# SILK Steganography Scheme Based on the Distribution of LSF Parameter

Yanzhen Ren, Weiman Zheng, Lina Wang
Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education.
School of Cyber Science and Engineering, Wuhan University

***Abstract*—SILK, as a speech codec for real-time packet-based voice communications, which is widely used in many popular mobile Internet application, such as Skype, WeChat, QQ, WhatsApp, etc. It will be a novel and ideal carrier for information hiding. In this paper, a secure steganography scheme for SILK is proposed, which embeds secret message by modifying the LSF (Line Spectral Frequency) quantization indices based on the statistical distribution of LSF Codebook. The experimental results show that the auditory concealment of the proposed scheme is excellent, the decrease in PESQ is very small. The average hiding capacity can achieve 129 bps and 223 bps under the sampling rate of 8 kHz and 16 kHz respectively. More importantly, the proposed scheme has good statistical security. In this scheme, the statistical distribution of LSF Codebook is considered as a constraint condition to make the distribution of stego's codeword close to that of the cover audio. Under the steganlysis scheme which is referenced from the existing steganlysis scheme for G.723.1, the average correct detection rate is under 55.4% for both cover and stego audio. To the best of our knowledge, this is the first work to hide information in SILK. Based on the similar principle of speech compression, the method can be extended to other CELP codec, such as G.723.1, G.729, AMR, etc.**

## I. INTRODUCTION

Steganography[17] is an important technique for secret communication, which embeds confidential information into digital media, such as text, image, audio and video[7], [20] In recent years, with the increasing popularity of mobile communication technology, voice communication become more important in our daily life. SILK is a speech codec for real-time voice communications which is designed and developed by Skype[6] Company. It has been accepted as a IETF (Internet Engineering Task Force) standard. Now SILK is adopted as the speech codec in Opus[19] codec, which is a very popular interactive speech and audio codec. Due to its high coding efficiency and good adaptability to the limitation of the network, SILK is used in many popular social network applications, such as Skype, WeChat, QQ, WhatsApp, etc.

With the popularity of social network applications, SILK will be a novel ideal carrier to transport confidential information. To the best of our knowledge, there is still not any steganography scheme for SILK. However, Due to the similarity of the speech compression principle with the other CELP(Code Excited Line Prediction) codecs, such as G.723.1, G.729, AMR, etc, there are many existing steganography schemes can be referenced. In these steganography schemes, secret information is embedded by modifying the compression parameters during the encoding process. There are three main compression parameters in CELP codec, and they are suitable for embedding, including Fixed Codebook(FCB)[5], [14], [23], [18], Linear Prediction Coefficient(LPC)[13], [12], [22] and pitch delay[15], [2], [21], [9]. While Fixed Codebook represents the selective excitation pulse, Linear Prediction Coefficient represents the short-term prediction coefficient, and pitch delay represents the long-term prediction coefficient.

In SILK codec, the main compression parameters is almost similar with the CELP codecs. LPC is an important compression parameter in all these speech codecs which are based on the linear predictive model. In the encoding process of these codecs, the LPC coefficient are all converted to LSF (Linear Spectral Frequencies) parameters, and the LSF parameters are compressed and encoded. There are many steganography scheme based on the modification of LSF parameters. Liu[13] proposed a steganography method for low-bit-rate speech codec by replacing quantization index set in LSF, and proposing a division strategy based on the genetic algorithm. Xiao[22] proposed to embed secret message in low bit-rate speech stream by complementary neighbor vertices(CNV), which divide the codebook into groups, and modify vector quantization index of LPC coefficients in the same group, QIM based on Graph theory is applied to improve the codebook partition. Peng[16] proposed a steganography to construct 3D-Sudoku matrix in multiple vector quantization characteristics of the Line Spectrum Pair (LSP), which is related to LPC signal directly to improve the concealment and the hidden capacity. Chiang [3] proposed the scheme to divide the codebook partition based on codeword cluster. Liu[10] divided neighbor-indexed codewords into separated sub-codebooks combined with a suitable stego coding strategy and embedded secret message in LSF quantization process.

Although those steganography has good concealment and high hiding capability, the distribution features of the LSF parameter would be changed after embedding. Since that there are some effective steganalysis schemes emerged. Li[1] proposed to extract quantitative feature of LSF to detect the quantization index modulation(QIM) stagngraphy in G.723.1. Li[8][8] discovered that the correlation characteristics of split vector quantization (VQ) codewords of LSF parameter are changed after steganography, and proposed a steganlysis

This work is supported by the Natural Science Foundation of China(NSFC) under the grant NO.U1536114, NO.U1536204, and China Scholarship Council.

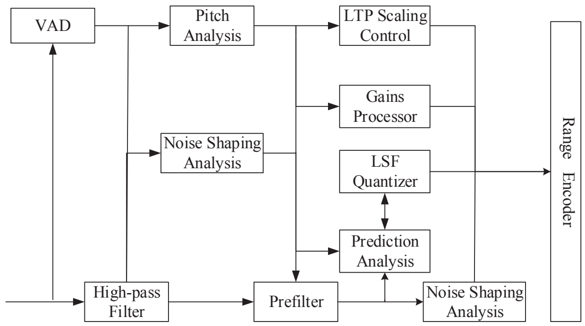



Fig. 1: SILK encoding diagram.

scheme base on Quantization Codeword Correlation Network (QCCN). The above steganlysis schemes are all effective, and have good detection accuracy rate.

In this paper, a secure steganography scheme for SILK ($SS\_DLSF$) is proposed, which embeds secret message by modifying the LSF (Line Spectral Frequency) quantization indices based on the statistical distribution of LSF codebook. The statistical distribution of LSF codebook is considered as a constraint condition to improve the statistic security of the proposed scheme.

The rest of the paper is organized as follows. The compression principle of SILK codec is introduced in Sect.2, and the distribution characteristics of LSF parameters are analyzed in detail. In Sect.3, the SILK steganography based on LSF codebook statistical distribution ($SS\_DLSF$) is introduced. Sect.4 presents the experimental results. Finally, the conclusion and future work is drawn in Sect.5.

