

RESEARCH

Open Access



DST approach to enhance audio quality on lost audio packet steganography

Qilin Qi*, Dongming Peng and Hamid Sharif

Abstract

Lost audio packet steganography (LACK) is a steganography technique established on the VoIP network. LACK provides a high-capacity covert channel over VoIP network by artificially delaying and dropping a number of packets in use to convey steganogram. However, the increasing loss of packets will hurt the quality of the VoIP service. The quality deterioration will not only affect the legitimate VoIP service but also constrain the capacity of the covert channel. Discrete spring transform (DST) is proven to be a way to eliminate the perceptual redundancy in the multimedia signal. In this paper, the DST is applied on the LACK so that the perceptual redundancy of the voice frames is suppressed. In this way, the less redundant VoIP frames with perceptual equivalent quality can be transmitted in a channel whose capacity is squeezed by the established covert channel. As a result, the VoIP perceptual quality can be maintained with the existence of the covert channel. Meanwhile, the proposed DST-based method demonstrates the possibilities in exploiting the perceptual space of the multimedia signal. The simulation results show that the DST on LACK achieves up to 24 % more capacity over the LACK scheme.

Keywords: DST, Perceptual quality, LACK, Steganography, VoIP

1 Introduction

Lost audio packet steganography (LACK), which was proposed in [1] and studied in [2–5], is an effective and high-capacity steganography scheme established over VoIP network. LACK takes advantages of the high data capacity of the VoIP data frame and its real-time feature. VoIP protocol is a popular technique for real-time voice communication through the Internet. The analog voice signal is sampled and packed in the VoIP voice frames to be transmitted over the IP network. In order to realize the real-time voice communication, VoIP protocol, which is also considered to be a real-time transport protocol (RTP), demands a very critical packet delay requirement. It cannot afford the delay tolerance level, which is admissible for the normal data packets since in real-time scenario, there is no time to wait for the delayed packets. As a result, a considerable amount of delayed packets will be dropped off at the receiver side. Therefore, a relatively high packet dropped-off rate is considered to be legitimate and necessary for RTP network. LACK simulates this behavior by generating many

delayed packets purposely. The receiver will drop those delayed packets without decoding the payload of those packets. Nevertheless, those packets could be used to establish a covert channel to transmit secret message to designated parties by replacing the payload of those packets with the steganogram. The intended receiver will decode those delayed LACK packets instead of dropping them. One of the reasons to use RTP network to implement LACK is that those delaying and dropping behaviors are normal in the RTP network. Therefore, the LACK packets will not draw much suspicion from the network monitors. Besides undetectability, another advantage of LACK steganography is its high capacity. The high capacity derives from the high packet loss rate allowed in the RTP network. A considerable number of packets can be used as a covert channel. On the other hand, compared to the protocol header-based covert channels where usually only few bits in the protocol header can be used as covert message, the entire payload of the packets can be used to convey steganogram in LACK covert channel which incurs a large increase to the channel capacity. Some other covert channel methods can be found in [6–11].

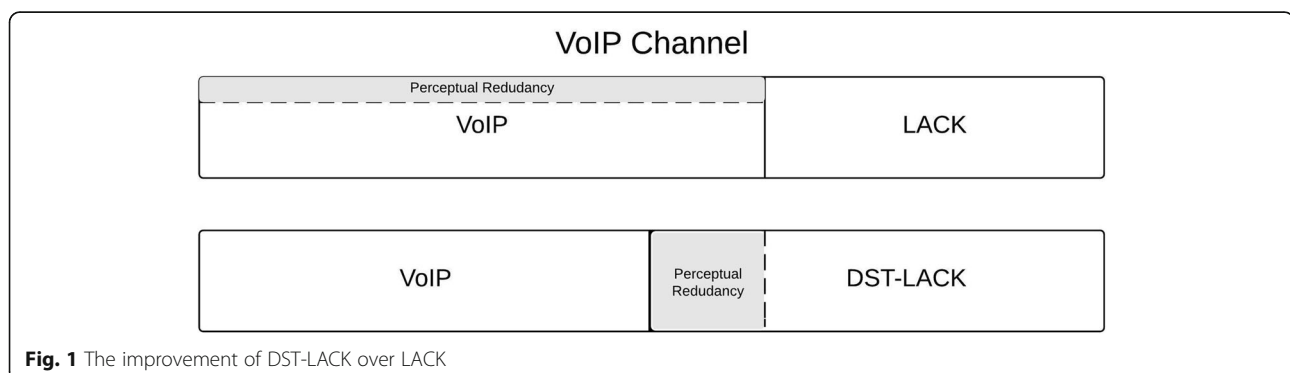
* Correspondence: qq@huskers.unl.edu
Department of Electrical and Computer Engineering, University of
Nebraska-Lincoln, Lincoln, USA

Though LACK provides a novel covert channel with high capacity and security over the VoIP network, it is limited by several factors. An important constraint of the LACK is that though it fits the voice applications where the frame sizes are small, the extension of its application to general types of multimedia services is limited. With the entire voice frame replaced by the secret message, it may not degrade the voice quality too much as long as the entire frame size is small. The quality requirement of the voice network also constrains the number of packets that can be used as covert channel. Meanwhile, the covert channel capacity is bound by the call duration distribution performance as well. Those constraints prevent the further increase of the covert channel. In this paper, discrete spring transform (DST) which is originally proposed as a multimedia steganography attack method is adopted to the LACK method so that both the covert channel capacity and the quality of service (QoS) of the voice network can be significantly improved. The proposed DST-LACK method provides a larger capacity covert channel while maintaining its undetectability.

