Speech information hiding algorithm based on complete binary tree dynamic codebook grouping (CBTDCG)

Wu Zhijun, Li Changliang, Sha Yongpeng, Zhang Chenlei
School of the Electronic Information and Automation, Civil Aviation University of China, Tianjin, 300300 China
Corresponding author: Zhijun Wu (e-mail: zjwu@cuc.edu.cn).

This work was supported in part by the National Natural Science Foundation of China under grant No. U1933108, the Key Program of Natural Science Foundation of Tianjin under grant No. 17JCZDJC30900, the National Science Foundation for Young Scientists of China under grant No. 61601467, and the Fundamental Research Funds for the Central Universities of China under grant No.3122018D007.

ABSTRACT In order to protect the data security and privacy of Voice over Internet Protocol (VoIP), a new algorithm for speech information hiding is proposed based on complete binary tree dynamic codebook grouping (CBTDCG). The algorithm uses the quantization index modulation (QIM) method to divide the dynamic codebook into two parts (corresponding to the embedded secret information bits ‘0’ and ‘1’ respectively), and adopts the complete binary tree to achieve the optimal grouping. The utility of complete binary tree makes the algorithm have the characteristics of fast search and high accuracy. The sender selects the corresponding code word according to the secret information bits for embedding the secret information into carrier speech. At the receiving end, the secret information bits are extracted according to the features of the code word vectors by the receiver. The experimental results show that the hidden capacity of the algorithm is larger than 450bps and the loss of the PESQ is less than 4%. Verification shows that the algorithm possesses a good robustness with a low bit error rate after being interfered by noise and the speech stream carrying secret information has better imperceptibility.

INDEX TERMS Enter key words or phrases in alphabetical order, separated by commas. For a list of suggested keywords, send a blank e-mail to keywords@ieee.org or visit http://www.ieee.org/organizations/pubs/ani_prod/keywrld98.txt. INTRODUCTION

I. INTRODUCTION
As one of the most innovative technologies in the history of communications, VoIP has brought great benefits. However, there are still many security risks and threats [1]. The main issues of VoIP include the security of communication data and the leakage of sensitive information [2].

The security risks of VoIP are mainly shown in three aspects [3]. First, it is based on the open architecture of the Internet, which uses the Transmission Control Protocol / Internet Protocol (TCP/IP). Second, it adopts an open protocol system, including Realtime Transport Protocol and Realtime Transport Control Protocol (RTP/RTCP), which could be attacked by hackers. Third, speech encoder: ilBC, G.729, G.723.1, etc. They are open protocols. Through the interception and analysis of RTP messages, hackers can obtain the conversation content of both parties. Therefore, the security risks of VoIP will inevitably cause various security issues to the network, cloud and mobile terminals [4].

In order to protect the privacy of user information and the security of communication data, many security technologies have been applied to network communication, especially data encryption. The traditional data encryption technology plays an important role in ensuring the information security. But the encrypted data can be cracked or violently broken, leading to a privacy disclosure. Information hiding is a form of secret communication. It transforms the focus of secret communication from signal to the channel. And it
turns the unknowable signal into invisible signal, which fundamentally hides the existence of signal. By using a hiding algorithm, the sender embeds secret information bits into the carrier and sends it to the network, which does not attract the attention of the unauthorized people [5]. When the carrier reaches the receiving end, the receiver extracts the secret information bits from the carrier, achieving the effect of the secret communication. As a specific application of information hiding technology, speech information hiding has gradually become a hot topic in recent years.

