A Detection Method of Subliminal Channel based on VoIP Communication

Huaizhou Tao, Yongfeng Huang
Tsinghua National Laboratory
of Information Science and Technology
Department of Electronic Engineering
Tsinghua University
Beijing 100084, China
taohuaizhou@gmail.com;
yfhuang@mail.tsinghua.edu.cn

Donghong Sun, Ping Hu
Network Research Center of Tsinghua University
Tsinghua University
Beijing 100084, China
sundh@cernet.edu.cn;
huping@tsinghua.edu.cn

ABSTRACT
With VoIP (Voice over IP) is widely applied in the Internet, Using the subliminal channel based on VoIP to transmit secret message has been a significant issue in Information Security. At present, many kinds of information hiding algorithms have been designed and applied in real-time VoIP communication, and the most widely used algorithm among them is the LSB (Least Significant Bit) hiding. However, it is difficult to detect the LSB matching steganography in real time. In response to this challenge, this paper raised a structure of system, which is able to capture the VoIP streams on the Internet and detect the real-time subliminal channel in them. We also designed a new steganalysis algorithm for the system which is aimed at the LSB matching steganography algorithm. It used the feature extraction method based on correlation and SVM (Support Vector Machine) classification. The experiment results demonstrated that while the embedding rate was over 50%, the new algorithm raised the accuracy of detecting the LSB matching by about 20% compared with existing algorithms. At the same time, the length of the samples cannot obviously affect the accuracy of the new algorithm; it can detect the information hiding in the samples whose length is only 3s. From this point of view, the new algorithm has good real-time performance.

Keywords
VoIP; Steganalysis; LSB matching

1. INTRODUCTION
The information hiding technology comes from the ancient steganography technology. The modern steganography is applied to embed the secret messages into normal carriers, so the secret messages will not be realized by others during the transfer process [1]. At the moment, the information hiding technology is widely used in covert communication, authentication, copyright protection and data integrity verification. The carriers used in information hiding almost covered all kinds of data type in the Internet.

IWIHC'14, June 3, 2014, Kyoto, Japan.

http://dx.doi.org/10.1145/2598908.2598910

In recent years, as the wide use of VoIP (Voice over IP) in the Internet, the steganography which uses audio signal as carrier developed rapidly, and nearly reached the stage of practical use [2]. The main advantages of VoIP steganography are large volume data, high real-time performance, and that the data produced in the communication process will not be saved [3]. These characteristics make the detection of VoIP steganography very difficult. The attacker must design a system to monitor and capture massive amounts of VoIP voice data passing the network nodes, and then detect if they are carrying secret messages in a short time. With more steganography algorithms are deployed in audio signal, the design of such a system becomes even more difficult.

The common coding method of audio data is PCM (Pulse Code Modulation), it applies the sampling and quantization to the analog audio signal and turn it into digit signal. To this kind of time domain coding method, the LSB (Least Significant Bit) substitution is very simple and effective [4]. It can convert the secret messages into the binary bit stream, and then replace the LSBs of the carrier's data. Because of the masking effect, it is difficult for human ears to distinguish the modified carrier and the original carrier. At the same time, this algorithm has the advantages of high capacity and low complexity. It has been widely applied in audio steganography.

However, the disadvantages of the LSB substitution have been revealed in recent years. The low complexity of it makes that the modification of the carrier data has some rules. Take the 8-bits PCM coding as example, and divide the carrier data points into (0, 1), (2, 3)… (254, 255) these 128 groups according to their values, we can find that if a data point is belong to the i-th group, it will not change its group after the LSB substitution. Generally, the secret message is converted into binary bit stream by encryption and coding, so we can treat the bit stream as an independent distribution sequence in which 0 and 1 have the same probability. After the 100% embedding of LSB substitution, the frequency of the same group's two values will be equal. Even though the embedding rate is reduced, the change of the data value's distribution will also reveal the steganography. In image information hiding, there has been some effective algorithms are proposed to detect the LSB substitution, for example, the RS (Regular-Singular Steganalysis) algorithm [5] and the SPA (Sample Pair Analysis) algorithm [6]. These algorithms also have high accuracy in detecting audio information hiding. On the other hand, there are new algorithms proposed based on feature extraction and classification to detect the LSB substitution [7] [8]. The safety of the LSB substitution has dropped dramatically.

