ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

CONCEALING THE SECRET MESSAGES IN INACTIVE FRAMES USING SOURCE CODEC

Dr. N.Saravana Selvam[1], S.Mohana Gowri[2], P.Dhivya[3]

Professor, Department of CSE, Sri Eshwar College of Engineering, Coimbatore, India[1]

Master of Engineering, Department of CSE, Sri Eshwar College of Engineering, Coimbatore, India[2,3]

Abstract: The paper makes use of high capacity steganography algorithm for producing high data embedding capacity. This is achieved by embedding data in low bit rate audio stream. The term "VoIP" describes the digitalization, compression and transmission of analog audio signals (in majority of speech signals) from a sender to a receiver using IP packets. Any system can be as client and any system made as server. When the client speaks the audio stream will be in the form of Analog signal. By using improved Voice Activity Detection (VAD) algorithm the analog signal is separated into active frame and inactive based on the value of energy and threshold. Active frames are of high bit rate and Inactive frames are of low bit rate. When the value of the energy is less than that of threshold then the signal is non-speech signal ,i.e. the signal is inactive then it comes under Inactive frame. Voice data will be embedded in active frame and text data will be embedded in inactive frame. High capacity steganography algorithm is used to embedded the data in inactive frame. G.723.1 codec is used to compress the audio stream. Then the voice data and text data is received at the receiver end in the form of analog signal. The communication in live so that it ensures security. Therefore this paper meets the requirements of information concealing, and satisfies the secure communication speech .

Keywords: Steganography, Inactive frames, Concealing, VoIP,Speech codec

I. INTRODUCTION

The goal of the research is to provide secure communication over VoIP using steganography. Simple requirements are internet connection and the IP address of the systems being used. In the early stages the communication over VoIP was not much secure. Steganography method for text data was used by making use of images to hide text data. This will be done either by using mail services or by giving directly to the receiver. Least Significant Bit(LSB) algorithms and Analysis By Synthesis(ABS) algorithms were used in early days for VoIP communication[5],[6]. The least significant bit is the bit position in a binary integer giving the units value, that is, determining whether the number is even or odd. ABS is the method of determining the parameters of a speech coder in which the consequence of choosing a particular value of a coder parameter evaluated by locally decoding the signal and comparing it to the original input signal.

This work proposes Improved Voice Activity Detection(VAD) algorithm where the audio frames will be detected and separated as active frame and inactive frame based on the value of energy and the value of the threshold. This algorithm is suggested for detecting inactive audio frames taking into packet loss account. Voice data will be embedded in the active frame where the energy level is more than the threshold level. Text data will be embedded in the inactive frame where the value of the energy is less than that of the value of the threshold. Recent studies have found that inactive frames have high embedding capacity compared to the active frames. Steganography is used by embedding text data in the inactive frame the two algorithms are used in this paper namely Embedding algorithm and Extracting algorithm as shown in Fig 1. Embedding algorithm is used to embed data and it will be used by the sender. Extracting algorithm is used to extract data and it will be used by the receiver. G.723.1 compresses voice audio in 30ms frames. The embedding data in various speech parameters led to different levels of concealment. Therefore perfect imperceptibility and high data embedding capacity is achieved.

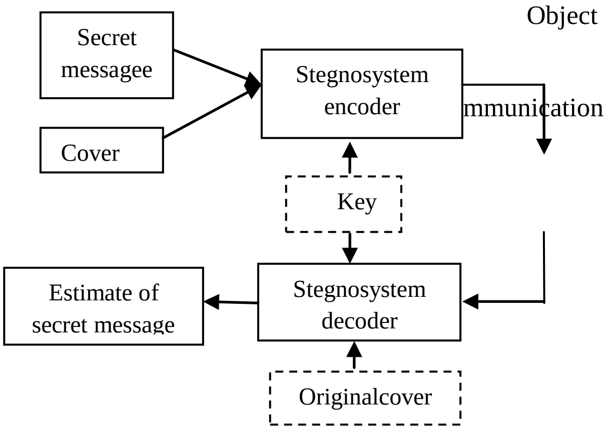

Fig.1 Steganography

1.1 Improved VAD Algorithm

In Voice over Internet (VoIP), Voice Activation Detection (VAD) could be a software system application that permits an information network carrying voice traffic over the net to sight the absence of audio and conserve information measure by preventing the transmission of "silent packets" over the network. Most conversations embody regarding five hundredth silence VAD (also known as "silence suppression") are often enabled to observe signals for voice activity so once silence is detected for a nominal quantity of our time, the appliance informs the Packet Voice Protocol[1] and prevents the encoder output from being transported across the network.