II. RELATED WORKS
For VoIP research, Keromytis conducted a classified survey of the existing papers and monographs. He proposed a work route to ensure the security of VoIP [6]. Srivatsa et al. developed a routing setup and maintenance protocol for VoIP. The protocol can meet the customized privacy protection requirements of users, and it has better performance and scalability [7]. Many achievements have been made in the field of information hiding. The information hiding methods have been summarized in paper [8]. The actual requirements of marking schemes and hiding system have also been described in paper [8]. Moulin et al. conducted a theoretical analysis of information hiding. He laid the theoretical foundation for the design of information hiding system and proposed the concept of hidden capacity [9]. In the field of speech information hiding, many achievements have also been made. According to the distribution of energy values, Wu et al. used the Least Significant Bit (LSB) replacement algorithm to embed secret information in the G.711 encoded speech stream, which achieved better speech quality, and the hidden capacity reaches 20kbps [10]. Wang proposed a method for real-time embedding using LSB algorithm [11]. Ma used the ABS method to embed secret information in the G.721 encoded speech stream [12]. Liu et al. used the multiple adjacent state parameters of the frame to achieve information hiding [13]. Huang studied the hiding algorithm of embedding secret information in the inactive frame and found that the inactive frame was more suitable for information hiding than the active frame [14].

The above research results involve encoders of low and medium rate. Most of the research uses general information hiding algorithms such as LSB and so on. Xiao et al. proposed a Complementary Neighbor Vertices (CNV) algorithm, which applied graph theory knowledge for the first time in the field of information hiding. The experimental results show that the proposed algorithm can be applied to speech stream encoded by G.723.1 or iLBC with the better concealment [15]. However, Huang proposed an effective detection method for CNV information hiding algorithm [16]. The above research methods are susceptible to detection based on statistical analysis. Consequently, an information hiding algorithm based on the dynamic codebook has been proposed in literature [17]. Experimental results show that the algorithm not only has high hidden capacity, but also with good concealment. Furthermore, in the aspect of resistance character statistics analysis, the algorithm is superior to the hiding algorithm based on static codebook. Based on the dynamic codebook quantization process in iLBC encoding, Yang Wanxia designed an information hiding algorithm. Experimental results show that the hiding algorithm can achieve the hidden capacity of 450bps at a carrier rate of 13.3kbps, and the PESQ value of the speech decreases by less than 6%, which has good concealment performance [18]. Yang Wanxia divides the dynamic codebook of the sub-frame into an odd group and an even group, and then searches for the best code word in the corresponding group according to the secret information bit. However, the time complexity of this method is high, resulting in a long time delay, which is not suitable for speech communication with high real-time requirements. Committed to reducing the computational complexity of the algorithm, this paper proposes an information hiding algorithm based on complete binary tree dynamic codebook grouping.

III. SPEECH INFORMATION HIDING ALGORITHM BASED ON ILBC ENCODING
In 2000, Global IP Sound developed an iLBC encoder for narrowband speech communication [19]. It belongs to the medium and low rate encoder and it can provide robust speech communication over IP. The iLBC encoder supports frames in two formats: for a 20ms frame, the bit rate is 15.20kbit/s. For a 30ms frame, the bit rate is 13.33kbit/s [3]. When the network environment is not good, IP packet delay is long or packet loss occurs, the Packet Loss Concealment (PLC) can effectively enhance the speech and ensure the communication quality. Therefore, the excellent iLBC has gradually become a widely used encoder in secret communication.

A. Based On The Principle iLBC Coding
iLBC is a model based on code book excitation linear prediction. A series of analysis is performed to obtain a set of parameters that can represent the phonetic characteristics of the speech. Then these parameters are packaged in a fixed format and sent to the network and the receiver recovers the speech according to the received characteristic parameters. Therefore, iLBC speech coding has gone through several coding steps, one of which is adaptive codebook coding. Aiming at
the characteristics of the adaptive codebook coding process of iLBC, a new information hiding algorithm based on dynamic codebook quantization is proposed.

(1) Code book storage structure: After the start state encoding is completed, the encoded data begins to be decoded. The codebook memory is filled according to the decoded LPC excitation signal, and its length is 85 or 147. When a sub-frame is encoded, the encoded data needs to be decoded immediately and sent to the codebook memory. So the data in the codebook memory are dynamically changing every time the next sub-frame is encoded.