The LSB substitution has been modified to improve the covertness. While embedding, if the LSB of the carrier data point is equal to the bit to be embedded, the value of data point will not be changed; if they are different, the value of date point will be increased or decreased by 1 at random. This modified algorithm is

called as the LSB matching steganography. The accuracy of existing detecting algorithms to the LSB matching is not enough high in practical use.

At present, the research of detecting LSB matching in audio information hiding has not much progress. But some new detecting algorithms are designed in image information hiding [9] [10]. As LSB matching makes the rule of the data value's change more complex, it is difficult to use some simple statistical data to deduce the embedding rate. These algorithms use the idea of feature extraction and classification. They group the feature vectors through extracting the correlation of data points in carrier, and classify the feature vectors to detect the steganography. This kind of algorithms cannot estimate the embedding rate, but raise the accuracy of detecting LSB matching.

This paper raised a detection system. This system is used to detect the subliminal channel in the VoIP communications passing the monitored network nodes. We also discussed the characteristics of audio LSB matching steganography and designed a new detecting algorithm based on feature extraction and classification. The new algorithm increased the accuracy of detecting the audio LSB matching effectively. It can be deployed in the system mentioned above because its high real-time performance.

The rest of paper is organized as following: The second chapter describes the principle of detection; the third chapter proposes the algorithm which is used to extract the feature vectors; the fourth chapter is about the experiment and the performance of the new algorithm; finally, the fifth chapter gives the conclusion and the future work.

2. PRINCIPLE OF DETECTION

In this paper, we focus on discussing the design and realization of the feature extraction and classification in this system.

The issue of steganalysis can be abstracted to a pattern classification problem [11]. It is assumed that there is a data set (*W*, *X*), in which *X* represents the time-domain coded audio data stream; *W* represents the type of data. *W* can only choose 0 or 1, 0 shows that this audio data stream is not embedded by secret message; conversely 1 shows that this data stream is embedded by secret message. From this perspective, a steganalysis algorithm can be regarded as a mapping *d* from *X* to *W*. It can estimate the type *k* of a new audio data stream *x*, its error rate can be calculated as:

$$P_E = P(d(X) \neq W) \tag{1}$$

The best detecting method has the minimum error rate; it can be defined as follow:

$$\begin{aligned} &P_E = \min_d P(d(X) \neq W) \Leftrightarrow \\ &d(x) = \arg\max_{k=0,1} P(W = k \mid X = x) \end{aligned} \tag{2}$$

However, it will be very complex in practical use that using formula (2) to get the mapping *d*. The sample *X* is not only influenced by *W*, but also influenced by other factors like noise, voice recording equipment and environment. To remove the additional influence before classification, *X* should be taken some preprocessing.

Other kind of influence becomes more serious in steganalysis. Because carrier *X* is made to carry innocent information and its main content is constituted by this innocent information. The secret message is only a little part of *X* and has a little effect to *X*. Such as in the audio data, the main content of *X* is filled by human voice, music or other sound, that makes it almost impossible to use formula (2) to judge the type W of the carrier *X*.

As it is difficult to estimate the type of carrier directly, to improve the practicality of the algorithm, we can divide the classification process into two steps. Firstly, extract the feature vector *C* which has higher relevancy with *W* from sample *X*, then judge the type *W* through the feature vector *C*, as follows:

$$C = g(X = x) \tag{3}$$

$$W = f(C = c) \tag{4}$$

g is a fixed function or algorithm and *f* is a kind of classifier. Training the classifier *f* with the samples whose type is known can make it predict the type of unknown samples, finally its results will approach the minimum error rate P_E. The performance of system is mainly decided by the correlation of the type *W* and the feature extracted by *g*. The following discussion is mainly around the feature extraction.

