VoIP Steganography and Its Detection—A Survey

WOJCIECH MAZURCZYK, Warsaw University of Technology

Steganography is an ancient art that encompasses various techniques of information hiding, the aim of which is to embed secret information into a carrier message. Steganographic methods are usually aimed at hiding the very existence of the communication. Due to the rise in popularity of IP telephony, together with the large volume of data and variety of protocols involved, it is currently attracting the attention of the research community as a perfect carrier for steganographic purposes. This article is a first survey of the existing Voice over IP (VoIP) steganography methods and their countermeasures.

Categories and Subject Descriptors: C.2.0 [**Computer-Communication Networks**]: General, Security and Protection

General Terms: Security, Algorithms

Additional Key Words and Phrases: Information hiding, steganography, VoIP, IP telephony, survey

ACM Reference Format:
Mazurczyk, W. 2013. VoIP steganography and its detection—A survey. ACM Comput. Surv. 46, 2, Article 20 (November 2013), 21 pages.
DOI: http://dx.doi.org/10.1145/2543581.2543587

1. INTRODUCTION

IP telephony or Voice over IP (VoIP) is a real-time service that enables users to make phone calls through IP data networks. It is one of the most important services of IP-based networks and is impacting the entire telecommunications landscape.

An IP telephony connection consists of two phases in which certain types of traffic are exchanged between the calling parties: signalling and conversation phases. During the first phase, certain signalling protocol messages, such as Session Initiation Protocol (SIP) messages [Rosenberg et al. 2002], are exchanged between the caller and callee. These messages are intended to set up and negotiate the connection parameters between the calling parties. During the second phase, two audio streams are sent bidirectionally. Real-Time Transport Protocol (RTP) [Schulzrinne et al. 2003] is most often utilised for voice data transport, and thus packets that carry the voice payload are called *RTP packets*. The consecutive RTP packets form an RTP stream.

Steganography is an ancient art that encompasses various information-hiding techniques, whose aim is to embed a secret message into a carrier (steganogram). Steganographic methods are aimed at hiding the very existence of the communication and therefore keep any third-party observers unaware of the presence of the steganographic exchange. Steganographic carriers have evolved through the ages and are related to the evolution of the methods of communication between people. Thus, it is not

This work was supported by the Polish Ministry of Science and Higher Education and Polish National Science Centre under grants 0349/IP2/2011/71 and 2011/01/D/ST7/05054.

Author's address: W. Mazurczyk, Warsaw University of Technology; Institute of Telecommunications, Warsaw, Poland, 00-665, Nowowiejska 15/19.

DOI: http://dx.doi.org/10.1145/2543581.2543587

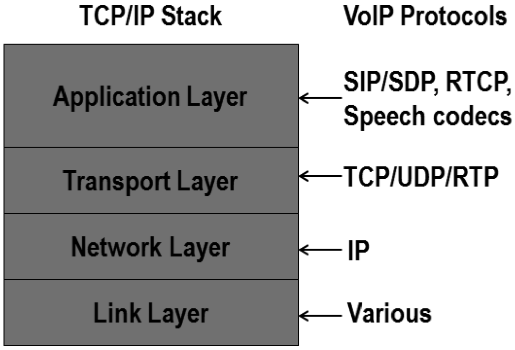

Fig. 1. VoIP stack and protocols.

surprising that current telecommunication networks are a natural target for steganography; in particular, IP telephony is attracting the attention of the steganography research community. It is because of the following features that IP telephony is a perfect carrier for steganographic purposes:

—It is very popular, thus its usage will not raise suspicions (i.e., it will not be considered as an anomaly itself).
—The large volume of VoIP data. The more frequent the presence and utilisation of such carriers in networks, the better their masking capacity, as hidden communications can pass unnoticed amongst the bulk of exchanged data.
—Potentially high steganographic bandwidth that can be achieved. For example, during the conversation phase of a G.711-based call, each RTP packet carries 20ms of voice; in this case, the RTP stream rate is 50 packets per second. Thus, even by simply hiding 1 bit in every RTP packet, we gain quite a high steganographic bandwidth of 50 bit/s.
—It involves the combined use of a variety of protocols. Thus, many opportunities for hiding information arise from the different layers of the TCP/IP stack. Hidden communication can be enabled by employing steganographic methods applied to the users' voice that is carried inside the RTP packets' payload, by utilising so-called well-known digital media steganography, or by utilising VoIP protocols as a steganographic carrier. This makes VoIP a multidimensional carrier.
—It is a real-time service, which induces additional strict requirements for steganographic methods and its detection (steganalysis) methods. It also simultaneously creates new opportunities for steganography (e.g., utilisation of excessively delayed packets that are discarded by the receiver without processing, because they cannot be considered for voice reconstruction).
—The VoIP calls are dynamic and of variable length, which makes VoIP-based steganography even harder to detect.

Presently, steganographic methods that can be utilised in telecommunication networks are jointly described by the term *network steganography* or, specifically when applied to IP telephony, by the terms *VoIP steganography* or *steganophony* [Lubacz et al. 2008]. These terms pertain to the techniques of hiding information in any layer of the TCP/IP protocol stack (Figure 1), including techniques applied to the speech codecs, or those that utilise the speech itself.

In general, steganography can be treated as a double-edged sword depending on who uses it and how. However, the ethical issues related to the utilisation of information-hiding techniques require consideration in a broader steganography context, which is

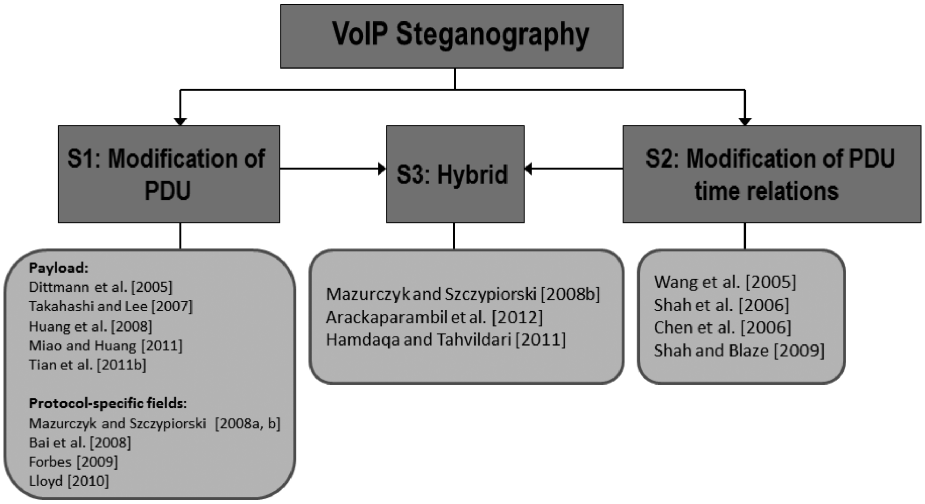

Fig. 2. VoIP steganography classification with exemplary methods.

beyond the scope of this article. The main application of network steganography is for providing a means of conducting clandestine communication. The purposes for establishing hidden communication can be varied; possible uses can fall into the category of legal or illicit activities. The illegal aspect of steganography starts from criminal communication, through confidential data exfiltration from guarded systems, cyber weapon exchange and control, up to industrial espionage. Legitimate uses include circumvention of Web censorship and surveillance, computer/network forensics, or copyright protection. Techniques can be used with VoIP to improve its resistance to packet losses and improve voice quality [Aoki 2003, 2004], extend communication bandwidth [Aoki 2007], or provide means for secure cryptographic key distribution [Huang et al. 2011a].

The expansion of TCP/IP networks opened up many possibilities for covert communication due to changes in the traditional circuit-switched networks paradigm; services/applications are created by the network users rather than by the network itself—the transport and control functions are not separated and can be influenced by the user. These possibilities are a consequence of the fact that network users can influence, and/or use the control of dataflow—the communication protocols—together with the service/application functionality of terminals to establish covert communication. That is why secret messages can be hidden, not only within ordinary noncovert (overt) messages as in traditional steganography and circuit-switched networks but also in the communication protocol's control elements, in effect by manipulation of the protocol's logic, or by the preceding combinations.