## II. Distribution characters of LSF in SILK

### A. Compression principle of SILK codec

SILK codec is designed for the real-time voice communications in internet. It has good coding efficiency and scalability, can handle adaptive coding technology of low-rate network. SILK operates on four different sampling rates: 8 kHz for narrowband, 12 kHz for medium band, 16 kHz for wideband and 24 kHz for super wide band. SILK codec has a variable bitrate that can be set between 6 and 40kbps. The compression principle of SILK is based on Liner Prediction. An overview of the SILK encoder is shown in Figure 1. When SILK encoder works, input signal is processed by a VAD (Voice Activity Detector) and a high-pass filter. The high-passed input signal is processed by the open loop pitch estimator, which compute to the pitch lags (per 5 ms) and voiced/unvoiced classification (per 20 ms). Then, the noise shaping analysis finds gains and filter coefficients used in the prefilter and noise shaping quantizer. The excitation signal is constructed by prediction analysis to get the short and long term prediction coefficients. And LSF Quantizer, LTP Scaling and Gains Processor are respectively to find the LSF quantization indices, LTP state scaling coefficient and processed gains. Finally, quantized signal is encoded by entropy coding, as final output of SILK encoder.

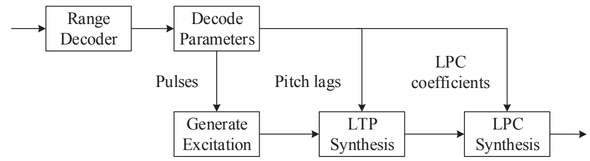



Fig. 2: SILK decoding diagram.

An overview of the decoder is shown in Figure 2. The range decoder decodes the encoded parameters, which includes three parts: pulse and gains for the excitation signal generation, pitch lags and LTP coefficients, and LPC coefficients. The pulses and gains for the excitation signal generation, as well as LTP and LSF codebook indices which are needed for decoding LTP and LPC coefficients, are used for LTP and LPC synthesis filtering to get the excitation signal respectively. So the embedding secret message must be in the three parts of parameters. This paper focus on the LSF quantization, and LSF(Linear Spectrum Frequency) vector is the representation of LPC parameters.

In the LSF quantization part, SILK encoder use the multi-stage vector codebook to achieve memory efficient and highly scalable. This structure quantize LSF vector to find LSF quantization indices. In the first stage the input is the LSF vector, and in any other stage s $>1$, the input is quantization error which is between input and codebook from the previous stage. In each stage, error sum between input and codewords is calculated as Equation 1:

$$S_e = \sum_{i=1}^{d}(x_i - cw_i) \tag{1}$$

while x represent the input in every stage, d is the LSF order, cw is the codeword value. In SILK codec, for different sampling rate, LSF vector has different orders. The LSF order is 10 when sampling rate is 8 kHz, and 16 in the other sampling rates. Besides, the codebook structure has 6 stages to quantize when the LSF order is 10; and 10 stages when the LSF order is 16. There are four codebooks with different sampling rate and frame type (voiced and unvoiced): CB0-16, CB0-10, CB1-16, CB1-10, where 0 and 1 represent voiced frame and unvoiced frame respectively, 10 and 16 represent the LPC order. Corresponding codebook is chosen according to the LPC order and frame type. Each codebook is divided into M sub-codebooks in order to quantize the input of every stage in LSF quantization, M represents the stages. Table 1 shows the codebook structure with 16 kHz sampling rate. Each sub-codebook has different numbers of codeword vector, every codeword vector has the same numbers of codewords as LSF order. The structure of the multi-stage vector codebook is shown in Figure 3.

The LSF vector quantization process is as follows: firstly, the error sums between LSF vector and each codeword vector in sub-codebook are calculated. Then the error sum of each codeword vector is sorted from small to large, and recorded the corresponding index of the codeword vector at the same

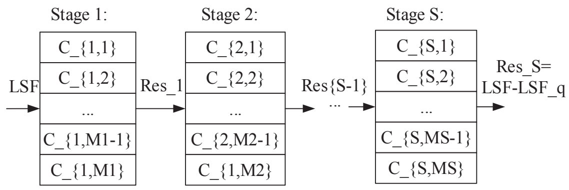



Fig. 3: Multi-Stage vector codebook in SILK.

TABLE I: Number of codeword vector with 16 kHz

| Stages/Groups | Unvoiced | Voiced |
|---|---|---|
| 0 | 32 | 128 |
| 1 | 8 | 16 |
| 2 | 8 | 8 |
| 3 | 8 | 8 |
| 4 | 8 | 8 |
| 5 | 8 | 8 |
| 6 | 8 | 8 |
| 7 | 8 | 8 |
| 8 | 8 | 8 |
| 9 | 8 | 16 |
| total | 104 | 216 |

time. After the error sum is sorted, then the error sum greater than the set threshold is discarded. And the error sum, as the input of next stage, is calculated with the codeword vector of next stage. The above steps are repeated until all stages are quantized. Finally, M indices which have the smallest error sum according M stages quantization is the output.

### B. Analysis of the distribution of LSF in SILK

The LSF vector is quantized by the multi-stage vector codebook, the result of indices are mapped to the codeword vector in the corresponding codebook. In order to design a steganography scheme with large hiding capacity, good auditory concealment, and good statistical security, the characteristics of LSF quantization is analyzed. Firstly, the modifiability of each stage vector of LSF parameter are analyzed to determine the suitable embedding domain of LSF parameters. Secondly, the usage probability of each codeword in each codebook are counted from a large number of normal audio samples to get the distribution of LSF parameter, which is the basic of the partition strategy in the proposed schemes.