(2) Construction of codebook: Fig.1 shows the generation of the base codebook from the dynamic codebook when encoding a sub-frame with a length of 40 samples. In Fig.1, a sliding window of length 40 slides over the code memory. Every time the window slides one grid to the left, the index increases by 1. When index = 0, the code word vector contains 40 excitation signals between 107 and 146. Correspondingly, the data in the window also changing, which in turn generates a series of new code word vector. All code word vectors are combined together to form the base codebook of the dynamic codebook.

(3) Search of codebook. The main steps of codebook search are as follows.

1. The weight vector of optimal matching perception needs to meet the following three conditions:
   a. Calculate metrics. As shown in formula (1): select the codeword vector in the codebook to maximize it
      \[
      \frac{(t \arg \mathbf{et} \cdot \mathbf{cbvec})^2}{|\mathbf{cbvec}|}
      \]  
      Formula (1) \(\mathbf{cbvec}\) represents the selected codebook vector, and target represents the target codebook vector.
   b. The absolute value of the chosen codebook vector \(\mathbf{cbvec}\) is less than 1. 3. The calculation formula of gain is shown in equation (2)
      \[
      \text{gain} = \frac{(t \arg \mathbf{et} \cdot \mathbf{cbvec})^2}{|\mathbf{cbvec}|}
      \]  
      c. In the first stage of matching, the dot product of the optimal codebook vector and the target vector must be greater than 0. That is, it satisfies equation (3).
      \[
      t \arg \mathbf{et} \cdot \mathbf{cbvec} > 0
      \]  
2. Quantify the gain at each stage. The gains in all three stages are quantified, using digits 5, 4 and 3 respectively.
3. Update the perceived weight target. Prior to the second and third stage of the search, the perceptively weighted target vector is updated by subtracting the weighted target vector from the selected target codebook times the corresponding quantized gain.

B. A Compression Domain Hiding Method Based On Complete Binary Tree

The dynamic codebook in the process of iLBC encoding consists of the base codebook and the extended codebook. When the length of the sub-frame is 40 samples, the dynamic codebook also includes augmented basic codebook and augmented extension codebook. The dynamic codebook has large space and contains many code words vectors, and it always maintains dynamic changes during the encoding process. According to the QIM method [20], proper grouping of dynamic codebook can realize information hiding with large hidden capacity and high concealment.

Binary tree is a finite set which contains \(n\) nodes. The structure of complete binary tree [21] is shown in Fig.2. In Fig.2, node A represents the root of the tree. The rest of the nodes are divided into two disjoint subsets \(T_1\) and \(T_2\). Between them, \(T_1 = \{B,D,E,H,I,J,...\}\), \(T_2 = \{C,F,G,...\}\). Both \(T_1\) and \(T_2\) are subtrees of a complete binary tree \(T\), and both of them are also a complete binary tree.

Since the construction of the complete binary tree is similar to the codebook grouping in speech information hiding, this paper proposes a speech information hiding algorithm based on the complete binary tree dynamic codebook grouping (CBTDCG). With the root node 0 of the complete binary tree as the boundary, the code word vector is stored in the node of the binary tree. Code word with an odd index number in the codebook space is stored in the left subtree node, and index numbers with an even
number is stored in the right subtree node. Finally, the dynamic codebook is grouped. The dynamic codebook space is large, contains many code word vectors, and always maintains dynamic changes during the encoding process. According to the principle of QIM hiding algorithm, proper grouping of dynamic codebook can realize information hiding with high hidden capacity and high concealment. As shown in Fig.3, the base codebook CB in the dynamic codebook can be divided into the following two groups.

\[
CB = \{CB_l, CB_r, 0\} \tag{4}
\]

Among them, CB_l represents the node in the left subtree of the complete binary tree, including the code word vector cbv0, cbv2, cbv3, cbv6, cbv7 and so on. CB_r represents the node in the right subtree of the complete binary tree, including the code word vectors cbv1, cbv4, cbv5 and so on.

**FIGURE 3.** Group dynamic codebook in a positive sequence

In order to enhance the ability of the algorithm to resist the steganalysis and enrich the arrangement of the codebook, it is necessary to further improve the way of placing code word vectors. By inverting the code word vectors in Fig.3, a complete binary tree as shown in Fig.4.