DST was first proposed as a way to attack the steganography embedded in the multimedia signal [12–17]. The basic idea of DST is to eliminate the perceptual redundancy of the multimedia signal [18–20]. The perceptual redundancy is defined as the part of the information in the multimedia signal that cannot be perceived by human beings. Unlike the traditional digital signal information theory, human being is unable to recognize the multimedia signal as accurate as it numerically is. In other words, there is a gap between the subjective perception and numerical values of the multimedia signal. DST is a transform that tries to exploit and reduce this gap as much as possible. This reduction must not harm the perceptual quality of the multimedia signal. So we can say that DST provides a perceptual equivalence of the original signal. This equivalent signal contains less gap between the subjective perception and the numerical values of the multimedia signal. As a result, there is less room for the steganography in this equivalence. So it

can be used for steganography attack. In addition, because of less redundancy, this equivalent signal can also be used to provide quality guaranteed service in a lower data rate channel. It should be noted that the equivalent signal can only preserve the perceptual quality; therefore, the theoretical information capacity could be reduced by DST. A real-time DST algorithm is proposed in [14] for a real-time voice processing. Compared to [12, 14, 15] where we proposed DST and presented as an effective method to attack steganography, in this paper, we proposed a DST-LACK scheme to embed the steganography in the VoIP streams. We proposed and studied DST in various literatures including the references [12, 14, 15]. However, in the previous literatures, the DST is used as a way to attack the steganography. In this paper, a totally different approach is proposed to enhance the steganography capacity.

A key factor to constrain the capacity of the LACK covert channel is the QoS of the VoIP network. The excessive delayed and dropped packets will hurt the quality of the VoIP service. Meanwhile, the reduced quality of the QoS service tends to reveal the existence of the covert channel. The covert channel utilization can be defined as the number of artificially delayed packets N_d over the number of total packets N_t in a certain time T which is expressed as $U(T) = \frac{N_d}{N_t}$. The utilization is bounded by the QoS requirement of the VoIP channel and undetectability requirement of the covert channel. In order to improve the utilization and therefore to improve the capacity of the covert channel, more packets are expected to be processed. The utilization can be significantly improved by the proposed DST scheme given those requirements are still satisfied. DST will eliminate the perceptual redundancy in the voice stream. As a result, the voice stream can be fitted in a suppressed channel with the perceptual quality preserved. On the contrary, the allowable packet dropped-off rate can be lifted with the same quality requirement as in the normal LACK implementation. The improvement of the covert channel capacity is shown in Fig. 1. In the original



LACK scheme which is shown in upper figure, the perceptual redundancy is distributed in the VoIP channel. In the DST-LACK scheme, the distributed perceptual redundancy represented as gray area in Fig. 1 is squeezed together for the use of DST-LACK channel.

In the implementation of DST over LACK, an additional multi-layer buffer is involved. The DST is implemented in and around the packets which are going to be dropped. As the DST has to be implemented in the physical level, a multi-layer buffer is required. The real-time DST guarantees that the voice frames can be correctly packed in the new packets after DST.

In this paper, two schemes are proposed. One of the schemes is straightforward; the real-time DST is run on the VoIP packets without any additional adjustment. The DST parameters are randomly assigned. The advantage of this scheme is its simplicity and compatibility. It works directly on the existing LACK algorithm. Another scheme, which is more complicated, alters the DST parameters according to automatic quality control system. The automatic quality control is realized by an objective voice perceptual quality evaluator. The quality controller cooperatively controls the DST parameters and LACK insertion rates. The insertion rate (IR) was the measure used by the author who proposed the LACK in [2]. It is a key measurement for evaluating the throughput of the steganography bits in the VoIP data stream. It measures the number of steganography bits carried by the VoIP data in unit time (bits/s). Compared to the other steganography methods, LACK features with extremely high IR because the entire VoIP packets are used as the steganography transmission. This scheme provides a larger covert channel capacity. However, the LACK and DST algorithms have to be integrated in order to achieve this. The complexity is traded off by the improved performance.

In addition to allowing more packets to be used as covert channel, DST-LACK offers a flexible requirement for the frame size. Since the DST voice stream will convey the information in the packets whose payloads are substituted by the steganogram, the frame size is allowed to be larger without affecting the quality of the voice frame. Furthermore, the numerical results also show that the call duration distribution performance is improved for the DST-LACK method which means that DST-LACK is more difficult to be detected in the VoIP network. The necessary trade-off of the DST-LACK method is the processing delay and hardware overhead involved by the DST buffer.

The rest of paper is organized as follows. In Section 2, LACK and DST are reviewed, and their features and relations are analyzed in order to show the potentials to incorporate them together. Section 3 focuses on one of the DST on LACK scheme without quality control; this

scheme proves the possibility to utilize DST on LACK and shows the capacity gains by adding DST on LACK. Section 4 proposes a DST on the LACK scheme with quality control; it not only improves the capacity of the LACK but also has the LACK adaptive over different quality settings. Section 5 shows the numerical results of the two proposed DST over LACK schemes compared with the conventional LACK scheme. Section 6 concludes this paper and provides the overview of the future work.

2 LACK and DST model and analysis

2.1 LACK

LACK is proposed in [1]. The primary advantage of LACK over other covert channel methods [21–23] is its high capacity for covert communication. Unlike most of the covert channel methods which make use of some bits in the protocol header, the entire frame can be used as covert channel in the LACK method. The implementation of the LACK is on the TCP/IP layer.

