Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **1**, January **2017**

Energy and Entropy Based Features for WAV Audio Steganalysis

Fatiha Djebbar

College of Information Technology
UAE University
Al Ain, UAE
fdjebbar@uaeu.ac.ae

Beghdad Ayad

Engineering College
Emirates Aviation University
Dubai, UAE
beghdadayad@gmail.com

Received May, 2016; revised September, 2016

Abstract. *Digital steganalysis techniques attempt to detect hidden information in digital media. The rising interest in steganalysis is attributed to the growing number of steganography algorithms and the threats they represent. This article presents a combined maximum entropy energy approach for audio steganalysis. First, the audio signal is divided into four energy-based regions: noise, low, medium and high; then entropy is computed from each region. Finally, a support vector machine is applied to the collected features for discovering the hidden data in audio signals. Active speech level algorithm is used to capture energy fluctuation in audio streams. The paper shows that the extracted features from separate energy-based regions of the signals have significantly improved detection accuracy of hidden messages. Our work includes comparisons with current state-of-the-art audio steganalysis techniques. The experimental results show that our method achieves up to 96.7% correct for an embedding rate of 25% or above of stego-signals produced by S-tools4, Steghide and Hide4PGP while using a much smaller feature set.*

Keywords: Information hiding, Audio steganalysis, Audio steganography, Signal processing, Maximum entropy, Energy.

1. **Introduction.** Digital steganography is a nascent but rapidly flourishing technique that has emerged as a prominent source of data security. Digital steganography techniques exploit the characteristics of digital media by utilizing them as carriers (covers) to hold hidden information. Covers can be of different types including image [1], audio [2, 3, 23], video [5], text [6], and IP datagram [7]. Steganography entails the undetectable modification of a multimedia file to embed data, in contrast to encryption which relies on rendering this data unreadable to a third party [8]. Steganography techniques have found their way into various and versatile applications. However, some of these applications are pernicious [9, 10]. Attempting to detect the presence of hidden message, is the primarily objective of a steganalyst. Steganalysis algorithms are regarded as "attacks" against steganography algorithms. These attacks could be significantly challenging especially when the only information available is the stego-file.

Steganalysis techniques in compressed and uncompressed audio format have been actively investigated in the last decade. Most steganalysis method presented lately are based on learning to differentiate between cover- and stego-audio signals. The learning process is performed by a machine learning such as, a support vector machine (SVM) on a dataset fed with statistical properties (features) extracted from the cover and stego-audio signals. The right choice of these features reinforces the discriminatory power between the cover- and stego-audio signals.

As the widespread use of WAV audio signal, various steganographic methods have been suggested and inevitably many steganalysis schemes have been developed to attack them. Authors in [11] have presented a WAV audio steganalysis algorithm to capture irregularities between cover and stego signals' spectrograms in high quality recorded speech. The spectrogram of each frame in the signals is calculated using Short-Time Fourier Transform (STFT). The collected features are classified through a non-linear SVM. According to the authors, this approach is more suitable for high-bit rate audio steganography such as, LSB methods and Hide4PGP. The use of audio quality measures for audio steganalysis was proposed by [12]. The authors used audio quality measures (i.e, signal-to-noise ratio, Log likelihood) to distinguish between the stego-signal and its de-noised version (used as an estimate to the cover-signal). ANOVA test [13] and sequential floating search [14] were used to select the most appropriate measures to better detect the presence of hidden messages. In order to classify a signal as stego or cover, a linear regression and support vector machines (SVM) classifiers were trained using the selected audio quality measures. [15] proposed content-independent distortion measures as features for the classifier design to improve the latter method. They proposed to use a single reference signal (common to all tested signals) instead of creating a reference signal via a de-noised version of the stego-signal. Authors in [16] extracted features from the histograms of both statistical moments and frequency domain of the tested audio signal. Comparatively, [17] have used only higher order statistical moments of histogram and frequency histogram for the signal and its wavelet sub-bands. The same principle in selecting the features was followed by [18], but the signal reference was a self-generated signal via linear predictive coding. [19] proposed an algorithm based on Mel-cepstrum to detect embedded messages. Authors in [20] combined MelCepstrum feature with temporal derivative-based spectrum analysis.

To detect hidden information in MP3 audio files, [23] presented a detection method for MP3Stego [22] based on the differential statistics of quantization step and in [24] by exploiting recompression calibration-based feature of the number of bits in the bit reservoir. In AMR compressed audio [25] used the joint probability of same pulse position matrix as feature.

Although previous research on audio steganography has managed to detect hidden data in various audio formats, it mostly relied on the change of the intrinsic properties of the audio signals (e.g., mel cepstrum, linear prediction coefficients (LPC), audio distortion measures, etc.) to distinguish between stego- and cover-audio signals. In addition, up to our knowledge most of the current literatures compute these properties over full-band audio signals, a process that could dilute the embedding error's effect on the stego-audio signal.