**FIGURE 4.** Group dynamic codebook in a reverse sequence

In the process of embedding secret information bits, the arrangement of the code word vectors is determined by the control information bits. The optimal fixed code word index is defined as the control information bit in this paper. Because it generates by the start state encoding shows randomness, which exactly meets the requirements of control information bits for randomness. If the control information bits meet \(\text{index mod } 2 = 0\), the complete binary tree generates in a positive sequence. If the control information bits meet \(\text{index mod } 2 = 1\), the complete binary tree generates in a reverse sequence. After constructing the complete binary tree, in order to find the code word vector that satisfies the optimal matching search condition, the sender performs a recursive search on the complete binary tree based on the embedded secret information bit. When the secret information bit is ‘0’, a recursive search is performed on the left subtree of the complete binary tree. On the contrary, when the embedded secret information is ‘1’, a recursive search is done on the right subtree of the complete binary tree. After the search is completed, the best index and other parameters are packaged in a fixed format and sent to the network, and finally the speech information hiding is realized.

**C. The Implementation Of The Embedding Process**

Fig.5 shows the process of embedding secret information bits of this paper. The embedding process is shown below.

i. The encryption of the secret information. The plaintext secret information \(S = \{b_1, b_2, ..., b_n\}\) is encrypted and the ciphertext \(E = \{e_1, e_2, ..., e_n\}\) is obtained.

ii. Divide speech carriers into frames. Assuming that the input signal has \(N\) frames, \(F = \{f_i | i = 0, ..., N\}\), \(f_i\) represents the \(i\)th frame.

iii. Select the start state and encode, then normalize the residual signal.

iv. Build dynamic codebook. The \(i\)th frame \(f_i\) in the speech stream \(F\) has 5 sub-frames except the start state. It is \(f_i = \{f_{1i}, f_{2i}, f_{3i}, f_{4i}, f_{5i}\}\). The construction process and steps of the dynamic codebook are as described above. The dynamic codebook is \(CB = \{CB_1, CB_2, CB_3, CB_4, CB_5\}\), where \(CB_i\) represents the dynamic codebook of the \(i\)th speech frame. \(CB^{cbv}_i\) represents the dynamic codebook of the \(K\)th sub-frame in the \(i\)th speech frame.

v. Construct a complete binary tree \(T\) according to the control information bits. Dynamic codebook \(CB^{cbv}_i\) becomes \(CB^{cbv}_i = \{CB^{cbv}_i, 0\}\). \(CB^{cbv}_{\text{cbv}}\) represents the dynamic codebook that contains the code word vector in the left subtree \(T_1\), and \(CB^{cbv}_{\text{cbv}}\) represents the dynamic codebook that contains the code word vector in the right subtree \(T_2\).

vi. Embed the secret information bits. Search for the corresponding code word based on the value of \(e_i\) in the encrypted secret information.

\[
\begin{align*}
\text{BW}_{\text{cbv}}^{cbv} &\in CB^{cbv}_{\text{cbv}}, \quad e_i = 0 \\
\text{BW}_{\text{cbv}}^{cbv} &\in CB^{cbv}_{\text{cbv}}, \quad e_i = 1
\end{align*}
\tag{5}
\]
If the secret information bit $e_i = 0$, search for the optimal matching code word vector in the codebook group $CB^{i,k}_{x}$—$CB^{i,k}_{y}$. If the secret information bit $e_i = 1$, search for the optimal matching code word vector in the codebook group $CB^{i,0}_{x}$—$CB^{i,0}_{y}$. In the corresponding codebook, the target vector and the code word vector are used for optimal matching. Finally, the best index and optimal gain are found and quantified. After the secret information bit is embedded in the current sub-frame, the target vector is updated, and the secret information is shifted back by one bit and the secret information bit becomes $e_{i+1}$. $BW^K_i$ represents the optimal code word vector corresponding to the $K^{m}$ sub-frame in the $i^{m}$ speech frame.

vii. Repeat steps iv, v and vi for a second level search of the code word vector.

viii. Repeat steps iv, v and vi for a third level search of the code word vector.

ix. If all the secret information bits have been embedded, then the remaining carriers are encoded by iLBC. Otherwise, it returns to the step iii.

