Detection of covert Voice over Internet Protocol communications using sliding window-based steganalysis

Detection of Covert VoIP Communications using Sliding Window Based Steganalysis

Huang Yongfeng [1], Zhang Yuan [1], and Shanyu Tang [2]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084. yfhuang@tsinghua.edu.cn

[2] London Metropolitan University, London N7 8DB, UK. s.tang@londonmet.ac.uk

Abstract

In this paper we describe a reliable and accurate steganalysis method for detecting covert Voice over IP (VoIP) communication channels. The proposed method utilises a unique sliding window mechanism and an improved RS algorithm for VoIP steganalysis, which detects the presence of Least Significant Bit (LSB) embedded VoIP streams. With the mechanism, the detection window moves forward one packet or several packets each time to screen VoIP streams. The optimum detection threshold for the proposed detection metric is computed by modelling the distributions of the new metric for stego and cover VoIP streams. Experimental analysis reveals the proposed method improves the detection time significantly, utilizing less memory resources for VoIP steganalysis, thereby enabling real-time detection of stego VoIP streams. The proposed method also provides a significant improvement on precision in detecting multiple covert VoIP channels when compared to the conventional RS method.

Key words: covert communications, VoIP, steganalysis, sliding window.

1. INTRODUCTION

Modern steganography is the art of covert communication, which hides the existence of secret messages embedded in cover objects over the public channel. Steganography in static cover objects, such as plaintext, image files with BMP or JPEG format, and audio files in WAV or MP3 format, has been explored extensively [1-3]. See [4] for a good survey of such techniques. Recently network protocols and media streams like Voice over IP (VoIP) were used as cover objects to embed secret messages [5]. Dittmann et al, for example, reported the design and evaluation of steganography for VoIP [6], indicating possible threats as a result of embedding secret messages in such a widely used communication protocol.

Steganalysis is the science of detecting messages hidden using steganography. The goal of steganalysis is to distinguish stego objects (objects containing a secret message) from cover objects with little or no knowledge of steganographic algorithms. The simplest method to detect steganographically-encoded packages/files is to compare them to known originals. Comparing the package against the original file will yield the differences caused by encoding the payload – and, thus, the payload can be extracted. Nowadays, steganalysis becomes increasingly important in computer forensics, for tracking and screening documents/audios/videos that are suspect of criminal and terrorism activities, and for information security to prevent leakage of unauthorized data.

There have been some attempts to develop the steganalysis techniques for detecting steganography in static cover objects such as text, image or audio files. A number of image and audio steganalysis methods have been reported in the literature [7-12]. In contrast with image and audio steganalysis, steganalysis in media streams (like VoIP) is largely unexplored so far.

Several audio steganalysis approaches, such as the works of Ozer et al [13], Johnson et al [14], and Ru et al [15], shall be mentioned here as related work. These approaches can be placed in two groups. One is based on the distances computed between a real signal and a self-generated reference signal [15] via linear predictive coding (LPC), benefiting from the very nature of the continuous wave-based audio signals like G.711 PCM (Pulse-code Modulation). Ozer et al suggested a de-noising function for generating reference signals [13]. The other group relies on a statistical model for normal and 'abnormal' behaviour. For example, Johnson et al [14] presented a statistical model that consists of errors in representing audio spectrograms using a linear basis. The linear basis was constructed from a principal component analysis (PCA) and the classification was conducted using a non-linear SVM (support vector machine).

Voice over IP enables the digitalisation, compression and transmission of analogue audio signals from a sender to a receiver using IP packets. As the size of the used network and the distance between the communicating parties are of little relevance for transmission, VoIP is used for worldwide telephony such as Skype. With the upsurge of VoIP applications available for commercial use in recent years, VoIP is becoming one of the most interesting cover objects for steganography [5, 6, 16]. VoIP streams are dynamic chunks of a series of packets that consist of IP headers, UDP headers, RTP headers, and numbers of audio frames. Those headers and frames have a number of unused fields, providing plausible covert channels and thus giving scope for steganography.

Kraetzer and Dittmann's work [17] is relevant to our work. They suggested a steganalysis method of detecting VoIP steganography based on Mel-Cepstrum, which is an excellent feature vector for representing the human voice and musical signals in the field of speech and speaker detection. In their work a feature based statistical model for the behaviour of the channel over time was computed and compared by χ^2-testing against standard distributions. However, they assumed the detection probability of 52% was the lower boundary for discriminating features. So far there are still few steganalysis methods available for detecting VoIP steganography. This was the reason that led us to propose this work in the first place.

Unlike image and audio steganalysis, VoIP steganalysis generally starts with a pile of suspect data streams (which is often quite large), but little information about which of the streams, if any, contain a secret message. There would appear to be two modes that can be used to determine whether a covert VoIP communication occurs. One is to perform an off-line steganalysis using an extremely powerful computer after all packets have been captured and stored in a buffer; the other is to conduct real-time detection of covert VoIP streams. The former has an obvious disadvantage because it needs a large buffer to store all streams. In this study we focus on the latter mode, i.e. seeking ways to detect stego VoIP streams among all the streams in a rapid and real-time manner.

Fridrich’s RS algorithm is regarded as an effective steganalysis method to detect Least Significant Bit (LSB) replacement steganography in colour and grayscale images [7]. Dumitrescu et al [8] suggested another steganalysis algorithm for LSB replacement embedding, Sample Pair Analysis (SPA). Both image steganalysis algorithms seem to follow the same methodology. Dumitrescu also claimed the algorithms are applicable to images, but also to other cover objects in principle. However, the effectiveness of the conventional RS algorithms for VoIP steganalysis is unknown, and this study is to address this issue.

The rest of this paper is organised as follows. In Section 2 a sliding window based steganalysis method is proposed for detecting stego VoIP streams, including a steganalysis model for VoIP steganography, sliding window detection, VoIP steganalysis algorithms, and the optimum detection threshold. Section 3 presents experimental results for performance evaluation, such as detection accuracy, detection window lengths, computational complexity and detection time. Finally, Section 4 concludes with a summary and directions for future work.

2. SLIDING WINDOW BASED STEGANALYSIS

2.1 Steganalysis Model for VoIP Steganography

The ‘prisoner’ problem is a de facto model for a covert communication channel [18]. Prior work on steganography assumes Alice (sender) and Bob (receiver) share a private key or a public/private key pair and a public function; this function takes the key and stego-object as inputs and outputs the secret message. Alice sends Bob a transmitted object which may either be a cover-object or a stego-object, and the adversary (Wendy) is free to examine all transmitted objects between Alice and Bob and must decide whether such transmissions include a hidden message.

Figure 1 shows a steganalysis model for VoIP steganography on the Internet, in which Alice and Bob use VoIP software to communicate secretly, so the VoIP stream transmitted between them contains a secret message, and such a stego VoIP stream is denoted by s_i. Other users like John and Smith use the software for normal communications where VoIP streams do not contain any secret message, and they are denoted by s_1, s_2, … s_{i-1}, s_{i+1}, … s_j. As a warden, Wendy monitors the router connected with the network so as to detect the stego VoIP stream, s_i, from all streams, $S = \{s_1, s_2, \ldots s_i, \ldots s_{j,} \ldots\}$. Assuming the stego stream, s_i, is composed of N_i packets, then $s_i = \left(p_i^1, p_i^2, \ldots, p_i^{N_i}\right)$, where p_i^j is the jth RTP packet in the stream.

In the steganalysis model for VoIP steganography, Wendy captures all the streams passing through the router, identifies VoIP streams among all the streams, and then distinguishes stego VoIP streams from ‘innocent’ VoIP streams. As Wendy’s task is time-consuming, rapid and near ‘real-time’ detection of stego VoIP streams should be considered in order to make the detection process more practicable.

For simplicity, Wendy’s task is assumed to make a decision on two choices for each VoIP stream. One choice is the null hypothesis H_0, i.e. no steganography happens. The other choice is the alternative hypothesis H_1, i.e. steganography occurs. In general, VoIP streams contain no hidden messages, so the non-existence of steganography is regarded as the conservative hypothesis which requires evidence to reject.

2.2 Sliding Window Detection

To achieve real-time detection of VoIP steganography, it needs to rapidly select packets from VoIP streams for steganalysis. In a real-time mode, steganalysis may be conducted against only part of packets in a VoIP stream rather than the entire stream, as the processed packets in the stream are discarded when successive packets arrive. Based on this idea, we propose a steganalysis method using sliding windows to improve the detection performance of stego VoIP streams. The definitions below introduce the sliding window method for VoIP steganalysis.

Definition 1, the detection window, *W*, is a set of packets that are sampled for detecting the secret message hidden in a VoIP stream.

Definition 2, the detection window length, *L*, represents the number of packets that the detection window contains.