On the sender side, LACK is implemented in two steps. In the first step, some random packets in the voice stream are selected. It should be noted that the maximum probability of one packet being selected is limited to satisfy the quality requirement of the voice service. The payload of the selected packets is replaced by the steganogram which refers to the secret message to be sent for the party of interest. The second step will hold those packets for a while to make sure the packets will be considered as late in the receiver side. The time for which a packet is held depends on the size of the receiver de-jitter buffer. Since the VoIP service is a time-sensitive service, very small delay is allowed for each packet, and the receiver de-jitter buffer will not be too large. The artificially delayed time must be greater than the de-jitter buffer size. However, it must be kept as small as possible to avoid detection.

The artificially delayed packets will contribute to the total packet loss rate whose tolerance depends on different codecs. Generally speaking, 1 to 5 % loss rate is acceptable for certain codecs. In this paper, the tolerable packet loss rate is increased by involving the DST in the voice frames. Consequently, the capacity of the covert channel is increased since more packets are allowed to be replaced by the steganogram. Another concern of LACK performance is the call duration distribution. When LACK is applied in a normal VoIP call, the call duration is affected by additional lost packets caused by LACK. The distribution of the call duration is a key measurement to detect whether a covert channel is established in the VoIP network or not. In order to preserve the call duration distribution after LACK, the insertion rate (IR) (bits/s) will be limited as well. DST is able to reduce space between the numerical value and the perceptual effect of the voice signal. So DST can

make the LACKed VoIP voice stream perceptually equivalent to the non-LACK voice stream without adding more call duration. In other words, the space compressed by the DST offsets the additional space consumed by the LACK covert channel. The improved undetectability by the DST can also be used to increase the capacity of the covert channel.

2.2 DST model and analysis

The motivation of the DST is to find the gap between the numerical value and perceptual effect of the multimedia signal. The gap refers to the numerical change of the digital multimedia signal which is not reflected to the perceptual effect. Under the same range of numerical difference, the change of the perceptual effect of the multimedia signal highly depends on the way to make those changes. DST is a generic way to minimize the perceptual change in the same numerical difference level. It helps eliminate some redundancy in the multimedia signal as long as the perceptual quality is the only concern. This condition is not always true, especially in the security and medical areas where the exact numerical value of the image must be maintained. However, it is not the case for the VoIP application, where the quality of the service is directly assessed by the human being who is taking part in the VoIP call. In fact, the digital values of the voice stream are changed dramatically from the sender to the receiver. As long as numerical accuracy is not important in the application, DST is able to make some extra-capacity for covert channel without causing perceptual quality deterioration. Some of the basics of DST are introduced below. The details of DST implementation can be found in our previous work [12].

Conceptually, a one-dimensional DST which works on the audio signal can be considered as a variable-density rate sampling operation. The continuous audio signal is sampled at a continuous dynamic sampling rate. A density function associated with DST can be defined as

$$D = d(t) \quad t \geq 0, d(t) > 0 \quad (1)$$

where D is the density of the sampling points on the audio signal $x(t)$ on the time axis. In order to make this operation unnoticeable to the audience compared to the original signal, two critical requirements for the density function are

1. The density function D must be continuous and differentiable in the time domain;
2. When $t_{i+1} - t_i \leq T$, $\int_{t_i}^{t_{i+1}} d(t) dt \approx 1$ where (t_i, t_{i+1}) is an arbitrary time span of the audio signal.

The first condition prevents the singularities in the audio signal that extremely deteriorate the audio

quality. The second condition requires that the audio signal remains in the same length as the original signal within a given time span. The threshold T determines the quality of the audio signal. The larger the threshold, the worse the audio signal quality would be. Usually the threshold can be selected based on the specific quality requirement and application scenario.

An important measurement related to the density function is,

$$S(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{T_0}^{T_0 + \Delta t} [d(t) - 1]^2 dt \quad (2)$$

It indicates the density change rate of the signal. The rate is bound by the quality of the audio signal as well. The integral form of the change rate can be called the accumulated signal change range which is expressed as

$$C(t) = \int_0^t [d(t) - 1]^2 dt \quad (3)$$

The DST problem can be then generalized as

$$\max_{d(t)} C(t) \quad (4)$$

which is subject to quality requirement. The optimized $d(t)$ should be able to receive the maximum signal change range among all other functions forms within the constraint of the quality.

Based on the density function, the DST can be expressed as

$$x[n] = \hat{x} \left(\int_0^n \frac{t}{f_s(t)} dt \right) \quad n = 0, 1, 2, \dots \quad (5)$$

where

$$f_s(t) = f_0 d(t)$$

An implementation of the DST in the digital form is the block-based DST. DST is a transform to variably squeeze and/or stretch the signal stream while the entire perceptual quality of the signal can be preserved. It is different from traditional re-sampling techniques which may greatly hurt the signal. DST localizes the squeeze-and-stretch process so that the effects of the change cannot be enlarged to the extent that is perceivable to human beings. Block-based DST is one of the DST implementations which is simple but effective. It is also used in the first scheme proposed in the next section to enhance LACK performance. In block-based DST, the digital signal is divided into several blocks whose size can be identical or different. The DST block parameter a_i is applied to a

block i . The processing in each block can be expressed as

$$y_i[k] = x \left[N_{i-1} + \left(k - N'_{i-1} \right) \frac{1}{a_i F_s} \right] \quad k = N'_{i-1}, \dots, N'_i \quad (6)$$

where x is the interpolated digital original signal, F . The block i ranges from (N_{i-1}, N_i) . It should be noted that new block boundaries (N'_{i-1}, N'_i) could be different from the original boundaries because the number of samples in each block could change. The boundaries of the block after DST will be progressively changed depending on the aggregated effects from all the previous blocks. To assure the perceptual quality of the signal after DST, the DST parameters can be modeled in different ways. When DST is used to attack steganography, the parameters could be randomly assigned in order to make the attack unrecoverable. To attain a better quality requirement, the parameters can be determined by a quality feedback model. The real-time DST, which is going to be used in the second scheme, adopts the automatic quality feedback model.

