SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks (2011)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.388

SPECIAL ISSUE PAPER

On steganography in lost audio packets

Wojciech Mazurczyk, Józef Lubacz and Krzysztof Szczypiorski*

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

ABSTRACT

This paper presents a new hidden data insertion procedure based on the estimated probability of the remaining time of the call for the steganographic method called lost audio packets (LACK) steganography. LACK provides hidden communication for real-time services such as voice over IP. The analytical results presented in this paper concern the influence of LACK's hidden data insertion procedures on the quality of voice transmission and the resistance to steganalysis. The proposed hidden data insertion procedure is also compared with previous steganogram insertion approaches on the basis of estimating the remaining average call duration. Copyright © 2011 John Wiley & Sons, Ltd.

KEYWORDS

VoIP; LACK; network steganography; performance analysis

***Correspondence**

Krzysztof Szczypiorski, Institute of Telecommunications, Warsaw University of Technology, Nowowiejska 15/19, Warsaw 00-665, Poland.
E-mail: ksz@tele.pw.edu.pl

1. INTRODUCTION

Lost audio packets (LACK) steganography is a steganographic method that modifies both Real-time Transport Protocol (RTP) [1] packets and their time dependencies, and is intended for a broad class of multimedia real-time applications such as IP telephony. This method utilises the fact that for usual multimedia communication protocols such as RTP, excessively delayed packets are not used for the reconstruction of transmitted data at the receiver; that is, the packets are considered useless and discarded.

Lost audio packets can be characterised by the following features: steganographic bandwidth, undetectability and steganographic cost. Steganographic bandwidth describes how much secret data that we are able to send using a specific method per time unit. Undetectability is defined as an inability to detect a steganogram inside certain carriers. The most popular way to detect a steganogram is to analyse statistical properties of the captured data and compare it with the typical properties of that carrier. Steganographic cost characterises the degree of degradation of the carrier caused by the steganogram insertion procedure. The steganographic cost depends on the type of carrier and, if it becomes excessive, leads to easy detection of the steganographic method. For example, if the method uses voice packets as a carrier for steganographic purposes in IP telephony, then the cost is expressed in conversation degradation. As another example, if the carrier is certain fields of the protocol header, then the cost is expressed as a potential loss in that protocol functionality.

It should be emphasised that the hidden data insertion procedures introduced and analysed in this paper can be utilised not only by decent LACK users who use their own voice over IP (VoIP) calls to exchange covert data, but also by intruders who are able to covertly send data using third party VoIP calls (e.g. an effect of earlier successful attacks by using trojans or worms or by distributing modified versions of a popular VoIP software [2,3]). This trade-off is typical in steganography and requires consideration in a broader steganography context, which is beyond the scope of this paper.

In this paper, we investigate LACK, which was originally proposed in [4] and studied in [5]. This paper is an extension and continuation of the previous work presented in [6].

The contributions of this paper are as follows:

- Detailed analysis of the LACK performance issues and of dependence of the insertion procedure on estimated VoIP call quality (Sections 3 and 4).
- Extension of the previously proposed hidden data insertion procedure based on estimated remaining average call duration by considering also influence of the estimated call quality (Section 5.2).
- Introduction of a new hidden data insertion procedure based on the estimated probability of the remaining time of the call (Section 5.3). In addition, for this procedure, influence of the estimated call quality is considered. For both methods, LACK performance results are presented.
- Comparison of both of the presented procedures for steganogram insertion in LACK (Section 5.4).

The rest of the paper is structured as follows. In Section 2, the basics of LACK functioning and detection is presented.

In Section 3, LACK performance issues involved in using the method are discussed. In Section 4 the dependence of the hidden data insertion rate $IR(t)$ on the estimated call quality is investigated. In Section 5, two methods for determining $IR(t)$ based on estimated call duration are presented, analysed and compared. Section 6 concludes our work and indicates possible future research.

2. LACK BASICS

The detailed description of LACK functioning is as follows (see Figure 1). At the transmitter, one packet is selected from the RTP stream, and its voice payload is substituted with bits of the steganogram (1). Then, the selected audio packet is intentionally delayed before transmitting (2). If an excessively delayed packet reaches a receiver unaware of the steganographic procedure, it is discarded (3) because, for unaware receivers, the hidden data is 'invisible'. However, if the receiver knows about the hidden communication, the receiver extracts the payload instead of deleting the packet (4). Because the payload of the intentionally delayed packets is used to transmit secret information to receivers aware of the procedure, no extra packets are generated.

Lost audio packets is a TCP/IP application layer steganography technique and is rather easy to implement. This ease of implementation is due to the fact that RTP is usually integrated in telephone endpoints (softphones) so that access to RTP packet generation and modification is easier to perform compared with the case of lower-layer protocols such as IP or UDP.

Steganalysis of LACK is hard to perform because packet loss in IP networks is a 'natural phenomenon', and therefore, intentional losses introduced by LACK are not easy to detect, if kept on a reasonable level. Potential LACK steganalysis methods include the following:

- Statistical analysis of lost packets for calls in a sub-network. This type of steganalysis may be implemented with a passive warden [7] (or some other network node), based, for example, on information included in Real-time Transport Control Protocol (RTCP) reports (the cumulative number of packets lost field) exchanged between users during their communication or by observing RTP stream flows (packets' sequence numbers). If for some of the observed calls the number of lost packets is higher than average (or some chosen threshold), this criterion may be used as an indication for the potential use of LACK.
- Statistical analysis is based on the VoIP calls duration. If the call duration probability distribution for a certain sub-network is known, then statistical steganalysis may be performed to discover VoIP sources that do not fit to the distribution (the duration of LACK calls may be longer compared with non-LACK calls as a result of introducing steganographic data).
- An active warden [7] that analyses all RTP streams in the network (Synchronization Source identifier and fields: sequence number and timestamp from RTP header) can identify packets that are already too late to be used for voice reconstruction. The active warden may erase their payload fields or simply drop them. A potential problem that arises in this case is avoiding elimination of delayed packets that may still be used for conversation reconstruction. The size of the jitter buffer at the receiver is, in principle, unknown to the active warden. If an active warden drops all delayed packets, then it will potentially drop packets that still can be useful for voice reconstruction. In effect, the quality of conversation may deteriorate considerably.

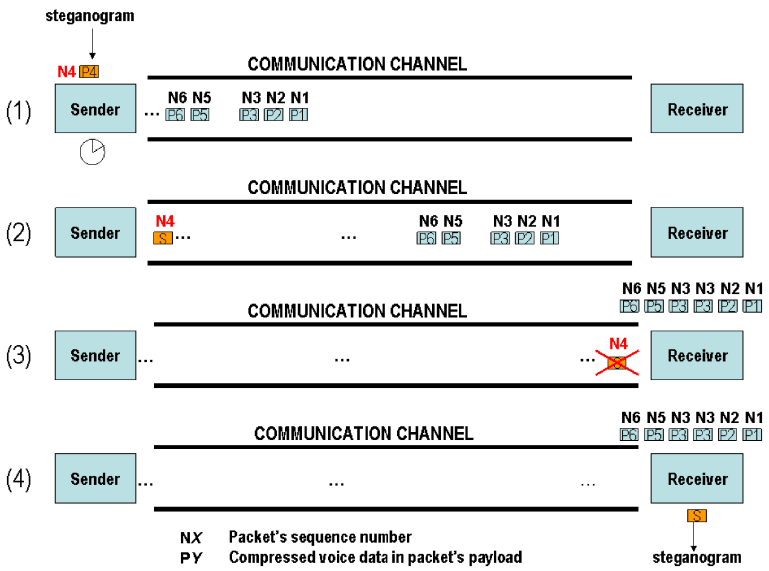

Figure 1. The idea of lost audio packets.

Moreover, steganographic calls are not only affected; non-steganographic calls are also 'punished'.

3. LACK PERFORMANCE ISSUES

The performance of LACK depends on many factors, such as the details of the communication procedure (including the type of codec used, the size of the voice frame, and the size of the receiving buffer) and on the network quality of service (QoS) (the packet delay, packet loss probability and jitter). We discuss these issues next.