For a stream s$_i$, if the last received packet is p_i^k , the current detection window is denoted by

$$\begin{cases} \left(p_i^{k-L+1}, p_i^{k-L+2}, \ldots, p_i^{k-1}, p_i^k\right), k \geq L \\ \left(p_i^1, p_i^2, \ldots, p_i^{k-1}, p_i^k\right), \qquad k < L \end{cases}$$

When the next packet p_i^{k+1} arrives, the detection window slides forward one packet. In the perspective of temporal dimension, the detection window moves forward one packet each time, and the new detection window is denoted by

$$\begin{cases} \left(p_i^{k-L+2}, p_i^{k-L+3}, \ldots, p_i^{k}, p_i^{k+1}\right), k \geq L \\ \left(p_i^{1}, p_i^{2}, \ldots, p_i^{k}, p_i^{k+1}\right), \qquad k < L \end{cases}$$

Figure 2 illustrates the sliding process of the detection window.

With the sliding window method, only part of packets in a VoIP stream needs to be buffered while the stream is being screened, thus the required buffer/storage size can be minimised. Another advantage of the method is that the detection time can be reduced significantly to an extent that rapid detection of stego VoIP streams would become possible. So the sliding window method would meet the requirement of real-time detection of VoIP steganography.

Another possible detection method is based on stream segmentation, in which the detection window slides forward *L* packets rather than one packet once as described above. The basic idea of the stream segmentation method is to divide a VoIP stream into multiple segments with each segment containing *L* packets, and detect the presence of the secret message by screening each segment.

The next step is to determine the minimum lengths of sliding windows at different levels of data embedding. The minimum sliding window length can be modelled by

$$f(\widetilde{P}_L) = \min\left\{ E[(\widetilde{P}_L - P)^2] \right\} \qquad (1)$$

where *E* represents the mean error, *P* is the practical (real) embedding rate in case of continuously embedding secret messages, and $\widetilde{P}_L$ is the estimated embedding rate when the sliding window length is more than *L* (numbers of packets).

2.3 VoIP Steganalysis Algorithms

The conventional RS algorithm for image steganalysis defines regular pixels (R) and singular pixels (S) in a cover image using the discrimination function and the flipping operation [7]. Flipping is a permutation of gray levels that entirely consists of 2-cycles. The principle of the RS steganalysis method is to estimate the four curves of the RS diagram and calculate their intersection using extrapolation. The general shape of the four curves varies with the cover image from almost perfectly linear to curved.

In VoIP steganography, the unused header fields are used to embed secret messages. By adopting an approach similar to the conventional RS algorithm for image steganalysis, we define regular header fields (R) and singular header fields (S) in VoIP streams. According to the operational rules detailed in [7], RS algorithm defines eight parameters, i.e. $Rp1$, $Rn1$, $Sp1$, $Sn1$ and $Rp2$, $Rn2$, $Sp2$, $Sn2$, which are the intersecting points of the four curves (R+, R-, S+, S-) of the RS diagram. All the parameters are calculated to decide whether there is an embedded message, and to estimate the embedding rate. The parameters of the curves are then determined from the points marked in Figure 3. The x-axis is the percentage of header fields with flipped LSBs; the y-axis is the relative number of regular and singular groups (header fields).

Let z be the non-embedding point after coordinate transformation, and it can be calculated using the following equation [7]

$$2(dp1 + dp2)\,z^2 + (dn1 - dn2 - dp2 - 3dp1)\,z + dp1 - dn1 = 0 \qquad (2)$$

where $dp1 = Rp1$ - $Sp1$, $dn1 = Rn1$ - $Sn1$, $dp2 = Rp2 - Sp2$, and $dn2 = Rn2 - Sn2$. And the embedding rate θ_p is given by

$$\theta_p = \frac{z}{(z - 0.5)} \qquad (3)$$

Let th_1 be the optimum detection threshold for the conventional RS algorithm. If $\theta_P < th_1$, it means the cover object contains no secret messages; otherwise, there is a secret message embedded in the cover object.

For VoIP steganalysis, one of the major problems is how to reduce the detection time so as to achieve real-time detection of stego VoIP steams. So the conventional RS algorithm needs to be modified to reduce computational complexity, utilising less memory resources, and make the algorithm applicable to VoIP steganalysis in a real-time manner.

The main purpose of VoIP steganalysis is to determine whether VoIP streams contain secret messages. Analysis of the curves of the RS diagram (Figure 3) reveals stego VoIP streams were already different from the original ‘innocent’ VoIP streams after $Rp1$ and $Rn1$ had been calculated. This means only $Rp1$ and $Rn1$ should be used for detection purposes, and thus, we suggest the normalized difference between $Rp1$ and $Rn1$, a statistical parameter, to be the embedding rate (the secret message length divided by the stego VoIP stream length), which is given by

$$\theta_r = \frac{Rn1 - Rp1}{Rn1 + Rp1} \qquad (4)$$

Let V_T be the optimum detection threshold for the VoIP steganalysis method. If $\theta_r < V_T$, there are no hidden messages; otherwise, VoIP streams contain secret messages.

Comparing equations (3) and (4) indicates the improved RS algorithm has less computational complexity than the conventional RS algorithm. This would help improve the performance of detection of VoIP steganography.

2.4 Optimum Detection Threshold

The optimum detection threshold for the above VoIP steganalysis method (V_T) is computed by modelling the distributions of stego and cover VoIP streams. The proposed mathematical model below requires some empirical data.

Let C be the mean loss value due to false detection, C_1 the loss of false positive (false alarm), and C_2 the loss of false negative (detection failure). $P(s_1)$ denotes the probability of VoIP streams not containing secret messages, and $P(s_2)$ denotes the probability that secret messages are embedded in VoIP streams. P_1(y) and P_2(y) are the probability density functions (pdf) of VoIP streams not containing secret messages and those containing secret messages, respectively; they are likelihood functions. The mean loss C is then given by

$$C = C_1 \cdot P(s_1) \cdot P(e \mid s_1) + C_2 \cdot P(s_2) \cdot P(e \mid s_2) \qquad (5)$$

where $P(e \mid s_1) = \int_{-\infty}^{V_T} p_1(y)dy$, $P(e \mid s_2) = \int_{V_T}^{\infty} p_2(y)dy$, and they denote the probability of false positive (false alarm) and the probability of false negative (detection failure), respectively.

Assuming the mean loss is the minimum value, then $\frac{\partial C}{\partial V_T} = 0$. Equation (5) becomes

$$C_1 \cdot P(s_1) \cdot p_1(V_T) = C_2 \cdot P(s_2) \cdot p_2(V_T) \qquad (6)$$

According to the central limit theorem, the likelihood functions, P_1(y) and P_2(y), should conform to Gaussian distributions. Their mean values and mean square deviations can be computed by using samples' mean values and mean square deviations. Let the mean values of P_1(y) and P_2(y) be μ_1 and μ_2, and the mean square deviations of P_1(y) and P_2(y) be σ_1 and σ_2, respectively. The probability density functions are given by

$$p_1(y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(y-\mu_1)^2}{2\sigma_1^2}\right] \qquad (7)$$

$$p_2(y) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right] \qquad (8)$$

Substituting equations (7) and (8) into equation (6) obtains

$$\frac{(V_T-\mu_1)^2}{2\sigma_1^2} - \frac{(V_T-\mu_2)^2}{2\sigma_2^2} = \ln\frac{C_1 P(s_1)\sigma_1}{C_2 P(s_2)\sigma_2} \qquad (9)$$

Assuming $\sigma_1 \approx \sigma_2$, and using σ to represent σ_1 and σ_2, the optimum detection threshold, V_T, is then given by

$$V_T = \frac{\mu_1+\mu_2}{2} + \frac{\sigma^2}{\mu_2-\mu_1}\ln\frac{C_1 P(s_1)}{C_2 P(s_2)} \qquad (10)$$

3. RESULTS AND DISCUSSION

3.1 Experimental Setup

Figure 4 shows a framework for sliding window based VoIP steganalysis. The steganalysis software consists of three modules, an IP packet capturer, a VoIP stream identifier, and a sliding window detector.

The packet capturer was used to collect all IP packets passing through the router and match streams with packets. The VoIP stream identifier was employed to identify VoIP streams among all streams. The sliding window detector was utilised to detect the existence of the secret message embedded in the packets of VoIP streams. Three modules were implemented on the Server with an Intel Xeon 3065 CPU, 2GB memory, and a 1000MB Ethernet interface card.

In the experiments, ten users in our laboratory employed Stego-Talk software to communicate securely with their counterparts at other universities over China Education and Research Network (CERNET). The VoIP steganalysis software was connected to the router that forwards packets between a LAN in the laboratory and the CERNET. We used G.711 speech codec and LSB steganography algorithms to embed secret messages in PCM audio payload at various embedding rates ranging from 0% to 100% in 5 percent increments. Experimental results are discussed in detail below.