*1) The Modifiability of Multi-stage LSF Parameter:*

In the SILK speech encoding process, different codebooks are used in LSF vector quantization according to the LPC order and frame type. Table 2 and Table 3 shows the the maximum value, minimum value of codeword at each stage in CB1-16 and CB0-16 respectively. From Table 2 and 3, we can find that the value of codewords in first stage is far greater than that of the other nine stages. The reason is that in the LSF quantization process, the first index represents the main features information of the LSF vector, the latter nine indices represents the error information between LSF vector and codeword vector, which will fix the error of first index information. Figure 4 shows the corresponding line chart for each codeword vector in each stage of CB1-16. There is an

TABLE II: Codeword Statistics of CB0-16

| stage(groups*order) | MAX | MIN |
|---|---|---|
| 0(128*16) | 31082 | 454 |
| 1(16*16) | 1158 | -1060 |
| 2(8*16) | 799 | -740 |
| 3(8*16) | 546 | -701 |
| 4(8*16) | 450 | -629 |
| 5(8*16) | 465 | -430 |
| 6(8*16) | 503 | -475 |
| 7(8*16) | 436 | -375 |
| 8(8*16) | 393 | -277 |
| 9(16*16) | 339 | -376 |

obvious statistical feature in Figure 4, the value of codewords increase regularly in turn from 1st order to 16th order which represents the linear growth in 0th sub-codebook, and the value of codewords are between 1012 and 31294 which is far greater than the value of the other stage. In other nine stages, the value of codewords are in a trend of fluctuations with a small range which is between -1000 and 1500. At the same time, from 1 stage to 9 stage, the value of codewords decreases, and the fluctuation range of codewords also gradually decreases, which tends to be relatively stable.

From the above analysis of multi-stage vector codebook quantization, it is possible to draw a conclusion that the 0th-stage index represents the main information of LSF vector. The modification of the 0th-stage index will introduce a great loss on speech quality. Therefore, the 0th-stage index couldn't be modified when embedding secret information. The quantization of LSF vector from 1st stage to 9th stage indicate the correction of the previous error, the error and correction range is getting smaller in the quantization process. Since the 1st-9th stage indices represent LSF quantization error, embedding secret information by modifying these indices will introduce less impact on the speech quality. While the range of 1st-9th stage codewords value gradually decrease and fluctuations trend to be stable, the impact on speech quality by modifying indices of latter stages is less than the previous stages.

Based on the above analysis, a basic experiment is carried out to test the modifiability of the 10 indices. The quality of the modified audio is evaluated by PESQ, which is one of the evaluation algorithms of voice quality proposed by the International Telecommunication Union (ITU). The PESQ is most relevant with subjective evaluation in objective evaluation algorithm of voice quality published by ITU. In this experiment, the last index is modified from the last stage, then the last two indices are modified, then the last three indices are modified, the indices are modified in turn and different numbers of indices are modified in each time. Meantime, 500 PCM audios are encoded by SILK codec and modified with the different numbers of indices in LSF vector quantization. The value of PESQ test is shown in Table 4. while the modified index $(n - m)$ represents the indices from $n$th stage to $m$th stage are modified with secret message. Table 4 shows that there is little decrease on speech quality when the 1st to 9th

TABLE III: The PESQ when modify different LSF quantization indices

| modified index | 9 | 8-9 | 7-9 | 6-9 | 5-9 | 4-9 | 3-9 | 2-9 | 1-9 | 0-9 |
|---|---|---|---|---|---|---|---|---|---|---|
| PESQ | 4.29 | 4.27 | 4.26 | 4.25 | 4.24 | 4.24 | 4.24 | 4.24 | 4.23 | 2.45 |

indices are modified, the PESQ is still above 4.23. However, there is a great influence on speech quality when the 0th index is modified, the PESQ is only 2.45. It shows that the 0th-stage index is the most important value in SILK LSF indices, it can't be modified, and is not suitable to be the hiding domain. The other stages indices are suitable to hide information by modifying the value of it, and will introduce little influence to the auditory concealment.

*2) The Distribution character of SILk LPC codebook:*

Although the existing LPC steganography schemes[13], [12], [22] embed secret messages by utilizing the original coefficient features through QIM-based codebook partition methods. These steganography schemes don't consider the feature of statistical distribution of codewords, which will change the statistical distribution of LSF parameters. Since that they are easy to be detected by the existing steganalysis scheme. The recall and precision of the existing steganalysis scheme[1] are all more than 90%. To solve this problem, the steganography scheme proposed in this paper consider the change of distribution character which is introduced by modifying of the LSF parameters, make the distribution character of the cover and stego are similar to ensure the security of steganography.

To statistic the distribution character of the LSF coefficients in SILK cover audio, 1000 PCM audios segment are encoded by SILK encoder. In the procedure of LSF vector quantization, the usage probability of each codeword vector are counted in each stage by Equation 2 and 3. The unvoiced frame and voiced frame are counted separately. To ensure the modified speech quality, main feature of original LSF coefficients need be reserved. Therefore, 0th stage codebook has not been counted, and the statistical distribution of 1st to 9th stage indices are analyzed.

$$C_i = \begin{cases} 1 & \text{x = i} \\ 0 & \text{x != i} \end{cases} \tag{2}$$

Figure 5 shows the boxplot of the usage probability of 1st, 2nd and 3rd stage LSF vector quantization of CB1-16 and CB0-16. The horizontal axis represents the 1st to 8th codebook vector and the the vertical axis represents the usage probability of each codeword vector in LSF quantization. In the boxplot, the upper and lower blue edge lines indicate the upper and lower quartiles of the data respectively. The middle red line indicates the median line. The upper and lower short blue lines indicate the abnormal value cutoff point. The red dots outsides the cutoff point range indicate outliers.

$$f_i = \frac{\sum_{i=1}^{s} C_i}{s} \tag{3}$$

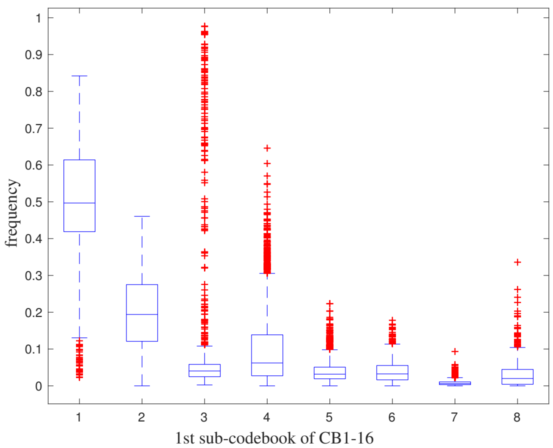
Figure 5 shows that there are some codeword vector with same distribution in each stage. Because the distribution character of codebook between cover and stego should be consistent after the secret message is embedded. Therefore, we proposed to make groups of codeword vectors with similar distribution in one stage. The steganography uses the groups with similar distribution of codeword vectors to embed secret message will not change the distribution character of codebook, and it will improve the security of steganography.