When quality control is involved, the voice quality of the VoIP service can be kept better. The quality of the voice is evaluated by an objective perceptual quality evaluation method [24] in a real-time manner. The structural similarity was originally proposed in [25] for image quality assessment. An expected score S_e can be set by using a base score S_b and the quality history of the signal. A formulation for the expected score can be expressed as

$$S_e = \begin{cases} S_b & T_l \leq S_{i-1} \leq T_u \\ (1-\beta)S_b - \beta S_{i-1} & S_{i-1} < T_l, S_{i-1} > T_u \end{cases} \quad (7)$$

where T_l and T_u are the predefined quality thresholds in the lower and upper bound, respectively. The expected score can be used to direct the DST parameters. The DST parameters are usually normally distributed as $a \sim \mathcal{N}(p, 1)$ where the means of the DST parameters, p , are reversely proportional to the expected scores.

In addition to apply DST for steganography attack, the proposed DST can be a high-level framework for exploring the numerical and perceptual gap existed in the multimedia data. Despite the common usage of the DST, this paper proposes a new algorithm in a new application environment. Besides proposing a new high-capacity steganography approach, by using DST in different applications, this paper explored the DST as a proper abstract model for the multimedia perceptual model.

3 DST on LACK without quality control

When LACK is applied, the legitimate VoIP service quality will be affected because of the increased delayed packets. One further reason that causes quality reduction is the loss of information in the packets replaced by the steganogram. The bandwidth of the VoIP channel is squeezed by the covert channel. In order to maintain integrity of the voice stream, one straightforward idea is to down-sample the entire voice stream to fit in the lowered bandwidth. Even though this helps to maintain the integrity of the voice stream, the quality of the voice is not improved because of the lowered sampling rate. From the information theory point of view, it is impossible to transmit the same voice stream in the same quality with a channel squeezed by the covert channel. One solution to this problem is to rearrange the perceptually insignificant parts of the signal into the dropped frames. On other hand, the perceptual quality of the voice stream is not lost because of the existence of the covert channel. A detailed way to implement the above scheme is to use the DST to voluntarily reduce those perceptual redundancies of the voice stream. DST dynamically resamples the voice stream in a variable sampling rate. The sampling rate is localized to not harm the perceptual quality of the signal. In fact, the basic idea for the DST is to make the distortion averagely distributed in the entire stream so that it is not noticeable to human beings.

The DST schemes usually keep the size of the digital signal. Nevertheless, in this implementation, the size of the digital signal is changed. In fact, the signal is allowed to change size in a smaller range. In a greater range, the signal size is still unchangeable. It can be expressed as

$$\int_{T_{\text{lack}}} d(t) dt < 1 \quad (8)$$

$$\int_{T_{\text{non-lack}}} d(t) dt \geq 1 \quad (9)$$

In the time range where LACK is present, the signal's DST density function tends to squeeze the size of the signal. In the time range without LACK, the density function compensates the size of the audio signal with a greater integral value. It not only compensates the size of the signal but also compensates the details of the signal lost in the area where LACK is sharing the channel. The condition given in the last section still holds with a greater threshold, T .

We only consider LACK delay in this section. For a general VoIP call, if the necessary number of the voice frames is N_0 , if the probability of one packet being delayed and dropped is p_d , and the size of the packet is

m , then the number of packets needed for this call M_p is

$$M_p = \left\lceil \frac{1}{1-p_d} \right\rceil \left\lceil \frac{N_0}{m} \right\rceil \quad (10)$$

For a normal voice signal, if the percentage of perceptual redundancy that can be eliminated by DST is p_r , then the voice stream that can be presented with N samples without losing perceptual quality is

$$N' = p_r N_0$$

Then the number of packets needed for this call will be

$$M_d = \left\lceil \frac{1}{1-p_d} \right\rceil \left\lceil \frac{N'}{m} \right\rceil = \left\lceil \frac{1}{1-p_d} \right\rceil \left\lceil \frac{p_r N_0}{m} \right\rceil \quad (11)$$

As a result, $M_p - M_d$ number of more packets can be used as covert channel. The probability that one packet can be dropped will be increased to