3. FEATURE EXTRACTION

Since the LSB matching algorithm mainly modifies the least significant bits of the carrier, we consider the correlation between the LSB and the second LSB as a kind of feature. $M_1(1:m)$ denotes the LSB of the carrier, and $M_2(1:m)$ denotes the second LSB. Here *m* is the length of the carrier audio sample, and *E* is the mathematical expectation. The covariance function is defined as

$$Cov(x_1, x_2) = E[(x_1 - u_1)(x_2 - u_2)] \tag{5}$$

Where $u_i = E(x_i)$.

*C*1 is defined as follows:

$$C1 = Cor(M_1, M_2) = Cov(M_1, M_2) / (\sigma_{M_1} \sigma_{M_2}) \tag{6}$$

Where σ_x is the standard deviation of the vector *x*.

In addition, we can extract the autocorrelation of the LSB vector. At First, define vector $X_{-k} = M_1(1:m-k)$, $X_k = M_1(k+1:m)$. Based on this, we can define the autocorrelation as

$$C(k) = Cor(X_{-k}, X_k) \tag{7}$$

Setting *k* to different values, we can get new features like $C2 = C(1)$, $C3 = C(2)$, $C4 = C(3)$, $C5 = C(4)$.

LSB matching also changes the value of the data points. The variable p_k denotes the histogram probability density of coverage at the intensity, k = 0, 1,…, N-1, for 8-bit PCM coding, N=256. The variable p_k' denotes the histogram probability density of adulterated audio at the intensity. Assuming the embedded message

is independent and identically distributed, and embed rate is r, $p_k^{'}$ is given as follows:

$$p_k^{'} = (1-\frac{r}{2}) * p_k + \frac{r}{4} * p_{k-1} + \frac{r}{2} * p_{k+1} \quad (8)$$

It is very difficult to accurately judge whether the carriers are embedded with secret messages or not, and to predict the embed rate r without the original carrier. However, LSB matching definitely changes the distribution density of the histogram. We can define correlation features from this. Firstly, define the histogram probability density vector H as:

$$H = (p_0, p_1, p_2, ..., p_{N-1}) \quad (9)$$

$$H_e = (p_0, p_2, p_4, ..., p_{N-2}) \quad (10)$$

$$H_o = (p_1, p_3, p_5, ..., p_{N-1}) \quad (11)$$

$$H_{-l} = (p_0, p_1, p_2, ..., p_{N-1-l}) \quad (12)$$

$$H_l = (p_l, p_{l+1}, p_{l+2}, ..., p_{N-1}) \quad (13)$$

The autocorrelation coefficients can be defined as:

$$C_H(l) = Cor(H_{-l}, H_l) \quad (14)$$

The new feature $C6$ is defined as:

$$C6 = Cor(H_e, H_o) \quad (15)$$

Set l = 1, 2, 3 and 4, other features are $C7 = C_H(1)$, $C8 = C_H(2)$, $C9 = C_H(3)$, $C10 = C_H(4)$.

The 10 correlation features above constitute the feature vector C 。

This algorithm realizes above methods to extract a feature vector from every carrier sample, and then input it into the classifier.

4. EXPERIMENTS AND RESULTS

4.1 Experiments System and Data Set

This paper designed a system structure which can detect the steganography based on VoIP stream-media in real time and real networks, which is showed in Figure.1. The system can capture and save the data packages of passing VoIP stream at any time, it can also reconstruct and analysis the saved data. This detecting system is constituted by 4 modules. First of all is the capture of VoIP stream, it monitors all passing data packages and captures the parts which belong to VoIP communication. The next module is the reconstruction and store of the stream; it separates the captured packages according to different VoIP streams and restores the original VoIP stream from them. This module is followed by the feature extraction module; the feature extraction module runs the algorithm to extract feature vectors from audio data. The last is the classification module, it uses the trained model to classify the extracted feature vectors; its judging results represent that if the VoIP stream is embedded by secret messages.