For example, secret information bit $e_i = 0$, the control information bits is ‘0’. The complete binary tree $T$ generates in a positive sequence, and a recursive search is performed on the left subtree of the complete binary tree. The best index 57 and the optimal gain 0.8 is quantified and packaged in a fixed format. Finally, the package is sent to the network.

D. iLBC Encoding Process Integrate The Algorithm Of Embedding Secret Information

After receiving the iLBC data packets containing the secret information bits, the receiving end analyses the iLBC data packets, extracts the relevant encoding parameters, reconstructs the start state, and determines the structure of the complete binary tree according to the optimal index of the fixed codebook, and then constructs the codebook memory. Finally, the corresponding complete binary tree is constructed according to the codebook memory. The secret information bits are determined according to the position of the code word vector in the binary tree. After all the secret information bits are embedded, the sender does not stop the encoding immediately until all the speech carriers are encoded completely. Therefore, the index of the carrier that does not carry the secret information will exhibit randomness, and the receiver cannot judge whether there is secret information bit according to the position of the code word vector in the binary tree. So the extraction algorithm is a non-blind extraction algorithm and it is described in Fig.6.
The process of extracting secret information bits includes the following steps.

i. The receiver analyses the received iLBC data packets and gets the relevant parameters.

ii. Determine whether the sender embeds secret information bits and determines the number of bits of the secret information. If the secret information is embedded, go to step iii, otherwise, go to step vii.

iii. Reconstruct the start state.

iv. Construct the codebook memory according to the start state. Next judge the control information bits according to the obtained fixed code word vector index.

v. Build the dynamic codebook \( CB = \{ CB_1, CB_2, CB_3, CB_4, CB_5 \} \) according to the codebook memory in iv. \( CB_i \) represents the dynamic codebook of the \( i \)-th speech frame.

vi. If the control information bits meet \( \text{index} \mod 2 = 0 \), the complete binary tree \( T \) generates in a positive sequence. If the control information bits meet \( \text{index} \mod 2 = 1 \), the complete binary tree \( T \) generates in a reverse sequence.

vii. Extract secret information according to the code word vector. The secret information bits are judged according to the position of the code word vector corresponding to the code word vector index. If \( BW_i^k = CB_{k,I} \), \( e_i = 0 \). If not, \( e_i = 1 \).

viii. The extracted secret information is integrated to obtain the encrypted secret information \( E = \{ e_1, e_2, ..., e_n \} \).

ix. Use the secret key to decrypt the \( E \) obtained in step viii and integrate again. Next get the plaintext \( S = \{ b_1, b_2, ..., b_i \} \).

x. Extraction is over and the algorithm is over.

IV. Experiment And Analysis Of The Results

The extraction method adopted in this paper is a non-blind extraction method, which means that the receiving end and the sending end have the same speech bank. In order to create the speech bank, 200 men and women of 22-27 years of age were selected. Each person was recorded of 2 pieces of speech samples with the duration of 10s, one of which is in English and the other in Chinese. Therefore, there are 400 pieces of speech samples with the duration of 10s, including 100 male Chinese speech samples, 100 male English speech samples, 100 female Chinese speech samples, and 100 female English speech samples.

The performance of the hiding algorithm is mainly measured by the hidden capacity and security. Next section also evaluates the performance of the hiding algorithm proposed in this paper based on these two indicators.

A. EQUATIONSILBC ENCODING PROCESS INTEGRATES THE ALGORITHM OF EMBEDDING SECRET INFORMATION

In order to verify the performance of the hiding algorithm proposed in this paper, according to the encoding standards given in literature [3], debug codecs, add hiding module and extracting module. Establish the experimental environment in the LAN to complete the evaluation of the hiding algorithm (see Fig.7) [1].