$$p'_d = \frac{(M_p - M_d) + p_d M_p}{M_p} \quad (12)$$

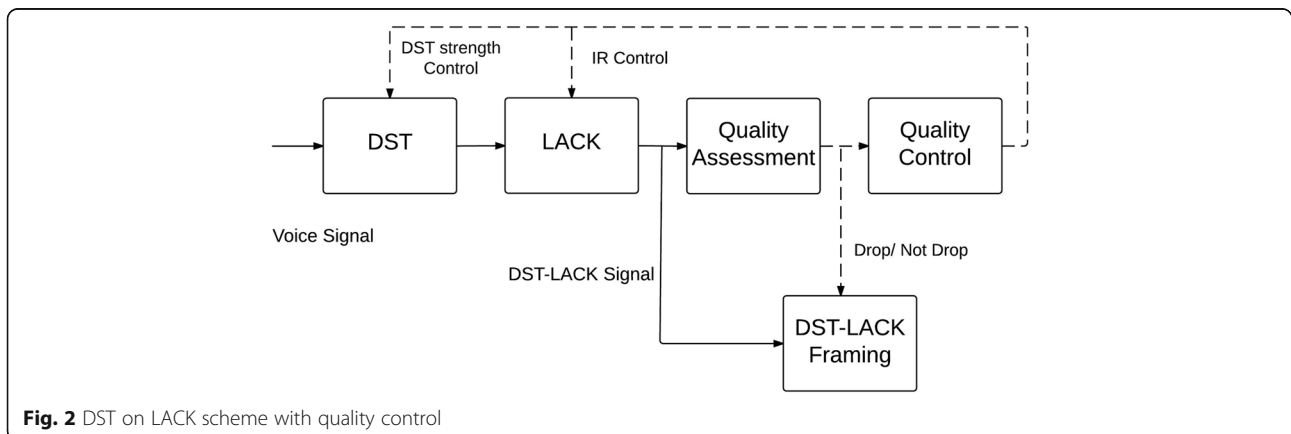
The utilization is also increased accordingly. In the first scheme, the DST parameters are chosen randomly. They are normally distributed. The expectation of the DST parameters is p_r . The real-time DST is applied on the voice stream. Each packet is considered as one block. In this section, the DST is simply applied on the existing LACK without modifying the LACK scheme itself. In fact, the timing of LACK insertion and the insertion rate of LACK can be optimized along with DST to achieve an improved channel capacity for the LACK covert channel.

4 DST on lack with quality control

In the last section, the probability of one package for LACK package is a constant. As a result, the insertion rate (IR) is constant. In this section, the insertion rate is assumed to be variable during the entire VoIP call. The variable IR will better adapt to the dynamic network environment. A higher IR is adopted when the VoIP channel is bearing lower channel noise and delay. In the DST-LACK scheme, the IR is dynamically monitored and adjusted based on automatic feedback quality control. The VoIP stream is evaluated by an objective audio quality evaluator periodically. A SSIM-based quality evaluator can be used in this application.

As we know, DST is not able to improve the LACKed VoIP stream unconditionally. The gap between the perceptual effect and the numerical value of the audio signal is limited. So the capacity of the DST is also limited to a certain extent. Though this capacity is difficult to be obtained explicitly, it can be argued that DST will become useless when the IR is higher than a threshold. The quality of the VoIP stream will inevitably deteriorate in this case. In Fig. 2, the scheme with quality control is shown. The quality assessment unit assesses the DST-LACK signal regularly and output a quality score. The quality score will direct the quality control unit to adjust the DST strength parameter and IR. The quality score will also determine if the output signal should be dropped or not. The feedback loop shown in Fig. 2 guarantees the output quality of the DST-LACKed frames.

A dual threshold empirical model is proposed based on the discussion above for quality control. When the quality score is lower than the first threshold T_1 , the DST process starts to slowly increase the DST parameters. It should be noted that the parameters are constrained in a certain range to prevent them from adversely hurting the VoIP quality. Either when the maximal allowable DST parameters are reached or when



the quality score is lower than the second threshold T_2 , the VoIP quality cannot then be improved further. The IR, at this time, must be dropped to a lower level to maintain the VoIP quality.

In first phase, DST is not applied, and the IR is increasing at a polynomial rate as

$$IR(t) = IR_0 + t^\lambda \quad \lambda \geq 2 \quad (13)$$

Once the quality score reaches the first threshold T_1 , the DST starts to operate, and the IR is expressed as

$$IR(t) = IR(t_i) + (t - t_i)^\xi \quad 1 > \xi > 0.5 \quad (14)$$

where t_i is the last time when the quality score was above T_1 which is obtained by periodically monitoring the quality of the VoIP stream. Once the quality is below the threshold, the IR remains the same. It should be noted that the IR is not decreased here and the burden to improve quality lies on DST. The DST gradually increases its parameters. Once the quality of the signal comes back above the threshold T_1 , DST temporarily suspends and locks its parameters. The proposed DST-LACK scheme is on the top of the LACK scheme, and the DST does not deteriorate the quality of the VoIP stream. On the contrary, the DST-LACK scheme provides a better quality level as demonstrated in Section 5, given the same embedding capacity. The better VoIP quality will make the covert channel more difficult to be discovered.

The quality score may also drop below the second threshold T_2 , in which case the quality is considered as unacceptable. This may be caused by many reasons including the excessive use of DST, too large IR, or channel deterioration. In this case, DST is halted and the IR returned to the initial value. The quality of the signal is kept being monitored for a certain period of time. If the quality of the signal cannot go back above T_1 , it means that the VoIP is experiencing a worse channel. In this case, the thresholds are automatically lowered and the IR is incremented as usual.

The timing selection of the LACK is also based on the quality monitoring in the VoIP channel. It selects the time range where the audio stream has more redundant capacities for DST to apply the LACK. It indicates the signal has potential for more modifications without causing noticeable distortions. A standard quality loss can be defined by applying DST on various kinds of audio signal. For a given parameter set with a standard quality loss of ϕ_0 , the stream is considered to be DST sensitive when the real-time quality loss is greater than ϕ_0 . At that moment, DST is not performed, and the LACK is performed with a lower IR. When a less quality loss is achieved by DST for a certain range of stream,

the DST and LACK can be applied immediately. The initial IR can be defined as

$$IR_0 = \begin{cases} 0, & \phi_0 < \phi \\ a + (\phi_0 - \phi)^r, & \phi_0 \geq \phi \end{cases} \quad (15)$$

5 Simulation results

In this section, simulation is carried out to show the performance of the DST-LACK scheme. The quality score is used to evaluate the perceptual quality of the VoIP audio stream. The metric method is the similarity structure [24]. The results show the improvements achieved by the DST over the LACK scheme. The experiments are conducted with two PCs connected over the Internet, the VoIP packages are transmitted between two packages, and the packages are processed by the Matlab.