Thus, in order to capture all variations in WAV audio streams due to embedding and to generate a set of meaningful features to the support vector machine we study continuous homogeneous energy-based audio stream segments. The proposed algorithm has two stages. First, each audio signal frame is classified based on its energy level as: noise, low, medium or high using active speech level algorithm (ASL), defined in ITU-T Recommendation P.56 [38]. Frames from same energy are grouped together to generate four energy-based regions of the audio signal (noise, low, medium and high). Second, maximum entropy is computed from each energy-based region to further enhance the discriminatory power of the classifier between stego- and cover-signal. As a result we are able to scan, analyze and show the impact of the embedding-process within all energetic regions of the audio signals. In our opinion, our maximum entropy energy audio steganalysis algorithm (EE-AS) is novel because segmenting the signal into energy-based regions has resulted in generating more meaningful features which reduced the false positive rate. The main contribution of this paper lies in the following:

1. Integrating energy and maximum entropy, a powerful duo, for features selection.
2. Capturing all possible variations in the audio streams resulting from data hiding. The features are extracted from full-band and from separate energy-based regions of the audio signal as it is commonly understood that, maximum entropy of the full-band spectrum by itself captures only the gross peakiness of the spectrum.
3. Our proposed method has a small features selection set size while achieving 97.67% detection rates for stego-audio signals.
4. Our scheme has lower false positive because instead of having one scan over the whole signal we reduced the margin of error by scanning multiple continuous homogeneous segments of the signal.

To assess the performance of the proposed method, LSBs-based audio steganographic software like: Steghide, S-Tools, Hide4PGP, found respectively in [26, 27, 28] are used to generate the stego-signals. EE-AS algorithm has a generalized design for detecting hidden data in audio signals, however, these software were adopted for practicality and usability reasons since they were used by the algorithm we are comparing our work to [20].

This paper is organized as follows: Section 2 presents a theoretical background for entropy and justifies its use in the present context. Energy as a key parameter in audio steganalysis is presented in Section 3 and in Section 4. Section 5 and Section 6 discuss the preprocessing steps to generate the features used to distinguish between cover- and stego-audio signals. In Section 7, classification results by SVM and evaluation study are revealed. Finally, we conclude this paper with a conclusion of our work in Section 8.

2. **Signal Entropy.**

The entropy of a signal is a measure of the amount of information a signal carries. The successful exploitation of entropy features in speech recognition [29] and voice activity detection [30] gave rise to their use in the context of signal and image processing. The application of the entropy concept for signal and image steganalysis is based on the fact that embedding techniques would modify the probability of bits of cover medium which in turn will change the entropy value. An audio signal is denoted $x(t)$ where (t =0,1,2,...,N-1). Similar to the additive noise model proposed in [31], a stego-signal is denoted s(t), which can be modeled by adding a noise or error signal $e(t)$ to the original signal $x(t)$; $s(t) = x(t) + e(t)$. The entropy of $x(t)$, $e(t)$ and $s(t)$ are denoted $H(x)$, $H(e)$ and $H(s)$, respectively. To uniquely identify the value of x, Shannons entropy [32] is employed. Small perturbations on the sample values of x produce smaller perturbations on the measured entropy. If the sample values of x are denoted by x_i then entropy is defined as follow:

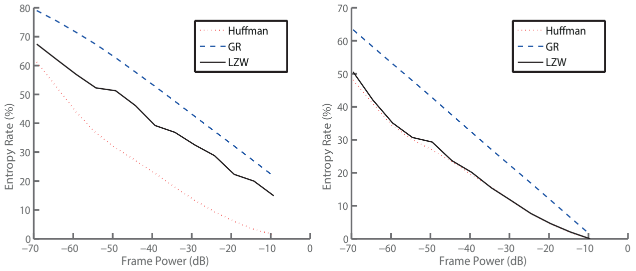
$$H(x) = -\sum_i p(x_i).ln(p(x_i)) \tag{1}$$

where $p(x_i)$ is the probability for the signal to take values x_i. The entropy of the sum of two independent discrete random variables (x, e), e.g., $H(s)$ is at least the minimum of their individual entropies [33, 34] given by:

$$max\{H(x), H(e)\} \leq H(s) \leq H(x) + H(e) \tag{2}$$

Thus, by adding more terms in the summation [33], the entropy can only increase.

3. **Maximum Entropy and Signal Energy.** Audio signal contains different amount of entropy and energy levels over time. Since stego-signal is additive [31], data hiding will change the entropy value. The effect of this change is related to changes in the signal energy. The proposed steganalysis algorithm exploits these irregularities in the stego-signals to detect steganography. Entropy-based algorithms in information theory literature such as Shannons entropy [32], Lempel-Ziv-Welch (LZW) complexity [35], Huffman [36] and Golomb-Rice coding (GR) [37], can be used to measure the irregularities in the tested signals. As a show case we used three distinct algorithms: LZW, Huffman and GR. These algorithms were chosen because of the differences in entropy values generated for the same audio stream. This feature leads to more informative data that can be used to enhance the classifier capability. Maximum entropy values shown in Figure 1 are computed from noise regions Figure (1a) and full-band signal Figure (1b). These Figures also illustrate the relation between entropy and signal energy, we can observe that entropy value increases as energy level decreases. Moreover, entropy values computed on full-band audio streams are not strong features on their own. Hence, the similarity between feature values collected from Huffman and LZW will not be of much help to the classifier. Consequently, time domain full-band entropy captures the gross distortion due to data hiding, for improved resolution, entropy features are also computed from energy-based regions: noise, low, medium and high of the audio signal. Audio division is carried out by ASL algorithm as presented in the next section.



(A) Noise region

(B) Full-band signal

Figure 1. Tradeoff between entropy and energy

4. **Energy-Based Signal Division.** Active speech level algorithm (ASL) is used to discriminate between noise, low, medium and high energy regions within the audio signal. ASL determines speech activity factor (Sp_l) which represents the fraction of time where the signal is considered to be active speech and the corresponding active level for the speech part of the signal [38]. The speech activity algorithm computes the speech energy value at each sample time (frame). To determine which frames belong to the high, medium, low and noise power classes, the active speech level Sp_l of the signal is first determined according to [38], then compared with a discrete set of thresholds as presented in Table 1. This thresholds set is chosen based on experimental considerations [39, 40] and they are specific to normalized audio files of our datasets. A detailed classification of the frames into the four power classes is presented in [39]. To build our audio signals database, we collected 1080 on-line audio files of 10 s length each from different types such as speech signals in different languages (i.e, English, Chinese, Japanese, French, and

Table 1. Datasets composition statistics in terms of audio signal parts and thresholds used to categorize the frames as noisy, low, medium or high.

