and Computer Science, South Dakota School of Mines and Technology; email: mengyu.qiao@sdsmt.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2011 ACM 1551-6857/2011/08-ART18 \$10.00

DOI 10.1145/2000486.2000492 <http://doi.acm.org/10.1145/2000486.2000492>

Steganalysis is the opposite of steganography, and aims at detecting and analyzing the hidden information in digital media. In the past few years, several steganalysis methods have been presented for detecting the information-hiding behaviors in multiple steganographic systems. Most of these methods focused on the detection of information-hiding in digital images. For instance, one of the well-known detectors, Histogram Characteristic Function Center Of Mass (HCFCOM) was successful in detecting noise-adding steganography [Harmsen and Pearlman 2003]. Another well-known method is to construct the high-order moment statistical model in the multiscale decomposition using a wavelet-like transform and then apply a learning classifier to the high order feature set [Lyu and Farid 2006]. Shi et al. proposed a Markov process-based approach to detect the information-hiding behaviors in JPEG images [Shi et al. 2007]. Based on the Markov approach, Liu et al. expanded the Markov features to the interbands of the DCT domains and combined the expanded features with the polynomial fitting of the histogram of the DCT coefficients, and successfully improved the detection of JPEG steganograms created by multiple hiding methods [Liu et al. 2008c, 2010]. Liu et al. also proposed neighboring joint density-based JPEG steganalysis [Liu et al. 2009a, 2011]. Other works in image steganalysis are found in the references [Farid 2002; Fridrich 2004; Holotyark et al. 2005; Liu and Sung 2007; Liu et al. 2008a, 2008b; Pevny and Fridrich 2007].

To detect the information-hiding in digital audio streams, Avcibas designed the content-independent distortion measures as features for classifier design [Avcibas 2006]; Ozer et al. constructed the detector based on the characteristics of the denoised residuals of the audio file [Ozer et al. 2006]; Johnson et al. set up a statistical model by building a linear basis that captures certain statistical properties of audio signals [Johnson et al. 2005]; Kraetzer and Dittmann recently proposed a Mel-cepstrum-based analysis to detect hidden messages [Kraetzer and Dittmann 2007]; Zeng et al. presented new algorithms to detect phase coding steganography based on analysis of the phase discontinuities [Zeng et al. 2007] and to detect echo steganography based on statistical moments of peak frequency [Zeng et al. 2008]. Of all these methods, Kraetzer and Dittmann's signal stream-based Mel-cepstrum audio steganalysis is particularly noteworthy since it is the first to utilize Mel-frequency cepstral coefficients—which are widely used in speech recognition—for audio steganalysis, and it delivers good performance and represents the state-of-the-art in audio steganalysis regarding the detection of several types of audio steganograms [Kraetzer and Dittmann 2007].

Meanwhile, to evaluate detection performance, most researchers take the information-hiding ratio as a major factor in evaluating steganalysis performance. Generally, for steganograms created using the same tool, we can expect higher detection accuracy with a higher information-hiding ratio. For image steganalysis, Liu et al. first introduced image complexity to enhance the framework for formal evaluation of detection performance [Liu and Sung 2007; Liu et al. 2008a, 2008b, 2009a]. The results demonstrate that detection performance is closely related not only to the information-hiding ratio but also to image complexity. In audio steganalysis, it is expected that a similar relation can be observed between detection performance and the audio's signal complexity.

In this article, we present an approach for audio steganalysis based on the Mel-cepstrum coefficients derived from the second-order derivative to improve Kraetzer and Dittmann's work. We also extract second-order derivative-based Markov transition probabilities as features. The relation between audio steganalysis performance and signal complexity is also studied experimentally. Our approach leads to dramatic improvements over the original signal-based Mel-cepstrum audio steganalysis and delivers high detection accuracy, even for audio streams with high signal complexity—while in such cases the original signal-based method works poorly.

The rest of the article is organized as follows. Section 2 presents second-order derivative-based Fourier spectrum and compares the characteristics of covers and steganograms. Section 3 describes second-order derivative-based Mel-cepstrum features for audio steganalysis. Section 4 details the

second-order derivative-based Markov approach. Signal complexity as a parameter for performance evaluation of audio steganalysis is introduced in Section 5, followed by experiments in Section 6 and discussion in Section 7. Section 8 concludes.

2. SECOND-ORDER DERIVATIVE-BASED SPECTRUM ANALYSIS

In image processing, the second-order derivative is widely used to detect isolated points and edges [Gonzalez and Woods 2008]. Exploiting its great usefulness in detecting various objects, we designed a scheme of second-order derivative-based audio steganalysis, the details of which are described as follows.