The sender and receiver are two PCs with the following parameters: Processor: Intel (R) Core i3-4370 CPU @ 3.80GHz 3.80GHz. Memory: 4GB Sound Card: Realtek High Definition Audio. System: Windows 7 Professional Edition Service Pack 1. The sender randomly selects a speech in the bank as the carrier to embed the secret information. The receiver determines carrier speech signal and decodes it, then selects the corresponding speech signal in the local bank for local encoding. Finally, the receiver extracts the secret information according to the characteristics and differences of the two parameters.

B. EVALUATION AND ANALYSIS OF HIDDEN CAPACITY

The hidden capacity, also known as the hidden rate, is the amount of secret information bits embedded in the speech carrier per second. A good hiding algorithm can guarantee good security and large hidden capacity, but usually these two indicators are contradictory. The large hidden capacity means that there are a number of secret
information bits embedded in the speech carrier, which will directly lead to the decline of the carrier, affect the imperceptibility of the secret information and increase the risk of being cracked. Therefore, the trade-off between security and hidden capacity is significant for speech information hiding algorithm. The hidden field of the speech information hiding algorithm is dynamic codebook. Therefore, the hidden capacity of the algorithm is directly related to the size of the codebook. According to the iLBC encoding standard, the dynamic codebook is used for encoding the remaining samples except the start state. Three times of searching are performed for each sub-frame, and one bit of secret information is embedded for each search. Therefore, each sub-frame is embedded with 3 bits of secret information bits. For the 30ms frame, the hidden capacity can be calculated by the equation (6). Similarly, for 20ms frame, the hidden capacity is 450bps.

$$C = \frac{1000}{30} \times 5 \times 3 = 500 \text{bps}$$ (6)

Compare the hidden capacity of the algorithm in this paper and other two algorithms, as shown in TABLE I. The first algorithm is the hiding algorithm whose dynamic codebook is based on a complete binary tree grouping. The second algorithm is the hiding algorithm whose dynamic codebook is based on the parity grouping [17–18]. The third is the CNV algorithm [15].

<table>
<thead>
<tr>
<th>Frame</th>
<th>The first</th>
<th>The second</th>
<th>The third</th>
</tr>
</thead>
<tbody>
<tr>
<td>30ms</td>
<td>500bps</td>
<td>500bps</td>
<td>200bps</td>
</tr>
<tr>
<td>20ms</td>
<td>450bps</td>
<td>450bps</td>
<td>150bps</td>
</tr>
</tbody>
</table>

It can be seen from TABLE I that algorithm 1 and algorithm 2 have the same hidden capacity because the two algorithms use the same hidden idea and hidden domain. Both of these algorithms belong to the hiding algorithm with large hidden capacity and high embedding rate. Algorithm 3 uses a fixed codebook grouping algorithm with a slightly smaller capacity, but also reaches 200bps. Comprehensive analysis, the proposed speech information hiding algorithm is a large capacity hiding algorithm. And the algorithm satisfies the hidden capacity demand of VoIP real-time secret communication.

C. SECURITY ANALYSIS
The security of the speech information hiding algorithm is mainly measured by the concealment and robustness of the secret information. Concealment is an indicator of the imperceptible ability of secret information in carrier. Robustness is the ability of the secret information to recover accurately when the information is interfered by the outside world.

C.1 Concealment Analysis

The change of the speech carrier after embedding the secret information can be intuitively judged from the time domain and frequency domain waveform diagrams. In order to verify the concealment of the algorithm, select a set of speech signals in the 400 speech samples of 10s duration and draw the time domain waveform and frequency domain waveform of the signal. As shown in Fig.8 and Fig.9. It can be seen from Fig.8 and Fig.9 that the changes of the time domain waveform and the spectrum domain waveform are small. From the perspective of signal processing, this hiding algorithm causes little speech fading and better concealment.