Power classes	Audio (%)	Speech (%)	Music (%)
	10800 sec	6200 sec	4600 sec
Noise	14.76	22.47	2.6
Low	15.54	22.56	6.5
Medium	50.03	39.17	63.64
High	19.67	15.8	27.26

Power classes	Threshold (dB)
noise	-45
Low	-35
Medium	-25
High	-15

Thresholds for power classes.

Arabic), and music (classic, jazz, rock, blues). All signals are sampled at 44.1 kHz and quantized at 16-bits. The statistics of the dataset are shown in Table 1. An example of audio-signal division process based on the set of the chosen thresholds is illustrated in Figure (2).

5. **Features Extraction.** The features extraction step starts by creating features vector representing the entropy value difference (H_{Dif}) between received (tested) audio-signals s_t and their self-generated reference version s_r. which are created by randomly modifying the first LSB layer in the temporal domain of the given signal using S-tools4, Steghide and Hide4PGP. Features extraction is achieved in three main steps detailed in the following subsections:

5.1. **Frames classification.** Each signal (tested and its reference) is divided into 4 parts based on the energy level. This process is detailed as follows:

1. The audio signal is split into M frames of 10 ms and N samples each, $s_t(m, N), \quad 1 \leq m \leq M$.
2. Compute Sp_l as in ITU-T P.56 [38] for each frame using the library tool voicebox [41].
3. Classify the frame as high, medium, low or noisy by comparing its Sp_l to the values shown in Table 1.
4. Reassemble the frames of the same category into one part as shown in Figure 2. At the end of the process, each audio file is divided into four energy-based regions: noisy, low, medium and high. The classification process is described in the following Algorithm.

5.2. **Entropy Computation.** Following signals division phase, maximum entropy (η_i) for each energy-based region is computed using three distinct entropy algorithms: Huffman, LZW and GR. This process results in 15 entropy values (η_i, i=1...15), three values for each energy-based region of the audio signal (noisy, low, medium, high and full band signal). The following example shows how to compute noisy region entropy values: η_1, η_2 and η_3.

Algorithm 1 Signal division to energy-based regions

```
INPUT: $s_t(m, N)$
OUTPUT: $s_{t_{\text{Noise}}}, s_{t_{\text{Low}}}, s_{t_{\text{Medium}}}, s_{t_{\text{High}}}$
for $m = 1$ to $M$ do
  if $Sp_l(s_t(m, N)) \leq -45$ then
    $s_{t_{\text{Noise}}} + \leftarrow s_t(m, N)$
  end if
  if $-45 \geq Sp_l(s_t(m, N)) \leq -35$ then
    $s_{t_{\text{Low}}} + \leftarrow s_t(m, N)$
  end if
  if $-35 \geq Sp_l(s_t(m, N)) \leq -25$ then
    $s_{t_{\text{Medium}}} + \leftarrow s_t(m, N)$
  end if
  if $-25 \geq Sp_l(s_t(m, N)) \leq -15$ then
    $s_{t_{\text{High}}} + \leftarrow s_t(m, N)$
  end if
end for
```

- $\eta_1 \leftarrow Huffman(s_{t_{\text{Noise}}})$
- $\eta_2 \leftarrow LZW(s_{t_{\text{Noise}}})$
- $\eta_3 \leftarrow GR(s_{t_{\text{Noise}}})$

Similarly, we calculate the entropy values of the reference signal to produce the entropy difference ($H_{\text{Dif}i}$, i=1...15). $H_{\text{Dif}i}$ is computed between similar energy regions of tested signal s_t and its reference s_r as shown in the following equation:

$$H_{\text{Dif}i} = \sqrt{|\eta_{ti}|} - \sqrt{|\eta_{ri}|} \tag{3}$$

The reason for presenting $H_{\text{Dif}i}$ by Eq.3 is that square root is not differentiable at 0. Hence, small η, $\sqrt{|\eta_{ti}|} - \sqrt{|\eta_{ri}|}$ may be much larger than $|\eta_{ti} - \eta_{ri}|$ and eventually will signify their impact in the classification process.

The features vector of each audio signal contains 15 coefficients:
$Features = H_{\text{Dif1}}, H_{\text{Dif2}}, H_{\text{Dif2}}, ..., H_{\text{Dif15}}$

The 15 H_{Dif} coefficients are the features set to be fed to the SVM classifier with an RBF kernel. Each H_{Dif} coefficient is computed by one of the entropy algorithms on an energy-level region of the audio signal (i.e, H_{Dif1} is the difference in entropy value measured by rar in the noisy regions of the tested signal and its reference). Thus, a features vector of 15 ratio coefficients is retrieved from each audio signal such as: H_{Dif1}, H_{Dif2},..., H_{Dif15}. The feature extraction process is further illustrated in Figure 3.