An audio signal is denoted $f(t)$ ($t = 0, 1, 2, \dots, N - 1$). The second-order derivative $D_f^2(\bullet)$ is defined as follows:

$$D_f^2(t) \equiv \frac{d^2 f}{dt^2} = f(t+1) - 2 * f(t) + f(t-1), \quad t = 1, 2, \dots, N-2. \quad (1)$$

Similar to the additive noise model proposed in the reference [Harmsen 2003], a stego-signal is denoted $s(t)$, which can be modeled by adding a noise or error signal $e(t)$ to the original signal $f(t)$,

$$s(t) = f(t) + e(t). \quad (2)$$

Second-order derivatives of $e(t)$ and $s(t)$ are denoted $D_e^2(\bullet)$ and $D_s^2(\bullet)$, respectively. Thus,

$$D_s^2(\bullet) = D_f^2(\bullet) + D_e^2(\bullet). \quad (3)$$

The Discrete Fourier Transforms (DFTs) of $D_s^2(\bullet)$, $D_f^2(\bullet)$, and $D_e^2(\bullet)$, are denoted F_k^s , F_k^f , and F_k^e , respectively.

$$F_k^s = \sum_{t=0}^{M-1} D_s^2(t) e^{-\frac{j2\pi}{M} kt} \quad (4)$$

$$F_k^f = \sum_{t=0}^{M-1} D_f^2(t) e^{-\frac{j2\pi}{M} kt} \quad (5)$$

$$F_k^e = \sum_{t=0}^{M-1} D_e^2(t) e^{-\frac{j2\pi}{M} kt} \quad (6)$$

Where $k = 0, 1, 2, \dots, M-1$ and M is the number of samples of the derivatives. We have

$$F_k^s = F_k^f + F_k^e. \quad (7)$$

Assume θ is the angle between the vectors F_k^f and F_k^e , then

$$|F_k^s|^2 = |F_k^f|^2 + |F_k^e|^2 + 2|F_k^f| \bullet |F_k^e| \bullet \cos \theta. \quad (8)$$

For most steganographic systems, the hidden message or payload do not depend on the cover, that is, $e(t)$, the signal that approximates the payload signal is irrelative to $f(t)$, the cover signal. Therefore, θ is an arbitrary value in the range $[0, \pi]$, the expected value of $|F_k^s|^2$ is calculated as follows:

$$E(|F_k^s|^2) = \frac{\int_0^\pi (|F_k^f|^2 + |F_k^e|^2 + 2|F_k^f| \bullet |F_k^e| \bullet \cos \theta) d\theta}{\int_0^\pi d\theta} = |F_k^f|^2 + |F_k^e|^2. \quad (9)$$

We have

$$E(|F_k^s|^2)/|F_k^f|^2 = 1 + |F_k^e|^2/|F_k^f|^2. \quad (10)$$

The expected value of the variance is obtained by the following equation:

$$E[(|F_k^s|^2 - E(|F_k^s|^2))^2] = \frac{\int_0^\pi 4|F_k^f|^2 \cdot |F_k^e|^2 \cdot \cos^2 \theta d\theta}{\int_0^\pi d\theta} = 2|F_k^f|^2 \cdot |F_k^e|^2. \quad (11)$$

Based on (10), the statistics of the spectrum from the cover signal $f(t)$ and that from the stego-signal $s(t)$ are different: the expected spectrum of the stego-signal is higher than that of the cover.

According to (11), the rate of the power change in different spectrum bands of the stego-audio is also different from the original cover. Generally, the cepstrum may be interpreted as information for the power change, which was first defined by Bogert et al. [1963]. Reynolds and McEachern showed a modified cepstrum called Mel-cepstrum for speech recognition [McEachern 1994; Reynolds 1992]. Recently, a signal-based Mel-cepstrum audio steganalysis was proposed [Kraetzer and Dittmann 2007].

Digital audio streams, especially speech audio clips, are normally band-limited; in other words, the magnitudes of their high-frequency components are limited. On the other side, regarding the low- and middle-frequency components, the power spectrum of audio signal (second-order derivative) is much stronger than the power spectrum of the error signal or hidden data (second-order derivative); that is, $|F_k^e|^2/|F_k^f|^2$ is almost zero. Based on (10), the difference between the spectrum of the cover and the stego-signal at low and middle frequency is negligible; however, the situation is very different at the high-frequency components. As frequency increases, $|F_k^e|$ increases, and $|F_k^f|$ may decrease, the change of the spectrum resulting from embedding hidden data is no longer negligible, hence the statistics extracted from the high-frequency components may be the clue to detecting the information-hiding behavior.

Figure 1 shows the spectra of the second-order derivatives of a cover (left) and the correlated stego-signal (right) over the whole frequency range (first row) and over the high-frequency region (second row). It clearly shows that the stego-signal has higher magnitude than the cover-signal in the derivative spectrum for high-frequency components.

We may directly take the derivative-based spectrum statistics in high-frequency regions as features for audio steganalysis. In real-world detection, however, the cover reference shown in Figure 1 is not available for steganalysis. Due to the fact that different audio streams have different spectrum characteristics, the detection derived from Eq. (10) may not be practical without a comparison to the original cover. In such case, Eq. (11) shows that the rate of power change in different spectrum bands of the stego-audio is quite different from the original. Based on Kraetzer and Dittmann's proposed signal-based Mel-cepstrum audio steganalysis, we designed a derivative-based Mel-cepstrum audio steganalysis, described in the following.