A considerable number of steganographic methods have been developed so far [Zander et al. 2007; Bender et al. 1996], and they cover all layers of the TCP/IP stack. Moreover, many of the previously proposed steganographic methods can be successfully adapted to VoIP traffic, and many new VoIP-specific methods have been introduced and are covered in this survey. However, if we look at VoIP from the perspective of traffic, then in general, every VoIP steganography technique can be classified into one of the three following groups (Figure 2) [Lubacz et al. 2008] based on what is used as the steganographic carrier:

—*S1:* Steganographic methods that modify the Protocol Data Unit (PDU)—network protocol headers or payload field. Examples of such solutions include (1) modifications of free/redundant headers' fields of IP, UDP, or RTP during conversation phase and (2) modification of signalling messages in, for example, SIP (using the same principle as in Murdoch and Lewis [2005]), or (3) modification of the RTP packets' payload by modifying the user content that it carries (e.g., by employing widely used Least Significant Bits (LSBs) modification method [Bender et al. 1996]), or simply by replacing user data [Mazurczyk et al. 2011]. Obviously, methods that rely on modification of both the header fields and payload are also possible.

—*S2:* Steganographic methods that modify PDU time relations, for example, by modifying PDU interpacket delay [Wang et al. 2005; Shah et al. 2006], by affecting the sequence order of PDUs (similarly like in Kundur and Ahsan [2003]), or by introducing intentional PDU losses (by adopting the solution from Servetto and Vetterli [2001]).

—*S3:* Hybrid steganographic methods that modify both the content of PDUs and their time relations. An example of such a solution is the Lost Audio Packets Stenography (LACK) method [Mazurczyk and Szczypiorski 2008b], which will be described in Section 3.

This classification quite precisely describes current information-hiding possibilities in VoIP. However, it must be noted that the main focus of the research community in the past decade was dedicated to the methods from the S1 group. Solutions from the S2 group are harder to deploy practically, because they usually offer low steganographic bandwidth and require synchronisation. Moreover, they can significantly degrade the quality of the call. Methods from the S3 group are rather recent discoveries, thus to date, there have not been many papers on this subject.

It must also be noted that there are a number of information-hiding techniques proposed for certain codecs, or audio files that could be also utilised for VoIP (for review of audio steganography methods, interested readers are referred to the book edited by Lu [2005]). In addition, there are a number of other existing steganographic methods (e.g., protocol steganography) that can be potentially utilized in the VoIP environment. However, many of these have not been evaluated for VoIP per se; thus, they are intentionally omitted. Therefore, it must be emphasised that in this article, we focus *only* on these existing steganographic and detection methods that have been proved feasible for VoIP.

This article is a first review of the practical, existing VoIP steganography methods and countermeasures available during the period from 2003 to 2012. The rest of the article is organised as follows. Section 2 introduces the fundamentals of network steganography and VoIP steganography in particular. Section 3 describes the existing steganographic techniques, whilst Section 4 presents possible countermeasures. Finally, Section 5 concludes this work.

2. STEGANOGRAPHY: TERMINOLOGY, COMMUNICATION MODEL, AND SCENARIOS

As mentioned in Section 1, the means of communication amongst people has evolved over time; from messengers, letters (e.g., where sympathetic ink was used to facilitate hidden data exchange), and telephones to computer networks; simultaneously, steganographic methods have evolved as well. Therefore, techniques of hiding information that utilise network protocols can be viewed as an evolutionary step of the hidden data carrier rather than some new phenomenon.

The scientific community has been using many terms, such as *steganography* [Petitcolas et al. 1999; Fridrich 2010], *covert channels* [Lampson 1973; DoD "Orange Book" 1985], or *information hiding* [Petitcolas et al. 1999], to describe the process of concealing information in the digital environment. This stems from the fact that the terms have not been introduced at the same time and because their definitions have

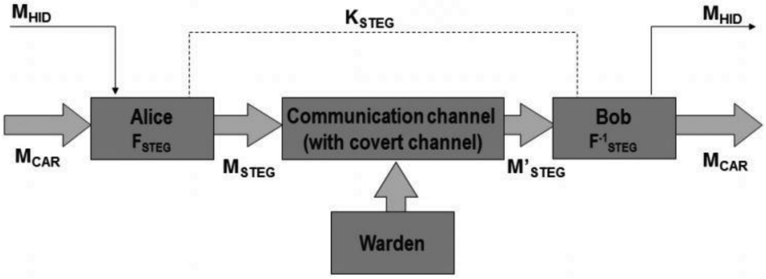

Fig. 3. Model for hidden communication.

evolved. However, it is our belief that drawing a distinction between steganography and covert channels in telecommunication networks environment is not necessary and that instead, one term—*network steganography*—should be used. We believe that in this case, steganographic methods are used to create a covert (steganographic) channel. Thus, the scope of network steganography encompasses all information-hiding techniques that can be applied in telecommunication networks to enable hidden data exchange.

The best example to prove this point is IP telephony services. Due to its nature, information hiding is possible within the speech that is carried inside the voice packets and in the modification of the protocols that enables it. The carrier is actually the same; the VoIP service and the methods used in both cases are inseparably bound to the process of communicating through the network. That is why, when considering steganography applied to the VoIP service, terms such as *VoIP steganography* or *steganophony* should be used.

Another argument against using two separate terms is that in the state of the art for steganography [Petitcolas et al. 1999], as well as for covert channels [Zander et al. 2007], the same model is referenced as a hidden communication model. It is the famous "prisoners' problem," which was first formulated by Simmons [1984] in 1983 (Figure 3). In this model, two prisoners—Alice and Bob—are jailed in separate cells and are trying to prepare an escape plan. The problem is that their communication is always passed through and inspected by the warden. If the warden identifies any conspiracy, he will put them into solitary confinement, so Alice and Bob must find a way to exchange hidden messages for their escape plan to succeed. The solution is to use steganography. By concealing a hidden message (M_{HID}) in an innocent-looking carrier (M_{CAR}), it is possible to achieve a modified carrier (M_{STEG}) that will raise no suspicions while traversing the communication channel. Thus, to be able to create a covert channel in which hidden data is exchanged, the utilisation of a steganographic method by the sender (F_{STEG} in Figure 3) and by the receiver ($F^{-1}{}_{STEG}$ in Figure 3) is always necessary. For Alice and Bob, the communication channel is also a covert channel that was created using a given steganographic method.

The similarity between the terms *steganography* and *covert channel* was also observed by other researchers in the field (e.g., by Fridrich [2010]).

For steganographic methods, it is usually assumed that there exists a secret stego-key (K_{STEG} in Figure 3) that is a form of shared secret between the communicating parties and allows building of the security of the steganographic system. In network steganography, knowledge of how the information-hiding technique operates can be also treated as a stego-key. Therefore, everything else related to the steganogram transmission can be known to the warden—the entity that performs detection (steganalysis). In particular, it:

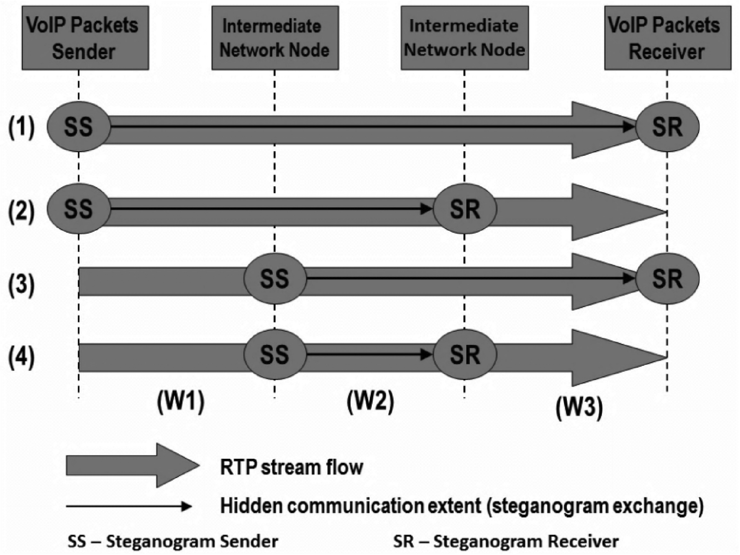
VoIP Packets Sender
Intermediate Network Node
Intermediate Network Node
VoIP Packets Receiver
(1)
(2)
(3)
(4)
SS
SR
(W1)
(W2)
(W3)
RTP stream flow
Hidden communication extent (steganogram exchange)
SS – Steganogram Sender
SR – Steganogram Receiver

Fig. 4. Model for hidden communication (based on Lucena et al. [2004] and Zander et al. [2007]).