Figure 4 shows entropy values computed for the cover-audio, stego-audio and their reference versions. The results can be summarized as follow:

1. Entropy values in higher energy regions (e.g., medium and high) are smaller than those of lower energy regions.
2. The cover and the stego signals are better discriminated in lower energetic parts of the audio signals.
3. Regardless of the energy level, the difference in entropy value between the cover and its reference signal 4a is always larger than that of the stego and its reference 4b. This observation is in fact the key to detecting the existence of hidden data.

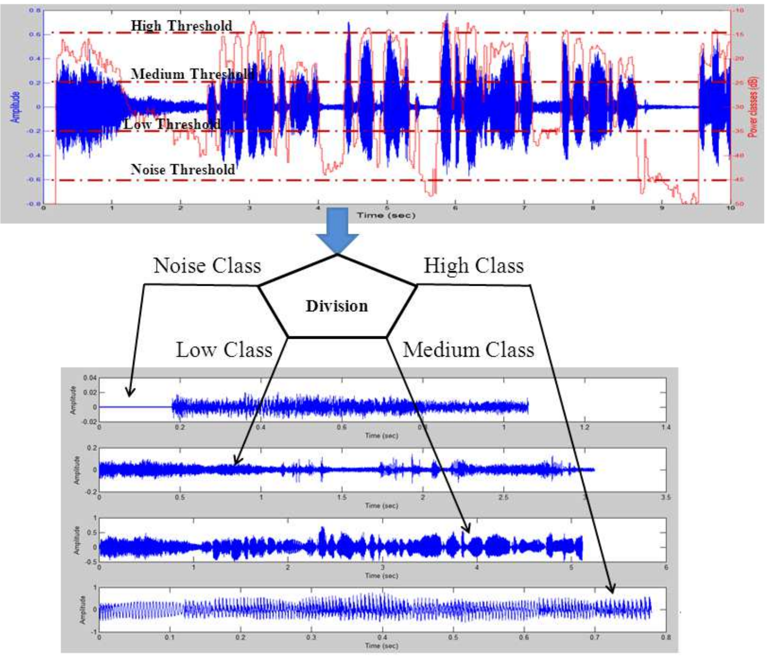


FIGURE 2. Speech signal division energy-based region using ASL. Temporal representation (blue curve) of a speech signal (left y-axis) and its division to different power classes (noise, low, medium and high) using the energy in (dB) per audio signal frame (right y-axis) computed by ASL (red curve) and classified to power classes using the thresholds set.

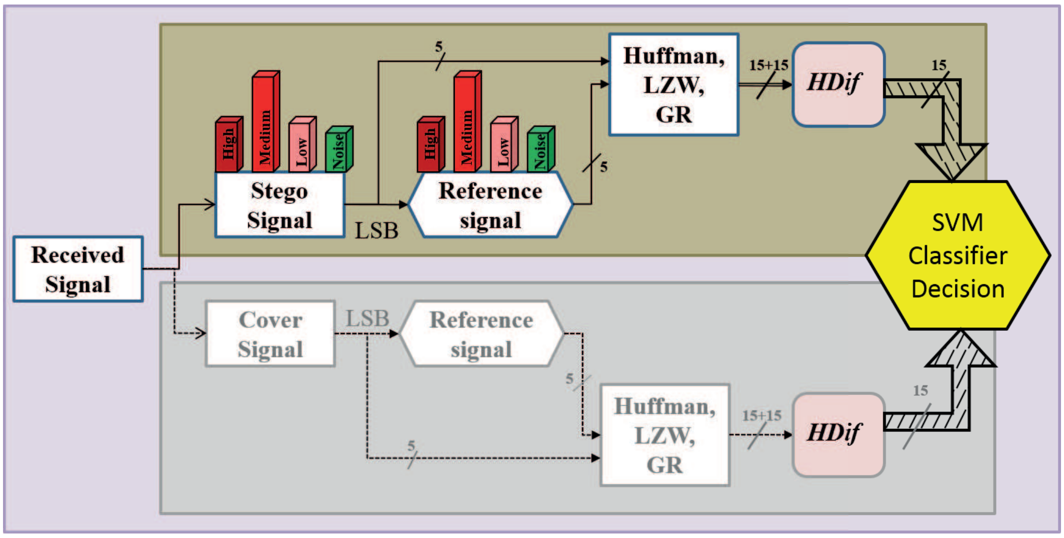


FIGURE 3. Features extraction workflow.

6. Features Classification.

6.1. **Datasets.** For each tested steganographic tool, two datasets are produced: training and testing (Tr and Ts). Each dataset contains 1080 WAV audio signals of 10 s length. All signals are sampled at 44.1 kHz and quantized at 16-bits. Each training and testing dataset contains 540 positive (stego) and 540 negative (cover) audio samples. We used on-line audio files from different types such as speech signals in different languages (i.e, English, Chinese, Japanese, French, and Arabic), and music (classic, jazz, rock, blues).

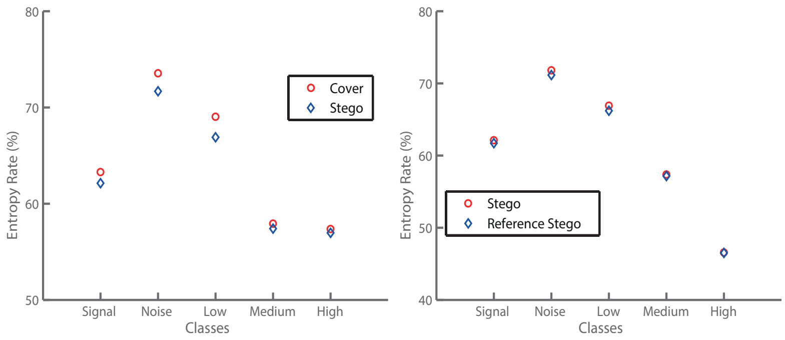


FIGURE 4. Entropy values of cover- (a), stego-signal (b) and their respective reference versions.