3. SECOND-ORDER DERIVATIVE-BASED MEL-CEPSTRUM

In speech processing, the Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. Mel-cepstrum is commonly used for representing the human voice and musical signals. Inspired by success in speech recognition, a signal-based Mel-cepstrum audio steganalysis was proposed [Kraetzer and Dittmann 2007], including the following two types of Mel-cepstrum coefficients:

(1) Signal-based Mel-frequency cepstral coefficients (MFCCs), $s_mel_1, s_mel_2, \dots, s_mel_M$, where M is the number of MFCCs; the value of M is 29 for a signal with a sampling rate of 44.1 kHz. MFCCs

Fig. 1. Spectra of the second-order derivatives of a cover signal (left) and the stego-signal (right). Both figures in the first row show half magnitude values due to symmetric characteristics of Fourier transforms decomposition [Liu et al. 2009b] ©2009 IEEE.

can be calculated by the following equation, where MT indicates the Mel-scale transformation:

$$\text{Signal_MelCepstrum} = FT(MT(FT(f))) = \begin{pmatrix} s_mel_1 \\ s_mel_2 \\ \dots \\ s_mel_M \end{pmatrix}. \quad (12)$$

(2) Signal Filtered Mel-frequency cepstral coefficients (FMFCCs), $sf_mel_1, sf_mel_2, \dots, sf_mel_M$. M is the number of FMFCCs. FMFCCs can be calculated by the following equation:

$$\text{Signal_FilteredMelCepstrum} = FT(\text{SpeechBandFiltering}(MT(FT(f)))) = \begin{pmatrix} sf_mel_1 \\ sf_mel_2 \\ \dots \\ sf_mel_M \end{pmatrix}. \quad (13)$$

In (13), the role of speech-band filtering is to remove the speech-relevant bands (the spectrum components between 200 and 6819.59Hz) [Kraetzer and Dittmann 2007].

To improve Mel-cepstrum-based audio steganalysis, we formulate the second-order derivative-based MFCCs and FMFCCs, obtained by replacing the signal f in (12) and (13) with the second-order

derivative $D_f^2(\bullet)$, the calculation is given by

$$\text{Derivative_MelCepstrum} = FT(MT(FT(D_f^2))) = \begin{pmatrix} d_mel_1 \\ d_mel_2 \\ \dots \\ d_mel_M \end{pmatrix} \quad (14)$$

and

$$\text{Derivative_FilteredMelCepstrum} = FT(\text{SpeechBandFiltering}(MT(FT(D_f^2)))) = \begin{pmatrix} df_mel_1 \\ df_mel_2 \\ \dots \\ df_mel_M \end{pmatrix} \quad (15)$$

Second-order derivative-based Mel-cepstrum coefficients, calculated by (14) and (15), form the first type of features in our detection.

4. SECOND-ORDER DERIVATIVE-BASED MARKOV APPROACH

The Markov approach has been widely used in different areas. In steganalysis, Shi et al. [2007] presented a Markov process to detect the information-hiding behaviors in JPEG images. Liu et al. expanded the Markov approach to the interbands of the DCT domains [Liu et al. 2008c]. Both of these JPEG steganalysis methods are based on the first-order derivative of the quantized DCT coefficients. Since second-order derivatives perform better than first-order derivatives in detecting isolated points and edges [Gonzalez and Woods 2008], we extend our previous work in audio steganalysis [Liu et al. 2008d, 2009b] and design a Markov approach for audio steganalysis based on second-order derivative of audio signals, described as follows:

An audio signal is denoted $f(t)$ ($t = 0, 1, 2, \dots, N-1$), the minimal interval of the magnitude is 1. The second-order derivative $D_f^2(t)$ ($t = 1, 2, \dots, N-2$) is defined in (1). The Markov transition probability is calculated as follows:

$$M_{D_f^2}(i, j) = \frac{\sum_{t=1}^{N-3} \delta(D_f^2(t) = i, D_f^2(t+1) = j)}{\sum_{t=1}^{N-3} \delta(D_f^2(t) = i)}. \quad (16)$$

Where $\delta = 1$, if its arguments are satisfied, otherwise $\delta = 0$. The range of i and j is $[-6, 6]$, so we have a 13×13 transition matrix, consisting of 169 features. Figure 2 shows the temporal magnitudes of a cover and the steganogram that was produced by using the Steghide algorithm [Hetzel and Mutzel 2005], and the Markov transition probabilities, respectively. Although the signals, shown in (a) and (c), are likely identical, the Markov transition probabilities, shown in (b) and (d), are apparently different; the difference is shown in Figure 2(e).

5. SIGNAL COMPLEXITY

For audio steganalysis performance, most researchers have conducted evaluation in terms of a information-hiding ratio or embedding strength. Generally speaking, for steganograms created by the same hiding method, a higher information-hiding ratio leads to better detection performance. Our work in image steganalysis [Liu and Sung 2007; Liu et al. 2008a, 2008b, 2009a, 2011] has demonstrated that taking the information-hiding ratio as the sole parameter is not sufficient for a complete and fair performance evaluation; this is because, at the same hiding ratio, different image complexities are associated with different detection accuracies in that higher image complexity leads to lower detection accuracy, and vice versa. We measured the image complexity by using the shape parameter of the generalized Gaussian distribution (GGD) of the discrete wavelet/cosine transform coefficients.

Fig. 2. The comparison of the temporal cover signal (a) and the steganogram (c); and Markov transition probabilities of the second-order derivatives, shown in (b) and (d). The difference of the transition probability between (b) and (d) is shown in (e).

We may employ the same metric of the GGD shape parameter to calculate the audio signal complexity. For a more efficient computation, we instead utilized the following formula involving the second-order derivative to measure the signal complexity:

$$C(f) = \frac{\frac{1}{N-2} \sum_{t=1}^{N-2} |D_f^2(t)|}{\frac{1}{N} \sum_{t=0}^{N-1} |f(t)|}. \quad (17)$$

Fig. 3. Audio signal samples with different measurements of signal complexity, $C(f)$.