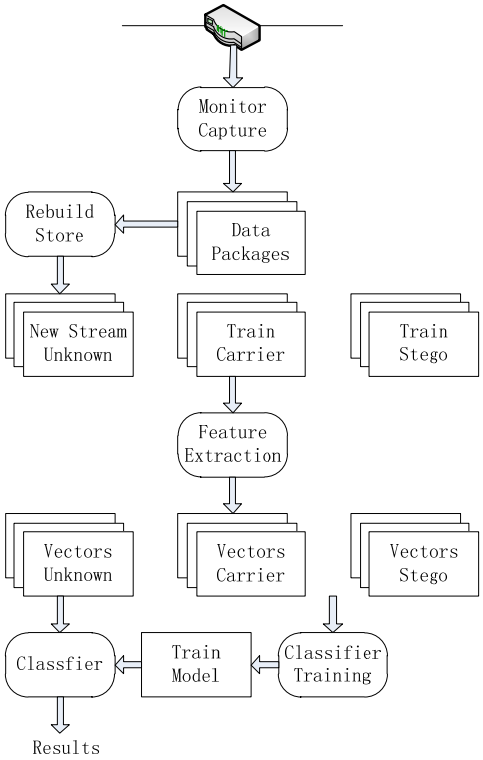

Figure 1. The structure of detecting system

The classifier applies the SVM[12] (Support Vector Model) algorithm to classify the vectors. We use the LIBSVM [13] as classifier to train and detect. The algorithm of feature extraction can divide audio samples into 10s long sections, and then extract the feature vectors from the sections (the length of the sections can be adjusted). If the feature extraction module already knows the type of the carrier sample, it can send the type and feature vector to classifier at the same time. In the process of training, the classifier can use the feature vectors with known type to train the model; and in the process of prediction, the classifier can compare the results with the known types to calculate the accuracy rate.

To prove the accuracy and adaptability of the detect algorithm, the data set for the experiment need to be large enough, and it also should have high diversity. In this paper, we choose the G.711.μ coding algorithm [14]; the sampling rate is 8 kHz. G.711.μ is a common coding format in VoIP communications; it has good performance on audio clarity and bandwidth occupation. The data set used in experiment can be divided into two groups according to the language, one is Chinese and the other is English, each group has 800 samples. In the experiment, we use 3/4 of the data set to train the classifier, and the other 1/4 to predict. The corresponding stego data set is generated by embedding 0 or 1 at random into original data set using LSB matching steganography, the embed rate can be modified. The stego data set is also divided into the group of training and the group of prediction with the same proportion, so we get all data sets used in the experiment.

4.2 Analysis of Detecting Accuracy

We use the RS algorithm and SPA algorithm to compare with our detecting algorithm. As the output of these two algorithms both are the embed rate of the secret message, we need apply a binary classification on the output of these two algorithms. Through calculating the output from the training data set, we can find a threshold, when the output rate is higher than it, we take it as a positive output, and otherwise we take it as a negative output. To get an appropriate threshold which makes the accuracy of the algorithm highest, its ROC (Receiver Operating Characteristic) curve should be drawn and studied. After find the best threshold of RS and SPA algorithms, the detecting accuracy of these 3 algorithms can be compared. The ROC curve of these two algorithms are showed in Figure 2.

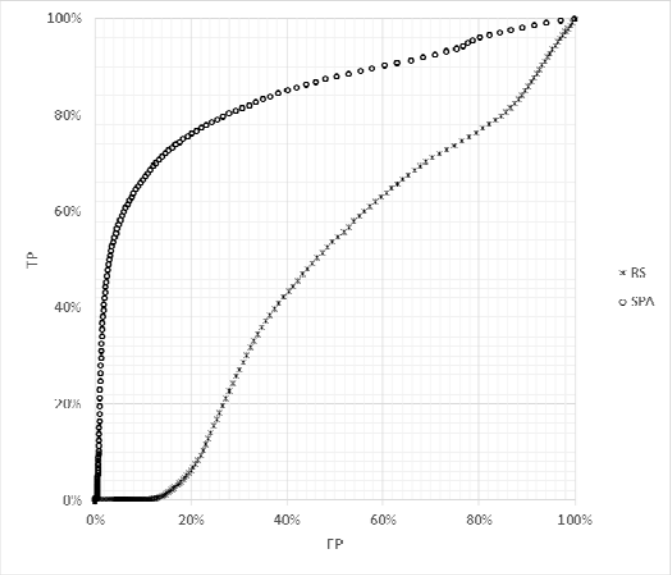

Figure 2. ROC curve of RS and SPA algorithms

In this experiment, we use seven different embed rate level, they are 100%, 90%, 80%, 70%, 60%, 50% , 40% , 30% , 20% and 10%. We calculated the accuracy of different algorithms under different rates. The results are shown as Table 1:

TABLE 1. DETECTING ACCURACY OF 3 ALGORITHMS

Algorithms	Embed rate(%)									
	100	***90***	***80***	***70***	***60***	***50***	***40***	***30***	***20***	***10***
Feature Extraction	98.6	98.6	98.6	98.6	98.6	98.4	94.9	78.3	67.8	55.0
RS	59.1	56.4	59.7	62.4	64.6	66.4	67.6	67.9	67.1	54.3
SPA	78.6	79.8	80.8	81.6	81.8	81.5	79.9	76.4	67.6	60.1

The results show that the algorithm designed in this paper can accurately detect the LSB matching steganography while the embed rate is higher than 30%. But if the embed rate decreases below 30%, the performance of this algorithm also decreases obviously. Compare with other algorithms, this algorithm has a 20% higher accuracy than SPA algorithm under high embed rate, at the same time, RS algorithm can hardly detect LSB matching steganography. The performance curve of these algorithms is showed in Figure 3 below:

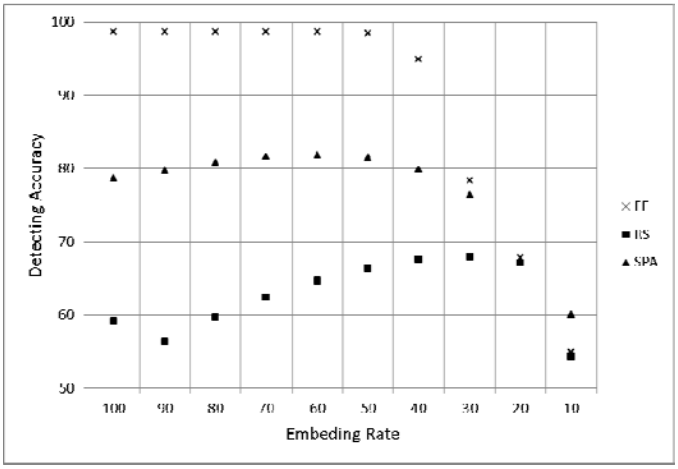

Figure 3. Detecting Accuracy of different algorithms

4.3 Analysis of Real-time Performance

VoIP communication has the characteristics of high volume and high real-time ability, and the audio data during the conversation will not be saved, that makes it very appropriate to apply steganography. On the other hand, when real-time detect the information hiding in VoIP communication; the detecting algorithm must have high response speed. It means that the algorithm can use very short samples to get reliable detect results.

So we change the length of the audio samples into 8s, 6s and 3s to constitute new data sets. Then we run and analysis the algorithm on new data sets to investigate the response speed of the algorithm. The results are shown in Table 2 below:

TABLE 2. DETECTING ACCURACY OF SAMPLES IN DIFFERENT LENGTH

Length (s)	Embed rate(%)									
	100	***90***	***80***	***70***	***60***	***50***	***40***	***30***	***20***	***10***
10	98.6	98.6	98.6	98.6	98.6	98.4	94.9	78.3	67.8	55.0
8	98.6	98.6	98.6	98.6	98.5	98.3	94.3	73.1	61.8	52.1
6	98.5	98.5	98.5	98.5	98.5	98.2	93.7	69.1	56.9	52.0
3	98.5	98.5	98.5	98.5	98.4	98.0	93.0	64.2	54.7	51.3

Table 2 shows that the length of samples has little effect on the performance of the feature extraction algorithm. Even the length of the audio samples decreases to 3s; the algorithm can also get a satisfactory result. It is proved that the algorithm had high response speed and good real-time performance.