—is aware that Alice and Bob can be utilising hidden communication to exchange data in a covert manner.
—has a knowledge of all existing steganographic methods, but not of the one used by Alice and Bob (this, as mentioned earlier, is assumed to be their stego-key).
—is able to try to detect and/or interrupt the hidden communication.

For VoIP steganography, four possible hidden communication scenarios may be considered, as illustrated in Figure 4. The first scenario (1 in Figure 4) is the most common: the sender and the receiver perform VoIP conversation while simultaneously exchanging steganograms. The conversation path is the same as the hidden path. For the next three scenarios (2–4 in Figure 4), only a part of the VoIP end-to-end path is used for the hidden communication as a result of actions undertaken by intermediate nodes; the sender and receiver are, in principle, unaware of the steganographic data exchange. The possible localisations of the warden are denoted as W1 through W3 and will be discussed in detail in Section 4.

Every network steganographic method can be described typically by the following set of characteristics: its steganographic bandwidth (also referred to as capacity typically for media steganography), its undetectability (also referred to as security in Fridrich [2010]), and robustness. *Steganographic bandwidth* refers to the amount of secret data that can be sent per unit of time when using a particular method. *Undetectability* is defined as the inability to detect a steganogram within a certain carrier. The most popular way to detect a steganogram is to analyse the statistical properties of the captured data and compare them with the typical values for that carrier. The last characteristic is *robustness*, which is defined as the amount of alteration that the steganogram can withstand without secret data being destroyed.

Of course, a good steganographic method should be as robust and hard to detect as possible while offering the highest bandwidth. However, it must be noted that there is always a necessary fundamental trade-off among these three measures.

Additionally, it is useful to measure the steganographic cost. It is a characteristic that belongs to the sphere of carrier fidelity and has direct impact on undetectability. It describes the degradation or distortion of the carrier caused by the application of the steganographic method, similar to the manner in which Mean-Square Error (MSE) or Peak Signal-to-Noise Ratio (PSNR) were utilized for digital media carriers. In the case of VoIP steganography methods, this cost can be expressed, for example, by providing a measure of the conversation quality degradation induced by applying a particular information-hiding technique.

3. VOIP STEGANOGRAPHY

IP telephony as a hidden data carrier can be considered a fairly recent discovery. However, existing VoIP steganographic methods stem from two distinct research origins.

The first is the well-established image, audio, and video file steganography (also referred to as digital media steganography) [Bender et al. 1996], which has given rise to methods that target the digital representation of the transmitted voice as the carrier for hidden data.

The second sphere of influence is the so-called covert channels (see Section 2) created in different network protocols [Kundur and Ahsan 2003; Murdoch and Lewis 2005] (a good survey on covert channels can be found in Zander et al. [2007]), also referred to in the literature as protocol steganography [Lucena et al. 2004]; these solutions target specific VoIP protocol fields (e.g., signalling protocol—SIP, transport protocol—RTP, or control protocol—Real-time Transport Control Protocol [RTCP]) or their behaviours.

In the following subsections we analyse VoIP steganography methods, taking into account the two research origins mentioned previously. In each case, chronology is retained.

3.1. Steganographic Methods Applied to Voice Payload

All information-hiding methods described in this subsection belong generally to the group S1 from classification presented in Figure 2 (modifying the payload of the VoIP PDUs).

3.1.1. Least Significant Bit–Based Methods Applied to Digital Voice Signal. Surprisingly, a lot of research effort is still devoted for improving LSB steganography. However, it must be noted that some of the techniques described next can be successfully applied also to other VoIP steganography methods to increase their undetectability, robustness, or steganographic bandwidth.

The first VoIP steganographic method to utilise the digital voice signal as a hidden data carrier was proposed by Aoki in 2003. The LSB steganography was utilised to provide the Packet Loss Concealment (PLC) method for G.711-based VoIP. Later, the PLC method was further improved [Aoki 2004, 2007], but the steganographic method and voice codec remained the same.

In 2005, Dittmann et al. presented the first VoIP steganography implementation prototype that also used the LSB method. This work was further extended and published in 2006, when Krätzer et al. demonstrated testing that proved a typical VoIP communication can be practically used for steganographic purposes.

Wu and Yang [2006] described the scheme of adaptive LSB, which for G.711-based speech calculates energy statistics to estimate the number of LSBs to be utilised as a hidden data carrier in each voice sample. The results proved that this approach performs better than simple LSB and offers a higher steganographic bandwidth (about 20Kbps) while introducing less degradation of the voice quality.

Druid [2007] described an implementation of SteganRTP. This tool employs the LSBs of the G.711 codec to carry steganograms and offers a reliable bidirectional covert communication channel that allows to exchange 1kB/s of secret data in single direction.

Wang and Wu [2007] also suggested using the LSBs of voice samples to carry secret communication; however, in their solution, the bits of the steganogram were coded using a low rate voice codec, such as Speex. Their prototype implementation is characterised by a small processing delay of about 0.257ms.

Takahashi and Lee [2007] presented a proof of concept LSB-based implementation—Voice over VoIP (Vo^{2}IP)—which is able to establish a hidden communication by embedding 8 kbit/s G.729-based compressed voice data into the regular Pulse Code Modulation (PCM)-based voice traffic.

Liu et al. [2008] found that the LSBs of each speech frame for G.729 can be replaced with secret data bits. The experimental results indicate that the method is perceptually transparent, whereas the steganographic bandwidth is relatively high (about 200 bit/s).

Huang et al. [2008] described how to use the LSB matching method to enable covert communication for VoIP. They developed a G.711-based prototype named Stega-Talk, which also offers the recovery of secret data lost due to network conditions. It is based on the Redundant Audio Data RTP payload type that is typically used to carry a Dual-Tone Multifrequency (DTMF) digit.

Tian et al. [2008] proposed the use of an LSB steganography-based system that employs a well-balanced and simple encryption of secret data. This system was evaluated for VoIP with G.729a speech coding using a proof-of-concept tool named StegTalk. The experimental results showed that the achievable steganographic bandwidth is in the range of 0.8 to 2.6 kbit/s and has a negligible effect on speech quality. Moreover, it met the real-time requirements of the VoIP service.

A real-time steganography system for VoIP was described by Tian et al. [2009a]. The main novelty of the proposed solution is not in the steganographic method used (LSB), but in utilising M-sequence encryption techniques to eliminate the correlation amongst secret messages, to increase the resistance against statistical steganalysis. Moreover, protocol steganography (usage of free/unused fields in protocols headers) is used to provide a novel synchronisation mechanism together with an RSA-based key agreement that ensures accurate restitution of the secret messages on the receiver side. This system was experimentally evaluated for 0.8 and 2.6 kbit/s steganographic bandwidth, which obtained a 0.3 and 1 quality drop in the Mean Opinion Score (MOS) scale (which is typically used for expressing the quality of VoIP calls), respectively, and the total embedding latency increased by about 4.7ms when 1MB of steganogram was transmitted. A similar LSB-based approach, relying on Adaptive VoIP Steganography (AVIS), was presented by the same authors [Tian et al. 2009b].

Xu and Yang [2009] proposed an LSB-based method dedicated to voice transmission using the G.723.1 codec in 5.3 kbit/s mode. They identified five LSBs of the Line Spectrum Pair Vector Quantization (LSP VQ) indices and used them to transmit hidden data; the method provided a steganographic bandwidth of 133.3 bit/s.

Tian et al. [2010] described a Dynamic Matrix Encoding Strategy (DMES) that dynamically chooses the size of each message group in a given set of adoptable message sizes. The purpose of DMES is to flexibly adjust the steganographic bandwidth and embedding transparency in accordance with the requirements of the users. Its main advantage is that it is also codec and cover independent.

An adaptive steganography scheme based on the smoothness of the speech block was introduced by Miao and Huang [2011]. By choosing lower embedding rates in the flat blocks and higher in sharp ones, the security of the method was improved. Such an approach outperforms the classic LSB-based methods in terms of voice quality. Experimental results showed that about 7.5 kbit/s of secret data can be sent covertly with a degradation of voice quality of less than 0.5 (expressed again on the MOS scale).

Tian et al. [2011a] presented an Adaptive Partial-Matching Steganography (APMS) approach. They introduced a Partial Similarity Value (PSV) measure to evaluate the partial matching between covert and overt messages. This allows an adaptive balance of the steganographic transparency and bandwidth. Additionally, they utilise triple

M-sequences to eliminate the correlation among secret messages, guide the adaptive embedding process, and encrypt synchronisation signalling patterns.