## III. The Proposed SILK steganography scheme $SS\_DLSF$

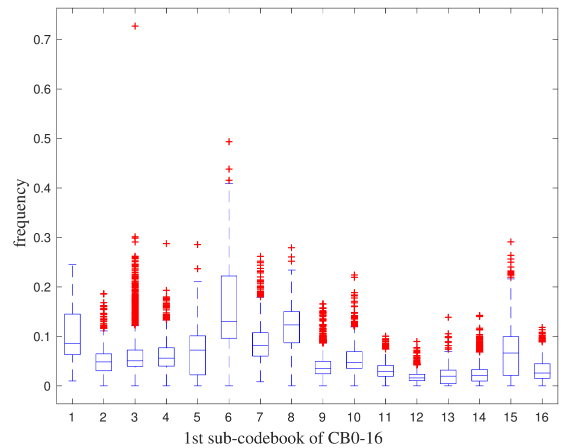
Based on the above analysis, a secure SILK steganography scheme based on distribution of LPC ($SS\_DLSF$) is proposed. The secret message is embedded by modifying LSF indices. The modifiability of indices in LSF quantization shows that the 0 index couldn't be changed in order to maintain speech quality, so the secret message will be embedded in the other 9 indices from 1st to 9th. Meanwhile, to maintain the statistical distribution of LSF codebook, the codeword vectors with similar distribution in one stage will be divided into a embedding group, and the secret message is embedded by choosing the codeword vector which represents the information of secret message in the corresponding embedding group. The statistical distribution of LSF codebook is trained by the statistical distribution experiment from large amount of cover audios.

There are four codebooks in LSF vector quantization, the number of codewords in each stage is 8 or 16. To embed the bit conveniently, the proper number of embedding group will be 2 or 4, which can be embedded in 1 bit or 2 bits secret message. The process to build the embedding group is described as follows. Firstly, the usage probability of each codeword vector is counted as Figure 5. In each codebook, the codeword vector is sorted from big to small of usage probability. Then, the difference between the usage probability of two adjacent is computed. If the difference between last usage probability and previous usage probability is lower than the default threshold, those two codeword vector can be arranged into same embedding group. Each codeword vector are traversed in turn. The number of each embedding group is limited in 1-4. Equation 4 shows the numbers of embedding secret message with corresponding codeword vectors in one embedding group. When there are 3 codeword vectors in one embedding group, 2 codeword vectors will be selected as the embedding group, the other codeword vector with the least usage probability will be removed.

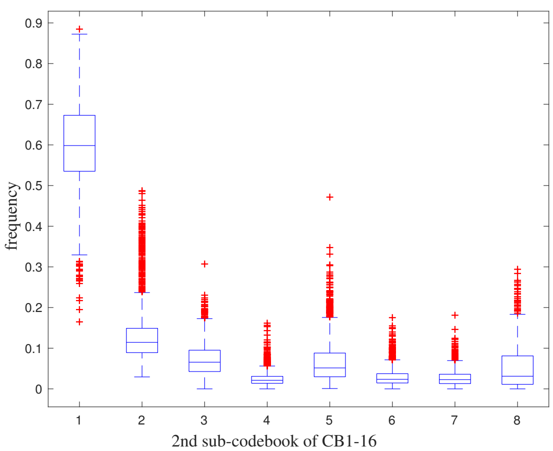
The threshold is important to the embedding group process, it should be proper to limit the number of codeword vectors in one group. If the threshold is too big, there are too many



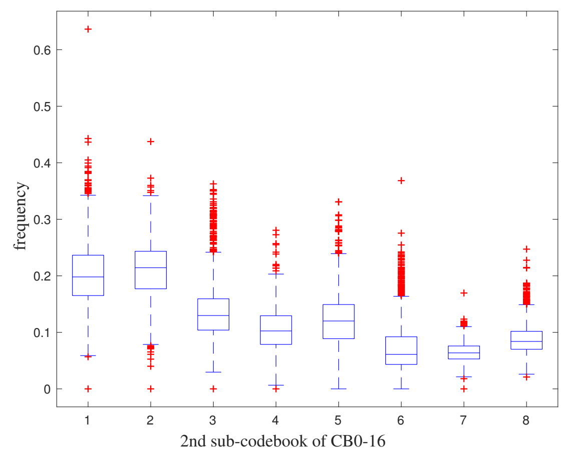

(a)



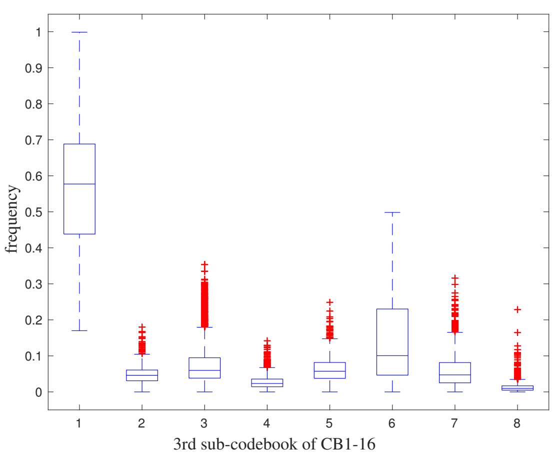

(b)