All stego-audio signals are generated by hiding data from different types: text, image, audio signals, video and executable files. The stego-signals produced by Hide4PGP with 25% maximal hiding capacity while in Steghide and Stools with 50%.

The datasets Tr and Ts consist of a matrix of $\{H_{\text{Dif}i}, l_i\}$,where $H_{\text{Dif}i}$ refers to 15 entropy features, and $l_i \in \{\pm 1\}$. The values +1 and -1 correspond to "Stego-audio" and "non Stego-audio" classes, respectively. The performance of the proposed steganalysis algorithm is measured by the ability of the system to recognize and distinguish between stego- and cover-audio signals.

6.2. **Classification Step.** The problem is basically formulated as a two-class, classification problem: both training and testing sets contain audio signals belonging to either cover- or stego-signal. This representation is combined with a powerful machine learning technique (SVM) [42], which is used to classify the two sets, and has proved to be very effective classifier in the last decade. SVMs, enjoy strong foundations in statistical learning theory, and have been successfully applied to data classification. The SVM algorithm addresses the general problem of learning to discriminate between positive and negative examples of a given class of n-dimensional vectors. SVM has been successfully applied to a number of applications ranging from bioinformatics [43], face detection [44], digital images steganalysis [45] and to digital audio steganalysis [15, 20]. In addition, SVM does not normally require complex parameters tuning. It minimizes the prediction error and generalizes well even for small training samples [43]. SVM classifier is used in conjunction with the Radial Basis Function (RBF) kernel [46].

In this study, we employed SVMs library tool [47]. The tuning parameters: $\gamma(> 0)$ the scaling parameter and $C(> 0)$ the regularization parameter which decides the trade-off between the training error and the margin of separation are set to 0.1 and 1 respectively. We prepared Tr and Ts datasets (listed in the previous section) for each steganography tool involved in this experiment: Steghide, S-Tools4, Hide4PGP v2.0 and our steganographic algorithm.

7. **Results and Discussion.** Our assessment of the performance of EE-AS on our datasets is based on two experiments. The first experiments assesses the recognition ability of our method in classifying stego- and cover-signals and it contains three scenarios. The second experiments compares the performance of our technique to other state-of-the-art audio-steganalysis methods. The mean and standard deviation (STD) of the length of 540 training and 540 testing audio samples datasets are listed in Table 2.

The accuracy of our predictions is measured by Precision (PP), Recall (R), F-measure, and the receiver operating characteristic (ROC). The precision is defined as the ratio of

Table 2. Mean and standard deviation of training and testing audio signals datasets

	Mean	**STD**
Training dataset	4.5510e-006	0.1228
Testing dataset	1.6519e-005	0.0637

$\frac{TP}{(TP+FP)}$. The recall is defined as $R = \frac{TP}{(TP+FN)}$. The F-measure combines precision and recall such as: $2.\frac{PP.R}{(PP+R)}$. The ROC is the fraction of true positives (TPR = true positive rate) versus the fraction of false positives (FPR = false positive rate). In this experiment, TP, FP and FN are defined in the contingency Table 3.

Table 3. The contingency table

	Stego-signal	**Cover-signal**
Stego classified	True positives (TP)	False Negatives (FN)
Cover classified	False Positives (FP)	True Negatives (TN)

The entries of the contingency table are described as follows:

- *TP*: stego-audio signal classified as stego-audio signal
- *TN*: cover-audio signal classified as cover-audio signal
- *FN*: stego-audio signal classified as cover-audio signal
- *FP*: cover-audio signal classified as stego-audio signal

It is possible that we might end-up with a high precision results (despite having a high number of FN), or a high recall (despite having a high number of FP). However, F-measure and the area under the ROC curve are two of the most popular computational methods to find a balance between false positives and false negatives.

7.1. **Performance Evaluation of EE-AS.**

7.1.1. *Scenario 1.* In this scenario, we assess the recognition ability of our method EE-AS in classifying 540 stego and 540 non-stego audio files (speech and music). The 15 features set are extracted from different energy-based regions: noisy, low, medium, high and full band audio signal. In Table 4, we record the overall accuracy where higher score values are interpreted as high detection rate. In this work, we were able to achieve high detection scores especially for Hide4PGP which shows almost 100% recognition rate. However it is pertinent to provide more analysis of the algorithm's performance and results.

Table 4. ROC and F-measure recorded from testing EE-AS on a 540 dataset of Audio signal

Hiding methods	**F-measure**	**ROC**
S-Tools	0.909	091
Steghide	0.791	0.795
Hide4PGP	1	1

7.1.2. *Scenario 2.* We assess the recognition ability of our method in classifying between 540 stego and 540 audio signals by using entropy features extracted only from full-band signal. The feature vector used contains only three elements and each of these elements is generated by one of the entropy algorithms (Huffman, LZW and GR). F-measure and ROC are listed in Table 4.

Table 5. ROC and F-measure recorded from testing EE-AS by using features vector of three elements extracted from full-band audio signals.

Hiding methods	**F-measure**	**ROC**
S-Tools	0.775	0.785
Steghide	0.578	0.63
Hide4PGP	0.985	0.985

Despite the small size of the feature set (3 elements) used, we managed to achieve good results, especially in detecting S-Tools and Hide4PGP steganography. Entropy features extracted from full-band signals are not strong on their own and have to be combined with features from multi energy regions (Table 4). Nevertheless, all these results show the high potential of entropy-based audio steganalysis.