5. CONCLUSIONS

This paper designed a new detection system. The function of this system is to detect the real-time subliminal channel based on VoIP communications on the Internet. At the same time, we proposed a new detecting algorithm to the LSB matching steganography in time-domain coding audio signals for this system. The new algorithm used feature extraction and classification to realize the detection of LSB matching. Comparing with existing algorithms, the new algorithm has higher accuracy. This algorithm also has high real-time performance; it can hold the accuracy when the length of audio samples reduces. These two advantages make the algorithm has high practical value. The method combined with

feature extraction and classification in this paper provided a new kind of thought to detect the LSB steganography in audio signals.

The future work first will be the research of the effects to the accuracy by choosing different features, and we will also discuss the situation in low embed rate to improve the performance of the algorithm. On the other hand, there is a point to discuss that if some characteristics of carrier will influence the accuracy of detection, like the language or SNR (Signal to Noise Ratio). It may help us to find the carrier which is more appropriate to embed secret message.

6. ACKNOWLEDGEMENTS

This work was supported in part by National High-tech R&D Program of China (863) [2011AA010704], China National Natural Science Foundation [No. 61271392], and Asia 3 Foresight Program of National Natural Science Foundation of China [No.61161140454].

REFERENCES

[1] Stephan Katzenbeisser; Fabien Petitolas, “Information Hiding Techniques for Steganography and Digital Watermaking”, EDPACS: The EDP Audit, Control, and Security Newsletter, Vol. 28, Iss. 6, pp. 1-2, Dec 2000.

[2] Yongfeng Huang; Jian Yuan; Shanyu Tang; C. Wang, “Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec”, IEEE Transactions on Information Forensics and Security, Vol. 6, No. 2, pp. 296-306, June 2011.

[3] Yongfeng Huang; Chenghao Liu; Shanyu Tang; Sen Bai, “Steganography Integration into Low-bit Rate Speech Codec”, IEEE Transactions on Information Forensics and Security, Vol. 7, No. 6, pp. 1865-1875, Dec 2012.

[4] Kratzer, C.; Dittmann, J.; Vogel, T.; Hillert, R., “Design and evaluation of steganography for voice-over-IP” Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on ,pp. 21-24, May 2006.

[5] Fridrich, J.; Goljan, M.; Rui Du, "Detecting LSB steganography in color, and gray-scale images," MultiMedia, IEEE, Vol.8, No.4, pp.22-28, Oct-Dec 2001.

[6] Sorina Dumitrescu; Xiaolin Wu; and Zhe Wang, “Detection of LSB Steganography via Sample Pair Analysis”, Information Hiding Lecture Notes in Computer Science, Vol. 2578, pp. 355-372, 2003.

[7] Micah K. Johnson; Siwei Lyu; Hany Farid, “Steganalysis of recorded speech”. Proc. SPIE 5681, Security, Steganography, and Watermarking of Multimedia Contents VII, 664, March, 2005.

[8] Huang, Y.F.; Tang, S.; Zhang, Y., “Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis,” Communications, IET , Vol.5, No.7, pp.929-936, May 2011.

[9] Qingzhong Liu; Andrew H. Sung; Bernardete Ribeiro; Mingzhen Wei; Zhongxue Chen; Jianyun Xu, “Image complexity and feature mining for steganalysis of least significant bit matching steganography”, Information Sciences, Vol. 178, Iss. 1, Pages 21-36, 2 January 2008.

[10] Pevny, T.; Bas, P.; Fridrich, J., “Steganalysis by Subtractive Pixel Adjacency Matrix”, Information Forensics and Security, IEEE Transactions on , Vol.5, No.2, pp.215-224, June 2010.

[11] Yongfeng Huang; Shanyu Tang; Chunlai Bao; Yau Jim Yip, “Steganalysis of compressed speech to detect covert voice over Internet protocol channels” , IEE/IEEE Journal, IET Information Security, Vol. 5, No. 1, pp. 26-32, March 2011.

[12] Cortes Corinna; Vapnik Vladimir N, “Support-Vector Networks”, Machine Learning, Vol. 20, pp. 273-297, Sept 1995.

[13] Chih-Chung Chang; Chih-Jen Lin, “LIBSVM: A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, Vol. 2, Iss. 3, pp.1-27, May 2011.

[14] ITU-T, “Pulse code modulation (PCM) of voice frequencies”, in ITU-T G.711, 1988.