![Time domain waveform before and after the secret message is embedded](image8)

![Frequency domain waveform before and after the secret message is embedded](image9)

At present, the most popular tool for evaluating the quality of speech is PESQ. In the field of speech information hiding, this value can reflect the change of the quality between the carrier which is embedded secret information and the carrier which is not [5]. In the theoretical range, the larger the PESQ value, the smaller the change of speech quality, the closer the synthesized speech is to the original speech, the better the concealment of the algorithm. This paper also uses the PESQ value to evaluate the quality of speech Fig.10 and Fig.11 show the changes of the quality caused by embedding secret information in the 30ms frame and the 20ms frame [24–25].
It can be seen from Fig.10 and Fig.11 that the hiding algorithm has little influence on the PESQ. When encoding the 20ms frame, the decline of the quality is slighter than the 30ms frame, which is consistent with the conclusion in the paper [16] that the 20ms frame is more suitable for speech information hiding than the 30ms frame in iLBC encoder [17-18]. The conclusion can be explained from the following two aspects: i. More 20ms frames are needed when the same number of secret information bits are embedded. The embedded secret information would not accumulate in the local area and the local influence to the carrier would also be weakened. This is one of the reasons that the PESQ in the 30ms frame reduces more. ii. when the iLBC encoder sequentially encodes sub-frames except the start state, the codebook is dynamically changing. So the changes in the front code word vector would be passed back, causing the distortion to be amplified continuously. The more times of it encodes, the more obvious of the distortion. The 30ms frame has 5 sub-frames and the 20ms frame has 3 sub-frame. Therefore, the distortion caused by embedding secret information in 30ms frame...
is greater than 20ms frame. Consequently, when the secret information bits are embedded in the 30ms frame, the PESQ value decreases more.

In order to further explore the concealment and stability of the hiding algorithm, the average worsening change in PESQ and variations in PESQ changes are calculated. As summarized in TABLE II and TABLE III.

<table>
<thead>
<tr>
<th>Category</th>
<th>The frame of 30ms</th>
<th>The frame of 20ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Binary tree grouping</td>
<td>Parity grouping [17-18]</td>
</tr>
<tr>
<td>Chinese Speech Man</td>
<td>0.13222</td>
<td>0.12655</td>
</tr>
<tr>
<td>Chinese Speech Woman</td>
<td>0.13841</td>
<td>0.18105</td>
</tr>
<tr>
<td>English Speech Man</td>
<td>0.12206</td>
<td>0.0965</td>
</tr>
<tr>
<td>English Speech Woman</td>
<td>0.15598</td>
<td>0.18089</td>
</tr>
</tbody>
</table>

In order to further explore the concealment and stability of the hiding algorithm, the average worsening change in PESQ and variations in PESQ changes are calculated. As summarized in TABLE II and TABLE III.

<table>
<thead>
<tr>
<th>Category</th>
<th>The frame of 30ms</th>
<th>The frame of 20ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Binary tree grouping</td>
<td>Parity grouping [17-18]</td>
</tr>
<tr>
<td>Chinese Speech Man</td>
<td>0.002024</td>
<td>0.003105</td>
</tr>
<tr>
<td>Chinese Speech Woman</td>
<td>0.002013</td>
<td>0.002673</td>
</tr>
<tr>
<td>English Speech Man</td>
<td>0.000165</td>
<td>0.002493</td>
</tr>
<tr>
<td>English Speech Woman</td>
<td>0.000247</td>
<td>0.005626</td>
</tr>
</tbody>
</table>

In order to further explore the concealment and stability of the hiding algorithm, the average worsening change in PESQ and variations in PESQ changes are calculated. As summarized in TABLE II and TABLE III.

TABLE II lists the comparisons of changes in PESQ between the proposed steganography algorithm and the algorithm presented in [17-18]. This paper analyses TABLE II from three aspects: i. the average value in PESQ between the original speech files and the stego speech files were so small, indicating that the proposed information hiding along with speech compression encoding had no or very little impact on the quality of the synthesized speech. ii. On average, data hiding had less effect on the PESQ values of the male speech samples. This is probably due to the fact that the pitch frequency of female speech has a greater range, and changes more quickly than male speech. iii. For 30ms frame, the average worsening changes in PESQ of CSW and ESW with the proposed algorithm were smaller than the algorithm in [17-18], those of CSM and ESM were bigger. For 20ms frame, the average worsening changes in PESQ of CSM, CSW and ESW with the proposed algorithm were smaller than the algorithm in [17-18], that of ESM was bigger.