Lost audio packets' steganographic bandwidth and resistance to detection can be influenced by the following elements:

- The number of intentionally delayed RTP packets.
- The delay of the LACK packets.
- Network QoS—packet delay, packet loss probability and jitter.
- Features of the transmission devices—specifically, the voice codec used (resistance to RTP packet losses and initial voice quality), the size of the RTP packet payload and the size of the jitter buffer.
- Hidden data insertion rate (*IR*)—number of bits the the steganogram carried in a unit of time (bit/s).

In general, the more hidden information is inserted into the data stream, the greater the chance that it will be detected, for example, by scanning the data flow or by some other steganalysis methods. Moreover, the more audio packets are used to send covert data, the greater the potential deterioration of the quality of the VoIP connection. Thus, the procedure of inserting hidden data has to be carefully chosen and controlled to minimise the chance of detecting inserted data and to avoid excessive deterioration of the QoS. For this reason, a trade-off among achieved steganographic bandwidth, call quality deterioration and resistance to detection is always required.

Let us assume that in a given moment of call t, the packet is chosen from the RTP packet stream with probability $p_L(t)$, and the network packet loss probability is $p_N(t)$. If p_T denotes the total acceptable probability of RTP packet losses, then, assuming independence of network packet losses from LACK choices, we get

$$p_T \leq 1 - (1 - p_N(t))(1 - p_L(t)) \tag{1}$$

which implies

$$p_L(t) \leq \frac{p_T - p_N(t)}{1 - p_N(t)} \tag{2}$$

which describes the admissible level of the RTP packet losses introduced by LACK.

Exemplary relationships among probabilities $p_L(t)$, $p_N(t)$ and p_T, are illustrated in Figure 2.

For example, if $p_T = 0.05$ and $p_N(t_\xi) = 0.02$, then $p_L(t_\xi) \leq 0.03$.

To guarantee that an audio packet will be recognised as lost by a receiver, the packet must be excessively delayed by the LACK procedure. To set this delay $d_L(t)$, the size of the receiver's jitter buffer must be taken into account. A jitter buffer is used to alleviate the jitter effect, that is, the variations in packet arrival times caused by queuing, contention and serialisation in the network. The size of the buffer is implementation dependent. This size may be fixed or adaptive, and is usually between 60 and 120 ms; the RTP packet will be recognised as lost when the delay is greater than the delay introduced by the jitter buffer. LACK users have to exchange information about the sizes of their jitter buffers before starting the steganographic procedure. To limit the risk of detection of the hidden data, the delay chosen by LACK users should be as low as possible.

The RTP packet delay at the transmitter exit is equal to

$$d_T(t) = d_D + d_K + d_E + d_L(t) \tag{3}$$

where $d_L(t)$ is the intentional delay of RTP packet introduced by LACK; d_D is the delay introduced by digital signal processor, which depends on the type of the codec and is equal usually from 2 to 20 ms; d_K is the delay

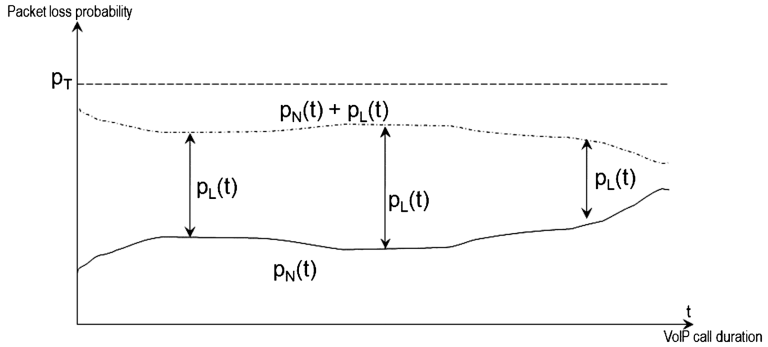

Figure 2. Lost audio packets influence on the total packet losses probability.

introduced by voice coding (typically under 10 ms); and d_E is the delay caused by encapsulation (from 20 to 30 ms).

As mentioned earlier, the value of the intentional delay $d_L(t)$ introduced by LACK must be carefully chosen. Together with $d_N(t)$ introduced by the network, the delay must be greater than the size of the jitter buffer (Figure 3), which is

$$d_T(t) + d_N(t) > t_B(t) \tag{4}$$

where $d_N(t)$ is the delay introduced by network and $t_B(t)$ is the size of the jitter buffer.

The jitter buffer can be of a fixed size or adaptive. For example, if the jitter buffer has a fixed size that is unchanged during the call and it does not consider network delay, then the delay at the transmitter output should be

$$d_T \geq t_B \tag{5}$$

and

$$d_L \geq t_B - d_D - d_K - d_E \tag{6}$$

Similar formulas can be derived for the adaptive jitter buffer case.

Additionally, to ensure high steganographic bandwidth and undetectability of LACK, it is necessary to observe network conditions while the call lasts. In particular, packet losses, delay and jitter introduced by the network must be carefully monitored because they have an influence on delay and packet losses that can be introduced by LACK without degrading the perceived quality of the conversation. Because LACK uses legitimate RTP traffic, LACK increases overall packet losses. Thus, the level of the lost packets used for steganographic purposes must be controlled and dynamically adapted.

Information about network conditions during the call can be provided to the transmitter, for example, with the use of sender report, receiver report [1] or extended report [8] that are defined in the RTCP. If packet losses, delays and jitter are not monitored during the call, then they can be determined on the basis of the historical statistical data related to the network quality. However, it should be noted that RTP packet losses introduced by a network can lead to a lowering of the LACK steganographic bandwidth if the lost packet is an RTP packet that contains a steganogram.

Lost audio packets steganographic bandwidth depends also on the codec used for VoIP conversation. Admissible levels of packet losses usually are in a range between 1% and 5%. For example, according to [17], maximum loss tolerance is 1% for G.723.1, 2% for G.729A and 3% for G.711 codecs. If a special mechanism to deal with lost packets at the receiver is utilised, for example, the packet loss concealment [18], then the acceptable level of lost packets, for example, for G.711 codecs, increases from 3% to 5%. The greater the codec resistance to packet losses is, the better the opportunity for achieving greater steganographic bandwidth for LACK. Thus, the amount of steganographic data that can be inserted by LACK, and in effect, the additional packet loss introduced by LACK, depends on the acceptable level of the total packet loss. For example, for the G.711 speech codec with a data rate of 64 kbit/s and a data frame size of 20 ms, if the packet loss probability introduced by the LACK procedure is 0.5%, then the theoretical hidden communication rate is 320 bit/s.

Another key element that influences LACK steganographic bandwidth and its resistance to steganalysis is the hidden data insertion rate $IR(t)$, which is defined as the number of steganogram bits carried in a unit of time during the call (bit/s). In general, the greater $IR(t)$ is, the greater the steganographic bandwidth and the greater the degradation in voice quality and the easier the steganalysis. $IR(t)$ is influenced by

- assumed, acceptable call quality;
- network conditions;
- the size of the steganogram and
- the duration of the call.

By applying the correct procedure for determining $IR(t)$, it is possible to control RTP packet losses and delays introduced by LACK without excessively affecting the call quality and the risk of being detected. This aspect was carefully analysed in Sections 3 and 4.

If LACK is used sporadically by a single user to transmit a small amount of hidden data, then utilising complex methods for determining $IR(t)$ is unnecessary because the chances of disclosure are very small and the effect on call

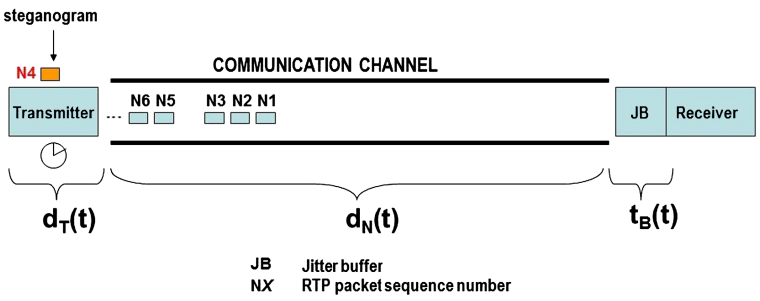

Figure 3. Elements of lost audio packets delay. RTP, Real-time Protocol.

quality is negligible. Complex variants of the $IR(t)$ calculation are important for cases in which LACK is used frequently by a single user or a group of users in certain network localisations.

