SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2015; **8**:3804–3814
Published online 29 June 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.1301

RESEARCH ARTICLE

On the undetectability of transcoding steganography

Artur Janicki, Wojciech Mazurczyk* and Krzysztof Szczypiorski

Institute of Telecommunications, Warsaw University of Technology, 15/19 Nowowiejska Str., 00-665 Warsaw, Poland

ABSTRACT

Transcoding Steganography (TranSteg) is a fairly new IP telephony steganographic method that is characterized by a high steganographic bandwidth, low introduced distortions, and high undetectability. TranSteg utilizes compression of the overt data to free space for the secret data bits. In this paper, we focus on evaluating different possibilities for TranSteg detection. Building on the previous works, we perform a wide analysis of different steganalysis methods to assess the possibility of TranSteg detection and identify the most 'undetectable' pairs of voice codecs. Copyright © 2015 John Wiley & Sons, Ltd.

KEYWORDS

IP telephony; network steganography; TranSteg

***Correspondence**

Mazurczyk, Wojciech, Institute of Telecommunications, Warsaw University of Technology, 15/19 Nowowiejska Str., 00-665 Warsaw, Poland.

E-mail: wmazurczyk@tele.pw.edu.pl

1. INTRODUCTION

Steganography relies on the embedding of secret data into an innocent-looking carrier of this message. Steganographic technique aims to hide the very existence of the communication so that any external observers remain unaware of the steganographic transmission. These solutions have been evolving throughout history, and now, naturally, packet networks are suitable targets for steganography. Steganography based on network protocols as the carrier for the steganographic communication is called *network steganography* [1]. Network steganography methods can be applied, for example, as a tool to circumvent oppressive government surveillance by providing a means to communicate covertly to avoid detection by current monitoring devices.

Currently, among the many diverse and complex services for IP networks, one of the most popular is IP telephony or Voice over IP (VoIP). From the network perspective, a typical VoIP call can be divided into two phases: signaling and conversation. During the first phase, the caller and the callee exchange certain signaling protocol messages, for example, of Session Initiation Protocol [2], to set up the connection and negotiate its parameters. During the second conversation phase, two voice streams based on a Real-Time Transport Protocol (RTP) [3] are sent in a bidirectional manner. Due to IP telephony's popularity and its traffic volume, it has increasingly attracted researchers' attention [4].

Typically, any information-hiding method can be evaluated by calculating a set of three characteristics: undetectability, steganographic bandwidth, and the steganographic cost. Undetectability is the inability to detect secret data within a hidden data carrier. Typically, detection is performed by analyzing the statistical properties of the captured traffic and then comparing them with the typical values for that carrier. Steganographic bandwidth describes the amount of secret information that can be sent per time unit for a given method. Finally, the steganographic cost illustrates the negative influence on the hidden data carrier caused by the steganographic technique.

Transcoding Steganography (TranSteg) is a fairly new steganographic technique that was originally introduced in [5]. The actual concept of the proposed method is significantly different from other steganographic techniques. In classical steganography, the covert data is usually compressed (because of the limited bandwidth of the steganographic channel), while in TranSteg, it is the overt data that has its size reduced to make space for the secret data. This is achieved through the transcoding (either in a lossy or lossless manner) of the speech data from a higher bit rate codec to a lower one, while, at the same time, minimizing the decrease in voice quality.

The TranSteg function is illustrated in Figure 1. First, RTP packets that carry the caller's voice are analyzed, and the codec originally utilized for voice encoding (hereinafter referred as the overt codec) is pinpointed by inspecting the payload type field in the RTP header (Figure 1(a)).

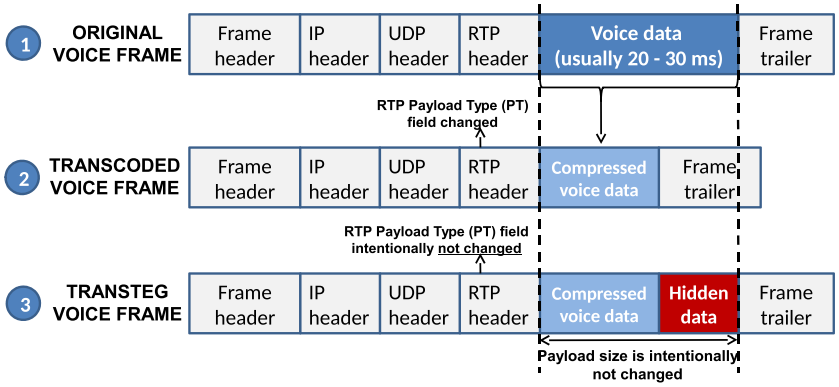

Figure 1. Frame carrying speech payload (1) encoded with overt codec, (2) typically transcoded, and (3) encoded with covert codec.

Typically, if (not steganographic) transcoding is realized, then the original voice frames are encoded using a different voice codec, which results in a smaller voice frame being achieved (Figure 1(b)). However, TranSteg selects a so-called covert codec for the originally utilized overt codec. The covert codec should yield a comparable voice quality but result in a smaller voice payload size than in the original. Thus, the voice stream is transcoded, but the larger, original speech payload size and indicator of codec type (in payload type field) are unchanged. As a result, the original voice is transcoded (using a covert codec) to a smaller size and is placed into the original payload field. Thus, the remaining free space can be populated with secret information (Figure 1(c)). Note that secret data can be spread across the payload field or interleaved with voice samples in a predetermined way (the selection of this algorithm is not in the scope of this work).

Transcoding Steganography performance depends mainly on the selection of overt and covert codecs. In an ideal situation, the *covert codec* should not have a significant impact on speech quality, when compared with the overt codec's quality, while resulting in the smallest possible speech payload size. On the other hand, the *overt codec*, when paired with the covert codec, should be able to achieve the largest possible payload size to give the highest possible steganographic bandwidth and should be chosen from the most popular codecs utilized for IP telephony, to avoid suspicion.

In [6], the most popular codecs for IP telephony were evaluated to establish which pairs of codecs should be chosen for transcoding to minimize the negative influence on the hidden data carriers while maximizing the obtained steganographic bandwidth. From this analysis, it was determined that the choice of covert codec depends on whether priority is given to higher steganographic bandwidth or better speech quality. From the evaluated overt/covert codec pairs, 10 were recommended as optimal for TranSteg purposes.