$C(f)$ measures the ratio of the mean absolute value of the second-order derivative to the mean absolute value of the signal. We may of course adopt several different metrics for signal complexity, $C(f)$ is introduced here as our measure, as it can be computed much faster than, say, GGD, and still captures all essential elements of measures for signal complexity. Figure 3 shows six audio signal samples with different complexity values of $C(f)$. If we hide the same message into these different audio clips, the expectation of detection performance ought to be different: it should be easier to detect information-hiding in the audios with lower signal complexity. This indeed will be validated by experimental results in Section 6.

6. EXPERIMENTS

6.1 Experimental Setup

We have 19380 mono 44.1 kHz 16-bit quantization in uncompressed, PCM coded WAV audio files, covering digital speeches and songs in several languages e.g., English, Chinese, Japanese, Korean, and several types of music (jazz, rock, blue). Each audio has the duration of 10 seconds. We produced audio steganograms by hiding different messages into these audio files. The hiding tools/algorithms include Hide4PGP V4.0, available at <http://www.heinz-repp.onlinehome.de/Hide4PGP.htm>; Invisible Secrets, available at <http://www.invisiblesecrets.com/>; LSB-matching [Sharp 2001]; and Steghide [Hetzl and

Mutzel 2005]. The hidden data includes voice, video, image, text, executable codes, random bits, and so on, and the hidden data in any two audio files are different. The amounts of audio steganograms are: 19380 produced by using Hide4PGP with 25% maximal hiding; 17158 and 17596 by Steghide with maximal and 50% maximal hiding; 18766 and 19371 by Invisible Secrets with maximal and 50% maximal hiding; 19000 and 19000 by using LSB-matching with maximal and 50% maximal hiding, respectively.

Additionally, we have 6357 mono 44.1 kHz 16-bit quantization in uncompressed, PCM coded WAV audio files, and most are online broadcast in English. Each audio has the duration of 19 seconds. We produced the same amount of the watermarking audio files by hiding randomly-produced 2 hexadecimal or 8 binary watermarking digits in each audio (maximal hiding) with the use of spread spectrum audio watermarking [Kirovski and Malvar 2003], which displays solid robustness against traditional signal processing, including arbitrary limited pitch-bending and time-scaling.

6.2 Statistics of Mel-Cepstrum and Markov Transition Features

We compared the statistics of signal-based Mel-cepstrum features [Kraetzer and Dittmann 2007], which contain two types of Mel-cepstrum coefficients, MFCCs and FMFCCs, totaling 58 features, with second-order derivative-based Mel-cepstrum features, described in (14) and (15), and second-order derivative-based Markov transition features, calculated by (16), in different signal complexities. We roughly divided all cover and steganogram audio files into four categories according to their signal complexity values: low complexity ($C < 0.04$); middle complexity ($0.04 \leq C < 0.08$); middle-high complexity ($0.08 \leq C < 0.12$); and high complexity ($C \geq 0.12$).

Figure 4 lists the F scores of one-way analysis of variances (ANOVA) [Hill and Lewicki 2005] of the features extracted from audio covers and Steghide steganograms with maximal hiding, and LSB-matching audio steganograms with 50% maximal hiding, respectively. The F scores shown in Figure 4 indicate that: signal-based Mel-cepstrum features are not as effective as second-order derivative-based Mel-cepstrum features; and second-order derivative-based Markov transition features are superior to both signal-based and derivative-based Mel-cepstrum features. It is expected that the detection performance by using the Markov transition features would be the best, followed by derivative-based Mel-cepstrum features and signal-based Mel-cepstrum features. Regarding the statistical significance under different categories of signal complexity, for Mel-cepstrum features, the F scores under low signal complexity are much higher than those under middle, middle-high, and high signal complexities; for derivative-based Markov transition features, the F scores under middle to high complexities are significantly noticeable, although the values drop a little on average with respect to those under low complexity. It is expected that the detection in the category of low signal complexity would be much better than that in other categories of signal complexity with the use of Mel-cepstrum features; the detection in all categories of signal complexity would be satisfactory with the use of the Markov transition features.

6.3 Comparison of Signal- and Derivative-based Audio Steganalysis

We compare signal-based Mel-cepstrum audio steganalysis (S-Mel) with 58 Mel-cepstrum coefficients [Kraetzer and Dittmann 2007], with second-order derivative-based Mel-cepstrum steganalysis (2D-Mel) with the 58 features described in (14) and (15), second-order derivative-based Markov approach (2D-Markov) with the 169 features calculated by (16), and combined derivative-based detection containing all features described in (14), (15), and (16), abbreviated as 2D-MM, in the four categories of signal complexity: low complexity ($C < 0.04$); middle complexity ($0.04 \leq C < 0.08$); middle-high complexity ($0.08 \leq C < 0.12$); and high complexity ($C \geq 0.12$). In Kraetzer and Dittmann's work, signal-based Mel-cepstrum coefficients and several other statistical features form an AMSL Audio Steganalysis Tool

Fig. 4. One-way ANOVA F scores of signal-based Mel-cepstrum features (first column, a and d), second-order derivative-based Mel-cepstrum features (second column, b and e) and Markov transition features (third column, c and f) under each category of signal complexity to separate 3000 Steghide steganograms and 3000 LSB-matching steganograms from 3000 covers, shown in the left and the right, respectively. The Y-label gives the F score and X-label is the number of features.