An insightful overview of the general techniques that can be applied to VoIP steganography methods to make their detection even more difficult was introduced by Tian et al. [2011b]. Additionally, they proposed three new encoding strategies based on digital logic. All techniques were evaluated for LSB-based steganography and proved to be effective.

Another adaptive LSB-based steganography approach—AVIS—was proposed by Xu et al. [2011]. AVIS has two components: Value-based Multiple Insertion (VAMI), which is responsible for dynamically selecting multiple bits based on the VoIP vector value, and Voice Activity Detection Dynamic Insertion (VADDI), which dynamically changes the embedding intervals to make detection harder. The approach was implemented for G.711-based VoIP, and the results prove that it is less detectable than a classic LSB method whilst achieving a steganographic bandwidth of about 114B/s, introducing acceptable delay and degrading the voice from 0.1 to 0.4 on the MOS scale.

Liu et al. [2012] adopted least significant digits rather than LSBs to hide secret data. This approach can increase around 30% of steganographic bandwidth while introducing lower steganographic cost than the classic LSB method.

3.1.2. Other Methods. Besides LSB-based information-hiding approaches, other effective methods that utilize speech as a secret data carrier have been proposed. These solutions are generally based on the following:

- Phase coding [Takahashi and Lee 2007; Nutzinger and Wurzer 2011]
- Quantization Index Modulation (QIM) technique [Xiao et al. 2008]
- Spectrum techniques in transform domain [Takahashi and Lee 2007; Nutzinger et al. 2010]
- Echo-hiding technique [Takahashi and Lee 2007]
- Analysis by synthesis (ABS)-based scheme [Ma et al. 2007; Wu et al. 2009]
- Discrete Fourier Transform (DFT) technique [Deng et al. 2008]
- Speech codec-specific approaches [Aoki 2008, 2009, 2010; Geiser et al. 2008; Nishimura 2009; Huang et al. 2011b]

Besides the LSB method, in 2007 Takahashi and Lee considered the feasibility of other methods that can be utilised in VoIP steganography, such as the Direct Sequence Spread Spectrum (DSSS), Frequency-Hopping Spread Spectrum (FHSS), and echo hiding. They also provided experimental results of steganographic bandwidth, which was about 20 bit/s for all three techniques, while still maintaining good voice quality robustness. Nutzinger et al. [2010] further improved DSSS by creating a hybrid steganographic algorithm that combined DSSS with frequency hopping and bit-rate variation. These authors also implemented a prototype and stated that the proposed scheme has none to negligible influence on the speech quality.

Ma et al. [2007] adopted an ABS algorithm [Wu et al. 2003] to enable hiding of 2.4 kbit/s Mixed-Excitation Linear Prediction (MELP) speech in G.721-based overt speech. The obtained steganographic bandwidth was estimated up to 8 kbit/s. In addition, ABS was utilized by Wu et al. [2009] to build an information-hiding scheme based on Linear Predictive Coefficients (LPCs) as the secret data carrier by means of LPC substitution. The experimental results for four speech codecs—G.721, GSM, G.728, and G.729—proved that the proposed approach is characterised by high steganographic bandwidth (from 800 to 3,600 bit/s) while offering good undetectability, robustness, and ability to perform in real time. The results were also compared with four traditional information-hiding technologies—LSB, echo hiding, phase encoding, and spectrum transform—and were shown to be superior.

A steganographic method based on the characteristics of PCM μ-law (PCMU), in which the 0-th speech sample can be represented by two codes due to the overlap (namely +0 and −0), was proposed by Aoki [2008]. This redundancy is then exploited to embed hidden data into the speech. The technique is lossless; thus, there is no voice degradation. The obtained steganographic bandwidth of this method was estimated to be in the range of 4.4 to 24 kbit/s, depending on the extent of the background noise level. The work was later extended and published in 2009 by Aoki, who studied lossless steganography techniques for G.711 PCMU and DVI Adaptive Differential Pulse Code Modulation (DVI-ADPCM) codecs. The proposed techniques also exploit the redundancy of the folded binary code employed in these codecs for embedding a steganogram into the speech without quality degradation. The improved technique offers steganographic bandwidth for G.711-based calls from 24 to 400 bit/s and from 0 to 8 bit/s for ADPCM-based calls, again depending on the background noise level. In addition, the semilossless variant of this technique for increased steganographic bandwidth was considered in 2010 by Aoki.

The PLC algorithm intended mainly for wireless VoIP systems, which rely on the specific side information that is communicated via a covert channel within the speech, was presented by Geiser et al. [2008]. To enable hidden transmission, the Algebraic Code-Excited Linear Prediction (ACELP) codebook (or Fixed Codebook [FCB]) is partitioned into subcodebooks that uniquely identify the selected secret message's bits. Prototype Adaptive Multirate (AMR)-based implementation was also described, and a 2 kbit/s steganographic bandwidth was achieved experimentally.

An interesting algorithm for QIM that can be utilised for low bit-rate speech streams was introduced by Xiao et al. [2008] and later named Complementary Neighbour Vertices-Quantisation Index Modulation (CNV-QIM) [Li and Huang 2012]. It is based on dividing the codebook into two parts, each representing 0 and 1, respectively. Moreover, the relationship between codewords is considered using the Complementary Neighbour Vertices (CNV) algorithm. This guarantees that every codeword is in the opposite part to its nearest neighbour, thus giving a bound of distortion. Experiments for Internet Low Bit Rate Codec (iLBC) and G.723.1 speech codecs proved that the proposed method is effective, as it only slightly decreases the speech quality whilst providing a steganographic bandwidth of 100 bit/s.

Deng et al. [2008] proposed a concept of Covert Speech Telephone (CST) that is intended to provide secure covert voice communication. This approach utilises a robust DFT watermarking scheme to hide covert speech in the G.711-based VoIP stream. The main innovation is to use speech recognition to efficiently reduce the size of the secret information, encrypt it, and hide it in the existing overt VoIP call.

An interesting study is described in Nishimura [2009], where the author proposed hiding information in the AMR-coded stream by using an extended quantisation-based method of pitch delay (one of the AMR codec parameters). This additional data transmission channel was used to extend the audio bandwidth from narrow-band (0.3 to 3.4kHz) to wide-band (0.3 to 7.5kHz).

Nutzinger and Wurzer [2011] introduced a novel approach to speech phase coding. Typically, the original phase values are replaced with some random data. In the proposed scheme, the original phase values are preserved to retain high voice quality. The algorithm embeds secret data by introducing a configurable phase difference between the mean of selected chunks of the phase spectrum. Experimental results prove that the proposed method can achieve up to 12.5 bit/s while introducing almost no degradation on the voice signal, good robustness, and security.

A high-capacity steganography technique based on the utilisation of inactive frames of G.723.1 speech codec was introduced by Huang et al. [2011b]. The authors proved that the inactive frames of VoIP streams are more suitable for data embedding than the active ones, thus more hidden data can be embedded in them with the same imperceptibility. They then proposed a steganographic algorithm in different speech parameters of the inactive frames for G.72k3.1 codec with 6.3 kbit/s bit rate. Experimental results show that the solution is imperceptible and that a high steganographic bandwidth up to 101 bits per frame is achieved.

Table I summarises the most important characteristics of the steganographic methods outlined in Section 3.1.

3.2. Steganographic Methods Applied to VoIP-Specific Protocols

3.2.1. Methods That Modify Protocol Data Unit Time Relations (S2). Utilisation of VoIP-specific protocols as a steganogram carrier was first presented by Wang et al. [2005] (later also described in Chen et al. [2006]). The authors proposed embedding a 24-bit watermark into the encrypted stream (e.g., Skype call) to track its propagation through the network, thus providing its de-anonymisation. The watermark is inserted by modifying the interpacket delay for selected packets in the VoIP stream. Authors demonstrated that depending on the watermark parameters chosen, they were able to achieve a 99% true-positive and 0% false-positive rate while maintaining good robustness and undetectability. However, they achieved steganographic bandwidth of only about 0.3 bit/s, which is enough for the described application but rather low to perform clandestine communication.

Shah et al. [2006] inspected the use of injected jitter into VoIP packets to create a covert channel, which is intended to exfiltrate users' keyboard activity (e.g., authentication credentials). The authors proved that such an attack is feasible even when the VoIP stream is encrypted.