Later, in [7], lightweight traffic analysis, which relies on monitoring the first byte of the payload in each of the VoIP packets, was developed. The motivation behind this approach was the fact that some codecs begin their payload with a control sequence (e.g., a mode ID). Therefore, with this approach, the detection of a potential mismatch between the declared and actual voice codec can be discovered. Experimental results revealed that some codecs are fairly easily detected using this simple analysis. However, some of the codec pairs could still not be easily detected. That is why in [8], a novel steganalysis method based on GMM models and Mel-Frequency Cepstral Coefficient (MFCC) parameters was proposed, implemented, and tested. Experimental results revealed that many codec pairs can be detected with an average detection probability of more than 85%.

The work described in this paper aims to analyze the undetectability of TranSteg in a wide spectrum. In other words, on the basis of the recommendations from previous work, we want to evaluate different possible detection methods to establish the most suitable one. Therefore, the contributions of the paper are as follows:

(1) to evaluate TranSteg undetectability by creating histograms of MFCC parameters for normal and abnormal traffic, and test various classifiers (Bayesian Networks, Decision Trees, C4.5, Support Vector Machines (SVMs), Multilayer Perceptron (MLP), and AdaBoost) based on these data; and
(2) to recommend the most suitable steganalysis method and the most covert pair of codecs for TranSteg purposes.

The article is structured as follows. In Section 2, the related work on IP telephony steganography detection is reviewed. Section 3 presents the assumed threat model for various possible hidden communication scenarios. In Section 4, the experimental methodology and obtained results are described. Finally, the last section summarizes our work.

2. RELATED WORK

This section presents an overview of existing work in two areas: (i) the methods of double compression detection for

digital media steganography (i.e., in images, audio, and video) and (ii) IP telephony steganalysis techniques.

2.1. Double compression detection

To detect TranSteg in some scenarios presented in detail in the next section, it is possible to look for artifacts caused by transcoding. Discovering the existence of double compression has been a subject of numerous analyses, but only for digital images (e.g., [9,10]), digital audio (mostly wideband MP3 files) [11,12], and video signals [13,14].

2.2. VoIP detection techniques

Many steganalysis methods have been proposed so far to enable the detection of covert communication for IP telephony. However, currently, specific and practically applicable steganalysis methods are not widespread (if used at all). In this section, we consider only the detection methods that have been evaluated and proved feasible for this multimedia service. It must be emphasized that many so-called audio steganalysis methods were also developed for detection of hidden data in audio files (so-called audio steganography). However, we consider these techniques beyond the scope of this paper.

Statistical steganalysis for least significant bits (LSB)-based VoIP steganography was proposed by Dittmann *et al.* [15]. They proved that it was possible to detect hidden communication with almost a 99% success rate on the assumption that there are no packet losses and the steganogram is unencrypted/uncompressed.

Takahasi and Lee [16] described a detection method based on calculating the distances between each audio signal and its de-noised residual when using different audio quality metrics. Then, an SVM classifier was utilized for detection of the existence of hidden data. This scheme was tested on LSB, direct sequence spread spectrum, frequency-hopping spread spectrum, and echo-hiding methods, and the results obtained show that for the first three algorithms, the detection rate was about 94% and for the last, it was about 73%.

A Mel-Cepstrum-based detection, known from speaker and speech recognition, was introduced by Kraetzer and Dittmann [17] for the purpose of VoIP steganalysis. On the assumption that a steganographic message is not permanently embedded from the start to the end of the conversation, the authors demonstrated that detection of an LSB-based steganography is efficient with a success rate of 100%. This work was further extended in [18] employing an SVM classifier. In [19], it was shown for an example of VoIP steganalysis that channel character-specific detection performed better than that when the channel characteristic features were not considered.

Steganalysis of LSB steganography based on a sliding window mechanism and an improved variant of the previously known Regular Singular algorithm was proposed by Huang *et al.* [20]. Their approach provided a 64% decrease in the detection time over the classic Regular Singular algorithm, which makes it suitable for VoIP. Moreover, experimental results prove that this solution was able to detect up to five simultaneous VoIP covert channels with a 100% success rate.

Huang *et al.* [21] also introduced a steganalysis method for compressed VoIP speech that was based on second-order statistics. To estimate the length of the hidden message, the authors proposed to embed hidden data into the sampled speech at a fixed embedding rate, followed by embedding other information at a different level of data embedding. Experimental results showed that this solution makes it possible not only to detect hidden data embedded in a compressed VoIP call but also to accurately estimate its size.

A steganalysis that relies on the classification of RTP packets (as steganographic or non-steganographic ones) and utilizes specialized random projection matrices that take advantage of prior knowledge about the normal traffic structure was proposed by Garateguy *et al.* [22]. Their approach was based on the assumption that normal traffic packets belong to a subspace of a smaller dimension (first method) or that they can be included in a convex set (second method). Experimental results showed that the subspace-based model proved to be very simple and yielded very good performance, while the convex set-based one was more powerful but more time consuming.

Arackaparambil *et al.* [23] analyzed how, in distribution-based steganalysis, the length of the window of the detection threshold and in which the distribution was measured should be depicted to provide the greatest chance of success. The results obtained showed how these two parameters should be set for achieving a high rate of detection, while maintaining a low rate of false positives. This approach was evaluated based on real-life VoIP traces and a prototype implementation of a simple steganographic method.

A method for detecting Complementary Neighbor Vertices-Quantisation Index Modulation steganography in G.723.1 voice streams was described by Li and Huang [24]. This approach developed two models, a distribution histogram and a state transition model, to quantify the codeword distribution characteristics. With these two models, feature vectors for training the classifiers for steganalysis can be obtained. The technique was implemented by constructing an SVM classifier, and the results showed that it can achieve an average detection success rate of 96% when the duration of the G.723.1 compressed speech bit stream was less than 5 s.

Two specific approaches were evaluated for TranSteg steganalysis [7,8]. First, in [7], lightweight traffic analysis that relies on monitoring the first byte of the payload of each VoIP packet was utilized. The motivation behind this approach was the fact that some codecs begin their payload with a control sequence (e.g., a mode ID). Therefore, with this approach, the detection of a potential mismatch between the declared and actual voice codec can be discovered. Experimental results revealed that some codecs were fairly easily detected using this simple analysis. Thus,

for instance, if the Speex codec is chosen as a covert one, additional actions are required, such as bit randomization, to make detection less likely.