7.1.3. *Scenario 3.* For the third scenario, we further investigate the recognition ability of our algorithm when the dataset contains only speech or music signals. The aim of this experiment is to put more emphasis on the behavior of the proposed algorithm when music-audio signals are used to convey hidden data versus those of speech. We split the dataset into two sets A (460 speech signal) and B (460 music signal). Each set is further split into 230 stego- and 230 cover-signal to create a training and testing dataset for speech and music. A set up similar to that described in scenario 1 is employed. In Table 6 we show ROC and F-measure. The statistical characteristics of speech- and music-signals are given in Table 1. Overall, these results show the high discriminatory power of energy-based statistics features on the detection accuracy. The features extracted from speech signals offer better accuracy compared to music, therefor speech signals are more vulnerable to our steganalysis method. This is due to the fact that the percentage of lower energy regions in speech (Table 1) is higher compared to that of music audio-signals.

Table 6. Music versus Speech energy-entropy audio Steganalysis

Hiding methods	**Audio Type**	**F-measure**	**ROC**
S-Tools	Speech	0.94	0.94
	Music	0.885	0.885
Steghide	Speech	0.797	0.805
	Music	0.763	0.775
Hide4PGP	Speech	1	1
	Music	0.9999	0.9999

7.2. **Comparison of Time Complexity and Detection Accuracy.** We tested another state-of-the-art audio-steganalysis methods, 2D-Mel [20] and we compared to our methods based on time complexity and classification rates. Additionally, we also tested 2D-Mel with a reduced feature set based on filtered MFCCs (29 FMFCCs) coefficients instead of combined 58 FMFCCs and MFCCs stated in [20]. The test is carried out to achieve two main objectives. First, to reduce the computation time of 2D-Mel. Second, to fully align with Liu's proposal where he proved that FMFCCs are the main source of embedding error. F-measure and ROC values of the primary experiment on a dataset of 540 training and 540 testing audio-signals are summarized in Table 7. Higher scores in the table correspond to a more accurate detection performance.

TABLE 7. F-measure and ROC recorded from testing EE-AS, 2D-Mel on a 540 dataset of Audio signal

Hiding methods	Steganalysis Method	F-measure	ROC
S-Tools	2D-Mel	0.706	0.725
	FMFCCs	0.885	0.745
	EE-AS	0.909	0.91
Steghide	2D-Mel	0.63	0.67
	FMFCCs	0.712	0.79
	EE-AS	0.791	795
Hide4PGP	2D-Mel	0.854	0.855
	FMFCCs	0.887	0.855
	EE-AS	1	1

The results registered in Table 7 show that EE-AS has better accuracy reflected by higher F-measure and roc. In addition, testing 2D-Mel steganalysis using only FMFCCs coefficients as features has resulted in a significant downsizing in the classification feature set (29 features instead of 58) and hence has improved the performance of 2D-Mel. These results could be explained as follows:

- The information extracted from FMFCCs are sufficient and more informative for the classification process.
- The 29 MFCCs features have added more noise than valuable information to the classification.

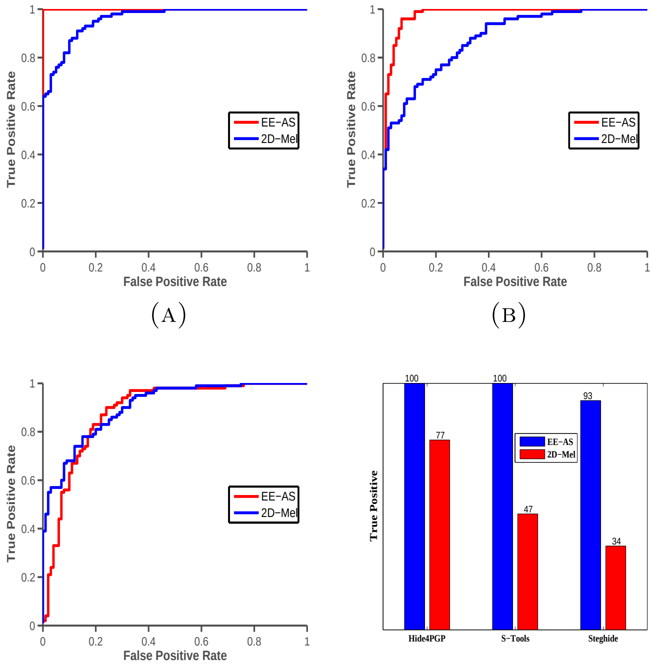
It is further found that EE-AS has better time complexity than 2D-Mel. The entropy features of N samples takes a maximum time complexity of $O(N \log N)$ [48], meanwhile the computation of mel cepstrum as indicated in Eq.4 requires $O(N^2(\log N)^2)$ knowing that FFT time complexity is $O(N \log N)$ [49].

$$MFCCs = FFT(MT(FFT(D_s^2))) = \begin{pmatrix} f_{mel1} \\ f_{mel2} \\ \cdots \\ f_{mel29} \end{pmatrix} \tag{4}$$

In addition to the complexity time, the computational time needed in the training stage in EE-AS is smaller, since, we only used 15 features compared to the 58 features in 2D-Mel.