A new kind of information-hiding technique called *interference channel* was introduced by Shah and Blaze [2009], which creates external interference on a shared communications medium (e.g., wireless network) in order to send hidden data. They describe an implementation of a wireless interference channel for 802.11 networks that is able to successfully transfer a steganogram over data streams (with a rather low steganographic bandwidth of 1 bit per 2.5s of the call) and that is proven to be especially well suited for VoIP streams.

3.2.2. Methods That Modify Protocol Data Unit–Protocol-Specific Fields (S1). Mazurczyk and Kotulski [2006a] proposed the use of steganography in unused fields in the RTP protocol headers and digital watermarking to embed additional information into RTP traffic to provide origin authentication and content integrity. The necessary information was embedded into unused fields in the IP, UDP, and RTP headers, and also into the transmitted voice. The authors [2006b] later further enhanced their scheme by also incorporating an RTCP functionality without the need to use a separate protocol, thus saving the bandwidth utilised by the VoIP connection.

A broader view on network steganography methods that can be applied to VoIP, to its signalling protocol—SIP with Session Description Protocol (SDP) [2008a]—and to its RTP streams (also with RTCP) was presented by Mazurczyk and Szczypiorski [2008b]. They discovered that a combination of information hiding solutions provides a capacity to covertly transfer about 2,000 bits during the signalling phase of a connection and about 2.5kbit/s during the conversation phase. In 2010, Lloyd [2010] extended [Mazurczyk and Szczypiorski 2008a] by introducing further steganographic methods for SIP and SDP protocols and by performing real-life experiments to verify whether they are feasible.

Bai et al. [2008] proposed a covert channel based on the jitter field of the RTCP header. This is performed in two stages: firstly, statistics of the value of the jitter field in the current network are calculated. Then, the secret message is modulated into

Table I. Summary of the VoIP Steganography Methods Applied to Voice Payload

	New concept for VoIP environment	New application for VoIP	Extension of existing VoIP method	Improvement of:		
				Undetectability	Robustness	Steganographic bandwidth
Wang and Wu [2007]	+	–	+	N/A	N/A	N/A
Dittmann et al. [2005], Aoki [2004, 2007]	–	–	+	N/A	N/A	N/A
Aoki [2008], Deng et al. [2008], Xiao et al. [2008], Ma et al. [2007], Xu and Yang [2009], Huang et al. [2011b]	+	–	–	N/A	N/A	N/A
Aoki [2003], Geiser et al. [2008], Nishimura [2009]	+	+	–	N/A	N/A	N/A
Takahashi and Lee [2007]	+	+	+	–	–	–
Tian et al. [2008], Tian et al. [2009a, 2009b], Miao and Huang [2011], Tian et al. [2011a], Tian et al. [2011b]	–	–	+	+	–	–
Tian et al. [2010]	–	–	+	+	–	+
Liu et al. [2008]	+	–	+	+	–	+
Wu and Yang [2006], Huang et al. [2008], Xu et al. [2011], Liu et al. [2012]	–	–	+	+	–	+
Aoki [2008, 2010]	–	–	+	–	–	+
Wu et al. [2009]	–	–	+	+	+	–
Nutzinger and Wurzer [2011]	+	–	+	+	+	–

the jitter field according to the previously calculated parameters. Utilisation of such modulation guarantees that the characteristics of the covert channel are similar to those of the overt channel.

Forbes [2009] proposed an RTP-based steganographic method that modifies the timestamp value of the RTP header to send steganograms. The method's theoretical maximum steganographic bandwidth is 350 bit/s.

Real-life experiments with VoIP steganography on a Session Border Controller (SBC), which was acting as a gatekeeper at the borders of trust, were performed by Wieser and Röning [2010]. They were trying to establish whether SBC has some countermeasures against information-hiding techniques based on SIP and RTP protocols. The results showed that it was possible to achieve a high steganographic bandwidth of even up to 569 kB/s.

Huang et al. [2011a] described how to provide efficient cryptographic key distribution in a VoIP environment for covert communication. Their proposed steganographic method is based on utilisation of the NTP field of the RTCP's Sender Report (SR) as a hidden data carrier and offered steganographic bandwidth of 54 bit/s with good undetectability.

The Transcoding Steganography (TranSteg) method, which relies on the compression of the overt data in a payload field of RTP packets in order to make free space for a steganogram, was introduced by Mazurczyk et al. [2011]. In TranSteg for a chosen voice stream, a codec that will result in a similar voice quality, although for a smaller voice payload size than the original, is found. Then, the voice stream is transcoded. At this stage, the original voice payload size is intentionally unaltered, and the change of the codec is not indicated. Instead, after placing the transcoded voice payload, the remaining free space is filled with hidden data. The resulting steganographic bandwidth that was obtained using proof-of-concept implementation was 32 kbit/s while introducing delays lower than 1ms and still retaining good voice quality. The work was further extended by analysing the influence of speech codec selection on TranSteg efficiency [Janicki et al. 2012a]. One of the interesting findings was that if the pair G.711/G.711.0 codecs are utilized, TranSteg introduces no steganographic cost and it offers a remarkably high steganographic bandwidth—on average about 31kbps.

In 2012, Tian et al. experimentally evaluated steganographic bandwidth and undetectability of two VoIP steganography methods proposed earlier by Huang et al. [2011a] (LSBs of NTP timestamp field of RTCP protocol) and by Forbes [2009] (LSBs of timestamp field of RTP protocol). The authors utilized the Windows Live Messenger voice conversations system and proved that by using the first approach, steganographic bandwidth of 335 bit/s could be obtained, and by using the second, 5.1 bit/s could be obtained. The latter method is also harder to detect.

3.2.3. Hybrid Methods (S3). In the work of Mazurczyk and Szczypiorski [2008b], a novel method—LACK—was introduced; it was later described and analytically analysed in Mazurczyk and Lubacz [2010]. LACK relies on the modification of both the content of the RTP packets and their time dependencies. This method takes advantage of the fact that in RTP, excessively delayed packets are not used for the reconstruction of the transmitted data by the receiver; that is, the packets are considered useless and therefore discarded. Thus, hidden communication is possible by introducing intentional delays to selected RTP packets and by substituting the original payload with a steganogram. Practical evaluation based on the LACK prototype was also presented by Mazurczyk [2012], where the method's impact on the quality of voice transmission was investigated. The concept of LACK was further extended by Hamdaqa and Tahvildari [2011] by providing a reliability and fault tolerance mechanism based on a modified (k, n) threshold based on Lagrange Interpolation. Results demonstrated that the complexity of steganalysis is increased. The “cost” for the extra reliability is a loss of some fraction of the steganographic bandwidth.

Arackaparambil et al. [2012] described a simple VoIP steganography method in which chosen RTP packets’ payloads are replaced with a steganogram, and the RTP header’s sequence number and/or timestamp fields are intentionally changed to make them appear as if they were excessively delayed by the network. This solution can be treated as a variation of the LACK method described previously.

Table II summarises the most important characteristics of the steganographic methods outlined in Section 3.2.

4. VOIP STEGANALYSIS

For network steganography, as well as for digital media steganography, there is still no universal, “one size fits all” detection solution; thus, steganalysis methods must be adjusted precisely to the specific information-hiding technique. That is why it is hard to compare existing VoIP steganalysis methods. However, it can be noted that as a classifier, typically Support Vector Machine (SVM) is utilized, which is state-of-the-art solution for steganalysis in general.

Let us consider again the hidden communication scenarios from Figure 4. For network steganography, there are three possible localisations for a warden (W1 to W3), as mentioned in Section 2. 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 (centralised warden) or multiple locations (distributed warden). In general, the localisation and number of locations in which the warden is able to inspect traffic greatly influences the effectiveness of the detection method.

For example, in scenario (1) from Figure 4, if a warden operates in an overt/covert sender’s Local Area Network (LAN), then some steganographic methods (e.g., those that utilise the packets’ delay) can be trivial to detect. This is because the anomaly introduced into the packets’ stream will be easily spotted near the transmitter. However, if a warden is present only in the sender’s or receiver’s LAN, then for scenario (4), hidden communication will remain undiscovered. Similar conclusions can be drawn for a reverse case and for other communication scenarios.

Moreover, if a distributed warden is used that has access to the same traffic flow in several network localisations, then the warden’s effectiveness is likely to increase. Depending on the communication scenario, steganographic modification to the network traffic can be spotted by simple comparison in two distinct locations. For example, if scenario (4) is utilised and a steganographic method like LSB is used, then by comparing the packets’ payload in two different locations (e.g., in the sender’s LAN and on some intermediate node in the external network), it is possible to uncover steganographic traffic modification. However, it must be emphasized that in practice, realisation of a distributed warden is hard to achieve, especially for VoIP.