In the simplest scenario, the $IR(t)$ value can be fixed and constant during the call and can be calculated as $IR = S/T$, where S is the size of the steganogram and T is the predetermined duration of the call. A simple alternative is also possible by choosing the constant IR and making the call last as long as the time it takes for the whole steganogram to be sent (the duration of the call is then equal to $T = S/IR$).However, the obvious disadvantage of such an approach is the lack of relationship between $IR(t)$ and the voice quality, and the resistance to steganalysis.

$IR(t)$ can also be set for the duration of the call based on statistical data (e.g. averages) on RTP packet losses and the quality of the calls. However, this method of setting the duration is not the proper solution for LACK because the method does not include potential changes in network conditions during the call and does not account for the relationship between $IR(t)$ and the size of the steganogram.

Methods for determining $IR(t)$ based on the current conversation quality, the size of the steganogram and the duration of the call are considered in the following sections.

4. DEPENDENCE OF THE $IR(T)$ ON THE ESTIMATED CALL QUALITY

In this section, we focus on the dependence of the insertion rate IR on the estimated call quality resulting from packet loss. The call quality may be expressed in terms of subjective and objective quality measures. Objective measures are usually based on algorithms such as the E-Model [9], Perceptual Analysis Measurement System (PAMS) or Perceptual Evaluation of Speech Quality (PESQ) [10]. The objective measures can be transformed into subjective quality measures. In our analysis, we shall use the subjective measure mean opinion score (MOS) [11], which, according to [12], can be related to packet loss probability p_N, as follows

$$MOS_N(t) = \alpha \cdot \exp(\beta \cdot p_N(t)) + \gamma \tag{7}$$

where α, β and γ are network/service-type dependent parameters; for Skype telephony, the parameters were evaluated to be [12]: $\alpha = 3.0829$, $\beta = -4.6446$ and $\gamma = 1.07$.

Because LACK introduces additional packet loss, p_L, p_N should be substituted with $p_N + p_L$ in the aforementioned equation.

$$MOS_L(t) = \alpha \cdot \exp(\beta \cdot p_N(t) + p_L(t)) + \gamma \tag{8}$$

Figure 4 shows the dependence of MOS on p_N for different values of p_L assuming values for α, β and γ that are estimated for Skype telephony.

The drop in call quality due to utilisation of LACK can be expressed as

$$\begin{aligned} \Delta MOS(t) &= MOS_N(t) - MOS_L(t) \\ &= \alpha \cdot \exp(\beta \cdot p_N(t)) \cdot (1 - \exp(\beta \cdot p_L(t))) \end{aligned} \tag{9}$$

Let IR_Q denote the call quality dependent hidden data insertion rate expressed as the MOS score. In general, IR_Q can be

- fixed during the VoIP call and determined on the basis of historical, statistical data on call quality, or
- dynamically adjusted, during the call, to the current estimation of voice quality.

In the rest of this subsection, we consider both cases described earlier.

4.1. Determining IR_Q on the basis of historical, statistical data on call quality in a given network

Let us assume that the MOS probability distribution for a considered network in which LACK is to be used is known. Figure 5 presents the MOS probability distribution for a VoIP network based on experimental data from [13].

Given η, the minimum acceptable call quality $MOS*$

$$P(MOS > MOS*) > \eta \tag{10}$$

Thus, based on Equation 8, the upper limit of p_L may be expressed as

$$p_L = \frac{\ln\left(\frac{MOS* - \gamma}{\alpha}\right)}{\beta} - p_N \tag{11}$$

If N_P is the number of RTP packets generated in a unit of time and P_P is the length of an RTP packet data field (in bits), then

$$IR_Q \leq p_L \cdot N_P \cdot P_P \tag{12}$$

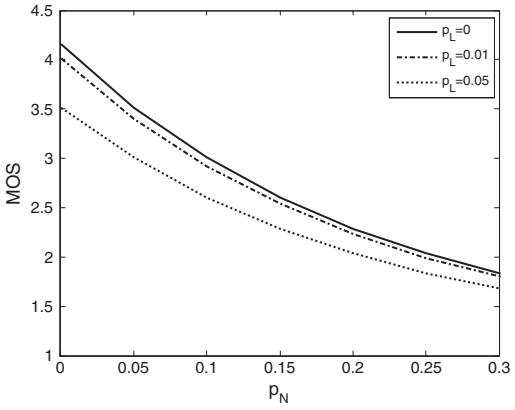

Figure 4. Mean opinion score (MOS) dependence on p_N and p_L for Skype telephony.

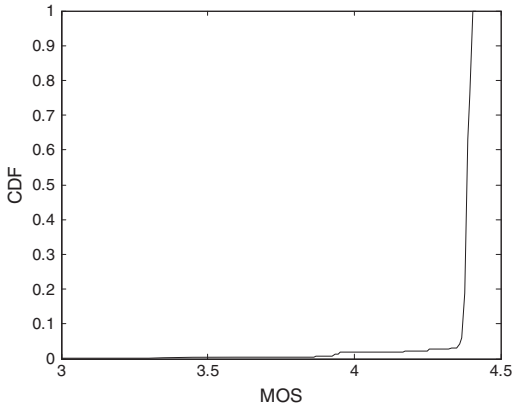

Figure 5. Mean opinion score probability distribution (experimental data [13]). CDF, cumulative distribution function.

4.2. Determining *IR*(*t*) on the basis of the current estimation of voice quality

An alternative to the approach described earlier is to adjust $IR(t)$ on the basis of the online measurement of network parameters such as network losses, as well as delays and jitter effects, which affect voice quality during the call. Such an approach would require online exchange of information on voice quality parameters between the sender and the receiver, for example, with the use of the RTCP (sender reports and receiver reports [1] or extended reports [8]). RTCP reports are exchanged by default every 5 s; however, these reports can be sent more frequently if they are required (i.e. if network parameters change often). With this information, the estimated current voice quality, $MOS_E(t)$, is calculated.

For a given upper limit of acceptable voice quality MOS^* during the call, whether $MOS(t) \geq MOS^*$ is verified. If this condition is fulfilled, then

$$IR_Q(t) \leq N_P \cdot P_P \left(\frac{\frac{\ln(MOS_E(t)-\gamma)}{\alpha}}{\beta} - p_N(t) \right) \qquad (13)$$

In any other case, $IR_Q(t) = 0$.

Dynamically adjusting $IR_Q(t)$ to current voice estimation can be troublesome and can cause instabilities. Thus, a more practical approach is to utilise average values for given periods.

5. DEPENDENCE OF THE *IR* ON THE ESTIMATED CALL DURATION

In the following analysis, we consider the dependence of the hidden data insertion rate *IR* for a particular call on the elapsed time of that call; that is, we consider the *IR* that is time dependent. As shown in our analysis, such a time-dependent *IR* procedure allows for decreasing the *IR* during the call duration, compared with the *IR* at the call initiation time. In effect, the negative influence of LACK on QoS can be decreased, and the resistance to steganalysis can be increased, especially for call duration distributions that have a coefficient of variation much greater than 1. Available experimental data concerning VoIP call duration distributions seem to indicate that this action is realistic for real-life VoIP calls. Our goal in this section is to express *IR* with the coefficient of variation for possibly a wide range of call duration distributions.

5.1. Voice over IP call duration probability distribution

For PSTN, the call duration probability distribution was well known as a result of extensive experimental research. For many decades, the exponential distribution was assumed to be a good enough approximation for engineering purposes. VoIP is a relatively new service, and thus, little reliable experimental data is available; hence, in many research papers concerning IP voice traffic (e.g. [14–16]), an exponential call duration is still assumed. Current experiments prove, however, that this assumption is far from realistic.

Birke *et al*. [13] captured real VoIP traffic traces (about 150 000 calls) from FastWeb, an Italian telecom operator. The call duration probability distribution obtained is reproduced in Figure 6, indicated with a solid line. To illustrate qualitatively the degree to which the experimental results differ from an exponential distribution, a broken line is used in Figure 6. As can be seen, the differences are substantial, and no straightforward approximation of the experimental data with standard distributions is available.

The experimental data from [13] yields an average call duration $E(D) = 117.31$ s and a standard deviation $\sigma(D) = 278.74$; thus, the coefficient of variation $C_V = \sigma(D)/E(D) = 2.37$ (for the exponential distribution $C_V = 1$).