Figure 3 shows the experiment results with the normal Internet environments; the package dropping rate is 5 %. The sub-figures show three different VoIP phone calls. The first is the normal conversation, the second is the classical music, and the third is the natural noise. Figure 4 shows the VoIP streams with highly distorted network environments. The package dropping rate is over 30 %. The three sub-figures use the same different audio subjects. In Figs. 3 and 4, the IRs are constant and the graphs show the achievable LACK channel capacity in a certain VoIP quality level. It demonstrates that, in the same quality requirement level, DST-LACK provides up to 24 % capacity increase. Those figures also show that a higher capacity can also be achieved even with a higher quality requirement for the DST-LACK method.

Figure 5 shows performance in terms of quality score. When the IR is set to be identical for DST-LACK and LACK methods, the DST-LACK can have a better quality score over the LACK method. It means a higher perceptual quality VoIP stream can be achieved when the DST is added to the LACK method. In Fig. 6, where the IRs are variables, the result indicates the capacity of the LACK channel. The real-time IR rate is shown in Fig. 6. It shows that a higher IR is achieved with the same quality threshold for the DST-LACK scheme.

6 Conclusion

LACK is a proven method to establish covert channel over the VoIP network. The best feature of the LACK is that it is extremely difficult to be detected. In the same time, the entire VoIP packet can be used for the covert channel which enables a relatively high-capacity covert channel communication. In this paper, DST is applied on the LACK steganography method. By differentiating