Now, let us consider again the VoIP steganography classification introduced in Section 1 (Figure 2). To verify the feasibility of methods from groups S2 and S3, experiments on real-life VoIP connections were conducted by Mazurczyk et al. [2010]. The results obtained show that from group S2, only the method that introduces intentional losses is practically applicable, but it offers low steganographic bandwidth (<1 bit/s). The other methods that employ reordering, or modification of interpacket delays, are impractical and easy to detect. Reordering of the RTP packets was never witnessed during the experiments, and interpacket delays varied so much because of the network conditions that applying such steganographic techniques would be a difficult task and would result (if they worked at all) in a very low steganographic bandwidth. For

Table II. Summary of the VoIP Steganography Methods Applied to VoIP-Specific Protocols

	New concept for VoIP environment	New application for VoIP	Extension of existing VoIP method	Improvement of:		
				Undetectability	Robustness	Steganographic bandwidth
Wang et al. [2005], Shah et al. [2006], Mazurczyk and Kotulski [2006a, 2006b], Shah and Blaze [2009], Huang et al. [2011a]	+	+	–	N/A	N/A	N/A
Chen et al. [2006], Lloyd [2010], Arackaparambil et al. [2012]	–	–	+	N/A	N/A	N/A
Mazurczyk and Szczypiorski [2008a, 2008b], Bai et al. [2008], Mazurczyk et al. [2011]	+	–	–	N/A	N/A	N/A
Forbes [2009], Wieser and Röning [2010], Mazurczyk and Lubacz [2010]	–	–	+	–	–	+
Janicki et al. [2012a]	–	–	+	+	+	+
Hamdaqa and Tahvildari [2011], Tian et al. [2012]	–	–	+	+	–	–

methods from group S3, it was concluded based on the LACK example that it can offer potentially high steganographic bandwidth when it tries to mimic delay spikes, the characteristic formation of packets, which can lead to packet drops at the receiving end. This in turn is possible by intentionally causing such RTP packet sequences that will surely lead to jitter buffer (receiving buffer) losses by causing late packet drops, or jitter buffer overflows. Thus, an effective steganalysis method for LACK is still desirable.

As mentioned in Section 1, the main focus of the research community in the past decade has been dedicated to developing steganographic techniques from the S1 group (see Figure 2). This in turn has resulted in the increased number of steganalysis methods developed for this group. It must be emphasised that many so-called audio steganalysis methods were also developed for detection of hidden data in audio files (so-called audio steganography). However, in this article, we consider only those detection methods that have been evaluated and proved feasible for VoIP.

Statistical steganalysis for LSB-based VoIP steganography was proposed by Dittmann et al. [2005]. They proved that it was possible to be able to detect hidden communication with almost a 99% success rate under the assumption that there are no packet losses and the steganogram is unencrypted/uncompressed.

Takahashi and Lee [2007] described a detection method based on calculating the distances between each audio signal and its de-noised residual by using different audio quality metrics. Then, an SVM classifier is utilised for detection of the existence of hidden data. This scheme was tested on LSB, DSSS, FHSSS, 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 Krätzer and Dittmann [2007] for the purpose of VoIP steganalysis. Under the assumption that a steganographic message is not permanently embedded from the start to the end of the conversation, these authors demonstrated that detection of an LSB-based steganography is efficient with a success rate of 100%. This work was further extended by Krätzer and Dittmann [2008a] employing an SVM classifier. In another study [Krätzer and Dittmann 2008b], it was shown for an example of VoIP steganalysis that channel character–specific detection performs better than when channel characteristic features are not considered.

Steganalysis of LSB steganography based on a sliding window mechanism and an improved variant of the previously known Regular Singular (RS) algorithm was proposed by Huang et al. [2011c]. Their approach provides a 64% decrease in the detection time over the classic RS, which makes it suitable for VoIP. Moreover, experimental results prove that this solution is able to detect up to five simultaneous VoIP covert channels with a 100% success rate.

Huang et al. [2011d] also introduced the steganalysis method for compressed VoIP speech that is based on second statistics. In order to estimate the length of the hidden message, these authors proposed to embed hidden data into a 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 not only allows the detection of hidden data embedded in a compressed VoIP call but also accurately estimates its size.

Steganalysis that relies on the classification of RTP packets (as steganographic or nonsteganographic ones) and utilises specialised random projection matrices that take advantage of prior knowledge about the normal traffic structure was proposed by Garateguy et al. [2011]. Their approach is 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, whereas the convex set-based one was more powerful but more time-consuming.

Arackaparambil et al. [2012] analysed how in distribution-based steganalysis, the length of the window of the detection threshold, and in which the distribution is measured, should be depicted to provide the greatest chance for success. The results obtained showed how these two parameters should be set for achieving a high rate of detection whilst 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 and was proved efficient.

A method for detecting CNV-QIM steganography in G.723.1 voice streams was described by Li and Huang [2012]. This approach is to build the two models—distribution histogram and state transition—to quantify the codeword distribution characteristics. Based on these two models, feature vectors for training the classifiers for steganalysis are obtained. The technique is implemented by constructing an SVM classifier, and the results show that it can achieve an average detecting success rate of 96% when the duration of the G.723.1 compressed speech bit stream is less than 5s.

A steganalysis method for TranSteg based on Mel-Frequency Cepstral Coefficients (MFCC) parameters and Gaussian Mixture Models (GMMs) was developed and tested for various overt/covert codec pairs in a single warden scenario with double transcoding [Janicki et al. 2012b]. The proposed method allowed for efficient detection of some codec pairs—for example, G.711/G.726 with an average detection probability of 94.6%, Speex7/G.729 with 89.6%, or Speex7/iLBC with 86.3% detectability. Other codec pairs remained more resistant to detection—for example, for the pair iLBC/AMR, an average detection probability of 67% was achieved. Successful detection of TranSteg using the proposed steganalysis method requires at least 2s of speech data to analyse.

5. CONCLUSIONS

We are currently facing an almost ubiquitous proliferation of VoIP. Skype, VoIP-ready Android phones, and numerous VoIP ISPs and open source communities have resulted in a large number of users and a huge volume of traffic generated. This in turn, together with the number of protocols that form IP telephony, makes it a perfect steganographic carrier that can be exploited for good (e.g., to enrich the service) or bad (e.g., to enable confidential data exfiltration) intentions. Thus, it is not surprising that clandestine communication based on steganophony is becoming reality and soon can be widely utilized with good or bad intentions.

The main focus of the research community during the past decade has been mainly dedicated to those VoIP steganography methods that modify the speech exchanged between the calling parties, or the codec with which the speech is encoded. Surprisingly, a lot of research effort is still devoted for improving methods like LSB. Researchers have also targeted various VoIP-specific protocols like SIP with SDP, RTP, and RTCP as steganographic carriers. Little effort has been devoted to the deployment of the methods that modify the time relations between packets in the RTP stream, which is an important branch of the network steganography field. This is mainly because such solutions are rather hard to deploy practically, as they usually offer low steganographic bandwidth and require synchronisation. However, for certain applications like flow watermarking that do not require high steganographic bandwidth, they were proved efficient and feasible [Wang et al. 2005]. An interesting research direction could be development of hybrid methods that modify both VoIP PDUs and their time relations, as it has been proven that such methods offer decent steganographic bandwidth while still being undetectable [Mazurczyk and Szczypiorski 2008b]. Moreover, general techniques that improve undetectability are currently closely analysed [Tian et al. 2011b].

Of course, this survey does not exhaust all of the possibilities of hidden communication that can be used with VoIP. More work will surely continue to appear on information hiding in the carried speech, as well as in the VoIP-specific protocols. Obviously, some of existing steganographic methods for other carriers could be successfully “transferred” to VoIP environment (e.g., Yu et al. [2007] and Ling et al. [2011]). However, each method must be carefully evaluated due to specific requirements for IP telephony as a real-time service.

At the same time, the steganalysis methods are present, but they are not universal and efficient enough to be practically deployed in the telecommunication network environment to perform real-time detection. As with steganographic methods, some existing steganalysis solutions could be utilized for VoIP steganography (e.g., Gianvecchio and Wang [2011] and Jia et al. [2009]); however, as in the previous case, their efficiency was not proved for real-time services.