3.2 Comparisons between Detection Methods

Comparisons in detection performance between three detection methods, Sliding window, Segmentation, and Entire stream (only one segment), were conducted and the results are shown in Figure 5. The x-axis is the length of the secret message continually embedded in VoIP streams. The experiments indicate that all the methods could not detect the existence of the hidden message when the secret message length was less than 1 second. As the secret message length increased up to 2 seconds, the sliding window method detected the existence of the hidden message. The segmentation method could not detect the hidden message until the secret message length was more than 4 seconds. The entire stream method could not detect the hidden message at all.

Analysis of the experimental results reveals the sliding window method is more accurate than the segmentation method in case of fewer packets containing secret messages. For example, if the length of the VoIP stream used to embed a message was less than 2 seconds at a 100% embedding rate, the segmentation method would not be able to detect the existence of the hidden message. The non-effectiveness of the segmentation method is probably due to the fact that the VoIP stream containing a secret message needs to be divided into at least two segments, so steganalysis of each segment would be less effective because of less hidden information.

To compare detection accuracy, the sliding window method and the segmentation method were used, respectively, to detect VoIP steganography in which secret messages were embedded in VoIP streams at a 20 percent embedding rate. The detection window length was 300 packets, and the segmentation length was also 300 packets. The improved RS algorithm was adopted in the experiments, and the detection threshold was set to 0.04. Figure 6 shows comparisons in detection precision between the sliding window method (top figure) and the segmentation method (bottom figure).

The experiments indicate that only the sliding window method could correctly detect the existence of hidden messages under the experimental conditions. The segmentation method failed to detect stego

VoIP streams because it needs a larger embedding message length threshold when compared to the sliding window method. So the sliding window method was chosen for the remaining experiments.

3.3 Lengths of Sliding Windows

To obtain the empirical length of sliding windows, a series of experiments were conducted to determine the minimum lengths of sliding windows at various embedding rates ranging from 5% to 100% in 5 percent increments, and the results are shown in Figure 7. Clearly the minimum length of sliding windows capable of detecting the existence of hidden messages decreased gradually with increasing embedding rate.

The relationship between detection accuracy and the length of sliding windows was used to determine the minimum length of sliding windows with reasonable detection accuracy. Practical false alarm rates and true detection rates [19] were measured at various lengths of sliding windows at 20% embedding rate, and the results are shown in Figure 8. Detection accuracy improved with an increase in sliding window lengths at the same false alarm rate, but computational complexity also increased correspondingly. Analysis of Figure 8 reveals the length of sliding windows should be approximately 300 packets in order to achieve 95% true detection with 5% false alarm.

3.4 Optimum Detection Threshold

As the likelihood function, P_2(y), varies with the embedding rate, experiments were designated in this study to establish the relationship between the embedding rate and the likelihood function. In the experiments, the embedding rate was set to 5%, the length of detection windows was 300 packets (each packet with a length of 160ms), and the results are shown in Figure 9. The results were then used to obtain the probability density functions for the VoIP streams not containing secret messages and those containing secret messages with 5% LSB being replaced, respectively.

In addition, the mean square deviation of the likelihood function has an impact on the performance of VoIP steganalysis. Figure 10 shows the mean square deviation of the likelihood function decreased as the detection window length increased.

Therefore, equation (10) and experimental data in Figures 9 and 10 enable ones to determine the optimum detection threshold at different levels of data embedding.

3.5 Improved RS Algorithm vs. Conventional RS Algorithm

To examine whether or not the improved RS algorithm would affect detection accuracy, a series of experiments were conducted by using the sliding window detection method with the improved RS algorithm or the conventional RS algorithm under the same experimental conditions.

The detection window lengths in the experiments were set to be 200, 250 and 300 packets in a G.711 codec, respectively; and the embedding rates were kept at 20%. The conventional RS algorithm and the improved RS algorithm were applied to the same detection window lengths. Both algorithms were compared in terms of detection performance, computational complexity and memory resources.

Figure 11 shows comparisons in detection accuracy between the improved RS algorithm using θ_r and the conventional RS algorithm using θ_p. The ROC curves for the algorithms show both steganalysis algorithms achieved similar detection accuracy. The experimental results indicate there were no obvious discrepancies in detection accuracy between using θ_r (the improved RS algorithm) and θ_P (the conventional RS algorithm) as the optimum detection threshold.

In view of computational time, both the improved and conventional RS algorithms are $o(n)$ degree since their computational complexity has a linear relation to the length of sampled stream data. Table 1 lists computational complexity and the detection time in ms for the improved and conventional RS algorithms. The results indicate both algorithms made correct judgments; however, the improved RS algorithm reduced the detection time from 292 ms to 105 ms, achieving up to 64% performance improvement over the conventional RS method, when encrypted secret messages were embedded in VoIP streams at 20% embedding rate.

Table 1 Comparisons in Detection Time

	Mean detection time (ms)	
Length of cover objects (bytes)	Improved RS algorithm	Conventional RS algorithm
46,560 (no embedding)	62	172
78,960 (20% embedding)	105	292

The improvement in detection time is due to the fact that the sliding window detection method using the improved RS algorithm avoids flipping all the data under the detection window, and it reduces the number of the parameters required to be calculated in Figure 3 from 8 to 2. So the size of the required buffer would be reduced by 50% when compared to the conventional RS method.

3.6 Multiple Covert VoIP Channels

The previous sections (3.2-3.5) detail experimental results for a single covert VoIP communication channel, i.e. only one secret message was embedded in VoIP streams each time. This section deals with multiple covert VoIP communication channels where various secret messages were embedded in different covert channels simultaneously.

The sliding window method using the improved or conventional RS algorithms was used to simultaneously detect multiple covert VoIP channels from all the streams passing through the router. Each covert channel contained a secrete message embedded in VoIP streams; different secrete messages were hidden in different covert VoIP channels. The detection precision of the sliding window method depends on a number of factors such as the detection window length, the used RS algorithm and the number of covert VoIP channels.

Figure 12 shows the detection percentage varied depending on the number of covert VoIP channels at 20% embedding rate. The results indicate the detection percentage decreased gradually as the number of covert VoIP channels increased. The detection accuracy of the conventional RS method reduced dramatically with increasing covert VoIP channels; the detection percentage was only 25% when there were eight covert VoIP channels. However, the sliding widow method using the improved RS algorithm achieved 100 percent detection until the number of covert VoIP channels reached five (i.e. five secret messages were embedded in five different covert channels simultaneously); this new method still attained over 65 percent of detection even there were eight covert VoIP channels. The experiments indicate that the proposed sliding window method provide a significant improvement on precision in detecting multiple covert VoIP channels when compared to the conventional RS method.

4. CONCLUSIONS

In this paper we have suggested a new approach to detecting covert channels in VoIP communications. Our proposed sliding window steganalysis method provides a 64 percent improvement on the detection time over the conventional RS method, thereby enabling rapid and real-time detection of stego VoIP streams. The proposed method has been proved to be suitable for detecting covert VoIP communications.

Experimental results have shown that the proposed sliding window method using an improved RS algorithm can correctly and simultaneously detect up to five covert VoIP channels with 100 percent detection accuracy. At this stage, the optimum detection window length is determined by modelling empirical data. Building a mathematical model for depicting the relationship between detection accuracy, efficiency and the detection window length is the subject of future work.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Foundation Theory Research of China (973 Program, No. 2007CB310806), and the National Natural Science Foundation of China (No. 60703053, and No. 60773140). The authors would like to thank anonymous reviewers for their valuable suggestions.

REFERENCES

[1] Peticolas, F.A.P., Anderson, R.J., and Kuhn, M.G.: 'Information Hiding – A Survey', IEEE Trans. Proc. Thy., 1999, 87, (7), pp. 1062-1078

[2] Artz, D.: 'Digital Steganography: Hiding Data within Data', IEEE Internet Computing, May 2001, pp. 75-80

[3] Bao, P., and Ma, X.: 'MP3-resistant music steganography based on dynamic range transform'. Proc. IEEE International Symposium on Intelligent Signal Processing and Communication Systems, Seoul, Korea, Nov. 2004, pp. 266-271

[4] Johnson, N.F., Duric, Z., and Jajodia S.: 'Information hiding: Steganography and watermarking-attacks and countermeasures' (Kluwer Acade. Publishers, 2000)

[5] Dittmann, J., Hesse, D., and Hillert, R.: 'Steganography and steganalysis in Voice-over IP scenarios:

operational aspects and first experiences with a new steganalysis tool set". SPIE and IS&T Proceedings, San Jose, California, USA, Jan. 2005, pp. 607-618

[6] Kratzer, C., Dittmann, J., Vogel, T., and Hillert, R.: 'Design and Evaluation of Steganography for Voice-over-IP'. Proc. IEEE International Symposium on Circuits and Systems, Kos, Greece, May 2006, pp. 4

[7] Fridrich, J., Goljan, M., and Du, R.: 'Detecting LSB Steganography in Colour and Grayscale Images', IEEE Multimedia, 2001, pp. 22-28