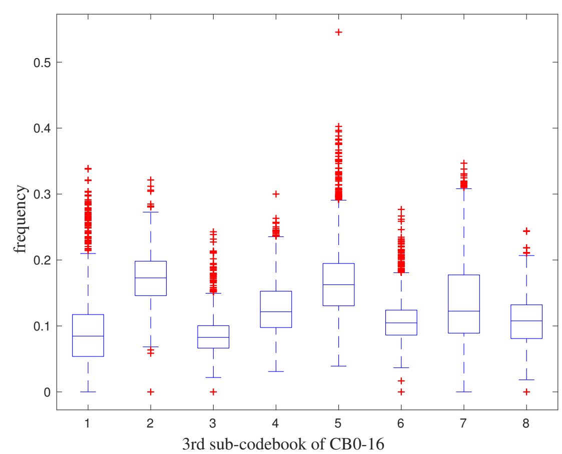

(c)




(d)




(e)




(f)

Fig. 5: The usage probability of the LSF codeword in cover

TABLE IV: The embedding group of 1st sub-codebook of CB1-16

| Group Number | Group Index |
|---|---|
| 1 | 2, 4 |
| 2 | 3, 5, 6, 8 |
| 3 | 1 |
| 4 | 7 |

TABLE V: The embedding group of 2nd sub-codebook of CB1-16

| Group Number | Group Index |
|---|---|
| 1 | 2, 3 |
| 2 | 5, 8 |
| 3 | 6, 7 |
| 4 | 1 |
| 5 | 4 |

TABLE VI: The embedding group of 1st sub-codebook of CB0-16

| Group Number | Group Index |
|---|---|
| 1 | 2, 3 |
| 2 | 4, 10 |
| 3 | 5, 15 |
| 4 | 9, 11 |
| 5 | 12, 13, 14, 16 |
| 6 | 1 |
| 7 | 6 |
| 8 | 7 |
| 9 | 8 |

TABLE VII: The embedding group of 2nd sub-codebook of CB1-16

| Group Number | Group Index |
|---|---|
| 1 | 1, 2 |
| 2 | 3, 5 |
| 3 | 4, 6 |
| 4 | 7, 8 |

codeword vectors in one group and the difference of statistical probability will be big. Otherwise if threshold is too small, the codeword vector couldn't be grouped, and the secret message couldn't be embedded. The usage probability is sorted from big to small, so the codeword vector with bigger probability can be reserved. According to the result of many experiments, the ideal threshold of sub-codebook with 8 codeword vectors is 0.35, and the ideal threshold of sub-codebook with 16 codeword vectors is 0.18. In our experiment, the embedding group is shown in Table 5, 6, 7, 8. For example, in Table 5, when the LSF vector quantization result of 1st sub-codebook of CB1-16 is 2 and the embedding secret message is 0, the LSF vector will not need be changed. When the embedding secret message is 1, the LSF vector should be changed to 4. And the codeword index 3, 5, 6, 8 in one group can be embedded in 2 bits as 00, 01, 10, 11 respectively.

$$M = \begin{cases} 1 & 0 \text{ bit} \\ 2 & 1 \text{ bit} \\ 3 & 1 \text{ bit, remove one} \\ 4 & 2 \text{ bit} \end{cases} \tag{4}$$

Based on the above analysis, the SILK steganography scheme based on LSF codebook statistical distribution feature is proposed, which is shown in figure 6. In the LSF vector quantization process, the corresponding codeword vectors with similar statistical distribution is used to embed secret message. If there is no similar codeword vectors in embedding group, the secret message can not be embedded. The procedure of embedding is as follows. Firstly, the LSF vector is quantified with Multi-stage vector codebook of SILK codec. Then the 1st to 9th indices are used to embedded secret message, the bits of secret message embedded depends on the embedding group of sub-codebook. In Figure 6, GofCw_4 indicates that there are 4 codeword vector in this embedding group, and 2 bits secret message can be embedded. GofCw_2 represents 2 codeword vectors, so 1 bit secret message can be embedded. And GofCw_1 represents there is no similar codeword vector and just one codeword vector in embedding group, the secret message couldn't be embedded. The extracting of the secret message is the inverse process of the embedding operation. In the extracting of the message, the embedding group is same as the embedding process.

## IV. EXPERIMENTS

To evaluate the performance of the proposed scheme $SS\_DLSF$, three experiments are carried out.

### A. Experiments Setup

1) Audio sequences: Audio databases are formed from 2000 original audios, 1000 audios of them are downloaded from Internet, and another 1000 audios are recorded with CoolEdit[4]. These audio samples contain digital speeches from different people and different languages, include English, Chinese, and Korean, and be converted to 2000 PCM audios with 8 kHz sampling and 16 kHz sampling rate by Adobe Audition, mono, signed 16 bit quantization, 30 seconds.

CDB: The cover database CDB include 2000 PCM audios with 8 kHz sampling rate and 2000 PCM audios with 16 kHz sampling rate are encoded by SILK codec respectively. The total number of cover SILK samples is 2000 * 2 = 4000.

2) Steganography Methods: the proposed steganography scheme $SS\_DLSF$ is used to embed secret information generated by pseudo-random sequence during the encoding process. So there are two stego database:

SDB1: The 2000 PCM audios with 8 kHz sampling rate to implement $SS\_DLSF$. All PCM audios from CDB are embedded with EBR of 10%, 20%, 30%, 50% and 100%, so the total amount of SILK audio steganmgrams is 2000 * 5 =

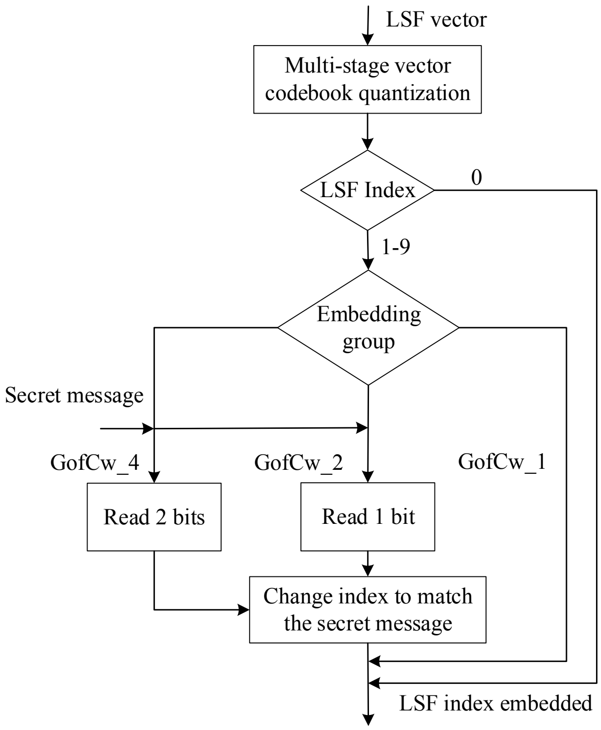



Fig. 6: Block diagram of the proposed steganography.