In [8], the analysis of TranSteg detectability was presented for a variety of scenarios and potential warden configurations. Particular attention was paid toward the very demanding case of a single warden located at the end of the VoIP channel. For this purpose, a novel steganalysis method based on the GMM models and MFCCs was proposed, implemented, and tested. Successful detection of TranSteg using the described method, for this scenario, requires at least 2 s of speech data to analyze (i.e., a hundred 20-ms VoIP packets).

The results showed that the proposed method allowed for efficient detection of some codec pairs (e.g., G.711/G.726) with an average detection probability of 94.6%, Speex7/G.729, with 89.6% detectability, or Speex7/iLBC, with 86.3% detectability. On the other hand, some TranSteg pairs remained resistant to detection using this method (e.g., the pair of iLBC/Adaptive Multi-Rate (AMR)) with an average detection probability of 67%, which we consider to be low. We found a correlation between steganographic cost of an overt/covert codec pair and detectability of TranSteg—usually, the lower the cost, the more difficult the detection of TranSteg. However, some results were surprising: the G.711/G.726 pair, with low steganographic cost (0.42 MOS), turned out to be relatively easy to detect. In contrast, the pair G.711/Speex7, offering a similar cost, proved to be resistant to steganalysis, with recognition accuracy as low as 63.3% and, what is more, with higher steganographic bandwidth. This confirms that TranSteg with properly selected overt and covert codecs is an efficient steganographic method if analyzed with a single warden.

3. THREAT MODEL

3.1. Hidden communication scenarios

Transcoding Steganography can be applied in four different communication scenarios (Figure 2). Scenario S1 is typically the most desired and common: the secret sender (SS) and the secret receiver (SR) set up an IP telephony call while exchanging secret messages (in an end-to-end manner). The hidden data path is identical to the conversation path. In the remaining three scenarios, only a fragment of the end-to-end path of the VoIP call is utilized for hidden communication purposes (S2–S4 in Figure 2). Therefore, in principle, the overt sender or (and) the overt receiver are unaware of the steganographic data exchange. The application of TranSteg for IP telephony allows the transfer of secret data while, simultaneously, still preserving the users' conversation.

In this paper, we focus on the worst-case scenario in terms of the speech quality for TranSteg—scenario S4, as it requires triple transcoding and TranSteg is responsible for two of them. For the other scenarios (S1–S3), the negative impact on voice quality would be significantly lower than presented in this paper.

The most important benefit of scenario S4 is its potential ability to utilize aggregated IP telephony traffic for covert communication purposes. If both SS and SR are able to exploit multiple VoIP calls, then the resulting steganographic bandwidth can be increased greatly.

In scenario S4, we assume that the SS and SR are capable of capturing and inspecting all the voice packets transmitted between the calling parties. In this case, both hidden parties behave similarly: they must transcode the data—

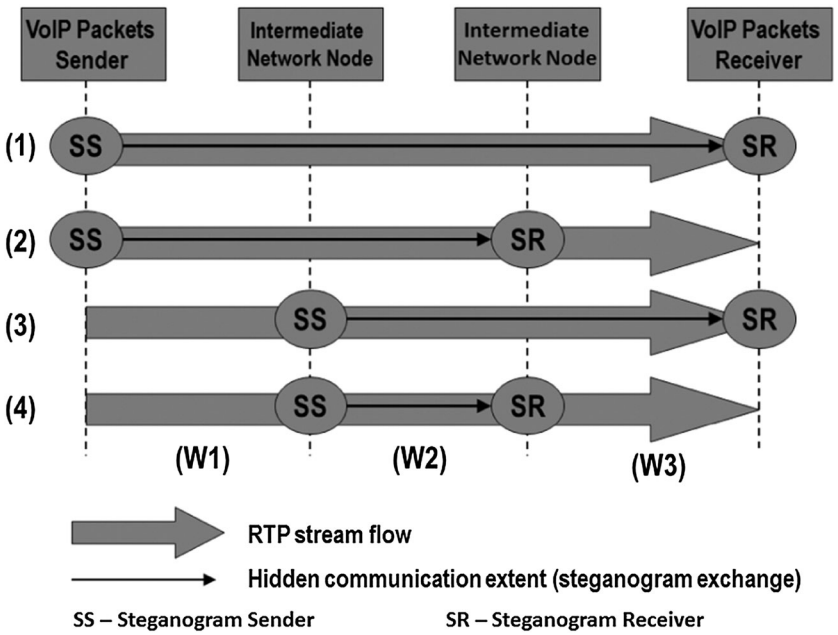

Figure 2. TranSteg hidden transmission scenarios.

hidden sender from overt codec to covert one and hidden receiver must conduct the reverse process. It must be noted that the secret data is transmitted only across part of the whole communication path and it never reaches the endpoints.

The SS first captures the incoming voice stream and transcodes the speech encoded with the overt codec to the covert codec. Then, the transcoded speech is inserted again into a voice packet, and the RTP packet's header is not modified. The resulting free space in the voice payload is replaced with secret data. Next, the encapsulated voice packets are sent to the receiver (SR), which also requires the lower layer protocols' checksums to be recalculated.

At the SR, the following steps are executed. First, the voice payload is analyzed, and the secret data bits from the consecutive RTP packets are extracted. Then, the speech payload is retranscoded (from the covert to the overt codec) and inserted once again in the voice packets. The RTP packet's header is not modified. Encapsulated voice packets are then transmitted to the original receiver (callee) after modifying the lower layer protocols' checksums.

3.2. TranSteg detection scenarios

It must be emphasized that currently for network steganography, as well as for digital media (image, audio or video files) steganography, there is still no universal 'one size fits all' detection solution, so steganalysis methods must be crafted precisely for the specific information-hiding technique.

Typically, it is assumed that the detection of hidden data exchange is left for the warden [25]. In particular, we assume the following: (i) it is aware that users can be utilizing hidden communications to exchange data in a covert manner; (ii) it has a knowledge of all existing steganographic methods, but not the one used by those users; and (iii) it is able to try to detect and/or interrupt the hidden communication.