[8] Dumitrescu, S., Wu, X., and Wang, Z.: 'Detection of LSB Steganography via Sample Pair Analysis', IEEE Transactions on Signal Processing, 2003, 51, (7), pp. 1995-2007

[9] Lyu, S., and Farid, H.: 'Detecting hidden messages using higher-order statistics and support vector machines', Lecture Notes in Computer Science, 2003, 2578, pp. 340-354

[10] Miche, Y., Roue, B., Lendasse, A., and Bas, P.: 'A feature selection methodology for steganalysis'. Proc. International Workshop on Multimedia Content Representation, Classification and Security, Istanbul, Turkey, September 2006, pp. 49-56

[11] Celik, M., Sharma, G., and Tekalp, A.M.: 'Universal image steganalysis using rate-distortion curves'. Proc. SPIE: Security and Watermarking of Multimedia Contents VI, San Jose, CA, Jan. 2004, pp. 467-476

[12] Fridrich, J.: 'Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes'. Proc. 6th Int. Workshop Information Hiding, Toronto, ON, Canada, May 2004, pp. 67-81

[13] Ozer, H., Avcibas, I., Sankur, B., and Memon, N.: 'Steganalysis of audio based on audio quality metrics'. Proc. SPIE Electronic Imaging Conf. On Security and Watermarking of Multimedia Contents, Santa Clara, Calif, USA, January 2003, pp. 55-66

[14] Johnson, M.K., Lyu, S., and Farid, H.: 'Steganalysis of recorded speech'. Proc. Security, Steganography,

and Watermarking of Multimedia Contents VII, San Jose, CA, USA, March 2005, pp. 664-672

[15] Ru, X.M., Zhang, H.J., and Huang, X.: ‘Steganalysis of audio: Attacking the steghide’. Proc. Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, China, August 2005, pp. 3937-3942

[16] Xiao, B., Huang, Y., and Tang, S.: ‘An Approach to Information Hiding in Low Bit-Rate Speech Stream’. IEEE GLOBECOM, IEEE, Dec. 2008, pp. 371-375

[17] Kraetzer, C., and Dittmann, J.: ‘Mel-cepstrum-based steganalysis for VoIP steganography’. Proc. SPIE, San Jose, USA, January 2007, pp. 650-661

[18] Simmons, G. J.: ‘The prisoners’ problem and the subliminal channel’, in Chaum D. (Ed.): ‘Advances in Cryptology: Proceedings of Crypto’83’ (Plenum Press, 1984), pp. 51-67

[19] Avcibas, I., Kharrazi, M., Memon, N., and Sankur, B.: ‘Image steganalysis with binary similarity measures’, EURASIP Journal on Applied Signal Processing, 2005, 17, pp. 2749-2757

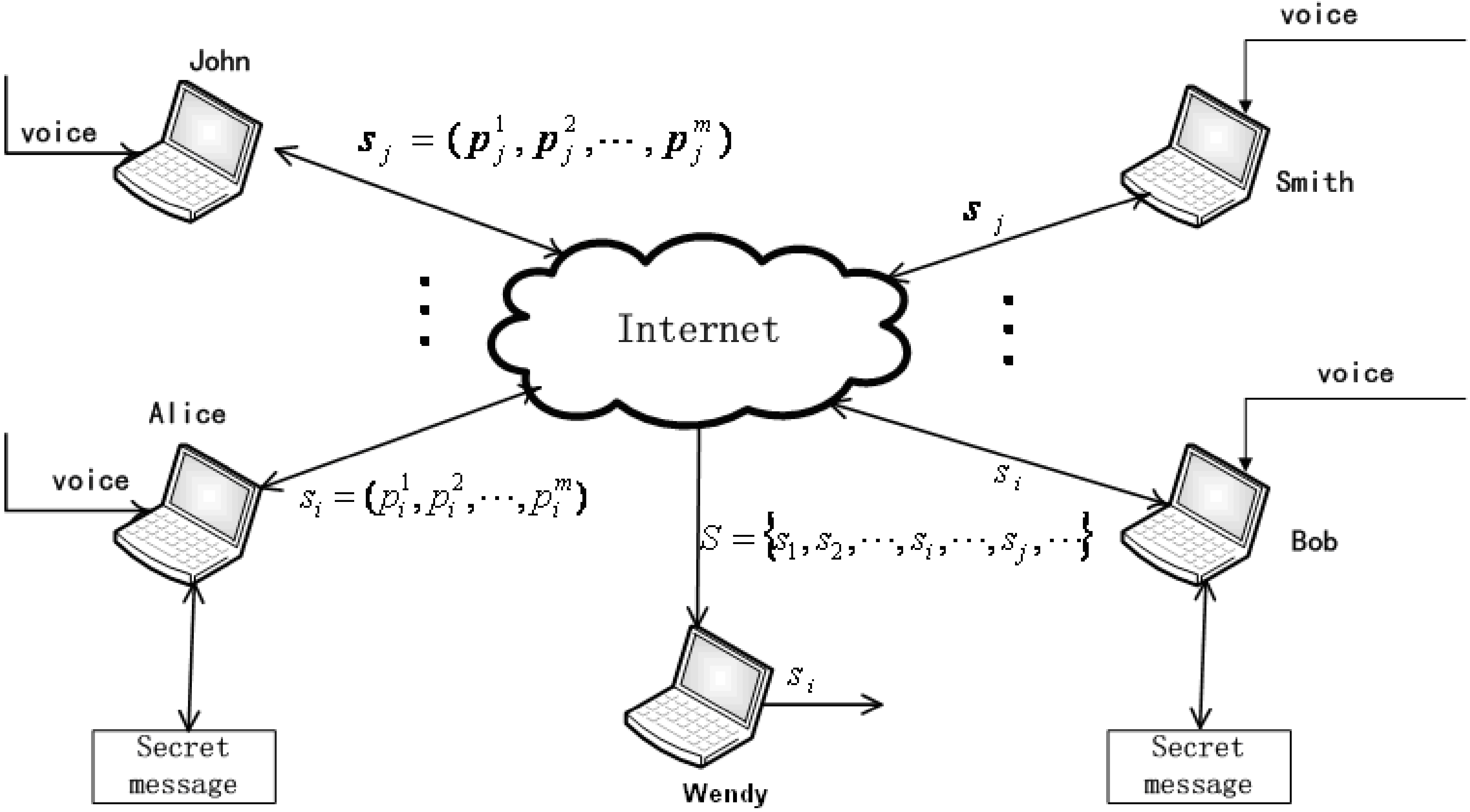
voice
John
voice
$s_j = (p_j^1, p_j^2, \cdots, p_j^m)$
Smith
s_j
Internet
voice
Alice
voice
$s_i = (p_i^1, p_i^2, \cdots, p_i^m)$
s_i
Bob
$S = \{s_1, s_2, \cdots, s_i, \cdots, s_j, \cdots\}$
s_i
Secret message
Wendy
Secret message