To achieve an analytic approximation of the experimental data, a combination of some standard distributions can be used, for example,

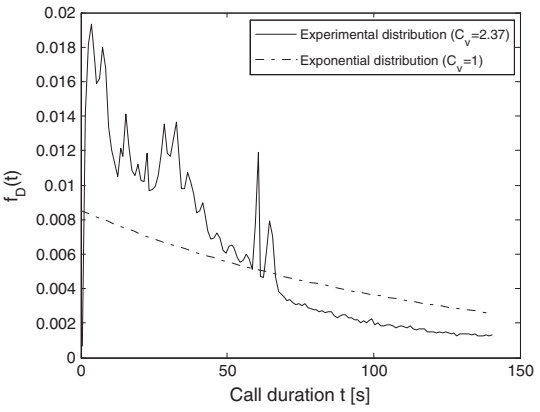

Figure 6. Voice over IP call duration—comparison of experimental and exponential probability distributions.

$$f_D(t) = \begin{cases} \frac{1}{1.55t\sqrt{2\pi}} e^{-\frac{(\ln(t)-3.8)^2}{4.805}} & \text{for } 0 \le t < 27.5 \\ 0.000114e^{-0.00114t} + 0.027252e^{-0.03028t} & \text{for } 66.5 < t \le 27.5 \\ \frac{1}{1.55t\sqrt{2\pi}} e^{-\frac{(\ln(t)-3.8)^2}{4.805}} & \text{for } 66.5 \le t \le 455 \end{cases} \quad (14)$$

The aforementioned analytic approximation is quite complex and is of little practical use for our purposes, that is, for establishing the dependence of the insertion rate *IR* on some simple enough characterisation of the call duration distribution.

Certainly, the experimental data presented are not representative for IP telephony in general. However, it proves that for different applications of VoIP, including steganographic applications, the call duration probability distribution is far from exponential.

A reasonably wide range of call distribution types can, however, be achieved and effectively analysed/used with the two-parameter Weibull distribution and appropriately chosen parameters: the shape parameter $k > 0$ and the scale parameter $\lambda > 0$. The complementary cumulative probability distribution function ($\bar{F}_D$) and the probability density function (f_D) are as follows:

$$\bar{F}_D(t;k,\lambda) = e^{-\left(\frac{t}{\lambda}\right)^k}$$
$$f_D(t;k,\lambda) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k} \quad (15)$$

The average call duration and the coefficient of variation C_V for this distribution are equal

$$E(D) = \lambda\Gamma\left(1+\frac{1}{k}\right)$$
$$C_V = \frac{\sqrt{\lambda^2\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]}}{\lambda\Gamma\left(1+\frac{1}{k}\right)} \quad (16)$$

The λ parameter was set to achieve the aforementioned experimental average call duration time $E(D) = 117.31$, and the k parameter was varied to obtain a wide range of C_V values. In Table I, the analysed values are summarised.

In Figure 7, the Weibull probability distribution is depicted for the parameters from Table I to illustrate the resulting wide range of distribution shapes. Note that for $k = 1$, the Weibull distribution equals the exponential distribution ($C_V = 1$); for $k = 2$, it becomes the Rayleigh distribution ($C_V = 0.52$); and for $k = 3.4$, it resembles the normal distribution ($C_V = 0.32$).

5.2. The dependence of *IR*(*t*) on the estimated remaining average call duration

The following method for determining $IR(t)$ was originally proposed in [6]. Here, the method is extended by considering the call quality, and the method is analysed in more detail.

For an arbitrary instant of a call, the average residual call duration is well known to be equal to

$$E(R) = \frac{E(D)^2}{2E(D)} \quad (17)$$

or, equivalently,

$$E(R) = \frac{C_V^2 + 1}{2} E(D) \quad (18)$$

Suppose that, at the beginning of a call, the insertion rate is set to $IR(0) = S/E(D)$, where S is the amount of data to be sent covertly. If $C_V > 1$, then $E(R) > E(D)$, which seems to be the case for VoIP real-world calls as indicated earlier, then, beginning from some arbitrary instant of the call, we may decrease the insertion rate to $IR = S/E(R)$, which is beneficial from the point of view of call quality and resistance to detection of the hidden data.

That discussed earlier indicates that it is reasonable to make the insertion rate dependent on the elapsed time of a call. It is nevertheless not practical to use the classical definition of residual call duration because this definition involves an arbitrary time instant and not the current call duration. We are rather interested in the expected call duration on the condition that it has already lasted for t units of time:

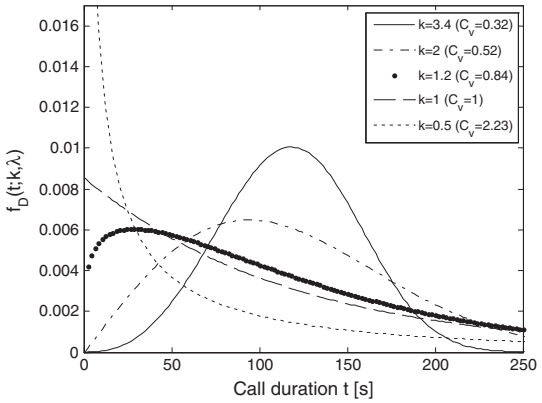

Figure 7. Weibull distribution for various k, λ and C_V.

Table I. Weibull distribution parameters k and λ and corresponding C_V values.

Weibull parameters	$k = 3.4$, $\lambda = 130.57$	$k = 2$, $\lambda = 132.37$	$k = 1.2$, $\lambda = 124.71$	$k = 1$, $\lambda = 117.31$	$k = 0.5$, $\lambda = 58.65$
C_V	0.32	0.52	0.84	1.00	2.23

$$E(D|D>t)=\frac{1}{P(D>t)}\int_{t}^{\infty}xf_{\mathrm{D}}(x)\mathrm{d}x \qquad (19)$$
$$=t+\frac{1}{F_{\mathrm{D}}(t)}\int_{t}^{\infty}\bar{F}_{\mathrm{D}}(x)\mathrm{d}x$$

for random variable D, which has values in the range $[0, \infty)$. This equation leads to the following estimations.

For $t=0$ $E(D|D>0)=E(D)$:

For every t,

$$E(D|D>t)\geq t \qquad (20)$$
$$E(D|D>t)\geq E(D)$$

because

$$E(D|D>t)\geq\frac{1}{P(D>t)}\int_{t}^{\infty}tf_{\mathrm{D}}(x)\mathrm{d}x=t$$

and

$$E(D|D>t)=t+\frac{1}{F_{\mathrm{D}}(t)}\int_{t}^{\infty}\bar{F}_{\mathrm{D}}(x)\mathrm{d}x\geq t+\int_{t}^{\infty}\bar{F}_{\mathrm{D}}(x)\mathrm{d}x=$$
$$=t+\int_{t}^{\infty}\bar{F}_{\mathrm{D}}(x)\mathrm{d}x-\int_{t}^{\infty}\bar{F}_{\mathrm{D}}(x)\mathrm{d}x\geq E(D) \qquad (21)$$

It is worth noting that for the exponential distribution, $E(D|D>t)=t+E(D)$.

Using the aforementioned estimations, it is possible to determine the set of admissible values for $E(D|D>t)$, which is illustrated in Figure 8.

The upper limit of $E(D|D>t)$ is as follows:

$$E(D|D>t)=\frac{1}{P(D>t)}\left(E(F)-\int_{0}^{t}xf_{\mathrm{D}}(x)\mathrm{d}x\right)\leq\frac{E(D)}{P(D>t)} \qquad (22)$$

For the two-parameter Weibull distribution considered in Section 4.1,

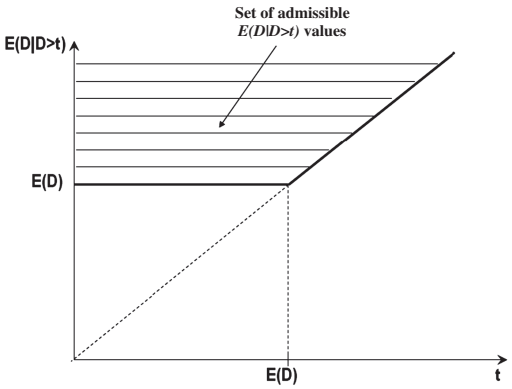

Figure 8. Admissible values for $E(D|D>t)$.

$$E(D|D>t)=t+\mathrm{e}^{\left(\frac{t}{\lambda}\right)^{k}}\int_{t}^{\infty}\mathrm{e}^{-\left(\frac{x}{\lambda}\right)^{k}}\mathrm{d}x \qquad (23)$$

and

$$E(D|D>t)\leq\mathrm{e}^{\left(\frac{x}{\lambda}\right)^{k}}\lambda\Gamma\left(1+\frac{1}{k}\right) \qquad (24)$$
$$E(D|D>t)\geq t$$
$$E(D|D>t)\geq\lambda\Gamma\left(1+\frac{1}{k}\right)$$

For the parameters chosen from Table I, we obtain the results shown in Figure 9. This figure shows also the $E(D|D>t)$ function for the experimental data presented in Figure 6.