Let us consider possible hidden communication scenarios (S1–S4 in Figure 2) as they greatly influence the detection possibilities for the warden. For IP telephony steganography, there are three possible localizations of a warden (denoted in Figure 2 as W1–W3). A node that performs steganalysis can be placed near the sender or receiver of the overt communication, or at some intermediate node. Moreover, the warden can monitor network traffic in single (centralized warden) or multiple (distributed warden) locations. In this paper, we assume a centralized warden—for consideration of a distributed warden please refer to [8].

For TranSteg-based hidden communication, we assume that the warden will not be able to 'physically listen' to the speech carried in the RTP packets because of the privacy issues related with this matter. This means that the warden will be capable of capturing and analyzing the payload of each RTP packet but will not be capable of replaying the call's conversation (its content); that is, we assume a scenario without a human-in-the-loop.

It is worth noting that communication via TranSteg can be thwarted by certain actions undertaken by the warden. Covert communication can be eliminated by applying random transcoding to every non-encrypted VoIP connection to which the warden has access (blind approach). Alternatively, only suspicious connections may be subject to transcoding. However, such an approach would lead to deterioration in the quality of the conversations. It must be emphasized that not only steganographic calls would be affected but the non-steganographic calls could also be degraded as well.

To summarize, the successful detection of TranSteg mainly depends on the following: (i) the location(s) at which the warden is able to monitor the modified RTP stream; (ii) the utilized TranSteg scenario (S1–S4); (iii) the choice of the covert and overt codecs; and (iv) whether encryption of the RTP streams is used.

If the warden is capable of inspecting traffic solely in a single localization (the most realistic assumption), three cases are possible:

(1) The warden analyzes the traffic that has not yet been subjected to transcoding caused by the TranSteg, and the voice is coded with the overt codec (scenarios S3 and S4 at localization W1). In that case, it is obvious that TranSteg detection is impossible.
(2) The warden analyzes the traffic that has been subjected to the TranSteg transcoding, and the voice is coded with the covert codec (e.g., scenario S1 at any localization; S2 at localizations W1 or W2).
(3) The warden analyzes the traffic that has been subjected to the TranSteg retranscoding, and the voice is again coded with the overt codec (scenarios S2 and S4 at localization W3). If a pair of lossy overt/covert codecs is used, the detection is not trivial as only the retranscoded, but encoded with an overt codec, voice signal is available.

In this paper, we focus on TranSteg detection for the worst-case scenario from the warden point of view (C3). We assume that the warden is capable of inspecting the traffic only in a single location. That is why we focus on the case where only the retranscoded voice is available and a lossy pair of overt/covert codecs was used (i.e., scenario S4 and localization W3).

It must be emphasized that especially for this scenario, TranSteg steganalysis is harder to perform than for most of the existing VoIP steganographic methods. This is because after the steganogram reaches the receiver, the hidden information is extracted and the speech data is practically restored to the originally sent one. As mentioned earlier, this is a huge advantage compared with existing VoIP steganographic methods, where the hidden data can be extracted and removed but the original data cannot be restored because it was previously erased due to the hidden data insertion process.

4. EXPERIMENTAL RESULTS

4.1. Influence of speech codes on TranSteg performance

As already mentioned, TranSteg performance depends strongly on the selection of the overt and covert codecs. This problem was discussed in detail in [6], where we experimentally measured steganographic bandwidth, that is, the difference between bitrates of the overt and covert codecs, and steganographic cost, that is, the decrease in quality caused by transcoding, for various pairs of speech codecs. We examined the overt codecs most commonly encountered in IP telephony: G.711, Speex, iLBC, and G.723.1. In addition, we analyzed the following codecs as potential candidates to become covert ones: AMR, G.726, G.729, GSM 06.10, and lossless G.711.0.

Because the detailed results were published in [6], let us recall here only the most important findings from those experiments. We determined that the pair G.711/G.711.0 provides no steganographic cost as G.711.0 is lossless. We classified this pair as 'Class 0'. As a consequence of its losslessness, its bandwidth varies depending on the speech data, so the value obtained during experiments (31.11 kbps; Table I) is a statistical mean. Nevertheless, on average, this pair is able to provide a high steganographic throughput at zero cost (this means that this TranSteg variant is practically impossible to detect).

The other pairs introduce certain steganographic costs, but if both codecs suit each other well, they can provide decent steganographic bandwidth at a moderate cost. Therefore, we recommended a few pairs with costs lower than 0.5, for example, G.711/AMR, G.711/Speex7, and iLBC/AMR (classified as 'Class 1' pairs).

Table I. Steganographic bandwidth (in kbps) and steganographic cost (in MOS) for a selection of overt/covert codec pairs [6].

Overt	Covert	Steganographic bandwidth (kbps)	Steganographic cost (MOS)
G.711	G.711.0	31.11	0.00
	G.726	32.00	0.42
	Speex7	39.40	0.35
	iLBC	48.80	0.59
	GSM06.10	51.00	0.86
	AMR	51.80	0.36
	G.729	56.00	0.74
	G.723.1	57.70	0.81
Speex7	iLBC	9.40	0.50
	GSM06.10	11.60	0.76
	AMR	12.40	0.43
	G.729	16.60	0.74
	G.723.1	18.30	0.74
iLBC	GSM06.10	2.20	0.58
	AMR	3.00	0.46
	G.729	7.20	0.74
	G.723.1	8.90	0.63

Other pairs with Speex7 and iLBC as the overt codecs result in higher steganographic costs; that is, a decrease in speech quality higher than 0.5 can be expected. If G.711 is the overt codec and a cost higher than 0.5 is allowed, the pair G.711/G.723.1 can provide a steganographic throughput of almost 58 kbps—these pairs were classified as 'Class 2' pairs. A summary of the most efficient codec pairs is presented in Table I.

4.2. Methodology

In this work, we decided to re-assess the TranSteg undetectability for various codec pairs. We verified the TranSteg detectability by analyzing histograms of MFCCs for the output speech.

The MFCC parameters are able to describe the spectrum of the speech in such a way that filter parameters of the speaker are considered while the source parameters (such as related to glottal excitation) are neglected. This is why these parameters have been previously used successfully (e.g., in speech-to-text systems). Knowing that lossy speech codecs affect these filter parameters and knowing that the MFCCs have already been used successfully in steganalysis [8,17], we decided to use them in the current study.