Fig. 1. A steganalysis model for VoIP steganography

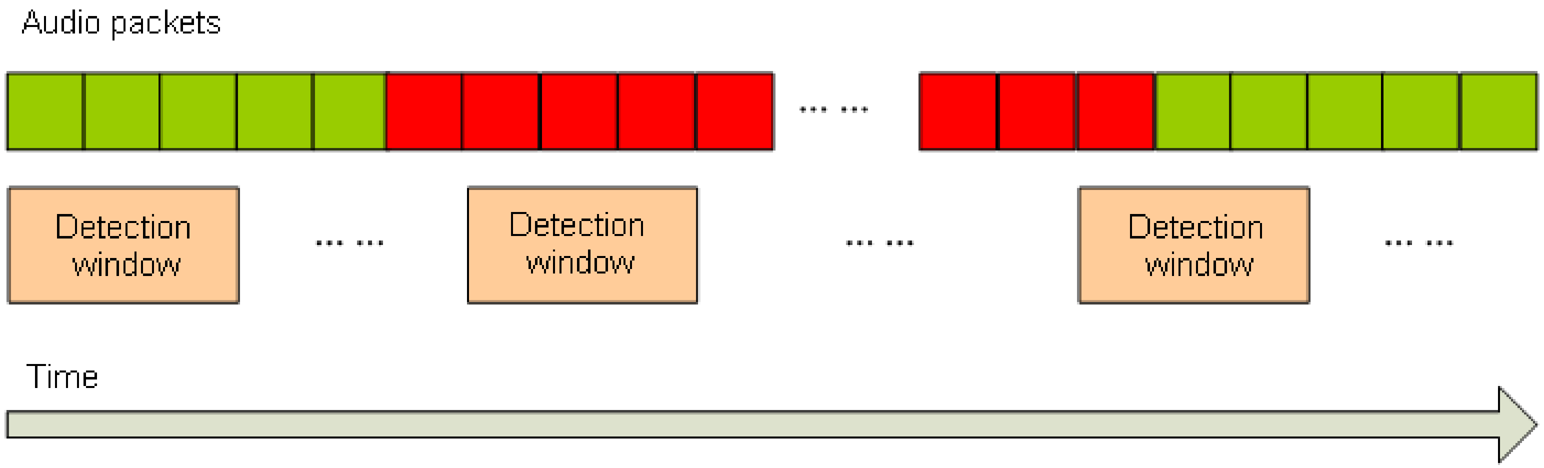

Fig. 2. The sliding process of the detection window

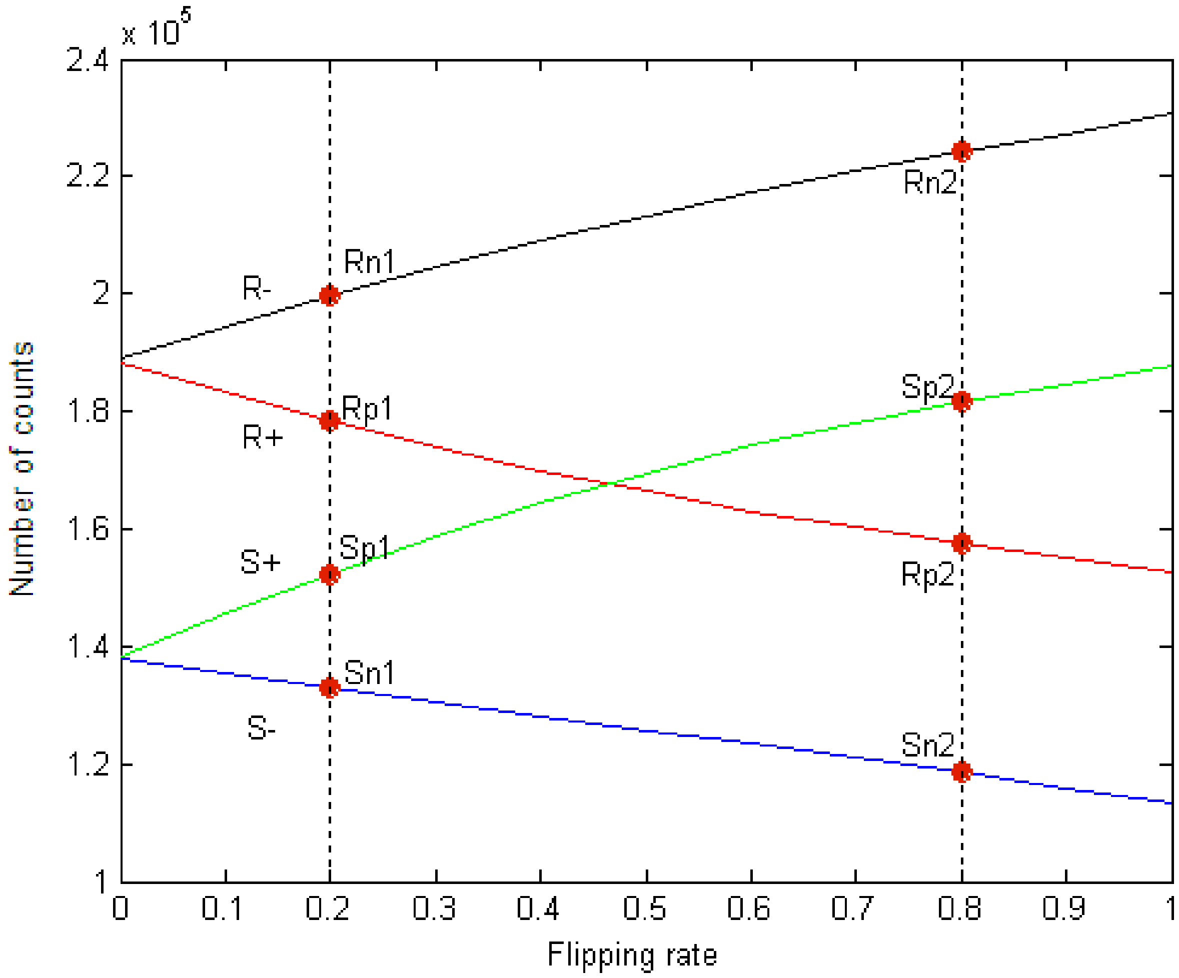

Fig. 3. RS-diagram of VoIP streams

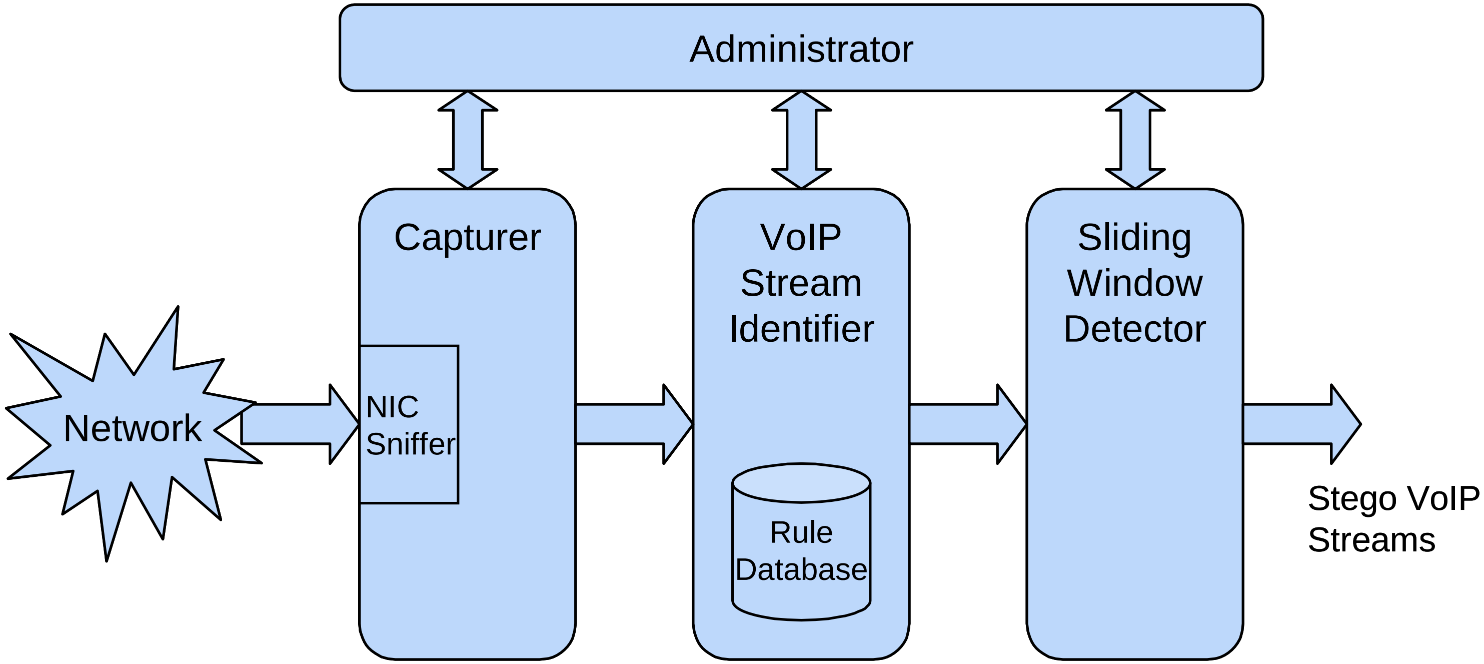