Table I. Testing Accuracy over 100 Experiments with Signal-Based Mel-cepstrum Audio Steganalysis (S-Mel), AAST, and Derivative-Based Audio Steganalysis Methods (Abbreviated as 2D-Mel, 2D-Markov, and 2D-MM)

Hiding method	Hiding ratio to maximum capacity	Signal complexity C	Testing accuracy				
			S-Mel	AAST*	2D-Mel	2D-Markov	2D-MM
Invisible	100%	low	97.8%	89.1	98.9	99.6	99.2
		middle	97.2	79.0	98.8	99.9	99.6
		middle-high	90.6	86.2	97.3	99.9	99.6
		high	76.4	65.9	91.5	99.9	99.6
	50%	low	93.8	78.2	96.7	99.0	98.0
		middle	89.0	71.2	96.5	99.2	98.9
		middle-high	78.7	74.9	88.9	99.1	98.7
		high	61.8	60.5	77.3	99.3	99.0
Hide4PGP	25%	low	97.8	86.1	98.9	99.6	99.2
		middle	97.2	80.1	98.9	99.9	99.7
		middle-high	90.6	86.2	97.4	99.9	99.7
		high	76.2	64.3	91.5	99.9	99.7
LSB matching	100%	low	97.8	86.8	98.9	99.5	99.2
		middle	97.2	80.1	98.9	99.8	99.6
		middle-high	90.8	87.1	97.3	99.7	99.6
		high	76.2	63.9	91.5	99.9	99.7
	50%	low	95.9	80.4	98.1	99.2	98.4
		middle	94.6	67.1	98.1	99.5	99.3
		middle-high	85.1	81.1	94.0	99.3	99.0
		high	66.1	60.1	84.8	99.6	99.4
Steghide	100%	low	97.0	89.6	98.6	97.6	97.7
		middle	96.4	81.8	98.6	98.6	98.6
		middle-high	87.4	83.6	96.2	98.6	98.3
		high	71.8	63.2	89.9	99.1	98.5
	50%	low	94.3	73.6	97.2	94.6	95.7
		middle	91.9	73.6	97.3	96.5	96.6
		middle-high	80.8	76.0	91.8	96.6	96.0
		high	64.0	59.8	84.4	98.2	97.1
Spread spectrum audio watermarking	100%	low	91.0	76.0	92.6	90.6	90.4
		middle	86.0	80.3	92.9	86.7	92.2
		middle-high	81.5	56.1	87.2	79.9	86.4
		high	67.5	51.0	70.7	85.3	81.8

*There are training failures with the use of AAST even when we adopt different kernels and kernel parameters. We calculate the accuracy of mean testing based on the results obtained from the correct learning models.

Set (AAST) were also tested in our experiments. To compare the detection performance, 100 experiments were performed on each feature set under each category of signal complexity in each detection. In each experiment, 30% of the audio files are randomly assigned to the training group and 70% are used for testing for steganalysis of Hide4PGP, Invisible Secrets, LSB-matching, and Steghide; 70% training, and 30% testing are randomly grouped in steganalysis of spectrum-spread audio watermarking. Support vector machines (SVM) with RBF kernels are used for classification. The results consist of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The classification accuracy is calculated as $w \times TP / (TP + FN) + (1 - w) \times TN / (FP + TN)$, where $w \in [0, 1]$ is a weighting factor. Without loss of generality, w is set to 0.5 in our experiments. Mean values for classification accuracy are listed in Table I. For comparing the five feature sets, the highest mean testing values are highlighted in bold.

Regarding the relation of detection performance to signal complexity—as shown in Table I—for signal and derivative-based Mel-cepstrum and AAST feature sets, as signal complexity increases, the detection performances generally decreases. However, there is no obvious performance deterioration

of the derivative-based Markov approach in high signal complexity. In a comparison of the five feature sets, second-order derivative-based Mel-cepstrum steganalysis improves the detection performance of signal-based Mel-cepstrum set in each category of signal complexity. Especially noticeable for detection of audio streams with high signal complexity, derivative-based Mel-cepstrum improves testing accuracy by about 15% to 20% for the steganalysis of Hide4PGP, Invisible Secrets, LSB-matching, and Steghide. Compared to signal-based Mel-cepstrum approaches, the second-order derivative-based Markov approach also gains significant advantage: the improvements are about 23% to 34% in detecting audio steganograms produced by Hide4PGP, Invisible Secrets, LSB-matching, and Steghide in high signal complexity. Additionally, the derivative-based Markov approach is better than the derivative-based Mel-cepstrum steganalysis for detecting Hide4PGP, Invisible Secrets, LSB-matching, and Steghide in high signal complexity. Although AAST includes all signal-based Mel-cepstrum features and several other statistical features, the detection performance is not as high as signal-based Mel-cepstrum audio steganalysis. Our study also shows that the standard deviation value of the testing results by using AAST is high; that is, the testing performance is not stable. We surmise that a statistical feature design of AAST is not ideal, which is verified by the statistical analysis of each individual feature in AAST.

We note that in steganalysis of steganographic systems, the derivative-based Markov approach takes the lead in testing accuracy, followed by derivative-based Mel-cepstrum method. However, in the steganalysis of audio watermarking, derivative-based Mel-cepstrum performs the best, except under high signal complexity. By combining the derivative-based Mel-cepstrum and Markov approaches, the testing results are very close to the best in each category of signal complexity; therefore, an effective detection system can be developed by incorporating both approaches.