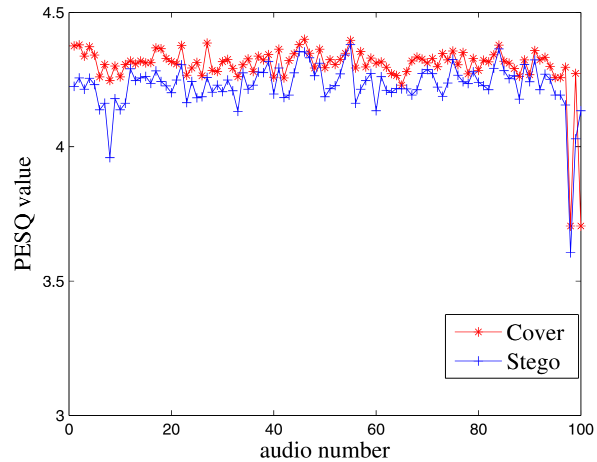



Fig. 7: The comparison of PESQ with 8 kHz sampling rate

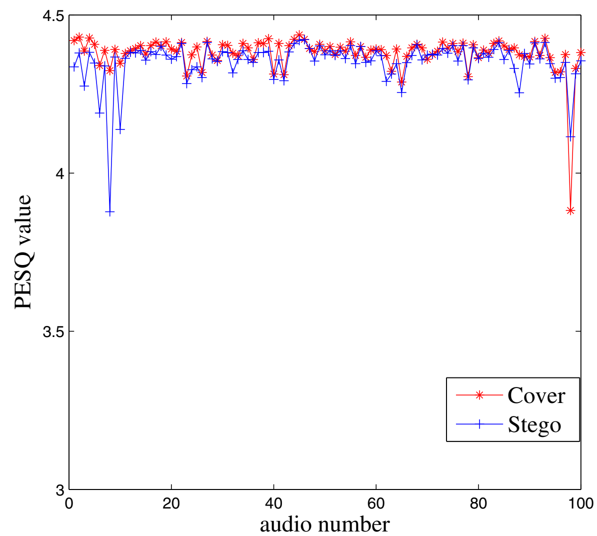



Fig. 8: The comparison of PESQ with 16 kHz sampling rate

10000.
SDB2: The 2000 PCM audios with 16 kHz sampling rate to implement $SS\_DLSF$. All PCM audios from CDB are embedded with EBR of 10%, 20%, 30%, 50% and 100%, so the total amount of SILK audio steganmgrams is 2000 * 5 = 10000.
3) Evaluation metrics: The metric to evaluate the performance of the proposed steganography scheme include auditory concealment, hiding capacity and statistical security.

Three experiments are carried out respectively. To evaluate the auditory concealment, PESQ is tested in experiment 1 for both SDB1 and SDB2. The hiding capacity of $SS\_DLSF$ is evaluated in experiment 2. In experiment 3, the statistical security of $SS\_DLSF$ are evaluated from two aspects. Firstly, The similarity of the statistical distribution of cover and stego's codebook are compared. Then, a steganalysis scheme which is referenced from the existing steganalysis scheme of G.723.1[1] is used to detect the stego generated from the proposed scheme $SS\_DLSF$.

## B. Experiments Results

Experiment 1: Since that the PESQ test just for wav audios, so all of the SILK audios from CDB, SDB1 and SDB2 are converted to wav audios. Figure 7 and 8 show the PESQ value of 100 audios which are chosen from the 2000 audios randomly. The experiment result shows that the PESQ value of cover and stego are between 4.0 and 4.5. The PESQ value of cover is slightly larger than stego. Table 9 shows the average PESQ value of 2000 audios of cover and stego with different embedding rate from CDB, SDB1 and SDB2. While SR represents the sampling rate, ER represents the embedding rate. When the sampling rate is 8 kHz, the PESQ value of stego with 10% embedding rate only decrease 0.006 compare to cover, and the PESQ value with 100% embedding rate only decrease 0.06. When the sampling rate is 16 kHz, the PESQ value of stego with 10% embedding rate only decrease 0.04, and the PESQ value with 100% embedding rate only decrease 0.2. The experiment result show that there is a little reduction in speech quality for stego audio. So the $SS\_DLSF$ have good auditory concealment.

Experiment 2: The hiding capacity of $SS\_DLSF$ is evaluated. By the embedding principle of $SS\_DLSF$, each audio has different embedding capability. In this experiment, the number of embedding bits of SDB1 and SDB2 with 100%