Figure 3 shows sample histograms of the first MFCC parameter for normal and abnormal (with TranSteg) transmissions for the G.711/G.726 codec pair. It clearly shows that for this pair of codecs, transcoding introduces changes in the MFCC values and, as a consequence, in their distributions.

In our approach, the training and evaluation procedure consisted of the following steps:

(1) Nineteen MFCC parameters were extracted from speech, for both normal (without TranSteg) and abnormal (with TranSteg) transmissions, for various codec pairs.
(2) Histograms were calculated for each of the MFCC parameters using 20 equally spaced bins.
(3) Frequency values in the histogram bins were normalized to sum up to unity.
(4) Normalized frequency values from all 19 MFCC parameters were stacked, thus forming supervectors with $19 \times 20 = 380$ elements.
(5) With the use of the created sets of supervectors, various classifiers were trained to classify both classes of supervectors (normal and abnormal), using the training speech data.
(6) The trained classifiers were tested on the evaluation speech data to determine the undetectability of various codec pairs.

The MFCC parameters were extracted every 10 ms using an overlapping window of 30 ms and a set of 26 mel-scale aligned triangle filters, which is a typical setup used in speech processing (e.g., in speaker recognition) [26].

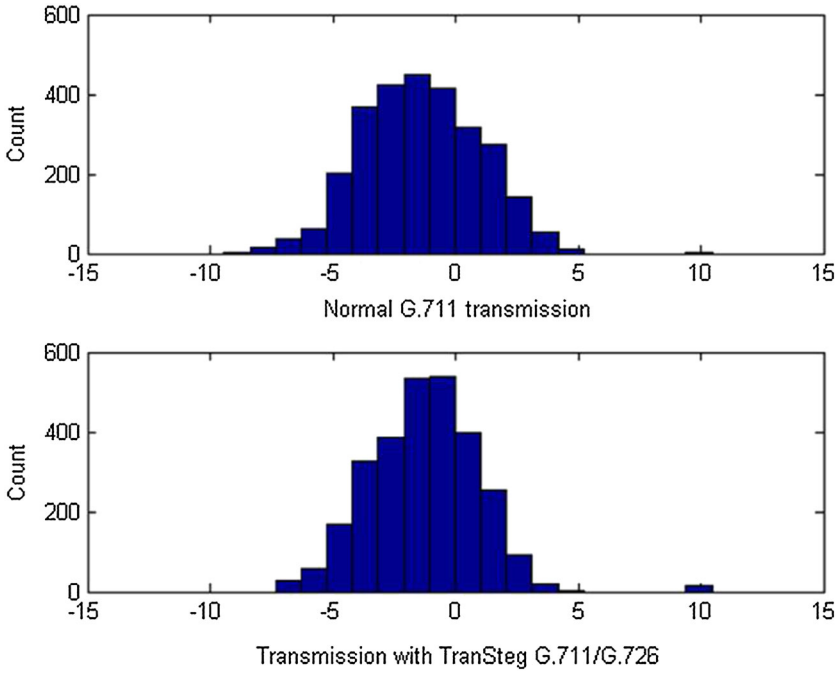

Figure 3. Sample histograms of MFCC1 coefficient for normal (upper) and abnormal transmission (lower).

For training, we used speech data from the TIMIT corpus [27]. We took 1600 sentences uttered by 200 different speakers of eight main American English dialects. For evaluation, we used data from four different corpora:

- TSP corpus [28], containing 1400 recordings originating from 24 US English native speakers;
- CHAINS database [29], with recordings originating from 36 Irish English speakers;
- CORPORA [30], with the recordings of 37 native Polish speakers each of them uttering over a hundred sentences in Polish; and
- AHUMADA [31], a database with recordings in Spanish originating from 104 native Spanish speakers.

To detect TranSteg, we tested a selection of widely used binary classifiers, such as Bayesian Networks (hereinafter called BayesNet), Decision Trees, C4.5 algorithm, SVMs, MLP, and AdaBoost. BayesNets are based on a directed acyclic graph representing probabilistic dependencies between data [32]. Classification with Decision Trees uses a tree with decision nodes and different costs associated with various paths. The C4.5 algorithm is another example of a decision tree, in which decision nodes are created on the basis of information entropy criterion. In this work, the J.48 Java implementation of this algorithm was employed.

Support Vector Machines are based on the support vectors concept originally proposed by Vapnik [33]. They are well known for their high generalization abilities and have, so far, been successfully employed in a variety of applications, ranging from steganalysis [24] to personality traits detection [34]. In this work, we used SVMs with polynomial kernel, using polynomials of the third degree. MLP is a widely used architecture for artificial neural networks. It consists of several layers of mathematical models of a neuron. Such a network with a sigmoid activation function and back-propagation algorithm was used successfully (e.g., in emotion recognition based on speech) [35]. Finally, we used the AdaBoost algorithm—a meta-classifier that iteratively uses other 'weak' classifiers, such as decision trees, and iteratively boosts their performance by exposing them to previously misclassified items [36].

As for the parameter selection algorithm, we used Sequential Forward Selection [37]. Figure 4 displays the statistics of the MFCC parameters, which were mostly

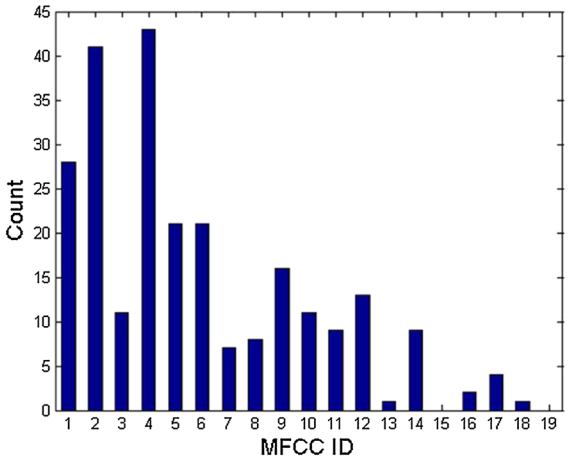

Figure 4. MFCC coefficients used in classification by classifiers reaching accuracy >80%.