In addition to the comparisons shown in Table I, the Receiver Operating Characteristic (ROC) curves using S-Mel, 2D-Mel, and 2D-MM are also given in Figure 5, for the steganalysis of Invisible (50% max-hiding), LSB-matching (50% max-hiding), Steghide (50% max-hiding), and the spread spectrum audio watermarking (abbreviated SSAW in the figure, max-hiding). Under the four categories of signal complexity (the ROC curves on Hide4PGP are similar to the curves on Invisible; to save space, the results are not included in Figure 5). Generally, the derivative-based Mel-cepstrum steganalysis outperforms the signal-based Mel-cepstrum approach, and the integration of derivative-based Mel-cepstrum and Markov approaches delivers the best detection performance, and the superiority is especially remarkable for steganalysis of audio streams with high signal complexity.

7. DISCUSSION

Second-order derivative-based methods have the advantage over the signal-based Mel-cepstrum audio steganalysis. Our explanation is that audio signals are generally band-limited, while the embedded hidden data is likely broadband, and most information-hiding inclines to randomly modify audio signals and tends to increase the high frequency information. Derivative-based detections first preprocess signals by extracting the derivative information, and it is relatively easy to expose the existence of hidden data. Consequently, derivative-based methods are more accurate in comparison with signal-based Mel-cepstrum audio steganalysis.

The derivative-based Markov approach obtains remarkable detection performance even in high signal complexity. On one hand, the range of i and j of the Markov transition feature, described in (16), is $[-6, 6]$. In other words, we extract the transition features from the smooth parts of audio streams, not from the audio streams in the temporal neighborhood with dramatic change or high complexity parts. Even when an audio is associated with high signal complexity, there are many smooth parts or subaudio streams with low signal complexity, and the difference between the magnitudes over the temporal neighborhood in these subaudio streams is not that big. Also, the Markov transition features are

Fig. 5. ROC curves for the steganalysis of Invisible Secret (50% max-hiding (a); LSB-matching (50% max-hiding (b); Steghide (50% max-hiding (c); and spread spectrum audio watermarking (abbreviated SSAW) max-hiding (d).

Fig. 5. Continued

Fig. 6. Spectrum of the second-order derivative of hidden data in a 44.1 kHz audio steganogram, shown in (a); spectrum of the detail wavelet sub-band of the same hidden data, filtered by using “db8”, shown in (b). The frequency shown in the x-axis (b) is reduced due to down-sampling of wavelet decomposition [Liu et al. 2009b] ©2009 IEEE.

correlated to these sub-audio streams. On the other hand, in most audio steganographic systems, payload embedding is not correlated to the audio signal; that is, these systems do not consider the signal complexity of the audio streams for adaptive hiding. Information-hiding also modifies the magnitude values in the subaudio streams with low and high signal complexity; in such case, Markov transition features extracted from the low complexity substreams obtain impressive detection accuracy in the audio signals with high signal complexity.

The advantage of the derivative-based Markov approach in steganalysis of spread spectrum watermarking is not so noticeable due to a different emphasis on watermarking that focuses on robustness against traditional signal processing. The merged derivative-based Mel-cepstrum and Markov approach still delivers good performance in different categories of signal complexity. Although AAST includes all signal-based Mel-cepstrum features and several additional statistical features, the detection performance is not as good as signal-based Mel-cepstrum audio steganalysis. It indicates that feature selection is also an important issue in steganalysis, which was conducted in our previous steganalysis study on digital images [Liu et al. 2008a, 2010].

In this article, the proposed steganalysis method was just tested on WAV uncompressed audio streams. To detect the information-hiding in the compressed domain. For example, for the steganalysis of MP3 audio streams, we utilize the statistics (mean, standard deviation, skewness, kurtosis) on the second-order derivative of the modified discrete cosine transform (MDCT) coefficients, and/or combine the statistics with the interframe MDCT statistics, an MP3-based audio steganographic system was developed successfully [Qiao et al. 2009].

We can use a high-frequency filter such as wavelet analysis instead of second-order derivative and then obtain the Mel-cepstrum features, which is also better than signal-based Mel-cepstrum audio steganalysis [Liu et al. 2009b]. In general, this alternative approach is not better than the second derivative-based Mel-cepstrum solution, which was verified by our experiments. Our analysis indicates that the application of a high-frequency filter such as “db” wavelet will produce the high-frequency signal that is similar to white noise, and that the spectrum is almost equally distributed over the entire frequency band. However, the second derivative suppresses the energy in low frequency and amplifies the energy in high frequency; the spectrum does not distribute equally over the entire frequency band. Figure 6(a) shows the spectrum of the second derivative of the hidden data, called error signal in the figure, in an audio steganogram. Figure 6(b) plots the spectrum of the detail wavelet subband of the same hidden data, filtered by using “db8”. Based on Eq. (11) in Section 2, as the error spectrum

increases, the expected value of the variance of the audio steganogram will prominently increase; that is, the rate of power change in different spectrum bands will change dramatically, since the Mel-cepstrum coefficients are used to capture the information for power change, in which case the advantage of the derivative-based Mel-cepstrum approach is noticeable.