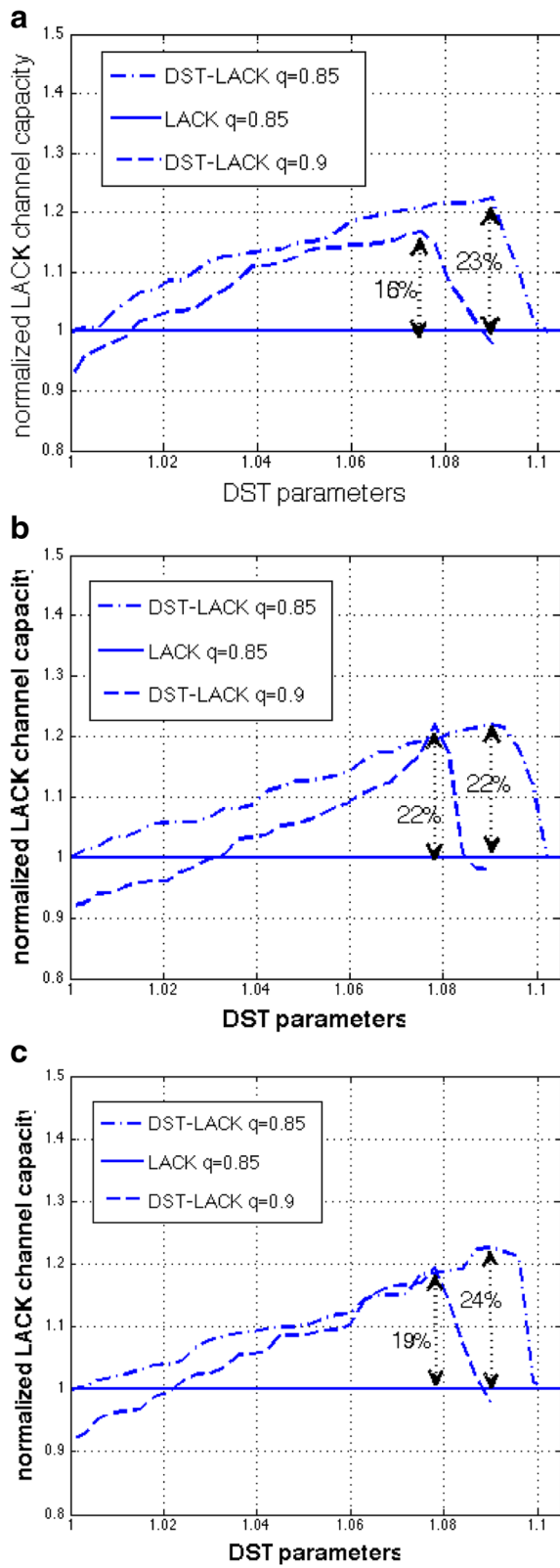


Fig. 3 a-c DST-LACK performance in terms of the capacity

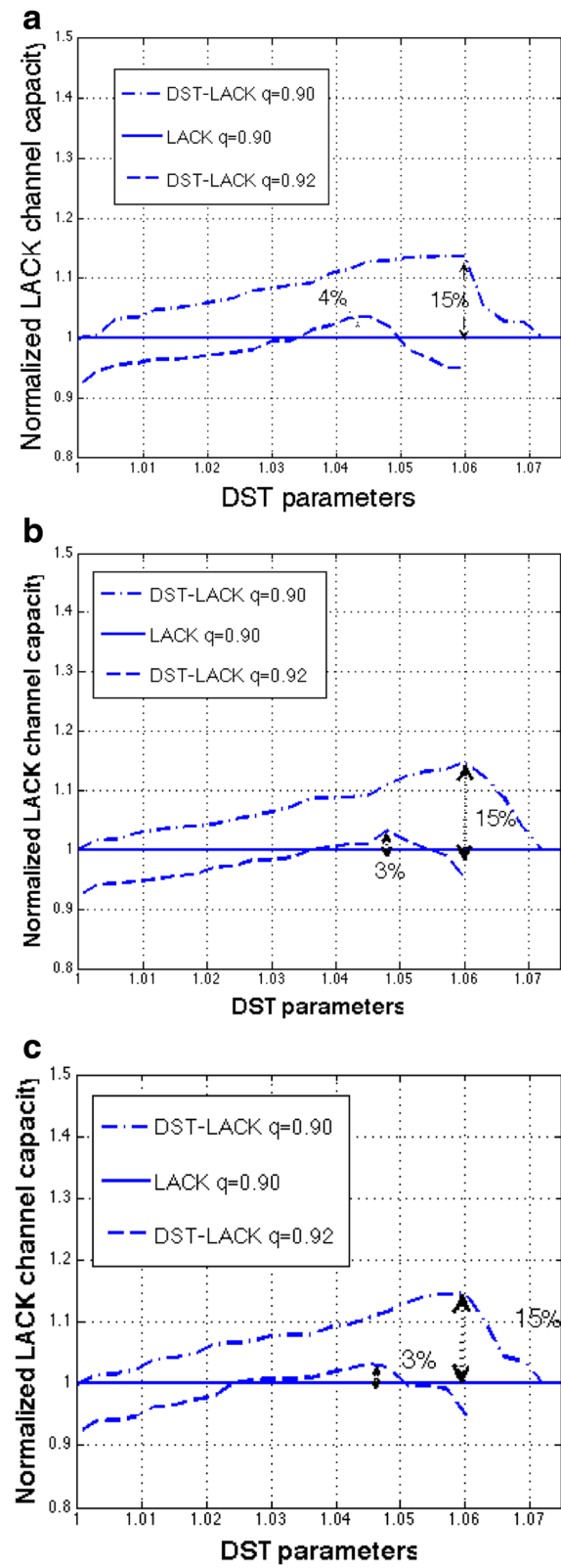


Fig. 4 a-c DST-LACK performance in terms of the capacity

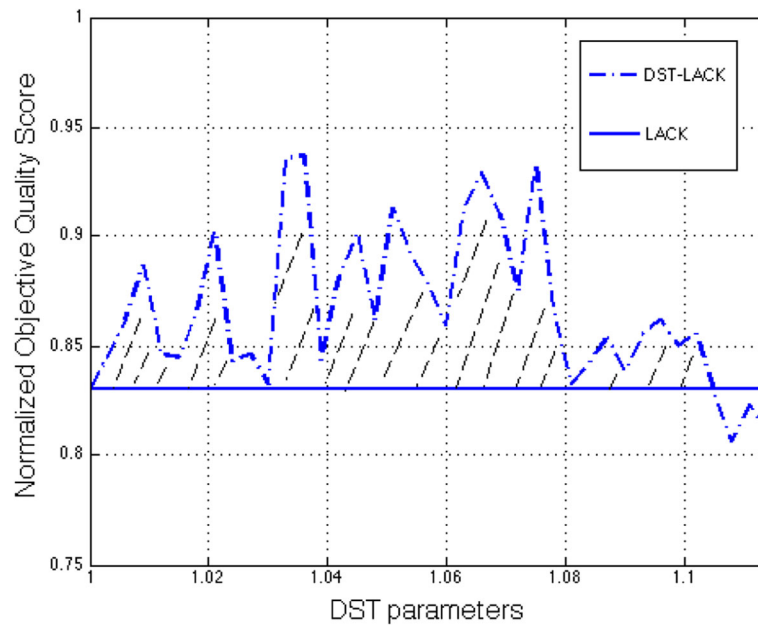


Fig. 5 DST-LACK performance in terms of the VoIP stream quality

the perceptual capacity and the numerical capacity of the VoIP data, DST further improves the capacity of the LACK channel over VoIP stream. At the same time, the quality of the VoIP stream is also improved with the existence of the LACK channel.

The simulation results show that up to 24 % capacity gain can be achieved with the same quality setting of the

conventional LACK. The results also show even with 5 % higher quality requirement, the DST over LACK still achieves up to 18 % capacity gain. A dynamic better quality score can also be achieved by adding the DST over LACK with the quality control scheme. The effectiveness of the DST working over LACK proves that DST, which is proposed to be a steganography attacking

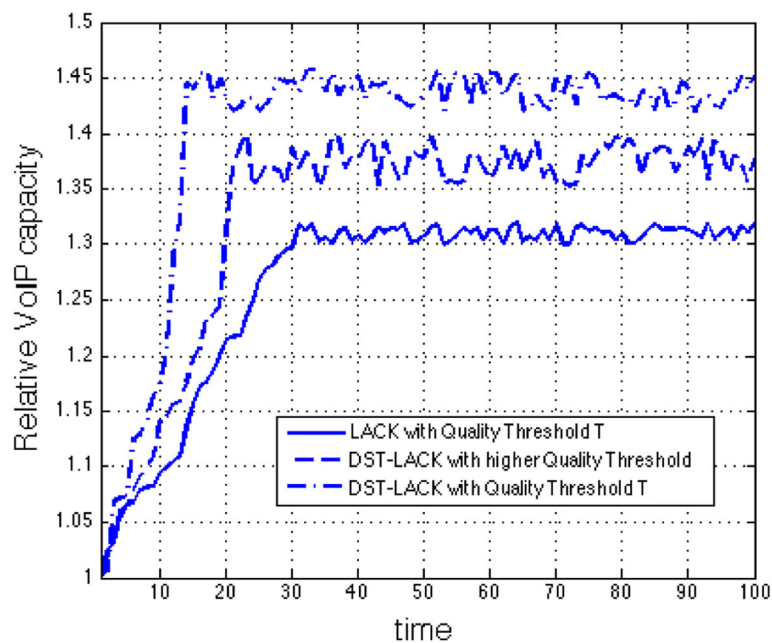


Fig. 6 DST-LACK with dynamic IR and quality control

method, can also be effective for improving certain steganography performance. Further study will focus on the theoretical limit of the improvement that the DST can provide for the LACK steganography.

