Volume 2, Issue 2, February 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Advanced Integrated Steganographic Approach With VOIP Communication

V.Ramesh kannan
Sri Ramakrishna engineering college,
Coimbatore, India
Rameshkannan.v@gmail.com

J.Selvakumar
Assistant Professor, Department of CSE (PG)
Sri Ramakrishna Engineering College
Coimbatore, India

ABSTRACT:Voice over IP (VOIP) is an upcoming technology that enables voice communication through the Internet. Packet-based network link shared between different connections, which give rise to interaction between various traffic types. This paper focus is on improving the data embedding capacity of steganography in low bit rate audio streams encoded by G711 source codec and to overcome the packet loss during the integration of hidden messages. Steganography in the inactive audio frames attains a larger data embedding capacity than that in the active audio frames under the same imperceptibility. The amount of data package increases when new services are offered and used via the internet. The large amount of data packages, the more information let through and new possibilities to hide information i.e., in the cover of something else, may be introduced. To identify the voice during the transmission whether the current audio frame is an active voice by comparing the energy of the frame with a threshold.

Index Terms—Audio streams, inactive frames, steganography, Voice over Internet Protocol (VoIP), packet loss.

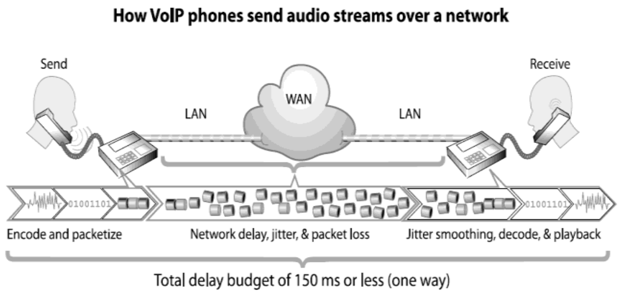
I. Introduction

VOIP phones use codec's to translate analog sound streams into digital packets for transmission. On the receiving end, the codec translates the packets back to analog. For two people to converse normally, all of this must happen in as close to real time as possible.

There have been several steganography methods of embedding data in audio streams. For example Wu *et al.* [2] suggested a G.711-based adaptive speech information hiding approach. Aoki [3] proposed a technique of lossless steganography in G.711 encoded speeches. Huang *et al.* [10] suggested an algorithm for embedding information in some parameters of the speech frame encoded by ITU G.723.1 codec, without leading to distinction between inactive voices and active voices.

For call setup, most enterprise VoIP solutions include one or more call managers, which are servers that set up calls between VoIP phones, and can also provide gateway connections to the Public Switched Telephone Network (PSTN) for calls outside the VoIP network. Typically, the call initiator contacts the call manager, which then rings the phone being called. Once the receiving party answers, the call manager provide a mechanism for the phones to negotiate codecs and connection parameters.

The connection itself is typically in the form of two full-duplex streams: a Real-time Transport Protocol (RTP) stream that carries the encoded audio, and an RTCP (Real-time Transport Control Protocol) stream to provide communications control. Once the call is set up, the call manager is no longer involved until the teardown phase, when the IP phones inform the call manager the call has been completed so the centralized call queue (a list of what phones are active) can be updated. Problems with VoIP audio quality are almost always related to network delay, jitter, and packet loss, or some combination of the three. It is common to see them together because they are both related to a general deterioration of network conditions. In any VoIP deployment, some delay is unavoidable. Codecs take time to encode/decode the audio stream, and even the fastest network medium is not instantaneous



Real-time voice communications are sensitive to delay and variation in packet arrival times. Codecs require a steady, dependable stream of packets to provide reasonable playback quality. Packets arriving too early, too late, or out of sequence result in jerky, jumbled playback. This phenomenon is called jitter.

II. VOIP Client and server Creation

a. VOIP server is an application that is made up the combination of two-application, server application (which runs on server side) and client application (which runs on client side).

b. This application is using for chatting in LAN. To start chatting you must be connected with the server after that your message can broadcast to each and every client.

c. For making this application we are using some core features like, collection, networking, I/O Streams and threading also.

d. In this application we have one server and any number of clients (which are to be communicated with each other).

III. A. Stream capturing and voice activity detection

The " **Stream Capturing** " is designed to involve buffering in the learning experiences by enabling them to record and play back the sent/Received voices in the context - by recording the voice, the data's can be verified for all the sent and received packets, specific sounds like the noises are eliminated in calls, then hearing them as part of the narrative, VOIP can benefit from increased Voice quality and reduced Delays. In this stream capturing module to capture the voice from the sender and to identify the noise. The voice activity detection (VAD) is used to check whether the current audio frame is an active by comparing the energy of the frame (Enr) with a threshold (Thr). As shown in (1)

$$\mathrm{VAD} = \begin{cases} 1, & \mathrm{Enr} \geq \mathrm{Thr} \\ 0, & \mathrm{Enr} < \mathrm{Thr}. \end{cases} \tag{1}$$

VAD = 0 means the frame is an inactive voice; otherwise, the frame is an active voice.