Both steganographic methods, with numerous ways to make them even more hidden, are being developed in parallel with the methods of detection. For every VoIP steganography method developed, a new steganalysis solution is sooner or later proposed. Then, even more stealthy approaches of previously known methods are developed, and their detection must once again catch up. This form of “arms race” will surely not come to an end in the foreseeable future.

ACKNOWLEDGMENTS

The author would like to thank Professor Krzysztof Szczypiorski from Warsaw University of Technology (Poland) for helpful comments and remarks.

REFERENCES

Aoki, N. 2003. A packet loss concealment technique for VoIP using steganography. In *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS'03)*. 470–473.

Aoki, N. 2004. VoIP packet loss concealment based on two-side pitch waveform replication technique using steganography. In *Proceedings of the IEEE Region 10 Conference (TENCON'04): Volume 3*. 52–55.

Aoki, N. 2007. Potential of value-added speech communications by using steganography. In *Proceedings of Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'07)*. 251–254.

Aoki, N. 2008. A technique of lossless steganography for G.711 telephony speech. In *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'08)*. Harbin, China, 608–611.

Aoki, N. 2009. Lossless steganography techniques for IP telephony speech taking account of the redundancy of folded binary code. In *Proceedings of the 5th International Joint Conference on INC, IMS and IDC*. 1689–1692.

Aoki, N. 2010. A semi-lossless steganography technique for G.711 telephony speech. In *Proceedings of the 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'10)*. 534–537.

Arackaparambil, C., Yan, G., Bratus, S., and Caglayan, A. 2012. On tuning the knobs of distribution-based methods for detecting VoIP covert channels. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS-45)*. Hawaii, 2431–2440.

Bai, L. Y., Huang, Y., Hou, G., and Xiao, B. 2008. Covert channels based on jitter field of the RTCP header. In *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'08)*. 1388–1391.

Bender, W., Gruhl, D., Morimoto, N., and Lu, A. 1996. Techniques for data hiding. *IBM. Syst. J. 35*, 3–4, 313–336.

Chen, S., Wang, X., and Jajodia, S. 2006. On the anonymity and traceability of peer-to-peer VoIP calls. *IEEE Netw. 20*, 5, 32–37.

Deng, Z., Shao, X., Yang, Z., and Zheng, B. 2008. A novel covert speech communication system and its implementation. *J. Electron. 25*, 6, 737–745.

DITTMANN, J., HESSE, D., AND HILLERT, R. 2005. Steganography and steganalysis in voice-over IP scenarios: Operational aspects and first experiences with a new steganalysis tool set. In *Proceedings of SPIE 2005: Volume 5681, Security, Steganography, and Watermarking of Multimedia Contents VII*. 607–618.

DoD "ORANGE BOOK." 1985. Trusted computer system evaluation criteria. Tech. rep. DOD 5200.28-STD. National Computer Security Center.

DRUID. 2007. Real-time steganography with RTP. Tech. rep. http://www.uninformed.org/?v=8&a=3&t=pdf.

FORBES, C. R. 2009. A new covert channel over RTP. Master's thesis, Rochester Institute of Technology. https://ritdml.rit.edu/bitstream/handle/1850/12883/CForbesThesis8-21-2009.pdf?sequence=1.

FRIDRICH, J. 2010. *Steganography in Digital Media—Principles, Algorithms, and Applications*. Cambridge University Press.

GARATEGUY, G., ARCE, G., AND PELAEZ, J. 2011. Covert channel detection in VoIP streams. In *Proceedings of the 45th Annual Conference on Information Sciences and Systems (CISS)*. 1–6.

GEISER, B., MERTZ, F., AND VARY, P. 2008. Steganographic packet loss concealment for wireless VoIP. In *Proceedings of the Conference on Voice Communication (SprachKommunikation)*. 1–4.

GIANVECCHIO, S. AND WANG, H. 2011. An entropy-based approach to detecting covert timing channels. In *IEEE T. Depend. Secure. Comput. 8*, 6, 785–797.

HAMDAQA, M. AND TAHVILDARI, L. 2011. ReLACK: A reliable VoIP steganography approach. In *Proceedings of the 5th International Conference on Secure Software Integration and Reliability Improvement (SSIRI'11)*. 189–197.

HUANG, Y., TANG, S., AND YUAN, J. 2011b. Steganography in inactive frames of VoIP streams encoded by source codec. *IEEE Trans. Inf. Forensics Security 6*, 2, 296–306.

HUANG, Y., TANG, S., AND ZHANG, Y. 2011c. Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis. *IET Commun. 5*, 7, 929–936.

HUANG, Y., TANG, S., BAO, C., YIP, Y. J., AND YAU, J. 2011d. Steganalysis of compressed speech to detect covert voice over Internet protocol channels. *IET Inf. Secur. 5*, 1, 26–32.

HUANG, Y., XIAO, B., AND XIAO, H. 2008. Implementation of covert communication based on steganography. In *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'08)*. 1512–1515.

HUANG, Y., YUAN, J., CHEN, M., AND XIAO, B. 2011a. Key distribution over the covert communication based on VoIP. *J. Electron. 20*, 2, 357–360.

JANICKI, A., MAZURCZYK, W., AND SZCZYPIORSKI, K. 2012a. Influence of speech codecs selection on transcoding steganography. In *Computing Research Repository (CoRR)*, abs/1201.6218, arXiv.org E-print archive. Cornell University, Ithaca, NY. http://arxiv.org/abs/1201.6218.

JANICKI, A., MAZURCZYK, W., AND SZCZYPIORSKI, K. 2012b. Steganalysis of transcoding steganography. In *Computing Research Repository (CoRR)*, abs/1210.5888, arXiv.org E-print archive. Cornell University, Ithaca, NY. http://arxiv.org/abs/1210.5888.

JIA, W., TSO, F. P., LING, Z., FU, X., XUAN, D., AND YU, W. 2009. Blind detection of spread spectrum flow watermarks. In *Proceedings of IEEE INFOCOM*. 2195–2203.

KRÄTZER, C., DITTMAN, J., VOGEL, T., AND HILLERT, R. 2006. Design and evaluation of steganography for Voice-over-IP. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'06)*.

KKRÄTZER, C. AND DITTMAN, J. 2007. Mel-cepstrum based steganalysis for VoIP steganography. In *Proceedings of the 19th Annual Symposium of the Electronic Imaging Science and Technology, SPIE and IS&T*. San Jose, CA.

KRÄTZER, C. AND DITTMAN, J. 2008a. Pros and cons of mel-cepstrum based audio steganalysis using SVM classification. In *Proceedings of the 9th International Conference on Information Hiding (IH'07)*. 359–377.

KRÄTZER, C. AND DITTMAN, J. 2008b. Cover signal specific steganalysis: The impact of training on the example of two selected audio steganalysis approaches. In *Proceedings of SPIE-IS&T Electronic Imaging: Volume 6819*.

KUNDUR, D. AND AHSAN, K. 2003. Practical Internet steganography: Data hiding in IP. In *Proceedings of the Texas Workshop on Security of Information Systems*.

LAMPSON, B. W. 1973. A note on the confinement problem. *Comm. ACM 16*, 10, 613–615.

LI, S. AND HUANG, Y. 2012. Detection of QIM steganography in G.723.1 bit stream based on quantization index sequence analysis. *J. Zhejiang Univ. Sci. C 13*, 8, 624–634.

LING, Z., FU, X., JIA, W., YU, W., AND XUAN, D. 2011. A novel packet size based covert channel attack against anonymizer. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM'11)*. 186–190.

Liu, L., Li, M., Li, Q., and Liang, Y. 2008. Perceptually transparent information hiding in G.729 bitstream. In *Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 406–409.

Liu, J., Zhou, K., and Tian, H. 2012. Least-significant-digit steganography in low bit-rate speech. In *Proceedings of the 47th IEEE International Conference on Communications (ICC)*. Ottawa, Canada, 1–5.

Lloyd, P. 2010. An exploration of covert channels within Voice over IP. Master's thesis. Rochester Institute of Technology, https://ritdml.rit.edu/bitstream/handle/1850/12241/PLloydThesis5-4-2010.pdf?sequence=1.

Lu, C. S. 2005. *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property*. Idea Group Publishing.

Lubacz, J., Mazurczyk, W., and Szczypiorski, K. 2008. Hiding data in VoIP. In *Proceedings of the 26th Army Science Conference (ASC'08)*. Orlando, FL.