The curves from Figure 10 may be approximated with good accuracy, as follows:

$$E(D|D>t)\approx 1.32C_{\mathrm{V}}+t\sqrt{C_{\mathrm{V}}}+0.59[\mathrm{min}] \qquad (25)$$

If $S_{\mathrm{R}}(t)$ is the amount of data remaining to be sent covertly at instant t of the call,

$$S_{\mathrm{R}}(t)=S-\int_{0}^{t}IR(x)\mathrm{d}x \qquad (26)$$

then the insertion rate at time t is

$$IR(t)=\begin{cases}\dfrac{S_{\mathrm{R}}(t)}{E(D|D>t)} & \text{for } IR(t)<IR_{\mathrm{Q}}(t)\\ IR_{\mathrm{Q}}(t) & \text{for } IR(t)<IR_{\mathrm{Q}}(t)\end{cases} \qquad (27)$$

where $IR_{\mathrm{Q}}(t)$ is calculated as described in Section 3.

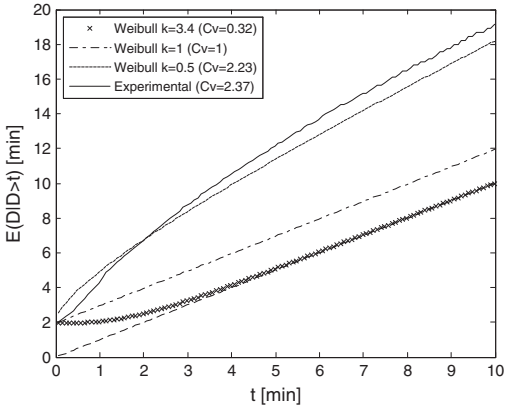

Figure 9. $E(D|D>t)$ for different Weibull distributions and for the experimental data distribution.

DOI: 10.1002/sec

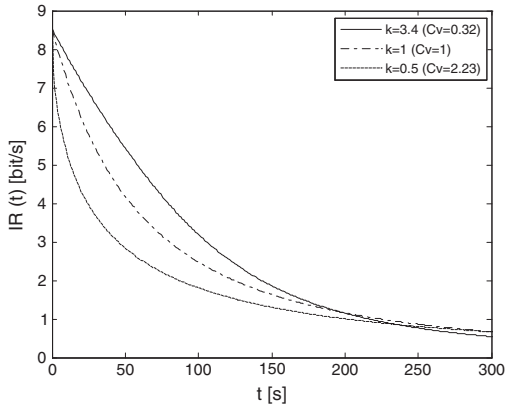

Figure 10. $IR(t)$ for chosen Weibull distributions, $S = 1000$ bits ($IR(t) < IR_Q(t)$).

With the results presented in Figure 9 and Equation 27, and assuming that $S = 1000$ bits, the $IR(t)$ functions for the chosen Weibull distributions are presented in Figure 10. For the sake of simplicity, we assumed that $IR(t) < IR_Q(t)$; that is, there are no limitations related to call quality. These limitations are considered in Figure 11.

Consider that if $IR(t) > IR_Q$ for $t < t'$, then

$$\int_0^{t'} IR(t)dt - t' IR_Q \tag{28}$$

describes this part of the steganogram, which will be sent if we do not consider the limitation $IR(t) < IR_Q(t)$ in the range $[0, t')$. Such an 'arrear' can be aligned by increasing $IR(t)$ for $t > t'$ (with the limitation that $IR(t) < IR_Q(t)$); this situation is illustrated in Figure 12 with the $IR(t) + IR^*(t)$ curve, which can, for example, be expressed as

$$IR^*(t) = \frac{\int_0^{t'} IR(t)\mathrm{d}t - t' \cdot IR_Q}{E(D|D > t')} \tag{29}$$

In Figures 12–14, the dependence of $IR(t)$ on the steganogram size under the limitation $IR(t) < IR_Q(t)$ is presented for given moments of the VoIP call (for the results obtained, we assumed the same probability distributions, and their parameters as in the previous calculations).

Figure 15 presents the total effect, or 'gain', from applying the procedure described earlier, which relates $IR(t)$ and $E(D|D > t)$ and which results from decreasing $IR(t)$ when compared with its initial value $IR(0)$. This effect was desired and was aimed at the following: as the call proceeds, the IR is adjusted (decreased) according to the expected remaining duration of the call, which is, as already mentioned, beneficial from the point of view of voice quality and resistance to steganalysis. In quantitative terms, the decrease in $IR(t)$ (notated by $X(t)$) is expressed by Equation 30 and the total gain (notated by Z) by Equation 31.

$$X(t) = IR(0) - IR(t) = \frac{S}{E(D)} - \frac{S - \int_0^t IR(x)\mathrm{d}x}{E(D|D > t)} \tag{30}$$

$$Z = \int_0^T X(t)\mathrm{d}t \tag{31}$$

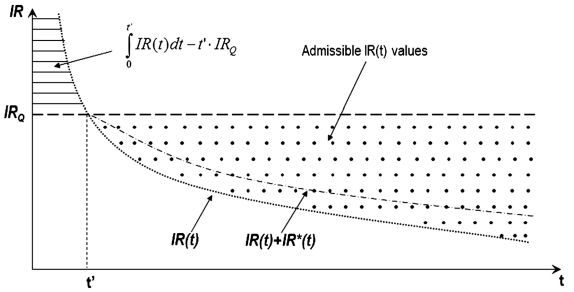

Figure 11. Relationship between $IR(t)$ and IR_Q.

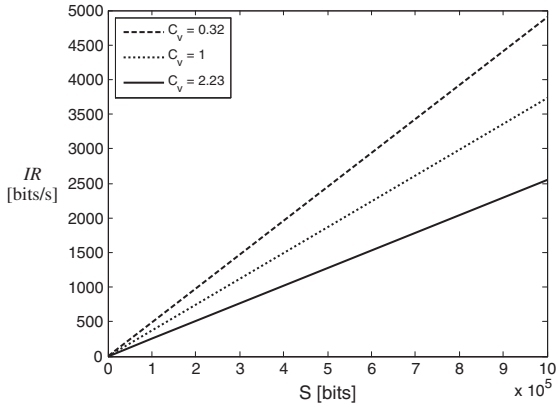

Figure 12. Dependence of $IR(t)$ on S, for $t = 60$ s and chosen C_V values.

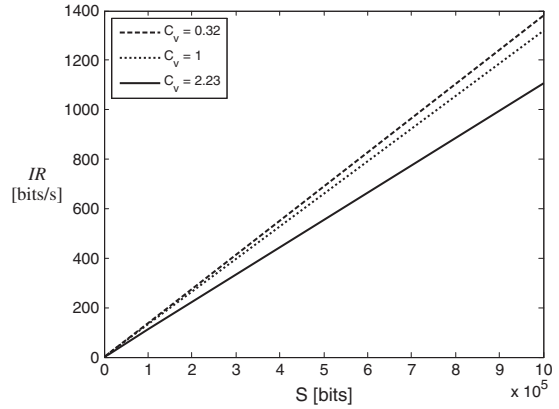

Figure 13. Dependence of $IR(t)$ on S, for $t = 180$ s and chosen C_V values.

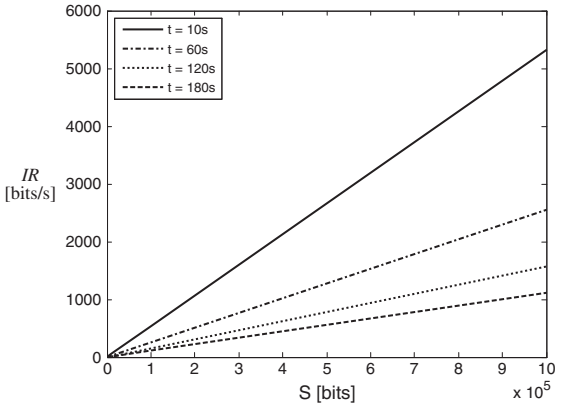

Figure 14. Dependence of $IR(t)$ on S, for chosen moments of VoIP call for $C_V = 2.23$.