B. Definitions of Inactive and Active Frames

The silence compression technique is an optional function for the source codec. In fact, most source codecs do not use silence compression in VoIP applications. All audio frames are encoded uniformly by using the normal encoding algorithm regardless of whether they are active voices or inactive voices. Thus two types of frames are outputted when the speech stream F is encoded by the source codec. For example, ITU G.711 codec encodes the speech into two types of frames, active frames and inactive frames, without using the silence compression algorithm.

Definition 1: The active frame F^{*A} is encoded by the source codec from the active voice of the speech. It is expressed as

$$f_i^{*A} = \varphi\left(f_i^A\right), \qquad i = 0, \ldots, N_1.$$

Definition 2: The inactive frame is encoded by the source codec from the inactive voice of the speech. It is expressed as

$$f_i^{*S} = \varphi\left(f_i^S\right), \qquad j = 0, \ldots, N_2.$$

As the speech is divided into inactive voices and active voices by VAD, all the voices are encoded uniformly by the source codec to form audio frames, in which inactive frames can be distinguished from active frames.

IV. Stegano Data Embedding & Extraction

Inactive frames are suitable for data embedding. All the speech parameters are sorted into three imperceptibility levels of steganography in terms of the distance of signal-to-noise ratio (DSNR), which is defined as the difference in signal-to-noise ratio (SNR) between the original speech and stego speech.

Close analysis of the data in Table shows the imperceptibility levels of steganography for different parameters of the inactive frames are widely different. So it is possible to choose different parameters and various parameter bits to embed data on demand of practical applications. In short, the parameters marked with level 1–2 are suitable cover objects for steganography.

TABLE VI
NUMBERS OF INACTIVE FRAMES OF 20 PCM SPEECH FILES

Group	Speech file name	File length (s)	Number of inactive frames (30 ms)	Average no of inactive frames
Group 1	MC1	10	183	138
	MC2	10	101	
	MC3	10	121	
	MC4	10	120	
	MC5	10	166	
Group 2	WC1	10	139	125
	WC2	10	125	
	WC3	10	110	
	WC4	10	147	
	WC5	10	106	
Group 3	ME1	10	69	52
	ME2	10	59	
	ME3	10	48	
	ME4	10	42	
	ME5	10	45	
Group 4	WE1	10	51	59
	WE2	10	53	
	WE3	10	56	
	WE4	10	66	
	WE5	10	69	

V. *Minimize packet loss in VOIP*

Delay jitter and packet loss are the two factors that affect the voice quality the most. The delay jitter manifests itself as packet loss in the dejitter buffer, which drops the late packets. It is impossible to remove packet loss from the network but it can definitely be minimized.

The three main problems occurring in real-time applications like Voice over IP (VoIP) are:

1) *End-to-End delay*: The total delay experienced by the packet from the sender till it reaches the receiver.
2) *Jitter:* The variation in packet interarrival time. The difference between when the packet is expected and when it is actually received is jitter.
3) *Packet loss:* Loss of voice packets from sender to receiver.

The total packet loss is composed of two elements: 1) packets lost over the network due to congestion, and 2) packets arriving late after their expected play out time that are discarded by the receiver. The jitter caused by variable delays in the network is ultimately translated into the effect of packet loss in the network, as the packets arriving after the play out time are considered as lost. Another factor that could affect the quality of VoIP is the choice of codec used to transform and compress analog signals into digital signals.

a. ***Replacing Lost Packets***

The impact of packet loss in different transmissions with respect to different codecs and then propose reconstruction strategies to recover lost information.

1. *Interleaving*

This technique distributes the effect of the lost packets in order to reduce the impact on quality. The information of a speech part is distributed in multiple packets. The data units are regrouped in a crossed form before transmission such that they are distributed, and at the receiver they are rearranged in their original form. Thus, instead of losing the whole packet small parts from distributed packets are lost.

2. *Repetition:*

Lost packets are replaced by copies of last received packets.

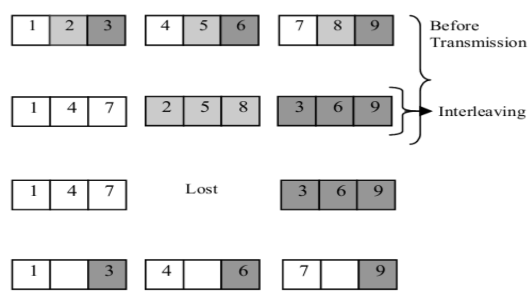


Figure 1. Example of Interleaving

3. *Simple Interpolation:*

Consists of interpolating (averaging) by using the packets after and before the lost packet.

4. *Interleaving with Repetition*

The data are interleaved before sending and then any missing part is substituted using the repetition technique at the receiver.

5. *Interleaving with Interpolation Calculation*

The interleaving technique is used before sending and then the receiver interpolates to replace any missing parts in the jitter Buffer.