Fig. 4. The framework for sliding window based VoIP steganalysis

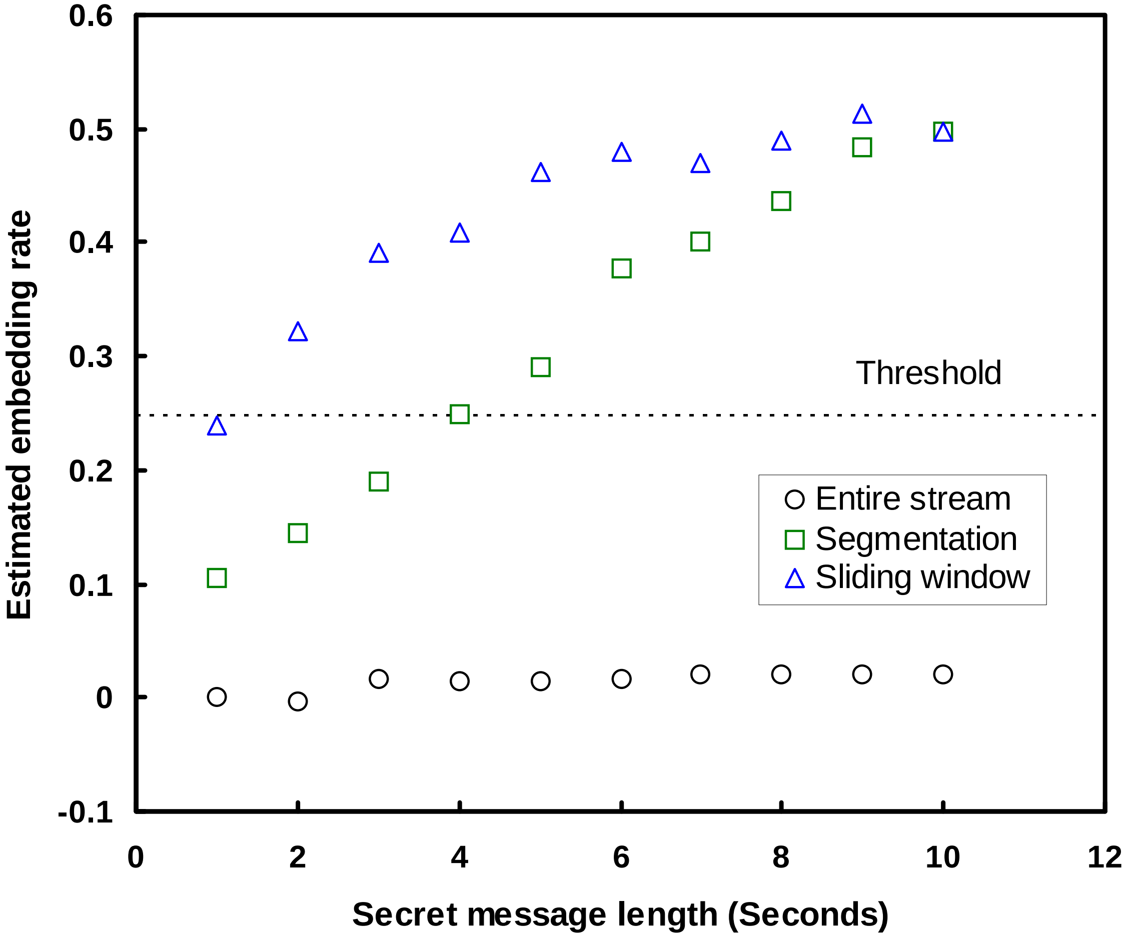

Fig. 5. Comparisons in performance between three detection methods

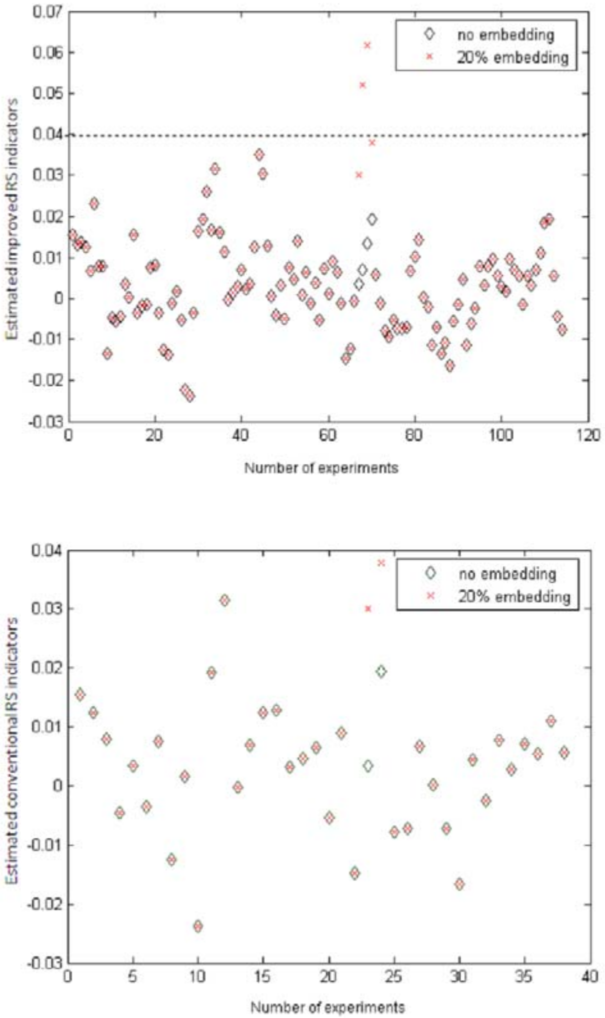

Fig. 6. Comparisons between the sliding window method (top figure) and the segmentation method (bottom figure)