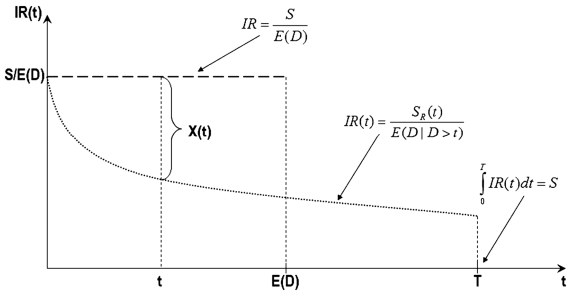

Figure 15. The effect of using $IR(t)$ based on $E(D|D > t)$.

$X(t)$ can be also related to the call quality expressed in the MOS scale, as follows. For a fixed, constant $IR = S/E(D)$, the call quality can be expressed as

$$MOS_{E(D)}(t) = \alpha \cdot \exp\left(\beta \cdot \left(p_N(t) + p_{E(D)}\right)\right) + \gamma \quad (32)$$

For the case of dependence of $IR(t)$ on $E(D|D > t)$, the call quality is

$$MOS_{E(D|D>t)}(t) = \alpha \cdot \exp\left(\beta \cdot \left(p_N(t) + p_{E(D|D>t)}(t)\right)\right) + \gamma \quad (33)$$

where $p_{E(D)}$ and $p_{E(D|D>t)}$ denote LACK packet loss probability for both of the aforementioned cases, respectively. That is, why call quality 'gain' equals

$$\Delta MOS_X(t) = \alpha \cdot \exp\left(\beta \cdot \left(p_N(t) + p_{E(D|D>t)}(t)\right)\right) - \alpha \cdot \exp\left(\beta \cdot \left(p_N(t) + p_{E(D)}\right)\right) \quad (34)$$

is because probabilities $p_{E(D)}$ and $p_{E(D|D>t)}$ can be expressed as follows

$$p_{E(D)} = \frac{IR(0)}{N_P \cdot P_P} \quad p_{E(D|D>t)}(t) = \frac{IR(t)}{N_P \cdot P_P} \quad (35)$$

Thus,

$$\Delta MOS_X(t) = \alpha \cdot \exp\left(\beta \cdot \left(p_N(t) + \frac{IR(0)}{N_P \cdot P_P}\right)\right) \times \left(\exp\left(\frac{-\beta \cdot X(t)}{N_P \cdot P_P}\right) - 1\right) \quad (36)$$

5.3. Dependence of *IR(t)* on the estimated probability of the remaining time of the call

Adjusting $IR(t)$ based on the estimated probability of the remaining time of the call is a proposed hidden data insertion procedure for LACK that has never been considered before.

In a previous subsection, we considered the problem of adjusting $IR(t)$ based on the estimated average call duration $E(D|D > t)$. In this section, we describe adjusting the $IR(t)$ based on $P(D > T|D > t)$, that is, the probability that the call will last longer than T under the condition that it already has lasted to $t \le T$:

$$P(D > T|D > t) = \frac{\bar{F}_D(T)}{\bar{F}_D(t)} \quad (37)$$

Hereafter, we analyse the dependence of $IR(t)$ on the T value, which results from fulfilling the condition $P(D > T|D > t) \ge \xi$, for a given t from the range $[0, \infty)$ and ξ from the range [0, 1]. As considered in this paper, the Weibull probability distribution is equal to the following:

$$P(D > T|D > t) = e^{\frac{-T^k + t^k}{\lambda^k}} \quad (38)$$

Thus,

$$T_\xi(t) \le \sqrt[k]{t^k - \lambda^k \ln\xi} \quad (39)$$

If the remaining hidden data left to be sent at moment t is $S_R(t)$, then

$$IR(t) = \frac{S_R(t)}{T_\xi(t) - t} \ge \frac{S_R(t)}{\sqrt[k]{t^k - \lambda^k \ln\xi} - t}, \text{ for } IR(t) < IR_Q(t) \quad (40)$$

$$IR(t) = IR_Q(t), \quad \text{for } IR(t) \ge IR_Q(t)$$

Figures 16–18 illustrates the $IR(t)$ curves for Weibull distributions for chosen C_V values, chosen ξ and $S = 1000$ bits of steganogram. We assumed that $IR(t) < IR_Q$. The problem related to limiting $IR(t)$ by $IR_Q(t)$ is analogous to

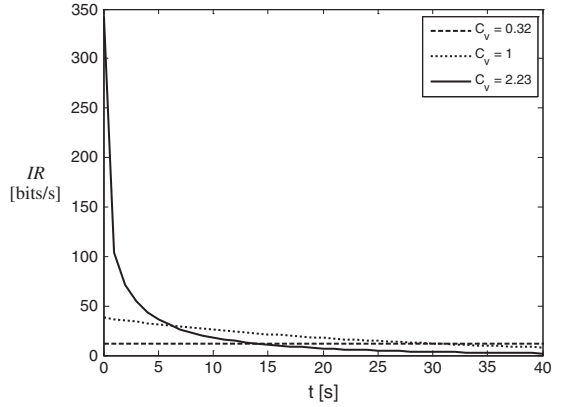

Figure 16. $IR(t)$ for chosen C_V values and $\xi = 0.8$.

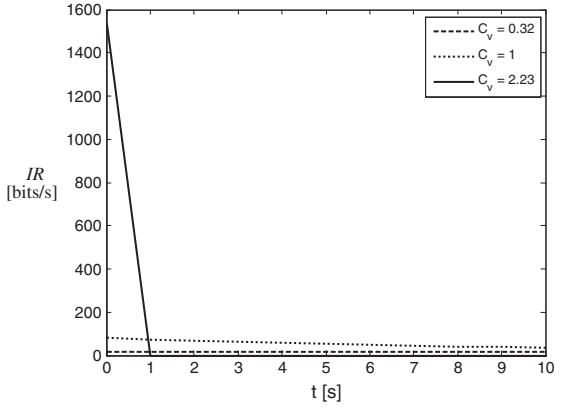

Figure 17. $IR(t)$ for chosen C_V values and $\xi = 0.9$.

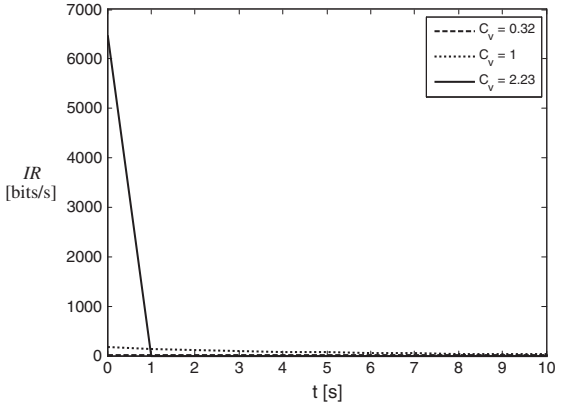

Figure 18. $IR(t)$ for chosen C_V values and $\xi = 0.95$.

the problem in the previous subsection (see Figure 11), as is the solution.

Figures 19–21 presents the dependence of $T_\xi(t)$ for the Weibull distribution and the chosen values of C_V and ξ.

The curves from Figure 19 can be approximated with good accuracy, as follows:

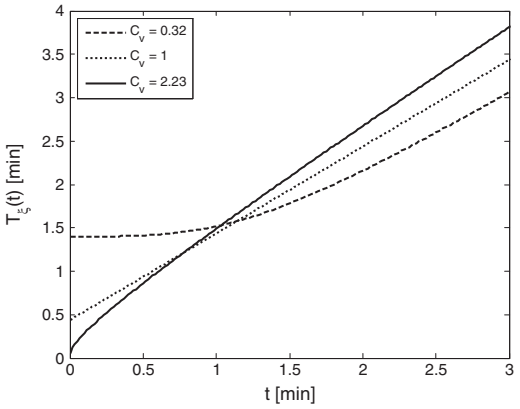

Figure 19. Dependence of $T_\xi(t)$ on t for chosen values $C_V = 0.32$, 1 and 2.23, and $\xi = 0.8$.