It should be noted that signal complexity may be measured in different ways. In addition to the signal complexity defined by (17) and the GGD shape parameter that was adopted in image steganalysis [Liu et al. 2008a, 2008b], entropy-based measurements can be used to measure the signal complexity. An audio signal and the second-order derivative are denoted by f and D_f^2 , respectively; the values of information entropy are expressed by $H(f)$ and $H(D_f^2)$ accordingly, in terms of a discrete set of probability $p(f)_i$ and $p(D_f^2)_i$,

$$H(f) = - \sum_i p(f)_i \log p(f)_i \quad (18)$$

$$H(D_f^2) = - \sum_i p(D_f^2)_i \log p(D_f^2)_i \quad (19)$$

Figure 7 compares the testing results of the Matthews Correlation Coefficient (MCC), which is generally used as a balanced measure, even when the classes are of very different sizes regarding the quality of binary classification, in the steganalysis of invisible steganograms with 50% maximal hiding capacity, using S-Mel, 2D-Mel, and 2D-MM feature sets with the complexity measurements $C(f)$, $H(f)$, and $H(D_f^2)$, respectively. Because $C(f)$, $H(f)$, and $H(D_f^2)$ have different values and ranges, these three types of measurements have been mapped to the same signal complexity space, shown by X-label values with mono-increasing from the left (low complexity) to the right (high complexity). The results also indicate that signal complexity is a significant parameter for the evaluation of steganalysis performance; derivative-based Mel-cepstrum steganalysis outperforms signal-based Mel-cepstrum audio steganalysis; the 2D-MM feature set exhibits the unbeatable superiority, especially in steganalysis of the signals with high complexity. Our study on other steganographic systems arrived at similar results.

Figure 8(a) and (b) shows the joint densities of $C(f)$ and $H(f)$ and $C(f)$ and $H(D_f^2)$, respectively. It roughly demonstrates that $H(f)$ and $H(D_f^2)$ increase while $C(f)$ increases. Although there are different ways to measure the signal complexity, the calculation of $C(f)$ has the advantage of low computational cost compared to entropy-based measurements.

8. CONCLUSIONS

In this article, we propose novel stream data-mining based on the second-order derivative to discover the existence of covert message in audio streams. We extract the Mel-cepstrum coefficients and Markov transition features of the second-order derivative and apply a support vector machine to the extracted features. Additionally, to allow a complete and fair evaluation of audio steganalysis performance, a metric for signal complexity is introduced, and we experimentally explore the relation of signal complexity to detection performance.

In comparison to a recently proposed audio steganalysis method, which is based on Mel-cepstrum coefficient-mining on signal streams, our method exhibits a prominent advantage in steganalysis of several types of audio steganograms under all categories of signal complexity. Especially remarkable is the fact that, in detecting steganography in audio streams with high signal complexity, while the method above (for comparison) does not perform well at all, our method delivers superior performance by merging second-order derivative-based Mel-cepstrum coefficients and Markov transition probability features.

Fig. 7. Steganalysis performance using S-Mel (a), 2D-Mel (b), and 2D-MM (c) features sets with the complexity measurements $C(f)$, $H(f)$, and $H(D_f^2)$.

Future work may include finding smaller feature sets; extending the steganalysis performance evaluation framework to include analysis of computational complexity; and building benchmark-testing sets to facilitate cross-validation of new results.

ACKNOWLEDGMENTS

We are grateful to Dan Ellis of Columbia University for his insightful discussions and invaluable suggestions, to Malcolm Slaney of Yahoo! research and Haojun Ai for their very nice discussions, and to Jana Dittmann and Christian Kraetzer for kindly providing us with the AAST document.

Fig. 8. Joint density of the complexity measurements $C(f)$ and $H(f)$ (a), $C(f)$ and $H(D_f^2)$ (b).

Special thanks to Mohan Kankanhalli, and the anonymous reviewers for their insightful comments and very helpful suggestions.