Voice activity detection (VAD) is extremely necessary for language applications like speech recognition, hands-free telecom and speech committal to writing. Once noise-free speech is noninheritable , a correct threshold set within the amplitude permits comparatively simple detection of the speech amount. However, real speech is distorted by background like computer-fans, air-conditioners and plenty of different surroundings sounds, particularly in distant-talking things. Inaccurate detection of the speech amount causes serious issues like degradation of recognition performance and deterioration of speech quality. it's thus extremely fascinating to develop a sturdy and reliable VAD technique.

Voice Activation Detection also can be accustomed forward idle noise characteristics (sometimes known as close or comfort noise) to a distant scientific discipline phonephone or entree. The universal normal for digitized voice is sixty four Kbps,which could be a constant bit rate whether or not the speaker is actively speaking, is pausing between thoughts, or is completely silent. While not idle noise giving the illusion of a relentless transmission stream throughout silence suppression, the auditor would be seemingly to suppose the road had gone dead. that the improved VAD algorithmic rule isn't laid low with packet loss, thereby guaranteeing the VAD result to be thought of between the sender and also the receiver. A new technique of police investigation active voices is then steered that's comparison the threshold with the residual energy of the frame instead of the energy of the frame as in Table 1.

Inactive frames, if Energy < Threshold
Active frames, if Energy > = Threshold
The design of a VAD algorithmic rule square measure

Speakers	ABS coder	G.729	Common Percentage
Female	42.71%	32.29%	25.00%
Male	52.08%	19.79%	28.13%
Total	47.40%	26.04%	26.56%

Table.1 Range of ABS

- Noise reduction stage.
- Then some properties square measure calculated from a region of the input.
- A classification rule is applied to classify the section as speech or non-speech – usually this classification rule finds once a price exceeds a threshold.

There could also be some feedback within the sequence, during which the VAD is employed to enhance the noise estimation within the voice reduction stage, or to vary the threshold. These feedback preference improves the VAD performance in non-stationary noise (i.e. once the noise varies a lot).Independently from the selection of VAD algorithmic rule, satisfactory compromise between having voice detected as noise or noise detected as voice . In these tough detection conditions it's usually desirable that a VAD ought to fail-safe Associate in nursing info, indicating speech that is detected once the choice is in unsure, to lower the possibility of losing speech segments. The largest downside within the detection of speech during this surroundings is that the terribly low signal/noise ratios (SNRs) that square measure encountered[7]. It should be not possible to tell apart between speech and noise exploitation easy level detection techniques once elements of the speech auditory communication square measure buried below the noise.

II. EXISTING SYSTEM

Streaming media, like voice over internet Protocol (VoIP) streams, area unit broadcast relive the web and delivered to end-users. Security remains one amongst the most challenges with this new technology. With the upsurge of VoIP applications on the market to be used in recent years, VoIP streams become one amongst the foremost attention-grabbing cover objects for contemporary steganography[2].

Digital steganography in low bit audio streams is usually considered a difficult topic within the field of information activity. However, VoIP area unit typically transmitted over low bit rate audio streams encoded by the supply codec like ITU G.723.1 codec to save lots of on network information measure. Low bit rate audio streams area unit less seemingly to be used as cover objects for steganography since they need fewer least vital bits than high bit rate audio streams[3]. Very little effort has been created to develop algorithms for embedding knowledge in low bit rate audio streams. The algorithms that were used area unit as follows

- LSB(Least Siginificant Bit) algorithmic program
- ABS(Analysis By Synthesis) algorithmic program

III. PROPOSED SYSTEM

The projected system of steganography algorithms have constrains on the information embedding capacity; that is, their knowledge embedding rates square measure too low to own sensible applications to cut back network information measure in VoIP applications, some supply codecs introduce silence compression throughout the inactive amount of audio streams. The silence compression technique has 2 components:

i) Voice Activity Detection (VAD)
ii) Comfort noise generator

The VAD is employed to make a decision whether or not this audio frame is a full of life voice by scrutiny the energy of the frame (enr) with a threshold (thr). The system additionally implements that the projected steganography formula is additional appropriate for embedding knowledge in inactive audio frames than inactive audio frames. However,the projected formula comes intosensible use in covert VoIP communications, and assure the integrity of hidden messages within the case of packet loss. The source codec employed in the projected system is G.723.1 and it takes the advantage of the prevailing system with less compression rate and packet loss. G.723.1 could be a results of a contest that ITU proclaimed with the aim to design a codec that might permit calls over twenty eight.8 and thirty three kbit/s electronic equipment links. There were 2 excellent solutions and ITU determined to use them each. as a result of that, it's 2 variants of G.723.1. They each operate audio frames of thirty milliseconds (i.e. 240 samples), however the algorithms disagree.The bitrate of the primary variant is sixty four kbit/s and therefore the MOS is three.9. The bitrate of the second variant is fifty three kbit/s with MOS=3.7. The encoded frames for the 2 variants square measure twenty four and twenty bytes long, respectively.