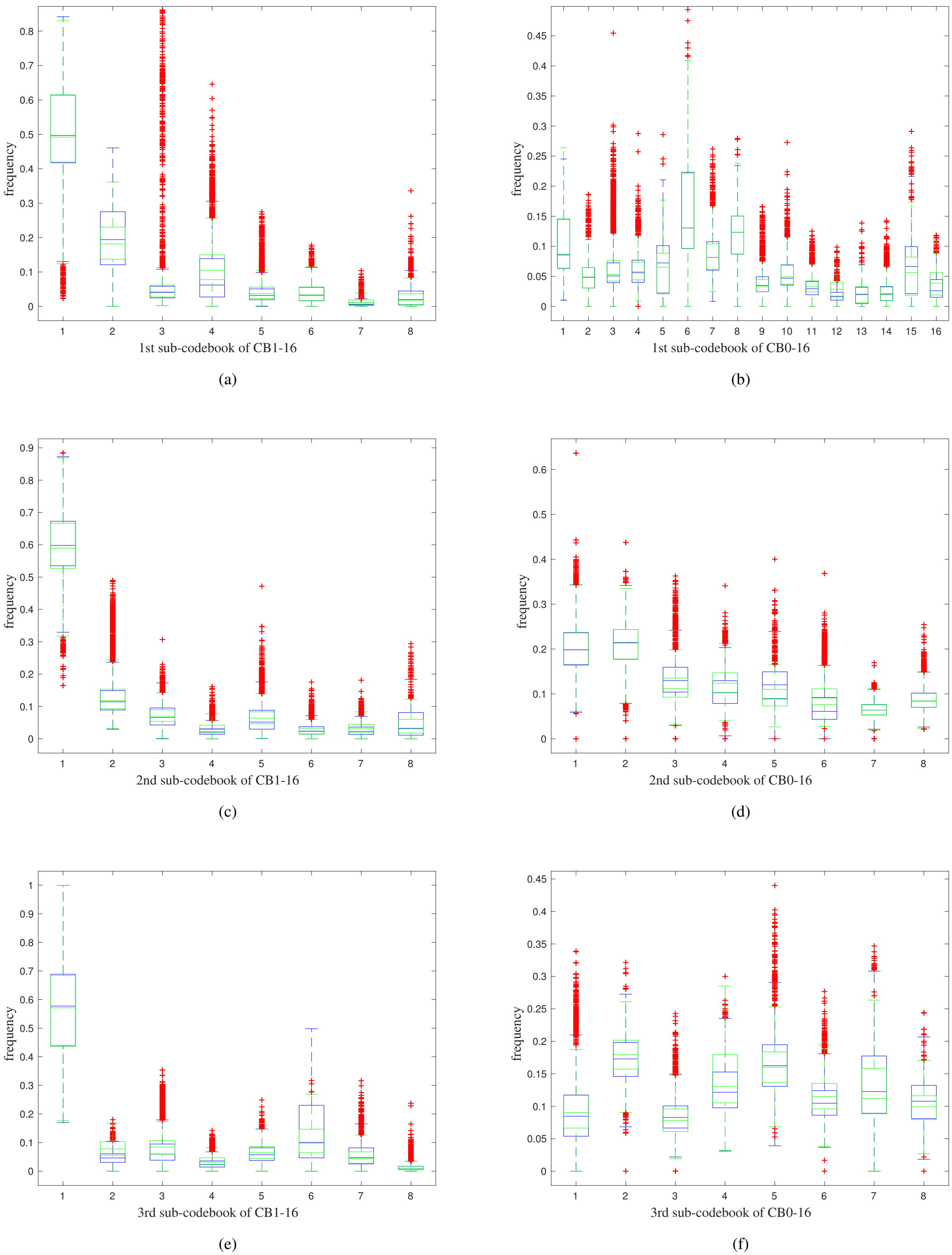



Fig. 9: The usage probability of the LSF codeword in cover and stego

TABLE VIII: The PESQ value of cover and stego

| SR/ER | Cover | 10% | 20% | 30% | 50% | 100% |
|---|---|---|---|---|---|---|
| 8 kHz | 4.33 | 4.29 | 4.27 | 4.25 | 4.23 | 4.14 |
| 16 kHz | 4.37 | 4.36 | 4.36 | 4.35 | 4.33 | 4.31 |

TABLE IX: The hiding capacity of the $SS\_DLSF$

| SR | Min | Max | Avg |
|---|---|---|---|
| 8 kHz | 117 bps | 146 bps | 129 bps |
| 16 kHz | 124 bps | 274 bps | 223 bps |

embedding rate are counted. The unit of hiding capacity is $bps$ which means the number of bits for per second. Table 10 shows the minimum , maximum and average embedding capacity of SDB1 and SDB2 with 100% embedding rate under 8 kHz and 16 kHz sampling rate.

Experiment 3: The statistic security of the proposed scheme $SS\_DLSF$ is evaluated. At first, The difference of distribution characteristic of the cover and stego with 100% embedding rate are compared. Figure 9 shows the usage probability of the LSF codeword vectors in 1st, 2nd and 3rd sub-codebook of CB1-16 and CB0-16, which is shown by boxplot. The blue one represents the usage probability of the codewords in cover audio, the green one represents that of the stego audio. Figure 9 shows that the distribution of the codeword vector in cover and stego are almost similar. It means that the scheme $SS\_DLSF$ will not introduce obvious feature change on frequency of codeword vector.

To evaluate the statistic security further, A steganalysis scheme Silk_SAlys which is referenced from the steganalysis scheme of G.723.1[1] is realized to detect stegos generated from scheme $SS\_DLSF$. In this steganalysis scheme, the usage probability of each LSF codeword vector of voiced and unvoiced frame are extracted and combined as the feature to detect the stegos. The distribution feature are calculated for training and classification. 50 percent of cover audios and stego audios from CDB, SDB1 and SDB2 with different sampling rate are randomly selected to train with the SVM classifier. The other half of the samples are tested by the previously trained models. And the audios with different embedding rate are trained to detect corresponding samples respectively. Such as cover audios and stego audios with 10% embedding rate are used to train models to detect the other half of cover audios and stego audios with 10% embedding rate. The evaluated metric are ACU, which is the average of True Positive Rate(TPR) and True Negative Rate(TNR). While TPR means the proportion that stego audios are judged as stego, and TNR means the proportion that cover audios are judged as cover. If the TNR and TPR is 50%, it means the steganography is perfect because the trained models couldn't distinguish cover and stego. Table 11 shows that the TNR, TNR and ACU of $SS\_DLSF$. Although the TPR increase with the embedding rate, they all between 49.3% and 56.4%. The ACU of 8 kHz sampling rate are all near 50% and 16 kHz sampling rate are all near 55%. It means that the proposed schemes has good statistic security.