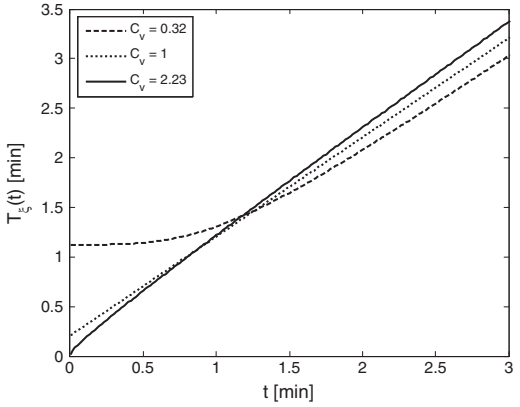

Figure 20. Dependence of $T_\xi(t)$ on t for chosen $C_V = 0.32$, 1 and 2.23, and $\xi = 0.9$.

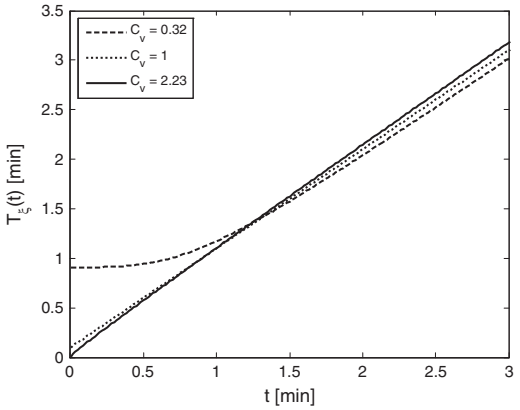

Figure 21. Dependence of $T_\xi(t)$ on t for chosen values $C_V = 0.32$, 1 and 2.23, and $\xi = 0.95$.

$$T_\xi(t) \approx -0.06C_V^2 + C_V(0.05t + 0.32) + 0.95t + 0.17 \quad (41)$$

Analogous approximations can be achieved for the other ξ values.

In Figures 22–24, the dependence of $IR(t)$ on the steganogram size for given moments of call is presented under the assumption $IR(t) < IR_Q(t)$. For the results obtained, we assumed the same probability distributions and their parameters as in the previous calculations.

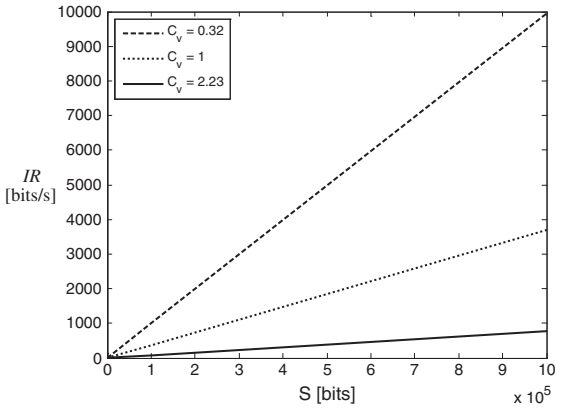

Figure 22. Dependence of $IR(t)$ on S, for $t = 60$ s and chosen C_V values.

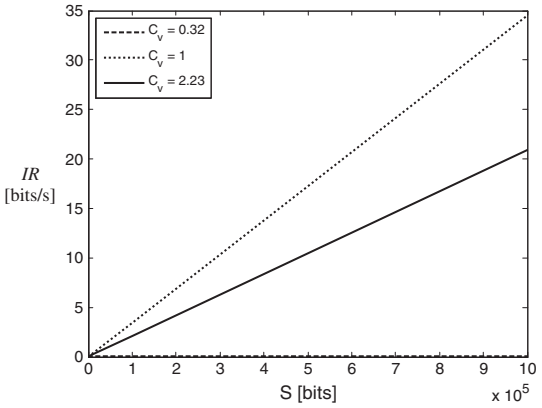

Figure 23. Dependence of $IR(t)$ on S, for $t = 180$ s and chosen C_V values.

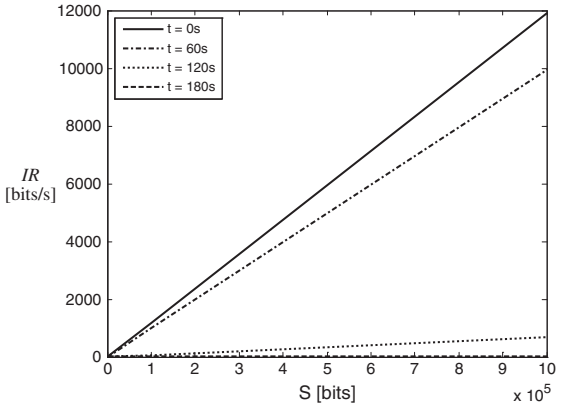

Figure 24. Dependence of $IR(t)$ on S, for chosen moments of call and $C_V = 0.32$.

5.4. Comparison of the methods of adjusting $IR(t)$ based on $E(D|D > t)$ and $P(D > T|D > t)$

In Figures 25–27, comparison of methods of adjusting $IR(t)$ for both methods presented in subsections 5.2 (based on $E(D|D > t)$) and 5.3 (based on $P(D > T|D > t)$) are presented for chosen parameters: $S = 1000$; $C_V = 0.32$, 1 and 2.23; and $\xi = 0.8$, 0.9 and 0.95. To simplify the comparison, we assumed that $IR(t) < IR_Q(t)$; thus, there are no limitations related to call quality.

With the figures presented earlier and the analyses carried out in the previous subsection, we can formulate the following conclusions: let $IR_{E(D|D>t)}(t)$ and $IR_{P(D>T|D>t)}(t)$ denote hidden data insertion rates for a method based on $E(D|D > t)$ and $P(D > T|D > t)$, respectively.

For the beginning of the call, $IR_{E(D|D>t)}(t) \leq IR_{P(D>T|D>t)}(t)$ ($t \leq t'$ and depends mainly on C_V). If $IR_Q(t) \leq IR_{E(D|D>t)}(t)$

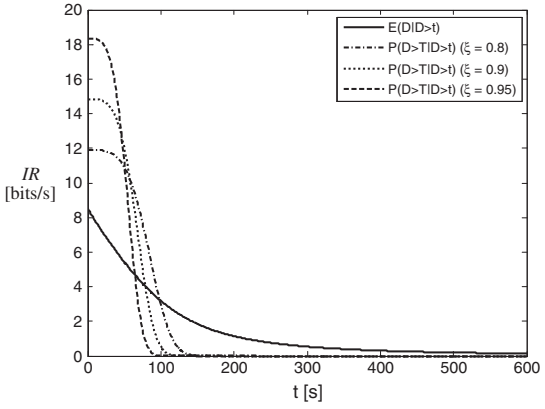

Figure 25. Comparison of methods for adjusting $IR(t)$ for $C_V = 0.32$ and $S = 1000$ bits.

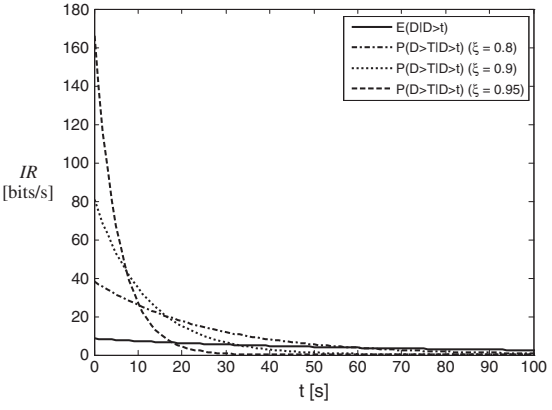

Figure 26. Comparison of methods for adjusting $IR(t)$ for $C_V = 1$ and $S = 1000$ bits.