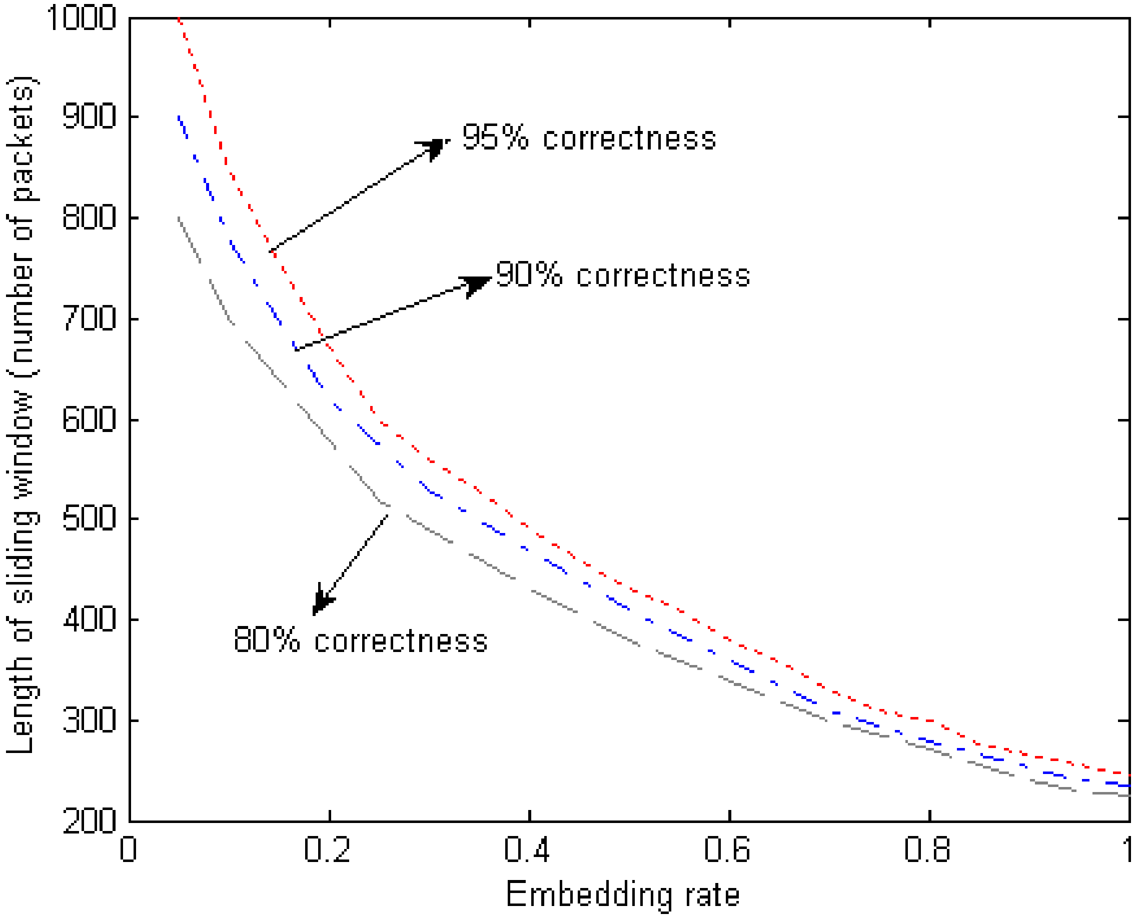

Fig. 7. The length of sliding windows vs. the embedding rate

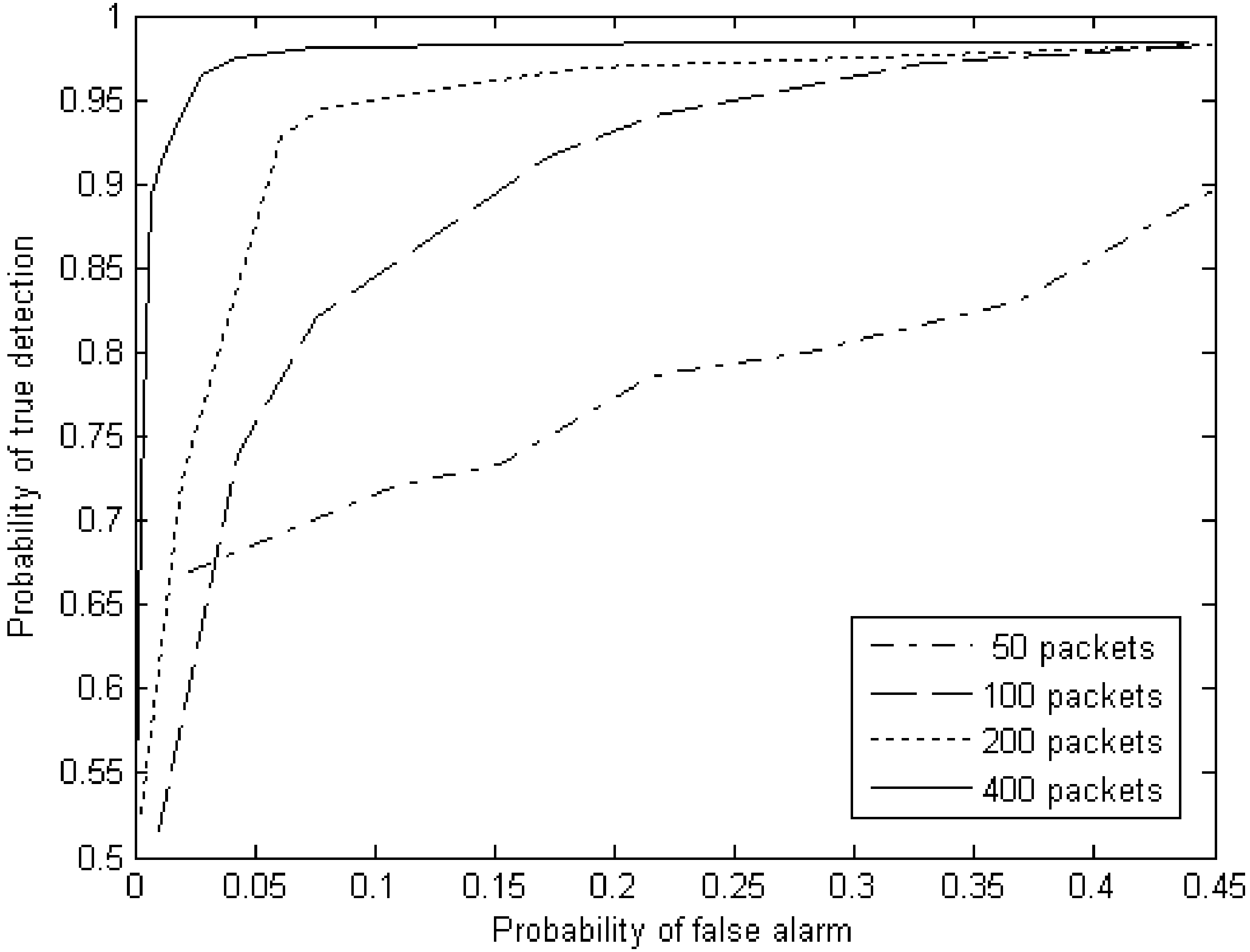

Fig. 8. ROC curves for various sliding window lengths at 20% embedding rate

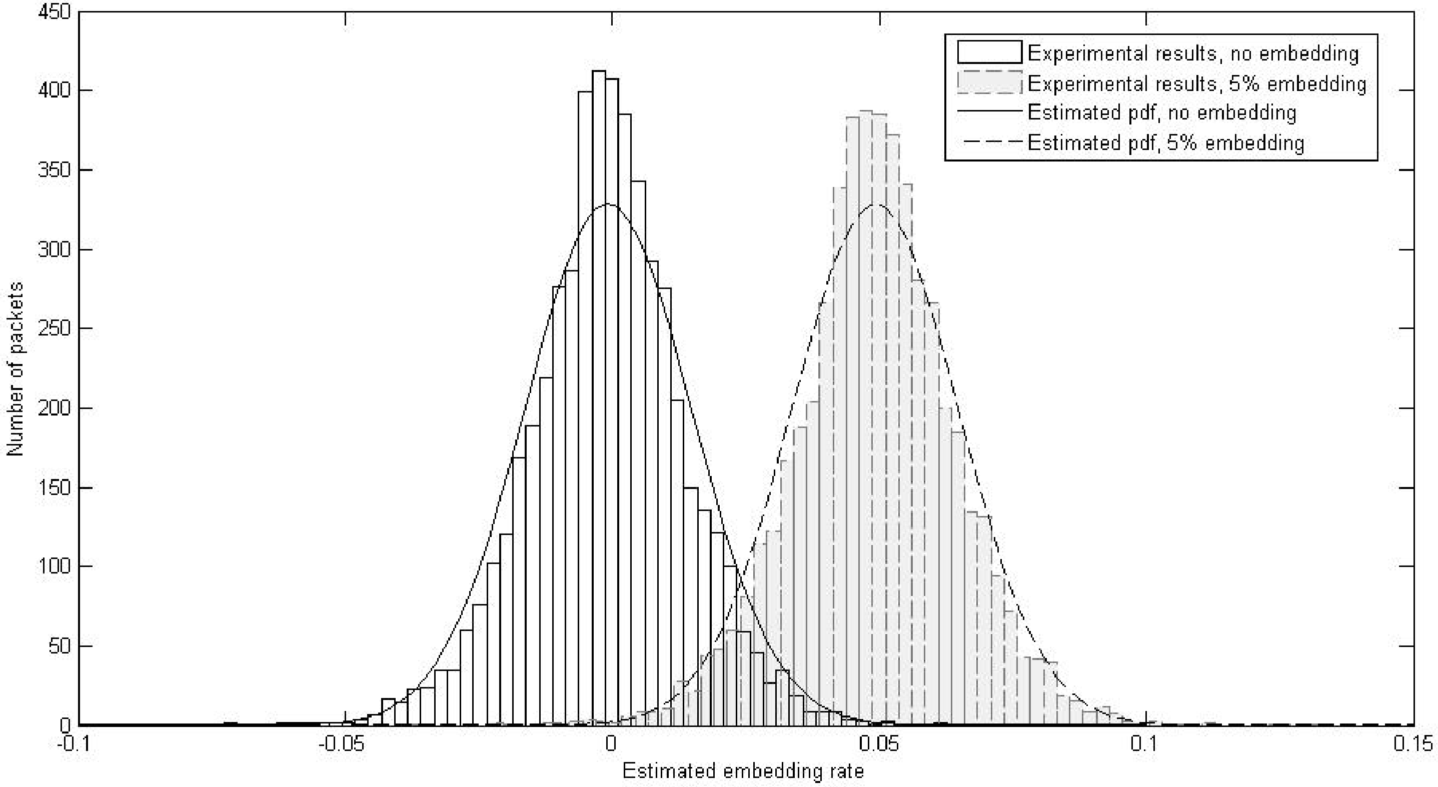

Fig. 9. The probability density functions