Further details on the behavior of each algorithm are represented in term of ROC curves in Figures (5a), (5b) and (5c). In each graph, a higher curve corresponds to a more accurate detection rate. While a lower curve corresponds to a less accurate detection rate. As an example, combined 2D-Mel produced a ROC of 85.5% for Hide4PGP while EE-AS have achieved a 100% accuracy. For Stools, 91% has been registered against 72.5%. Moreover, the proposed method offers a better accuracy with regards to stego-audio streams detection expressed by higher rate of true positive (TP) as shown in Figure (5d).

Using any of these performance measures, EE-AS method performs better than 2d-Mel. Thus, the features extracted by EE-AS are more informative for the classification process.



(a) (b) (c) (d)

Figure 5. ROC curves for EE-AS and 2D-Mel [20] tested by Hide4PGP (5a), Stool (5b), Steghide (5c) and performance comparison of EE-AS and 2D-Mel based on TP rates (5d)

Beside the overall accuracy out-performance, EE-AS has two more advantages over 2D-Mel. Firstly, the 2D-Mel method is computationally more intensive as the classification vector has 58 elements against 15 only required by our algorithm. Secondly, the proposed method offers a better accuracy with regards to stego audio files detection which is the main objective of steganalysis techniques. As an example, Figure 5d shows that stego files in our method are 100% detected when Hide4PGP and Stool are used versus 77 % and 47% respectively in 2D-Mel. For steghide, 93% accuracy has been registered against 34%.

8. **Conclusion and Future work.** In this paper we present an efficient audio steganalysis technique. Our method is based on the assumption that embedding techniques would modify the entropy and the energy content of audio signals. This assumption is justified by the additive property of stego-signal which can be modeled by adding noise or error-signal to the cover-signal. However, the main feature of this algorithm is to demonstrate that maximum entropy in conjunction with energy can be very powerful technique for audio steganalysis. The experimental results show that our method (EE-AS) applied on a large audio signal dataset achieved a true positive rate above 97.67% for embedding rate of 25% or above while improving the computational time. EE-AS requires reduced feature vector to detect changes in audio signal due to steganography with higher F-measure and ROC. The improved accuracy of our method follows from the integration of energy and entropy which are two powerful information content measures. In addition, the use of SVM classifier empowered the proposed algorithm, since it is based on strong foundations in statistical learning theory and proved to be effective classifier in the last decade.

Finally, the success of the proposed steganalysis method in detecting steganographic audio signals encouraged us to pursue future investigations to further minimize the features vector and the embedding rate, introduce signal complexity and extend our proposed method to other steganographic applications developed in transform and coded domains.

REFERENCES

1. A. Cheddad, J. Condell, K. Curran and P. Mc Kevit, Digital image steganography: Survey and analysis of current methods , *Signal Processing*, vol 90, issue 3, pp 727-752, Marsh 2010.
2. F. Djebbar, B Ayad, K Abed-Meraim, H Habib, Unified phase and magnitude speech spectra data hiding algorithm , *Security and Communication Networks, John Wiley and Sons, Ltd*, vol 6, issue 8, pp 961-971, 1 Aug, 2013.
3. F. Djebbar, K Abed-Maraim, D Guerchi, H Hamam, Dynamic energy based text-in-speech spectrum hiding using speech masking properties , *2nd International Conference on Industrial Mechatronics and Automation (ICIMA),* vol.2, pp 422426, China, 30-31 May 2010.
4. Driss Guerchi and Fatiha Djebbar, Narrowband Speech Hiding using Vector Quantization , *International Journal of Information and Communication Engineering*, pp 5-8, 2009.
5. R. Balaji,G. Naveen, Secure data transmission using video Steganography , *IEEE International Conference on Electro/Information Technology (EIT),* pp.1,5, 15-17 May 2011.
6. Shirali-Shahreza, M.; Shirali-Shahreza, S., Persian/Arabic Unicode Text Steganography , *SIAS Fourth International Conference on Information Assurance and Security,* pp.62,66, 8-10 Sept. 2008.
7. T. G. Handel, T. Maxwell Sandford II, Hiding Data in the OSI Network Model , Information hiding: first international workshop, *Cambridge, UK. Lecture Notes in Computer Science*, vol. 1174, pp. 23-38, Berlin Heidelberg New York: Springer- Verlag, 1996.
8. F. Djebbar, B. Ayad, K. Abed Meraim, H. Hamam, Comparative study of digital audio steganography techniques , *EURASIP Journal on Audio, Speech, and Music Processing,* pp 1-16, Dec 2012
9. Steganography in the news, `http://www.sarc-wv.com/news/steganography\_in\_the\_news/2010/`.
10. Federal news radio, `http://www.federalnewsradio.com/?nid=\&sid=1909760/`
11. K. M. Johnson, S. Lyu, and H. Farid, Steganalysis of Recorded Speech , *In Proceedings of the Conference on Security, Steganography and Watermarking of Multimedia (SPIE), San Jose, USA*, pp. 664-672, January, 2005.
12. H. Ozer, I.Avcibag , B. Sankur, et al., Steganalysis of Audio Based on Audio Quality Metrics , *Proceedings of SPIE, Santa Clara, CA, USA*, pp. 55-66, June 2003.
13. A.C. Rencher, Methods of Multivariate Data Analysis , *2nd Edition. John Wiley*, 2002.
14. P. Pudil, J. Novovicova, J. Kittler, Floating Search Methods in Feature Selection, *Pattern Recognition Letters*, pp. 1119-1125, November 1994.
15. Avcibas, Audio steganalysis with content independent distortion measures , *IEEE Signal Process Letter*, vol. 13, no. 2, pp. 92-95, 2006.
16. Yincheng Qi, Jianwen Fu, and Jinsha Yuan, Wavelet domain audio steganalysis based on statistical moments of histogram , *Journal of System Simulation*, vol 20, no. 7, pp. 1912-1914, April 2008.
17. G. Xuan, Y. Q. Shi, J. Gao, et al, Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions , *In Proceeding of Information Hiding Workshop*, pp. 262-277, 2005.
18. X. Ru, H. Zhang, X. Huang, Steganalysis of Audio: Attacking the Steghide , *In Proceeding of the Fourth International Conference on Machine Learning and Cybernetics*, pp. 3937-3942, 2005.
19. C. Kraetzer, J. Dittmann, Pros and cons of Melcepstrum based audio steganalysis using SVM classification , *Lecture Notes in Computer Science, 4567,* pp. 359377, 2008.
20. Q. Liu, A. H. Sung, M. Qiao, Temporal derivative-based spectrum and mel-cepstrum audio steganalysis , *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 359-368, 2009.
21. W. Zeng, R. Hu, H. Ai, Audio steganalysis of spread spectrum information hiding based on statistical moment and distance metric , *Multimedia Tools and Applications*, vol. 55, no. 3, pp. 525-556, December 2011
22. F. A. P. Petitcolas, MP3Stego, [Online]. Available: `http://www.cl.cam.ac.uk/fapp2/steganography/mp3stego/index.html`(2002).
23. D. Yan, R. Wang, X. Yu and J. Zhu , Steganalysis for MP3Stego using differential statistics of quantization step , *Digit. Signal Process.*, vol. 23, no. 4, pp. 1181-1185, 2013