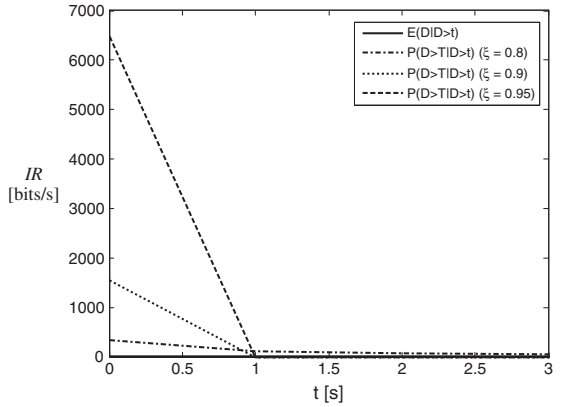

Figure 27. Comparison of methods for adjusting $IR(t)$ for $C_V = 2.23$ and $S = 1000$ bits.

in range $[0, t')$ for both methods, then we are witnessing a hidden data insertion 'arrear', which is smaller for the method based on $E(D|D > t)$. This 'arrear' must be aligned later during the call after the moment t', which requires increasing $IR_{E(D|D>t)}(t)$ and $IR_{P(D>T|D>t)}(t)$ for $t > t'$. However, the degree of increasing $IR_{E(D|D>t)}(t)$ is smaller than for $IR_{P(D>T|D>t)}(t)$, which is beneficial from the call quality and resistance to steganalysis points of view (if $IR_{P(D>T|D>t)}(t) \geq IR_f(t) \geq IR_{E(D|D>t)}(t)$ in the range $[0, t')$, then the method based on $E(D|D > t)$ does not introduce any 'arrear').

In time intervals in which $IR_{P(D>T|D>t)}(t) \geq IR_{E(D|D>t)}(t)$, the method based on $E(D|D > t)$ potentially has a lower negative influence on the call quality and resistance to steganalysis. On the other hand, in the time intervals in which $IR_{P(D>T|D>t)}(t) \leq IR_{E(D|D>t)}(t)$, the method based on $P(D > T|D > t)$ is, for the same reasons, potentially more valuable.

For greater values of $IR_{P(D>T|D>t)}(t)$ and $IR_{E(D|D>t)}(t)$, there is an increasingly greater potential steganographic bandwidth. Thus, from this point of view, the more favourable the method for given time intervals is, the greater the hidden data insertion rate. For this reason, if we consider LACK call quality and resistance to detection, it is more rational to utilise the method for adjusting $IR(t)$ based on $E(D|D > t)$. However, if we consider LACK steganographic bandwidth, then it is more advantageous to use the method based on $P(D > T|D > t)$.

Thus, the choice of the method for adjusting $IR(t)$ requires making a trade-off between the desired call quality, the resistance to steganalysis and the desired steganographic bandwidth. This trade-off depends on the context and application of LACK, which is why it cannot be established arbitrarily.

One must always take into consideration that mutual relationships between presented methods depend mainly on statistical properties of VoIP call duration and on C_V in particular. If we acknowledge that the presented experimental data (see Section 4.1) is representative for IP telephony, at least when it comes to the average and variance of the call duration, then only C_V values substantially greater than 1 should be considered. Thus, mutual relationships between $IR_{P(D>T|D>t)}(t)$ and $IR_{E(D|D>t)}(t)$ will be similar to those presented in Figure 27.

6. CONCLUSIONS AND FUTURE WORK

In this paper, the LACK steganographic method was subjected to a detailed performance evaluation. We have focused on two hidden data insertion rate IR procedures. The first procedure is based on estimating the remaining average call duration, and the second procedure is based on the estimated probability of the remaining time of the call. In addition, we have focused on the dependence of these procedures on estimated call duration and voice quality.

It was shown that the insertion rate may be effectively made dependent on the current call duration time and that this dependence can be expressed with good accuracy with the coefficient of variation of the call duration probability distribution. We have also derived analytical relations that enable making $IR(t)$ dependent on voice quality parameters. All of the derived formulae are simple and can be straightforwardly implemented. Comparison of both of the presented procedures was also included. This comparison showed that the choice of the method for adjusting $IR(t)$ requires making a trade-off between desired call quality, resistance to steganalysis and desired steganographic bandwidth.

The effectiveness of the resulting hidden data insertion procedures will depend on the accuracy of the estimated mean call duration, the coefficient of variation of the call duration and the probability distribution of voice quality for the network (sub-network), which is intended to be used for sending steganographic data with the LACK method. Thus, to evaluate realistically this effectiveness, more experimental data have to be gathered; nevertheless, the authors believe that the analysis presented in this paper indicates that LACK provides a good chance for high effectiveness.

Future work will include conducting experiments for LACK in real VoIP networks and assessing the practical steganographic bandwidth and resistance to detection for different network conditions, for types of jitter buffers and for voice codecs that can be achieved without excessively degrading the call quality.

ACKNOWLEDGEMENTS

This work was partially supported by the Polish Ministry of Science and Higher Education under Grants: N517 071637 and IP2010 025470.

The authors would like to thank R. Birke, M. Mellia, M. Petracca and D. Rossi from Politecnico di Torino (Italy) for sharing details of their VoIP experimental data.

REFERENCES

1. Schulzrinne H, Casner S, Frederick R, Jacobson V. RTP: A Transport Protocol for Real-time Applications, IETF, RFC 3550, July 2003.
2. Lian S (ed). *Multimedia Communication Security: Recent Advances*. Nova Publishers: Hauppauge, NY, April 2009.
3. Lian S, Zhang Y (eds). *Handbook of Research on Secure Multimedia Distribution*. IGI Global (formerly Idea Group, Inc): Hershey, Pennsylvania, March 2009.
4. Mazurczyk W, Szczypiorski K. Steganography of VoIP Streams, In Meersman R, Tari Z (eds). OTM 2008, Part II—Lecture Notes in Computer Science (LNCS) 5332, Springer-Verlag Berlin Heidelberg, Proc. of The 3rd International Symposium on Information Security (IS'08), Monterrey, Mexico, November 10–11, 2008, 1001–1018.
5. Mazurczyk W, Lubacz J. Analysis of a procedure for inserting steganographic data into VoIP calls, In Proc. of the PGTS '08—the Fifth Polish–German Teletraffic Symposium, Berlin, Germany, ISBN 978-3-8325-2047-2, Logos Verlag Berlin, pp. 119–128, October 6–8, 2008.
6. Mazurczyk W, Lubacz J. LACK—a VoIP steganographic method—In: Telecommunication Systems: Modelling, Analysis, Design and Management, Vol. 45, Numbers 2–3, 2010, ISSN: 1018-4864 (print version), ISSN: 1572-9451 (electronic version), Springer US, Journal no. 11235.
7. 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.
8. Friedman T, Caceres R, Clark A. RTP Control Protocol Extended Reports (RTCP XR), IETF RFC 3611, November 2003.
9. ITU-T, Recommendation G. 107, The E-Model, a computational model for use in transmission planning, 2002.
10. ITU-T, Recommendation. P.862, Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.
11. ITU-T, Recommendation. P.800, Methods for subjective determination of transmission quality, 1996.
12. T Hoßfeld, P Tran-Gia, M Fiedler. *Quantification of Quality of Experience for Edge-based Applications, 20th International Teletraffic Congress (ITC20)*, Vol. **4516**, Springer LNCS: Ottawa, Canada, June 2007; 361–373.
13. Birke R, Mellia M, Petracca M, Rossi D. Understanding VoIP from backbone measurements, 26th IEEE International Conference on Computer Communications (INFOCOM 2007), 6–12 May 2007, 2027–2035, ISBN 1-4244-1047-9.
14. Choi Y, Lee J, Kim TG, Lee KH. Efficient QoS scheme for voice traffic in converged LAN, Proceedings of International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'03), July 20–24, 2003, Montreal, Canada.
15. Miloucheva I, Nassri A, Anzaloni A. Automated analysis of network QoS parameters for voice over IP applications, D41—2nd Inter-Domain Performance and Simulation Workshop (IPS 2004).
16. Bartoli M *et al*. Deliverable 19: Evaluation of Inter-domain QoS Modelling, Simulation and Optimization, INTERMON-IST-2001-34123. URL: http://www.ist-intermon.org/overview/im-wp5-v100-unibe-d19-pf.pdf
17. Na S, Yoo S. Allowable propagation delay for VoIP calls of acceptable quality. In *Proc. of First International Workshop, AISA 2002, Seoul, Korea, August 1–2, 2002*, Vol. **2402/2002**, LNCS, Springer Berlin: Heidelberg, 2002; 469–480.
18. ITU-T Recommendation, G.711: Pulse code modulation (PCM) of voice frequencies, November 1988.