Security Comm. Networks 2015; **8**:3804–3814 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/sec

chosen during the parameter selection process. In this analysis, we considered only the classifiers and codec pairs where accuracy reached 80% or more. It turned out that the initial nine MFCC parameters were most useful for detecting TranSteg; however, among them, the third, seventh, and eighth MFCC parameters were far less frequently chosen by the parameter selection algorithm.

Figure 5 shows which histogram bins were the most helpful in TranSteg detection (if this detection was possible at all). According to these statistics, the middle bins (10th, 11th, and 12th) were the ones most often selected by the feature selection algorithm, which suggests that MFCC values around zero had the highest impact on the TranSteg detection. The fourth bins from the histograms of each MFCC parameter seem to play an important role as well.

All TranSteg detection and parameter selection experiments were run in the Weka environment [38]. As evaluation metrics, we used recognition accuracy, that is, the percentage of correctly detected instances of both classes (normal and abnormal) against the total number of instances in each of these classes.

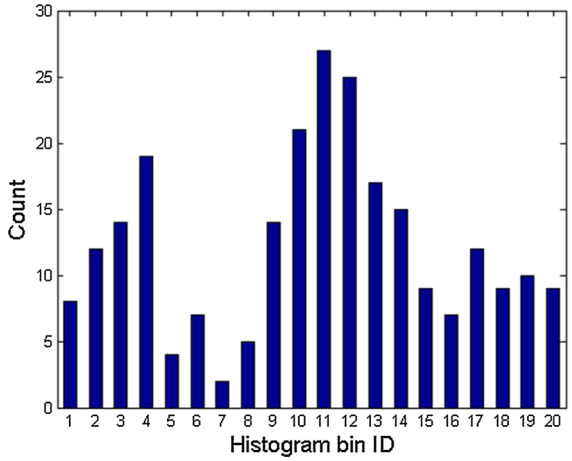

Figure 5. Count of histogram bins used in TranSteg detection based on results of the feature selection algorithm.

4.3. Detection results

Table II shows the results of the detection accuracy for various classifiers and various codec pairs. For comparison, we also displayed the results from our previous studies: accuracy of the GMM-based detector [8], denoted as 'GMM', and accuracy of the payload first byte analyzer [7], denoted as 'PFB security'. Security marked here as 'low' means that both overt and covert codecs can be easily detected by simple analysis of the first payload byte; 'fair' means that neither overt nor covert codec can be easily detected; 'high' security means that overt and covert codecs are easily confused by the codec payload first byte analyzer; 'OC risk' or 'CC risk' mean that either an overt or covert codec can be easily detected, if a warden has access to the W2 point in the transmission path.

Out of the tested classifiers, BayesNet, SVMs, and AdaBoost yielded the best results. However, most classifiers performed worse than the GMM-based classifier, with the exception of the BayesNet classifier, which was able to outperform GMMs in most of the pairs with iLBC as the overt codec and for the Speex7/AMR pair. This can also be seen in Figure 6.

G.711/Speex7 remained the pair with the highest undetectability based on the output signal; the iLBC/GSM06.10 pair was poorly detectable (less than 76% of accuracy). Compared with our previous study, the detectability of the other pairs with iLBC as the overt codec increased to around 85%. TranSteg using Speex7/G.729 and G.711/G.726 was fairly easily detectable by most of the tested classifiers.

Table II. Results of TranSteg detection accuracy (in percentages) for various classifiers and various codec pairs. Results in the GMM and PFB columns are quoted from our previous studies [7,8].

Overt	Covert	PFB security	GMM	BayesNet	DT	J48	SVM	MLP	AdaBoost
G.711	G.726	Fair	94.62	79.75	90.91	90.91	67.77	79.75	90.91
	Speex7	CC risk	63.31	54.55	55.37	54.13	62.81	55.37	55.79
	iLBC	Fair	82.91	81.41	58.26	61.98	76.03	71.90	59.92
	GSM06.10	High	86.10	50.41	58.26	63.22	50.41	50.83	53.31
	AMR	CC risk	81.64	59.92	62.40	63.64	52.89	60.33	63.64
	G.729	Fair	89.50	79.75	66.94	64.88	85.12	82.64	64.05
	G.723.1	Fair	82.02	60.74	67.36	58.68	73.55	73.55	61.98
Speex7	iLBC	OC risk	86.30	78.10	65.70	58.68	71.90	73.55	66.94
	GSM06.10	Low	80.64	52.48	54.96	60.74	60.74	64.05	53.31
	AMR	Low	75.94	80.17	72.31	68.18	56.61	70.66	63.22
	G.729	OC risk	89.62	89.26	71.49	70.25	88.02	86.78	80.58
	G.723.1	OC risk	79.17	76.86	53.31	57.44	61.57	62.40	75.21
iLBC	GSM06.10	CC risk	75.59	71.90	63.64	69.42	73.55	73.55	69.01
	AMR	CC risk	67.00	84.71	67.36	76.86	69.01	69.42	66.94
	G.729	Fair	71.54	83.88	71.49	66.53	73.55	75.62	78.93
	G.723.1	High	70.61	85.12	63.22	56.61	73.55	66.94	75.21

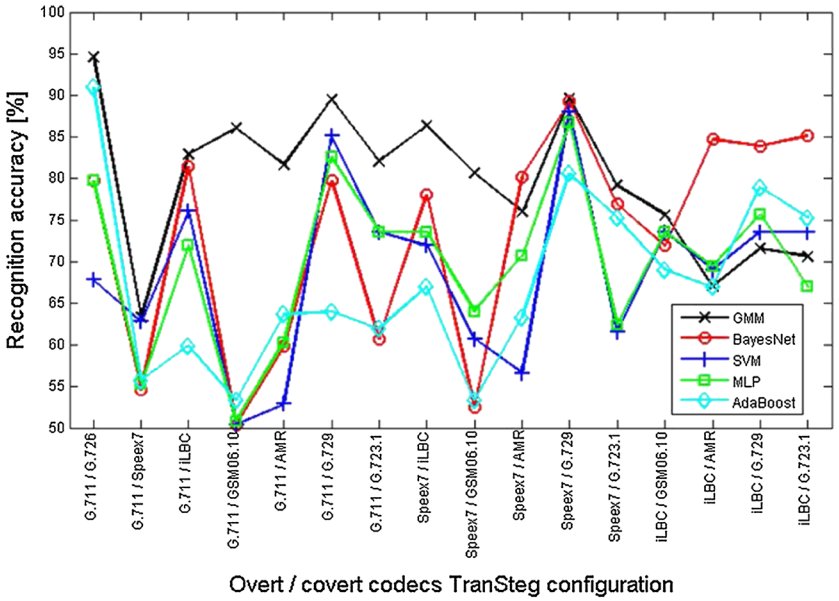

Figure 6. TranSteg detection accuracy for various codec pairs and the four best classifiers (GMM results added from [8]).