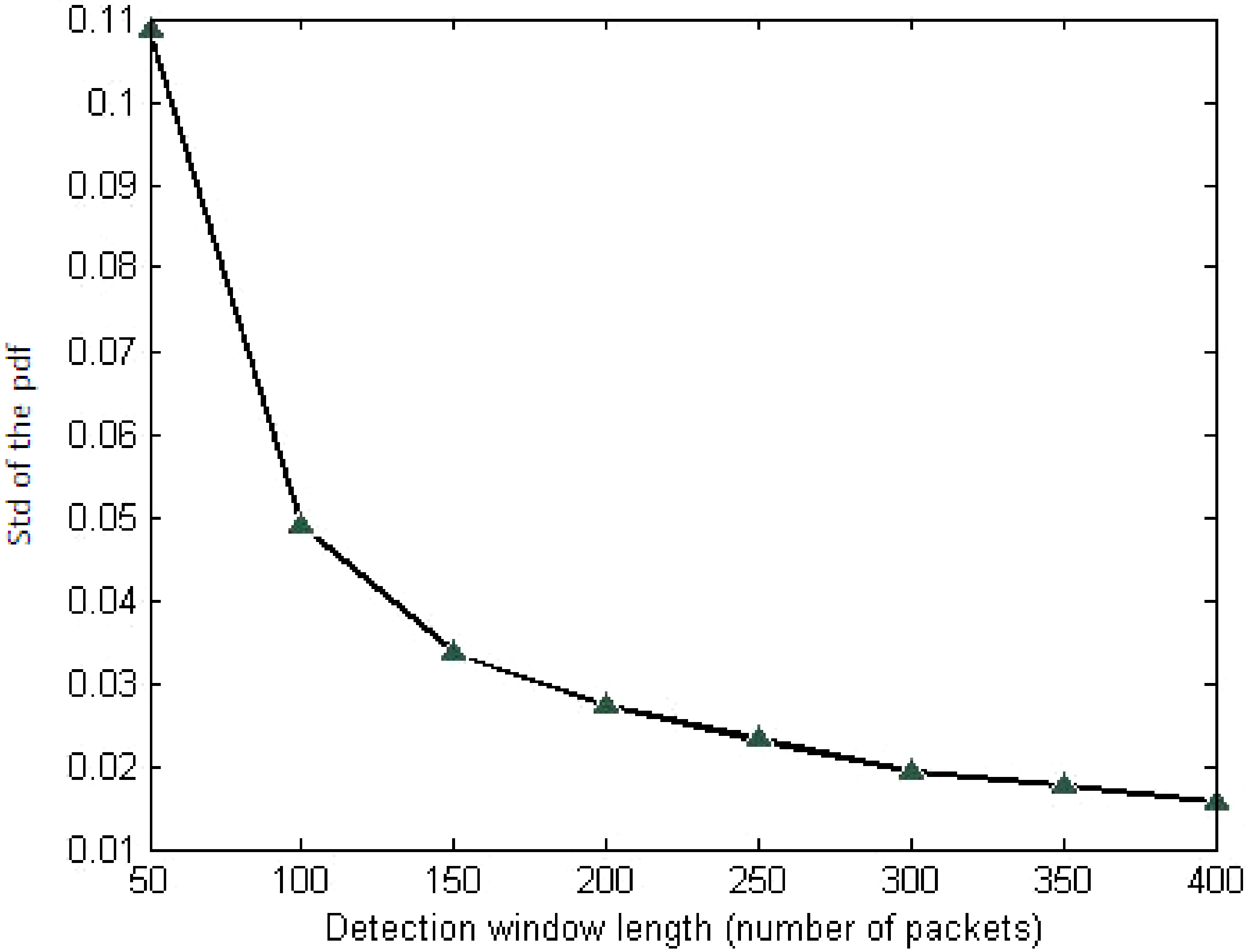

Fig. 10. The mean square deviation vs. the detection window length

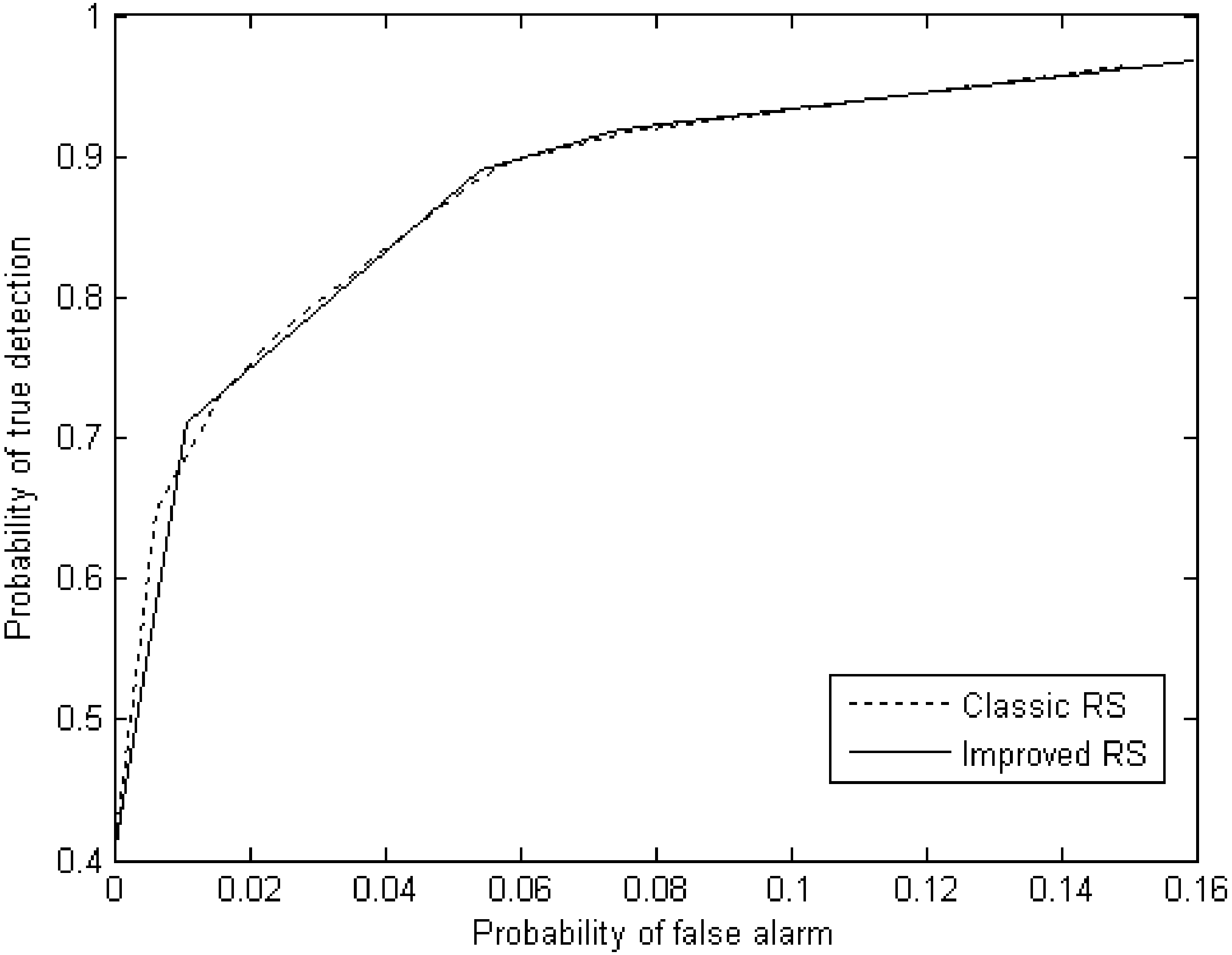

Fig. 11. ROC curves for the algorithms using θ_r and θ_p

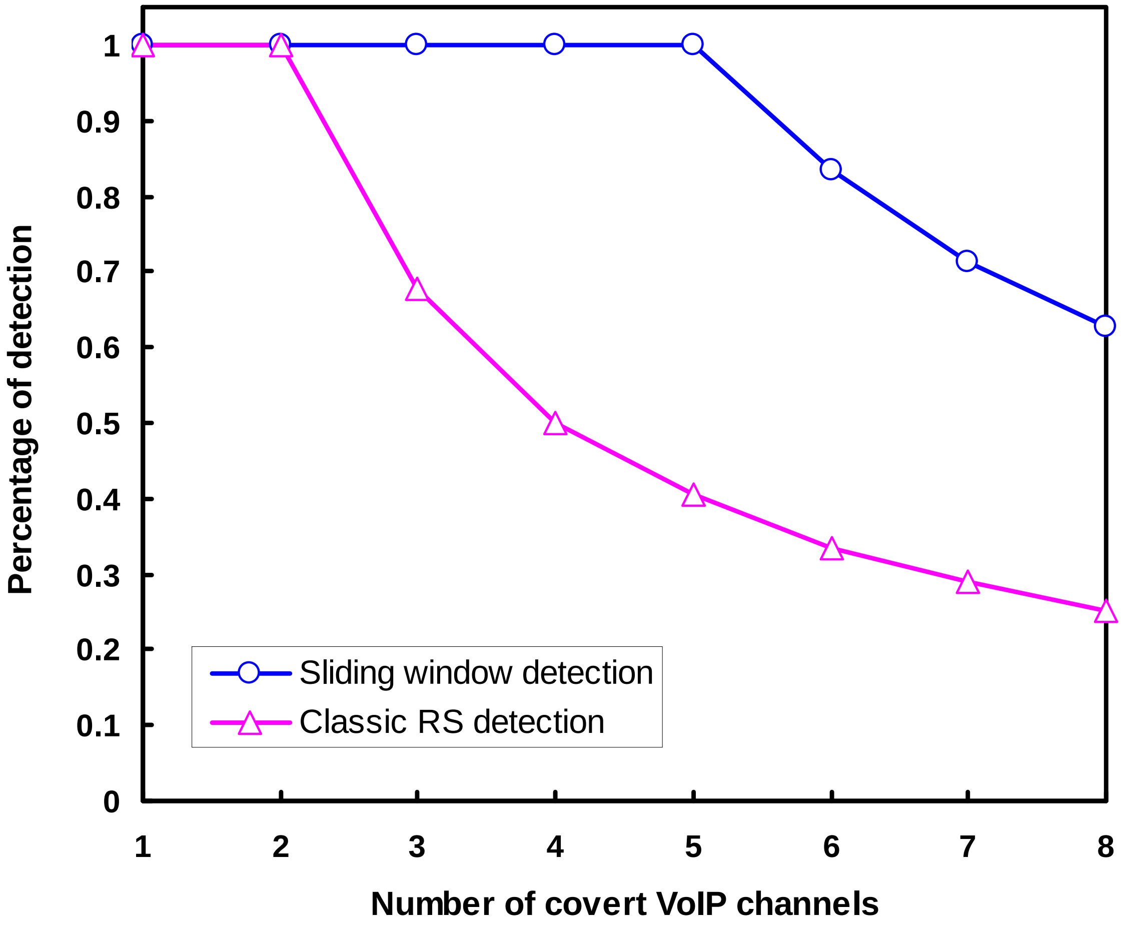

Fig. 12. Detection percentage vs. the number of covert VoIP channels