C.2 Robustness Analysis
The signal will be subjected to various kinds of outside interference in the process of network transmission and the noise is one of them. The additive noise on the physical circuit and the noise added by the attacker will affect the secret information [22]. In order to verify the stability of the secret information when the hiding algorithm is interfered by noise, this paper selects the `randn()` function in MATLAB to simulate random noise. Carry on the noise interference to the encrypted code stream and then use the extraction algorithm to extract the attacked secret information. Equation (7) illustrates the attack process

$$S' = S + \rho * N$$

(7)

Where $S$ represents the code stream carrying secret information, $\rho$ indicates the intensity of the attack, $N$ is the random noise and $S'$ is the code stream which has been interfered by noise.

Among the 400 speech signals described above, 100 secret information bits are embedded in all speech signals, and then carry on the noise interference to
them. The results of the attack are summarized in TABLE IV.

![Table IV](image)

Comparing to the attack results in the paper [23], the information hiding algorithm of this paper shows better robustness. The average bit error rate is close to 0 when the attack intensity is less than 0.1, which is lower than the average bit error rate in the paper [23]. In order to verify the bit error rate caused by intensity of the attack, the attack intensity is amplified 10 times and it is estimated from 0.1. It can be seen from TABLE IV that the algorithm has a very low bit error rate in the 30ms and 20ms frame. The bit error rate is still less than 10% when the attack intensity reaches to 0.5, indicating that the algorithm has good robustness and anti-noise interference capability [24–25].

### D. Real-time Analysis

The real-time nature of the algorithm is represented by the processing time of the speech frame. In the environment shown in Fig.7, the speech frame processing time was experimented with two computers connected by a 150 Mbps router. A piece of speech signal is randomly selected in the 400 speech samples of 10s, and the first 60 speech frames are selected for real-time testing. After the speech is processed by the hiding algorithm and the hiding algorithm in paper [17-18], the speech frame processing time on the two computers is counted and recorded. Statistics show that the average processing time of speech frames on desktop computers is between 6.38ms and 8.46ms, and the average processing time of speech frames on laptops is between 8.27ms and 10.19ms. The average time of speech frame processing using this algorithm is lower than the average processing time of algorithm in the paper [17-18], which is about 0.17ms. The experimental results show that the method of using the binary tree to group the dynamic codebook reduces the processing time of the speech frame. Therefore, the hiding algorithm of this paper is superior to the hiding algorithm proposed in [17-18] in real time.

### V. Conclusion

Aiming at the problem of privacy leakage and data security in VoIP, this paper puts forward a hiding algorithm whose dynamic codebook is based on the complete binary tree grouping. The method has less computational complexity and can maintain a small delay in speech communication which requires high real-time performance. The experimental results show that the hidden capacity and security of both the 20ms frame and 30ms frame in the iLBC encoder can reach a better level and can be applied to real-time speech communication. The hiding algorithm has a rather low bit error rate in the 30ms and 20ms frame. The bit error rate is still less than 10% when the attack intensity reaches 0.5, indicating that the algorithm has good robustness and anti-noise attack capability. However, the method is a non-blind extraction algorithm and it increases the risk of being cracked. In the future, the algorithm needs to be improved to implement the blind extraction of secret information.

### REFERENCES


Wu Zhi-Jun, born in 1965, Ph.D. professor. He received the BS and MS degrees in Information Processing from Xidian University, China, and Ph.D. degree in Cryptography from, Beijing University of Posts & Telecommunications, China. He was a professor in the department of Communication Engineering, Civil Aviation University of China. His research areas are secure communication based on speech information hiding, security in big data and cloud computing.

Li Chang-Liang, born in 1993, M.S. candidate. His research interests include Network and Information security.

Sha Yong-Peng, born in 1991, M.S. candidate. His research interests include Network and Information security.

Zhang Chen-Lei, born in 1994, M.S. candidate. His research interests include Network and Information security.