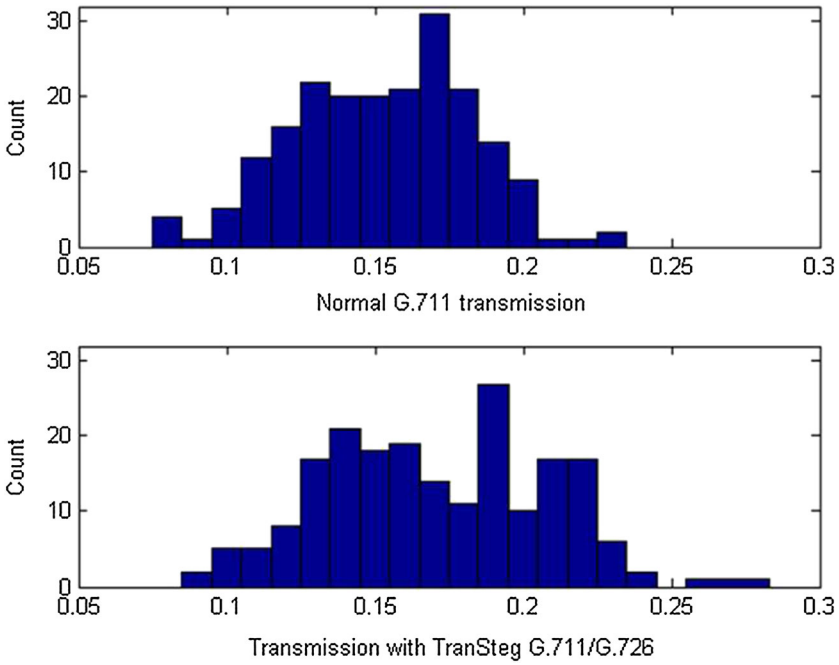

Figure 7. Comparison of histograms for one of the features (10th bin of MFCC3 histogram) selected for TranSteg detection. Top: normal transmission, bottom: TranSteg with G.711/G.726 pair used.

4.4. Discussion

The proposed steganalysis method proved to be able to detect TranSteg in certain circumstances. Figure 7 shows the distributions of one of the features from the feature vectors used for detection of abnormal transmission—it shows clearly that such differences in distributions can be in favor of TranSteg detectability.

Usually, the accuracy achieved using the proposed method was inferior to the method proposed earlier; however, there were some exceptions: the proposed method yielded higher detection accuracy (around 85%) for pairs with iLBC as the overt coded. Therefore, we have to re-assess the undetectability of these pairs and classify them as pairs with high detectability risks.

On the basis of undetectability and drawing conclusions from the current and previous studies [8], for effective and secure TranSteg, we recommend three pairs: G.711/Speex7, iLBC/GSM06.10, and G.711/G.711.0 (not analyzed in the current experiment because, as already mentioned, it is lossless, thus impossible to detect based on the output signal, i.e., at W3 point in the network).

If other scenarios were considered (e.g., S1, S2, or S3), the detectability of TranSteg would be even lower. In S4 (the worst-case scenario mostly examined in this study), we deal with triple transcoding, causing more distinct changes in the speech spectrum. Scenarios S1–S3 involve one or two transcodings, so probability of the detection based on the output speech would be lower.

Following the results from our previous study [7], we have to bear in mind that using Speex and G.711.0 codecs can pose a risk of being detected using a payload first byte analyzer, on the condition that the warden has access to the W2 point in the network. However, we claim that this risk can be easily mitigated by randomizing initial bytes of the payload or simply by removing them.

If the overt transmission is realized using Speex7, AMR is the best choice for the covert codec, although with a quite high risk of detection if a warden uses either the steganalysis method proposed in this study (with the BayesNet or MLP classifiers) or the GMM-based detector proposed in [8].

5. CONCLUSIONS AND FUTURE WORK

Transcoding Steganography is a novel steganographic technique that is applicable to multimedia services, such as VoIP. In this work, we discussed the problem of undetectability of this method. We proposed another method of steganalysis, based on the analysis of histograms of mel-cepstral parameters (MFCCs) and various binary classifiers, such as Bayesian networks or SVMs.

We ran several experiments and described their results. We also summarized the conclusions from previous studies and combined them with the outcomes of the current work. As a result, we recommend three codec pairs as the ones with the highest undetectability: G.711/Speex7, iLBC/GSM06.10, and G.711/G.711.0. It should be noted, however, that Speex7 and G.711.0 require payload modification to avoid easy detection based on the initial byte(s). We also believe that combined GMM-based and BayesNet-based classifiers can be effective in detecting TranSteg if 'unsecure' pairs of codec are used. This combined detection approach will be further pursued in future work.