3.1 Data Embedding and Extracting

Inactive frames square measure appropriate for information embedding. All the speech parameters square measure sorted into 3 physical property levels of steganography in terms of the space of S/N (DSNR), that is outlined because the distinction in S/N (SNR) between the initial speech and stego speech. Close analysis of the info in shows the physical property levels of steganography completely different for various parameters of the inactive frames square measure wide different. Thus it's attainable to decide on completely different parameters and numerous parameter bits to plant information on demand of sensible applications.

In short, the parameters marked with level 1–2 square measure appropriate cover objects for steganography. One set of techniques makes an attempt to estimate the noise and take away its effects from the target speech. whereas noise estimation will add low-to-moderate levels of slowly variable noise, it fails utterly in louder or additional variable conditions. A second approach utilizes noise models and makes an attempt to decrypt speech taking into consideration their presence. Again, model-based techniques will work for straightforward noises, however they're computationally advanced beneath realistic conditions and need models for all sources present within the signal.

3.2 Extracting and Embedding Algorithm

The embedding process in steganography over VoIP is divided into four steps

The speech with PCM format is divided into frames ,F={f1,f2,……..,fi}.Each frame is given into the VAD detector.The frame is marked with “A” if it is determined to be an active frame, otherwise marked with “S”.

$$\begin{cases} fi = f_i^{A}\text{ , if fi is an active frame} \\ fi = f_i^{S}\text{ , else} \end{cases}$$

Encoding all the frames by G.723.1 codec with 6.3kb/s.The resulting low bit rate audio stream containing acive and inactive frames is then outputted from the codec.
$F^{*} = \{f_i^{*A} , f_j^{*S}|i=1,\ldots..,N1,j=0,\ldots..,N2\}$

According to the frame type ,two different steganography algorithms are used to embed information in the frames

$$f_i^{-}= \Theta1(f_i^{*},S)= f_i^{*S}\ \Theta\ S,\text{if } f_i^{*}= f_i^{*S}$$
$$f_i^{-} = \Theta2(f_i^{*},S)= f_i^{*A}\ \Theta\ S,\text{if } f_i^{*} =f_i^{*A}$$

The inactive frames and active frames with hidden information are encapsulated in VoIP packets, which are transmitted over the Internet.
$P=\{pi|pi = \Theta f_i^{-}, i=1,\ldots.,n\}$

The extracting process in steganography over VoIP is divided into four steps

- The VoIP packets are received ,buffered and then decapsulated by the receiver. The decapsulation algorithm is described as
$F=\{fi|fi = \Theta^{-1}(pi), i=1,\ldots.,n\}$
- The buffered frames are copied to the decoding buffer and decoded into the PCM formatted audio stream
$F^{’}=\{f_i^{’A}, fi^{’S}|\ i=1,\ldots\ldots,N1,\ j=1,\ldots,N2\}$
- The active and inactive frames of the low bitrate audio stream are identified by the receiver
$F=\{fi|i = 1,\ldots,n\}$.

3.3 Active and Inactive frames

Client and also the server have to be compelled to communicate to share their information. The speech from the consumer aspect are within the variety of analog signal. Analog signal is separated into 2 frames particularly active frame and inactive frame. The active frame and inactive frame is separated supported the value of energy and threshold. once the value of the energy is larger than or capable the value of the edge then that individual frame are thought of as active frame. Once the value of the energy is a smaller amount than the value of the edge then it'll be thought of as inactive frame.

When the value of the energy is larger than or capable the value of the edge then that individual frame are thought of as active frame. Once the value of the energy is a smaller amount than the value of the edge then it'll be thought of as inactive frame.

The improved Voice Activity Detection(VAD) is employed to determine whether or not the present audio frame is an energetic voice by comparison the energy of the frame(enr) with a threshold(thr).

- Vad=1,enr>=thr ------------⟶ active frame

- Vad=0,enr Voice knowledge are present within the active frame. Text knowledge are present within the inactive frame. Steganography methodology is applied here by putting text knowledge within the inactive frame.