Authors' contributions

QQ developed the algorithm and conducted the experiments. DP proposed the initial idea and helped to develop the idea. HS gave the instructions on the experiment design and proof reading and revised the paper draft. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 3 November 2015 Accepted: 10 September 2016

Published online: 23 September 2016

References

1. W Mazurczyk, K Szczypiorski, Steganography of VoIP streams, in *On the move to meaningful Internet systems: OTM 2008* (Springer, Berlin, 2008), pp. 1001–1018
2. W Mazurczyk, Lost audio packets steganography: the first practical evaluation. *Secur. Commun. Netw.* **5**(12), 1394–1403 (2012)
3. W Mazurczyk, VoIP steganography and its detection—a survey. *ACM Comput. Surv. (CSUR)* **46**(2), 20 (2013)
4. M Wojciech, L Józef, LACK—a VoIP steganographic method. *Telecommun. Syst.* **45**(2–3), 153–163 (2010)
5. W Mazurczyk, J Lubacz, K Szczypiorski, On steganography in lost audio packets. *Secur. Commun. Netw.* **7**, 2602–2615 (2014)
6. J Harmsen, W Pearlman, Capacity of steganographic channels. *IEEE Trans. Inf. Theory* **55**(4), 1775–1792 (2009)
7. P Moulin, J O'Sullivan, Information-theoretic analysis of information hiding. *IEEE Trans. Inf. Theory* **49**(3), 563–593 (2003)
8. J Kodovsky, J Fridrich, Quantitative structural steganalysis of Jsteg. *IEEE Trans. Inf. Forensics Secur.* **5**(4), 681–693 (2010)
9. H-M Sun, C-Y Weng, C-F Lee, C-H Yang, Anti-forensics with steganographic data embedding in digital images. *IEEE J. Sel. Areas Commun.* **29**(7), 1392–1403 (2011)
10. M Li, M Kulhandjian, D Pados, S Batalama, M Medley, Extracting spread-spectrum hidden data from digital media. *IEEE Trans. Inf. Forensics Secur.* **8**(7), 1201–1210 (2013)
11. F Rezaei, T Ma, M Hempel, D Peng, H Sharif, An antisteganographic approach for removing secret information in digital audio data hidden by spread spectrum methods, in *Communications (ICC), 2013 IEEE International Conference on*, 2013, pp. 2117–2122
12. Q Qi, A Sharp, D Peng, Y Yang, H Sharif, An active audio steganography attacking method using discrete spring transform, in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, IEEE, 2013, pp. 3456–3460
13. Q Qi, A Sharp, Y Yang, D Peng, H Sharif, "Steganography attack based on discrete spring transform and image geometrization", *10th Wireless Communications and Mobile Computing Conference (IWCMC)*, 2014
14. Q. Qi, A. Sharp, D. Peng, H. Sharif, "Realtime audio steganography attack based on automatic objective quality feedback," *Secur. Commun. Netw.* (2014). in minor revision
15. A Sharp, Q Qi, Y Yang, D Peng, H Sharif, "Frequency domain discrete spring transform: A novel frequency domain stegano-graphic attack", *9th IEEE/IET International Symposium on Communication Systems, Networks and Digital Signal Processing, (CSNDSP14)*, 2014
16. A Sharp, Q Qi, Y Yang, D Peng, H Sharif, A novel active warden steganographic attack for next-generation steganography, in *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, IEEE, 2013, pp. 1138–1143
17. A Sharp, Q Qi, Y Yang, D Peng, H Sharif, A video steganography attack using multi-dimensional discrete spring transform, in *Signal and Image Processing Applications (ICSIPA), 2013 IEEE International Conference on*, IEEE, 2013, pp. 182–186
18. Y Huang, C Liu, S Tang, S Bai, Steganography integration into a low-bit rate speech codec. *IEEE Trans. Inf. Forensics Secur.* **7**(6), 1865–1875 (2012)
19. J Shikata, T Matsumoto, Unconditionally secure steganography against active attacks. *IEEE Trans. Inf. Theory* **54**(6), 2690–2705 (2008)
20. R Anderson, FAP Petitcolas, On the limits of steganography. *IEEE J. Sel. Areas Commun.* **16**(4), 474–481 (1998)
21. H Zhao, Y-Q Shi, Detecting covert channels in computer networks based on chaos theory. *IEEE Trans. Inf. Forensic Secur.* **8**(2), 273–282 (2013)
22. S Gianvecchio, H Wang, An entropy-based approach to detecting covert timing channels. *IEEE Trans. Dependable Secure Comput.* **8**(6), 785–797 (2011)
23. X Luo, E Chan, P Zhou, R Chang, Robust network covert communications based on TCP and enumerative combinatorics. *IEEE Trans. Dependable Secure Comput.* **9**(6), 890–902 (2012)
24. Z Wang, A Bovik, H Sheikh, E Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
25. S Kandadaï, J Hardin, C Creusere, Audio quality assessment using the mean structural similarity measure, in *Acoustics, speech and signal processing*, ICASSP 2008. IEEE International Conference on, March 2008, 2008, pp. 221–224

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com