TABLE X: The detection of $SS\_DLSF$ under the steganalysis scheme

| SR | ER | TNR | TPR | ACU |
|---|---|---|---|---|
| 8 kHz | 10% | 52.4% | 47.3% | 49.85% |
| | 20% | 47.6% | 52.4% | 50% |
| | 30% | 46.7% | 53.1% | 49.9% |
| | 50% | 45.4% | 53.3% | 49.35% |
| | 100% | 43.9% | 54.7% | 49.3% |
| 16 kHz | 10% | 58.4% | 50.2% | 54.3% |
| | 20% | 55.4% | 55% | 55.2% |
| | 30% | 52.3% | 59.1% | 55.7% |
| | 50% | 49% | 63.5% | 56.25% |
| | 100% | 47.3% | 65.5% | 56.4% |

## V. CONCLUSION

In this paper, a secure steganography scheme based on the modification of LSF parameters in SILK codec is proposed. The main contributions of this paper include: firstly, SILK is a very popular speech codec in social media today. To the best knowledge as we know, this is a first work to embed secret information into the compression parameter of SILK. This work will open up the research in this area. Secondly, the character of the LSF parameter in SILK is analyzed in detail, and the proposed scheme is based on the feature of this embedding domain. Thirdly, the statistical distribution of LSF Codebook is considered as a constraint condition, and the scheme is designed based on it to make the distribution of stego close to that of the cover audio. This is a novel method in the LSF steganography scheme, and can be extended to the steganography scheme of other CELP codec, such as G.723.1, G.729, AMR, etc. The experimental results show that the proposed scheme has good auditory concealment. The average hiding capacity can achieve 129 bps and 223 bps under the sampling rate of 8 kHz and 16 kHz respectively. More importantly, the proposed scheme has good statistical security. In the future, we will investigate the other embedding domain of SILK codec, such as pitch delay and quantization pulse, to find the statistic distribution of those compression parameters of the cover audio, to design the secure steganography scheme based on those distribution feature of the cover.

## REFERENCES

[1] Song bin Li, Huai zhou Tao, and Yong feng Huang. Detection of quantization index modulation steganography in g.723.1 bit stream based on quantization index sequence analysis. *Journal of Zhejiang University-Science C(Computers & Electronics)*, 13(8):624–634, 2012.

[2] Y. U. Chi, Liu Sheng Huang, Wei Yang, Zhi Li Chen, and Hai Bo Miao. A 3g speech data hiding method based on pitch period. *Journal of Chinese Computer Systems*, 2012.

[3] Yung-Kuei Chiang, Piyu Tsai, and Feng-Long Huang. Codebook partition based steganography without member restriction. *Fundamenta Informaticae*, 82(1):15–27, 2008.

[4] Garrick Chow. *Adobe Audition CS6 : essential training*. lynda.com, 2012.

[5] Bernd Geiser and Peter Vary. High rate data hiding in acelp speech codecs. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4005–4008, 2008.

[6] Soeren Jensen, Koen Vos, and Karsten Soerensen. Silk speech codec. 2010.

[7] Andrew D. Ker, Patrick Bas, Scott Craver, and Jessica Fridrich. Moving steganography and steganalysis from the laboratory into the real world. In *ACM Workshop on Information Hiding and Multimedia Security*, pages 45–58, 2013.

[8] Songbin Li, Yizhen Jia, and C. C. Jay Kuo. Steganalysis of qim steganography in low-bit-rate speech signals. *IEEE/ACM Transactions on Audio Speech & Language Processing*, PP(99):1–1, 2017.

[9] Cheng Haoa Liu, Sena Bai, Yong Feng Huang, Yia Yang, and L. I. Song-Bin. An information hiding algorithm based on pitch prediction. *Computer Engineering*, 39(2):137–140, 2013.

[10] Jin Liu, Hui Tian, Jing Lu, and Yonghong Chen. Neighbor-index-division steganography based on qim method for g.723.1 speech streams. *Journal of Ambient Intelligence & Humanized Computing*, 7(1):139–147, 2016.

[11] Jin Liu, Ke Zhou, and Hui Tian. Least-significant-digit steganography in low bitrate speech. 11(18):1133–1137, 2012.

[12] Peng Liu, Songbin Li, and Haiqiang Wang. Steganography in vector quantization process of linear predictive coding for low-bit-rate speech codec. *Multimedia Systems*, pages 1–13, 2015.

[13] Peng Liu, Songbin Li, and Haiqiang Wang. Steganography integrated into linear predictive coding for low bit-rate speech codec. *Multimedia Tools & Applications*, 76:1–23, 2016.

[14] Haibo Miao, Liusheng Huang, Zhili Chen, Wei Yang, and Ammar Al-Hawbani. A new scheme for covert communication via 3g encoded speech. *Computers & Electrical Engineering*, 38(6):1490–1501, 2012.

[15] Akira Nishimura. Data hiding in pitch delay data of the adaptive multi-rate narrow-band speech codec. In *International Conference on Intelligent Information Hiding & Multimedia Signal Processing*, pages 483–486, 2009.

[16] Xueshun Peng, Yongfeng Huang, and Fufang Li. A steganography scheme in a low-bit rate speech codec based on 3d-sudoku matrix. In *IEEE International Conference on Communication Software and Networks*, pages 13–18, 2016.

[17] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Information hiding-a survey. *Proceedings of the IEEE*, 87(7):1062–1078, 1999.

[18] Nikunj. V Tahilramani and Ninad Bhatt. Steganography in speech signal with enhanced multi-pulse excitation codevector with reduced number of bits. In *International Conference on Electrical, Electronics, Signals, Communication and Optimization*, pages 1–4, 2015.

[19] Jm. Valin, K. Vos, and T. Terriberry. Definition of the opus audio codec. 2012.

[20] Ziling Wei, Baokang Zhao, Bo Liu, Jinshu Su, Liyang Xu, and Erci Xu. A novel steganography approach for voice over ip. *Journal of Ambient Intelligence & Humanized Computing*, 5(4):601–610, 2014.

[21] Zhi Jun Wu, Wei Yang, and Yi Xian Yang. Abs-based speech information hiding approach. *Electronics Letters*, 39(22):1617–1619, 2003.

[22] B. Xiao, Y. Huang, and S. Tang. An approach to information hiding in low bit-rate speech stream. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM*, pages 1–5, 2008.

[23] Shufan Yan, Guangming Tang, and Yanling Chen. Incorporating data hiding into g.729 speech codec. *Multimedia Tools & Applications*, 75(18):11493–11512, 2016.