24. X. Yu, R. Wang and D. Yan , Detecting MP3Stego using calibrated side information features , *J. Softw.,* vol. 8, no. 10, pp. 2628-2636, 2013 (Feb. 2013). UnderMP3Cover. [Online]. Available: http://sourceforge. net/projects/ump3c/
25. Y. Z Ren, T. T. Cai, M. Tang, L. Wang, AMR Steganalysis Based on the Probability of Same Pulse Position , *IEEE Transactions on Information Forensics and Security*, vol.10, no.9, pp. 1801-1811, Sept. 2015
26. Steghide, `http://steghide.sourceforge.net/`
27. Stools Version 4.0, `http://info.umuc.edu/its/online_lab/ifsm459/s-tools4/`
28. Hide4PGP, `http://www.heinz-repp.onlinehome.de/Hide4PGP.htm`
29. H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, Spectral entropy based feature for robust asr , in Proc. ICASSP, May 2004, pp. 193196.
30. P. Renevey and A. Drygajlo, Entropy based voice activity detection in very noisy conditions , *in Proc.Eurospeech, USA*, Sept. 2001, pp. 18871890.
31. HARMSEN, Steganalysis of additive noise modelable information hiding. *Masters thesis, Rensselaer Polytechnic Institute*, Troy, NY. J. J. 2003
32. C.E. Shannon, A mathematical theory of communication , *Bell System Technical Journal*, vol. 27, pp. 379423, 623656, July, Oct. 1948.
33. Harremoes, P., Binomial and Poisson distributions as maximum entropy distributions, *IEEE Transactions on Information Theory*, vol.47, no.5, pp. 2039,2041, Jul 2001
34. T. Tao. Sumset and inverse sumset theory for Shannon entropy, *Combinatorics, Probability and Computing*, 19:603639, 2010.
35. A. Lempel and J. Ziv, On the complexity of finite sequences , *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 7581, 1976.
36. D.A. Huffman, A Method for the Construction of Minimum Redundancy Codes , *Proceedings of the I.R.E., September 1952*, pp 10981102.
37. Golomb, S.W.: Run-Length Encodings, *IEEE Transactions on Information Theory,* IT-12, pp. 399-401, July 1966.
38. ITU-T Recommendation P56, Telephone Transmission Quality: Objective Measuring Apparatus , March 1996.
39. E. Paajanen, B. Ayad, VV. Mattila, New objective measures for characterisation of noise suppression algorithms , *IEEE Workshop on Speech Coding, Proceedings* pp. 23-25, 2000.
40. E. Paajanen, VV. Mattila, Improved objective measures for characterization of noise suppression algorithms , *IEEE Workshop on Speech Coding, Proceedings.* pp. 77-79, 2002.
41. `http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html`
42. Cristianini N, Shawe-Taylor J, An introduction to Support Vector Machines Cambridge University Press , 2000.
43. N. Zaki, S. Wolfsheimer, G. Nuel and S. Khuri, Conotoxin Protein Classification Using Free Scores of Words and Support Vector Machines , *BMC Bioinformatics 2011*, vol. 12, pp. 217-227, (2011)
44. C. A. Waring and X. Liu, Face detection using spectral histograms and SVMs , *IEEE Transactions on Systems Man and Cybernetics*, vol. 35, no. 3, pp. 467-476, June 2005.
45. X. Y. Luo, D. S. Wang, P. Wang, F. L. Liu, A review on blind detection for image steganography , *Signal Processing*, vol. 88, no. 9, pp. 2138-2157, September 2008.
46. V. Vapnik, Statistical Learning Theory , *Hoboken, NJ: Wiley*, 1998.
47. `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
48. S. K. Shukla, M. V. Prasad, Lossy Image Compression, Domain Decomposition-Based Algorithms , *Springer Science and Business Media*, ISBN 1447122186, 9781447122180, 2011.
49. Beaudoin, N. and S. S. Beauchemin, A new numerical Fourier transform in d-dimensions, *IEEE Transactions on Signal Processing*, Vol. 51, No. 5, 1422-1430, 2003.