REFERENCES

1. Lubacz J, Mazurczyk W, Szczypiorski K. 2014. Principles and overview of network steganography, IEEE Communication Magazine, vol. 52, no. 5, May 2014
2. Rosenberg J, Schulzrinne H, Camarillo G, Johnston A. 2002. SIP: Session Initiation Protocol. IETF, RFC 3261, June
3. Schulzrinne H, Casner S, Frederick R, Jacobson V. 2003. RTP: a transport protocol for real-time applications. IETF, RFC 3550, July
4. Mazurczyk W. VoIP steganography and its detection—a survey, ACM Computing Surveys 2014; **46**(2): 1–21.
5. Mazurczyk W, Szaga P, Szczypiorski K. Using transcoding for hidden communication in IP telephony. In: *Multimedia Tools and Applications* 2012; **70**(3): 2139–2165. DOI: 10.1007/s11042-012-1224-8, June 2014.
6. Janicki A, Mazurczyk W, Szczypiorski S. Influence of speech codecs selection on transcoding steganography. *Telecommunication Systems: Modelling, Analysis, Design and Management* 2015. doi:10.1007/s11235-014-9937-9.
7. Janicki A, Mazurczyk W, Szczypiorski K. Evaluation of efficiency of transcoding steganography. *Journal of Homeland Security and Emergency Management* 2014; **11**(4): 555–578. DOI: 10.1515/jhsem-2014-0028. October 2014
8. Janicki A, Mazurczyk W, Szczypiorski K. 2014 Steganalysis of transcoding steganography. *Annals of Telecommunications* 2014; **69**(7): 449–460. DOI: 10.1007/s12243-013-0385-4
9. Pevny T, Fridrich J. 2008. Detection of double-compression in JPEG images for applications in steganography. *IEEE Transactions on Information Forensics and Security* 2008; **3**(2): 247–258
10. Wang J, Liu G, Dai Y, Wang Z. Detecting JPEG image forgery based on double compression. *Journal of Systems Engineering and Electronics* 2009; **20**(5): 1096–1103.
11. Liu Q, Sung A, Qiao M. Detection of double MP3 compression. *Cogn Comput* 2010; **2**:291–296.
12. Luo D, Luo W, Yang R, Huang J. 2012 Compression history identification for digital audio signal, In Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)
13. Xu J, Su Y, You X. 2012. Detection of video transcoding for digital forensics, audio, language and image processing (ICALIP), 2012 International Conference on, vol., no., pp.160,164, 16–18 July 2012
14. Wang W, Farid H. 2006. *Exposing Digital Forgeries in Video by Detecting Double MPEG Compression*, MM&Sec'06, 26–27 September 2006, Geneva: Switzerland.
15. Dittmann J, Hesse D, Hillert R. 2005. Steganography and steganalysis in voice-over IP scenarios: operational aspects and first experiences with a new steganalysis tool set. In: Proc SPIE, Vol 5681,

Security, Steganography, and Watermarking of Multimedia Contents VII, San Jose, 607–618.
16. Takahashi T, Lee W. 2007. An assessment of VoIP covert channel threats. In: Proc 3rd Int Conf Security and Privacy in Communication Networks (SecureComm 2007), Nice, France, 371–380.
17. Kräetzer C, Dittmann J. 2007. Mel-cepstrum based steganalysis for VoIP-steganography. In: Proc. of the 19th Annual Symposium of the Electronic Imaging Science and Technology, SPIE and IS&T, San Jose, CA, USA, February 2007
18. Kräetzer C, Dittmann J. Pros and Cons of mel-cepstrum based audio steganalysis using SVM classification. *Lecture Notes on Computer Science, LNCS* 2008; **4567**:359–377.
19. Kräetzer C, Dittmann J. 2008. Cover signal specific steganalysis: the impact of training on the example of two selected audio steganalysis approaches. In: Proc. of SPIE-IS&T Electronic Imaging, SPIE 6819
20. Huang Y, Tang S, Zhang Y. Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis. *IET Communications* 2011; **5**(7):929–936.
21. Huang Y, Tang S, Bao C, Yip YJ. Steganalysis of compressed speech to detect covert voice over Internet protocol channels. *IET Information Security* 2011; **5**(1):26–32.
22. Garateguy G, Arce G, Pelaez J. 2011. Covert channel detection in VoIP streams. In: Proc. of 45th Annual Conference on Information Sciences and Systems (CISS), March 2011, 1–6.
23. Arackaparambil C, Yan G, Bratus S, Caglayan A. 2012. On tuning the knobs of distribution-based methods for detecting VoIP covert channels. In: Proc. of Hawaii International Conference on System Sciences (HICSS-45), Hawaii, January 2012
24. Li S, Huang Y. Detection of QIM steganography in G.723.1 bit stream based on quantization index sequence analysis. *Journal of Zhejiang University Science C (Computers & Electronics)* 2012; **13**(8):624–634.
25. Fisk G, Fisk M, Papadopoulos C, Neil J. Eliminating steganography in Internet traffic with active wardens, 5th International Workshop on Information Hiding. *Lecture Notes in Computer Science* 2002; **2578**: 18–35.
26. Janicki A, Staroszczyk T. Speaker Recognition from Coded Speech Using Support Vector Machines, Text, Speech and Dialogue / Habernal Ivan, Matoušek Václav (red.), LNCS 6836, 2011, Springer Berlin Heidelberg, 2011; 291–298.
27. Garofolo J, Lamel L, Fisher W, Fiscus J, Pallett D, Dahlgren N, *et al. TIMIT Acoustic–Phonetic Continuous Speech Corpus*. Linguistic Data Consortium: Philadelphia, 1993.
28. Kabal P. *TSP Speech Database, Tech Rep, Department of Electrical & Computer Engineering*. McGill University, Montreal, Quebec, Canada, 2002.
29. Cummins F, Grimaldi M, Leonard T, Simko J. The CHAINS corpus: CHAracterizing INdividual Speakers. In *Proc of SPECOM' 06*. St Petersburg, Russia, 2006; 431–435.
30. Grocholewski S. *CORPORA—Speech Database for Polish Diphones, 5th European Conference on Speech Communication and Technology Eurospeech' 97*. Rhodes, Greece, 1997.
31. Ortega García J, González Rodríguez J, Marrero-Aguiar V. AHUMADA: a large speech corpus in Spanish for speaker characterization and identification. *Speech Communication* 2000; **31**:255–264.
32. Pearl J. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann: San Francisco, California, 1988.
33. Vapnik V. *The Nature of Statistical Learning Theory*. Springer-Verlag: New York, NY, 1995.
34. Górska Z, Janicki A. Recognition of extraversion level based on handwriting and support vector machines. *Perceptual and Motor Skills, AmSci* 2012; **114**(3): 857–869.
35. Rybka J, Janicki A. Comparison of speaker dependent and speaker independent emotion recognition. *International Journal of Applied Mathematics & Computer Science* 2013; **23**(4):797–808.
36. Freund Y, Schapire R. A short introduction to boosting. *Japanese Society for Artificial Intelligence* 1999; **14**(5):771–780.
37. Liu Y, Zheng YF. FS_SFS: a novel feature selection method for support vector machines. *Pattern Recognition* 2006; **39**:1333–1345.
38. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software: an update. *SIGKDD Explorations* 2009; **11**(1):10–18.