The speech communication that takes place between the consumer and also the server is live so it ensures security. there's no explicit server that keeps observation the conversations. Everything takes place solely between the shopper and also the server so the voice knowledge and also the text knowledge area unit sent and received only between the chosen consumer and chosen server so secure communication are achieved. Periods characterised by low rates of packet loss. The gap proportion is that the proportion of time that the decision tough low-rate packet loss; the gap density is that the actual proportion rate of packet loss throughout the gaps.

To reduce network information measure in VoIP applications, some supply codecs introduce silence compression throughout the inactive amount of audio streams. The silence compression technique has 2 components: Voice Activity Detection (VAD) and comfort noise generator. The VAD is employed to determine whether or not the present audio frame is an active voice by comparison the energy of the frame with a threshold, the system additionally implements that the projected steganography algorithmic rule is additional appropriate for embedding knowledge in inactive audio frames than inactive audio frames. The system projected a high-capacity steganography algorithmic rule for embedding knowledge within the inactive frames of low bit rate audio streams encoded by G.723.1 supply codec.

3.4 Speech Codec G.723.1

G.723.1 may be a dual rate speech computer user normal from International Telecommunication Union–Telecommunication standardization sector (ITU-T), for compressing the toll quality speech (8000 samples/second). compressing the speech or alternative audio signal element of transmission services will be done at a very low bit rate. the standard applications of this speech computer user square measure in telecom over packet networks, like Voice-over-Internet-Protocol (VoIP). This computer user has two bit rates, 5.3 and 6.3 Kbps. each bit rates share identical short analysis techniques for process the speech. For long-run analysis of speech, the algorithms used area unit completely different. For 5.3 Kbps computer user, Algebraic Code Excited Linear Prediction (ACELP) principles square measure used wherever as in sixty three Kbps computer user, Multi Pulse-Maximum Likelihood Quantization (MP-MLQ) techniques square measure used. The computer user works on a frame of 240 speech samples (30 msec). Besides, there's a glance prior to sixty samples (7.5 msec). Therefore the total recursive delay for the computer user is 37.5 msec.

REFERENCES

[1] Aoki.N "A band extension technique for G.711 speech using steganography," IEICE Transactions on Communications,Vol.E89-B, No.6, pp.1896-1898,June2009.
[2] Aoki.N "A packet loss concealment technique for VoIP using steganography based on pitch waveform replication", IEICE Transactions on Communications, Vol.J86-B, No.12, pp.2551-2560,2008.
[3] Kratzer.C, J. Dittmann, T. Vogel, and R. Hillert(2006) "Design and evaluation of steganography for VoIP", in Proc. IEEE Int. Symp.Circuits Syst., pp. 2397–3234,2008.
[4] Lu.Z.M, Yan.B, and Sun.S. "Watermarking combined with CELP speech coding for authentication", IEICE Trans. Inf. Syst., Vol. E88-D, No. 2, pp. 330–334,2007.
[5] Wu.Z, Gao.W, and Yang.W "LPC parameters substitution for speech information hiding", J. China Univ. Posts Telecommunication., vol. 16, no. 6, pp. 103–112,2007.
[6] Wu.Z, Yang.W, and Yang.Y "ABS-based speech information hiding approach",Electron. Lett., Vol. 39, No. 22, pp. 1617–1619,2009.
[7] Xiao.B, Y. F. Huang, and S. Tang(2008), "An approach to information hiding in low bit rate speech stream" ,in Proc. IEEE GLOBECOM 2008, pp. 371–375, IEEEPress,2011.

BIOGRAPHY

DR. N. SARAVANA SELVAM has obtained his Ph.D. in Computer Science and Engineering from Anna University, Chennai in the year 2013. He has obtained both of his Post Graduate degree, M.E.(Computer Science and Engineering) and Graduate degree B.E., (Electronics and Communication Engineering) from Madurai Kamaraj University (Tamilnadu, India). He is currently serving as Professor & Head of Department of Computer Science and Engineering at Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu. During his fifteen years of teaching profession, he shouldered a member of teaching, administrative and societal based assignments. He is a Life Member of ISTE, IAEng and IACSIT. Currently, he is specializing in the area of NetworkEngineering.

S.MOHANA GOWRI received her B.Tech(IT) Degree from Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamilnadu, India and pursuing M.E Degree from Sri Eshwar College of Engineering, Coimbatore, India. Her fields of Interest are Operating systems, Network Security and Software Engineering.

P.DHIVYA received her B.Tech(IT) Degree from P.A College of Engineering, Pollachi, Tamilnadu, India and pursuing M.E Degree from Sri Eshwar College of Engineering, Coimbatore, India. Her fields of Interest are Network Security, Operating systems and Data structures.