Lubacz, J., Mazurczyk, W., and Szczypiorski, K. 2010. Vice over IP. *IEEE Spectr*. 40–45.

Lucena, N. B., Pease, J., Yadollahpour, P., and Chapin, S. 2004. Syntax and symantics-preserving application-layer protocol steganography. In *Proceedings of the 6th International Conference on Information Hiding (IH'04)*. 164–179.

Ma, L., Wu, Z., and Yang, W. 2007. Approach to hide secret speech information in G.721 scheme. In *Proceedings of the Intelligent Computing 3rd International Conference on Intelligent Computing Theories and Applications (ICIC'07)*. 1315–1324.

Mazurczyk, W. 2012. Lost audio packets steganography: A first practical evaluation. *Int. J. Secur. Commun. Netw. 5*, 12, 1394–1403.

Mazurczyk, W. and Kotulski, Z. 2006a. New VoIP traffic security scheme with digital watermarking. In *Proceedings of the 25th International Conference on Computer Safety, Reliability and Security (SafeComp'06)*. 170–181.

Mazurczyk, W. and Kotulski, Z. 2006b. New security and control protocol for VoIP based on steganography and digital watermarking. In *Proceedings of the 5th International Conference on Computer Science—Research and Applications (IBIZA'06)*.

Mazurczyk, W. and Lubacz, J. 2010. LACK—a VoIP steganographic method. *Telecommun. Syst. Model Anal. Des. Manag. 45*, 2–3, 153–163.

Mazurczyk, W. and Szczypiorski, K. 2008a. Covert channels in SIP for VoIP signalling. In *Proceedings of 4th International Conference on Global E-Security*. London, United Kingdom, 65–70.

Mazurczyk, W. and Szczypiorski, K. 2008b. Steganography of VoIP streams. In *Proceedings of the On The Move Federated Conferences and Workshops: 3rd International Symposium on Information Security (IS'08)*. Monterrey, Mexico, 1001–1018.

Mazurczyk, W., Cabaj, K., and Szczypiorski, K. 2010. What are suspicious VoIP delays? *Multimed. Tools Appl. 57*, 1, 109–126.

Mazurczyk, W., Szaga, P., and Szczypiorski, K. 2011. Using transcoding for hidden communication in IP telephony. In *Computing Research Repository (CoRR)*, abs/1111.1250, arXiv.org E-print archive. Cornell University, Ithaca, NY. http://arxiv.org/abs/1111.1250.

Miao, R. and Huang, Y. 2011. An approach of covert communication based on the adaptive steganography scheme on Voice over IP. In *Proceedings of the IEEE International Conference on Communications (ICC'11)*. 1–5.

Murdoch, S. and Lewis, S. 2005. Embedding covert channels into TCP/IP. In *Proceedings of the 7th International Conference on Information Hiding (IH'05)*. 247–266.

Nishimura, A. 2009. Steganographic band width extension for the AMR codec of low-bit-rate modes. In *Proceedings of Interspeech 2009*. Brighton, UK, 2611–2614.

Nutzinger, M. and Wurzer, J. 2011. A novel phase coding technique for steganography in auditive media. In *Proceedings of the IEEE International Conference on Availability, Reliability and Security*. Vienna, Austria, 91–98.

Nutzinger, M., Fabian, C., and Marschalek, M. 2010. Secure hybrid spread spectrum system for stegnanography in auditive media. In *Proceedings of the IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 78–81.

Petitcolas, F., Anderson, R., and Kuhn, M. 1999. Information hiding—a survey. *Proc. IEEE 87*, 7, 1062–1078.

Rosenberg, J., Schulzrinne, H., Camarillo, G., and Johnston, A. 2002. SIP: Session Initiation Protocol. IETF, RFC 3261.

Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V. 2003. RTP: A transport protocol for real-time applications. IETF, RFC 3550.

Servetto, S. D. and Vetterli, M. 2001. Communication using phantoms: Covert channels in the Internet. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT'01)*.

Shah G. and Blaze, M. 2009. Covert channels through external interference. In *Proceedings of the 3rd USENIX Conference on Offensive Technologies*. Montreal, Canada, 3–3.

Shah, G., Molina, A., and Blaze, M. 2006. Keyboards and covert channels. In *Proceedings of the 15th USENIX Security Symposium*. 59–75.

Simmons, G. J. 1984. The prisoner's problem and the subliminal channel. In *Proceedings of CRYPTO '83*. Plenum Press, Santa Barbara, CA.

Takahashi, T. and Lee, W. 2007. An assessment of VoIP covert channel threats. In *Proceedings of the 3rd International Conference on Security and Privacy in Communication Networks (SecureComm'07)*. 371–380.

Tian, H., Guo, R., Lu, J., and Chen, Y. 2012. Implementing covert communication over voice conversations with Windows Live Messenger. *AISS 4*, 4, 18–26.

Tian, H., Zhou, K., and Feng, D. 2010. Dynamic matrix encoding strategy for voice-over-IP steganography. *J. Cent. South Univ. Technol. 17*, 1285–1292.

Tian, H., Zhou, K., Huang, Y., Feng, D., and Liu, J. 2008. A covert communication model based on least significant bits steganography in Voice over IP. In *Proceedings of the 9th International Conference for Young Computer Scientists*. 647–652.

Tian, H., Zhou, K., Jiang, H., Liu, J., Huang, Y., and Feng, D. 2009a. An M-sequence based steganography model for voice over IP. In *Proceedings of the IEEE International Conference on Communications (ICC'09)*. 1–5.

Tian, H., Zhou, K., Jiang, H., Liu, J., Huang, Y., and Feng, D. 2009b. An adaptive steganography scheme for Voice over IP. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'09)*. Taipei, Taiwan, 2922–2925.

Tian, H., Jiang, H., Zhou, K., and Feng, D. 2011a. Adaptive partial-matching steganography for Voice over IP using triple M sequences. *Comput. Comm. 34*, 2236–2247.

Tian, H., Jiang, H., Zhou, K., and Feng, D. 2011b. Transparency-orientated encoding strategies for Voice-over-IP steganography. *Comput. J. 55*, 6, 702–716.

Wang, C. and Wu, W. 2007. Information hiding in real-time VoIP streams. In *Proceedings of the 9th IEEE International Symposium on Multimedia (ISM'07)*. 255–262.

Wang, X., Chen, S., and Jajodia, S. 2005. Tracking anonymous peer-to-peer VoIP calls on the Internet. In *Proceedings of the 12th ACM Conference on Computer and Communications Security (CCS'05)*. 81–91.

Wieser, C. and Röning, J. 2010. An evaluation of VoIP covert channels in an SBC setting. In *Proceedings of the Security in Futures–Security in Change Conference*. 54–58.

Wu, Z. and Yang, W. 2006. G.711-based adaptive speech information hiding approach. In *Proceedings of the International Conference on Intelligent Computing: Volume Part I*. 1139–1144.

Wu, Z. J., Gao, W., and Yang, W. 2009. LPC parameters substitution for speech information hiding. *J. China Univ. Posts Telecomm. 16*, 6, 103–112.

Wu, Z. J., Yang, W., and Yang, Y. X. 2003. ABS-based speech information hiding approach. *Electron. Lett. 39*, 1617–1619.

Xiao, B., Huang, Y., and Tang, S. 2008. An approach to information hiding in low bit-rate speech stream. In *Proceedings of the Global Telecommunications Conference (GLOBECOM'08)*. 1–5.

Xu, E., Liu, B., Xu, L., Wei, Z., Zhao, B., and Su, J. 2011. Adaptive VoIP steganography for information hiding within network audio streams. In *Proceedings of the International Conference on Network-Based Information Systems*, 612–617.

Xu, T. and Yang, Z. 2009. Simple and effective speech steganography in G.723.1 low-rate codes. In *Proceedings of the International Conference on Wireless Communications and Signal Processing (WCSP'09)*. 1–4.

Yu, W., Fu, X., Graham, S., Xuan, D., and Zhao, W. 2007. DSSS-based flow marking technique for invisible traceback. In *Proceedings of the IEEE Symposium on Security and Privacy (SP'07)*. 18–32.

Zander, S., Armitage, G., and Branch, P. 2007. A survey of covert channels and countermeasures in computer network protocols. *IEEE Commun. Surv. Tutor. 9*, 3, 44–57.

Received March 2012; revised February 2013; accepted March 2013