Forward Error Correction and Concealment

This approach encompasses both loss correction and loss concealment algorithms. Loss Correction uses media-dependent and media-independent Forward Error Correction (FEC) techniques. FEC constitutes adding redundancy data to the normal voice stream to protect from packet loss. FEC introduces overhead in terms of the total amount of traffic on the network, but if the amount of redundancy is controlled then this approach can be used. In media-independent FEC, general protection codes like Reed Solomon or Viterbi are used to produce an extra protection packet that follows the protected set of voice packets. These codes do not depend on any particular underlying media characteristics, but introduce a higher delay which may not be tolerated by many applications, including VoIP. In media-dependent FEC, the sender uses a high-quality codec to create the voice samples and a lower quality codec to generate redundant bits that are added to every packet rather than being sent in their own separate packet. The receiving codec removes the redundancy. If the receiver must use that redundant data to substitute for a lost packet, the result is a lower-quality (but not missing) segment of voice. It introduces minimum delay but may introduce more computational processing delay as it is media dependent. Concealment techniques can be used to supplement FEC for even better lost-packet compensation. The most common concealment approaches include:

- Silence substitution is substitution of the lost frame by a silence frame of the same temporal frequency, but it can introduce noise if several of them are introduced.

- In noise substitution Gaussian noise frames are used to substitute for the missing frames. This produces better quality.
- In frame replication, missing frames are replaced by already present redundancy in the voice. This has low computational complexity and is efficient as more redundancy is expected to be present in the neighboring voice frames. It does not need large temporal size.

- Waveform substitution uses the frames prior to the lost frames and tries to use the most recent ones. It examines buffered frames and searches for the best match.

Optimized unequal error protection

Forward Error Correction schemes allocate equal amounts of error-control resources to each voice packet, irrespective of the perceived importance of a packet. This technique proposes signal-adaptive, unequal error protection in which certain packets are allocated more FEC protection than others, depending on their Perceived importance. Here, the basic unequal protection varies

the number of copies of a packet that are piggybacked onto subsequent packets: an adaptive Reed Solomon (RS) coding scheme provides only certain packets with RS protection. The error control resources that should be allocated to the packet are determined by anticipating what the packet loss concealment will do if the packet is lost, and calculating the expected distortion for different protection scenarios. This technique is based on Lagrangian optimization used in video communications to balance rate against distortion.

VI. CONCLUSION

In this paper, we have suggested a high capacity steganography approach for embedding secret information in inactive frames with low bit rate audio streams by using G711 codec. The audio streams are encoded into audio frames by using the codec and the voice activity detection has been identified using the Voice Activity Detection (VAD) algorithm. Moreover, packet losses have been detected during the integration of the hidden message and to resolve using replacing lost packets method.

VII REFERENCES

[1] Yong Feng Huang, Shanyu Tang,” Steganography in Inactive Frames of VoIP Streams encoded by Source Codec” ieee transactions on information forensics and security, vol. 6, no. 2, June 2011.

[2] Z.Wu and W.Yang, “G.711-based adaptive speech information hiding approach,” *Lecture Notes Comput. Sci.*, vol. 4113, pp. 1139–1144, 2006.

[3] N.Aoki,”A technique of lossless steganography for G.711 telephony speech,” in Proc. 2008 4th Int. Conf. Intelligent Inf. Hiding Multimedia Signal Process. (IIH-MSP), Harbin, Aug. 2008, pp. 608–611.

[4] L.Ma, Z. Wu, and W. Yang, “Approach to hide secret speech information in G.721 schemes” Lecture Notes Computer. Sci., vol. 4681, pp. 1315–1324, 2007.

[5] P.Chang and H. Yu, “Dither-like data hiding in multistage vector quantization of MELP and G.729 speech coding,” in Proc. Conf. Rec. 36th Asilomar

Conf. Signals, System. Computer. Nov. 2002, vol. 2, pp. 1199–1203.

[6] B.Xiao, Y.F.Huang, and S. Tang, "An approach to information hiding in low bit rate speech stream," in Proc. IEEE GLOBECOM 2008, Dec.2008, pp. 371–375, IEEE Press. Z.Wu, W.Yang, and Y.Yang, "ABS-based speech information hiding approach," Electron. Lett. vol. 39, no. 22, pp. 1617–1619, Oct. 2003.

[7] Z.Wu, W.Gao, andW. Yang, "LPC parameters substitution for speech information hiding," J. China Univ. Posts Telecommun., vol. 16, no. 6, pp. 103–112, 2009.

[8] C.Krätzer, J.Dittmann, T. Vogel, and R. Hillert, "Design and evaluation of steganography for voice-overip," in Proc. IEEE Int.Symp. Circuits Syst., May 2006, pp. 2397–3234.

[9] C. Bao, Y. F. Huang, and C. Zhu, Steganalysis of compressed speech," in *Proc. IMACS Multiconf. Computational Eng. Syst. Applicat.(CESA)*, Oct. 2006, pp. 5–10.

Authors

V.Ramesh Kannnan is a PG Scholar in a Department of CSE (PG) at Sri Ramakrishna Engineering College, Coimbatore. Email ID: rameshkannan.v@gmail.com

J. Selvakumar received the B.E degree in computer Science & Engineering from the Madras University in May 2001 and the M.E degree in Computer Science & Engineering from the Bharathiyar University in December 2002. He is currently working towards the Ph.D degree at the Anna University, Chennai. He is currently an Assistant Professor. His research interests in Requirements Engineering.