REFERENCES

- AVCIBAS, I. 2006. Audio steganalysis with content-independent distortion measures. *IEEE Signal Process. Lett.* 13, 2, 92–95.
- BOGERT, B., HEALY, M., AND TUKEY, J. 1963. The frequency analysis of times series for echoes: cepstrum, pseudoautocovariance, cross-cepstrum, and saphe cracking. In *Proceedings of the Symposium on Time Series Analysis*.
- FARID, H. 2002. Detecting hidden messages using higher-order statistical models. In *Proceedings of the 2002 International Conference on Image Processing (ICIP'02)*. 905–908.
- FRIDRICH, J. 2004. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In *Information Hiding*, Lecture Notes in Computer Science, vol. 3200, Springer, Berlin, 67–81.
- GONZALEZ, R. AND WOODS, R. 2008. *Digital Image Processing* 3rd ed. Prentice Hall, Englewood Cliffs, NJ.
- HARMSEN, J. J. 2003. Steganalysis of additive noise modelable information hiding. Master's thesis, Rensselaer Polytechnic Institute, Troy, NY.
- HARMSEN, J. AND PEARLMAN, W. 2003. Steganalysis of additive noise modelable information hiding. In *Proceedings of the SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents*. vol. 5020, 131–142.
- HETZL, S. AND MUTZEL, P. 2005. A graph-theoretic approach to steganography. In *Communications and Multimedia Security*, Lecture Notes in Computer Science, vol. 3677, Springer, Berlin, 119–128. The code is available at <http://steghide.sourceforge.net/>.
- HILL, T. AND LEWICKI, P. 2005. *Statistics: Methods and Applications*. StatSoft, Inc.
- HOLOTYAK, T., FRIDRICH, J., AND VOLOSHYNOVSKIY, S. 2005. Blind statistical steganalysis of additive steganography using wavelet higher order statistics. Lecture Notes in Computer Science, vol. 3677, Springer, Berlin, 273–274.
- JOHNSON, M., LYU, S., AND FARID, H. 2005. Steganalysis of recorded speech. In *Proceedings of the SPIE*. vol. 5681, 664–672.
- KIROVSKI, D. AND MALVAR, H. S. 2003. Spread spectrum watermarking of audio signals. *IEEE Trans. Signal Process.* 51, 4, 1020–1033. The audio watermarking hiding tool is available at <http://research.microsoft.com/en-us/downloads/885bb5c4-ae6d-418b-97f9-adc9da8d48bd/default.aspx>.
- KRAETZER, C. AND DITTMANN J. 2007. Mel-cepstrum based steganalysis for VOIP-steganography. In *Proceedings of the SPIE*. vol. 6505.
- LIU, Q. AND SUNG, A. H. 2007. Feature mining and neuro-fuzzy inference system for steganalysis of LSB matching steganography in grayscale images. In *Proceedings of the 20th International Joint Conference in Artificial Intelligence (IJCAI)*. 2808–2813.
- LIU, Q., SUNG, A. H., CHEN, Z., AND XU, J. 2008a. Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images. *Patt. Recogn.* 41, 1, 56–66.
- LIU, Q., SUNG, A. H., RIBEIRO, B., WEI, M., CHEN, Z., AND XU, J. 2008b. Image complexity and feature mining for steganalysis of least significant bit matching steganography. *Inf. Sci.* 178, 1, 21–36.
- LIU, Q., SUNG, A. H., RIBEIRO, B., AND FERREIRA, R. 2008c. Steganalysis of multi-class JPEG images based on expanded Markov features and polynomial fitting. In *Proceedings of the 21st International Joint Conference on Neural Networks (IJCNN)*. 3351–3356.
- LIU, Q., SUNG, A. H., AND QIAO, M. 2008d. Detecting information-hiding in WAV audios. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR)*. 1–4.

- LIU, Q., SUNG, A. H., AND QIAO, M. 2009a. Improved detection and evaluation for JPEG steganalysis. In *Proceedings of the 17th ACM International Conference on Multimedia (MM'09)*. ACM, New York, 873–876.
- LIU, Q., SUNG, A. H., AND QIAO, M. 2009b. Temporal derivative based spectrum and mel-cepstrum audio steganalysis. *IEEE Trans. Inf. Forensics Security* 4, 3, 359–368.
- LIU, Q., SUNG, A. H., QIAO, M., CHEN, Z., AND RIBEIRO, B. 2010. An improved approach to steganalysis of JPEG images. *Inf. Sci.* 180, 9, 1643–1655.
- LIU, Q., SUNG, A. H., AND QIAO, M. 2011. Neighboring joint density based JPEG steganalysis. *ACM Trans. Intell. Syst. Technol.* 2, 2, Article 16.
- LIU, Y., CHIANG, K., CORBETT, C., ARCHIBALD, R., MUKHERJEE, B., AND GHOSAL, D. 2008. A novel audio steganalysis based on high-order statistics of a distortion measure with Hausdorff distance. *Lecture Notes in Computer Science*, vol. 5222, Springer, Berlin, 487–501.
- LYU, S. AND FARID, H. 2006. Steganalysis using higher-order image statistic, *IEEE Trans. Inf. Forensics Security* 1, 1, 111–119.
- MCEachern, R. 1994. Hearing it like it is: Audio signal processing the way the ear does it. *DSP Applications*.
- OZER, H., SANKUR, B., MEMON, N., AND AVCIBAS, I. 2006. Detection of audio covert channels using statistical footprints of hidden messages. *Digital Signal Process.* 16, 4, 389–401.
- PEVNY, T. AND FRIDRICH, J. 2007. Merging Markov and DCT features for multi-class JPEG steganalysis. In *Proceedings of the SPIE Electronic Imag.* vol. 6505.
- QIAO, M., SUNG, A. H., AND LIU, Q. 2009. Steganalysis of MP3stego. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN'09)*. 2566–2571.
- REYNOLDS, D. 1992. A Gaussian mixture modeling approaching to text-independent speaker identification. Ph.D. dissertation, Department of Electrical Engineering, Georgia Institute of Technology.
- SHARP, T. 2001. An implementation of key-based digital signal steganography. In *Proceedings of the 4th International Workshop on Information Hiding*, Lecture Notes in Computer Science, vol. 2137, Springer, Berlin, 13–26.
- SHI, Y., CHEN, C., AND CHEN, W. 2007. A Markov process based approach to effective attacking JPEG Steganography. In *Information Hiding*, Lecture Notes in Computer Science, vol. 4437, Springer, Berlin, 249–264.
- VAPNIK, V. 1998. *Statistical Learning Theory*. Wiley, New York.
- ZENG, W., AI, H., AND HU, R. 2007. A novel steganalysis algorithm of phase coding in audio signal. In *Proceedings of the 6th International Conference on Advanced Language Processing and Web Information Technology*. 261–264.
- ZENG, W., AI, H., AND HU, R. 2008. An algorithm of echo steganalysis based on power cepstrum and pattern classification. In *Proceedings of the International Conference on Information and Automation*. 1667–1670.

Received August 2008; revised April 2010; accepted May 2010