International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 3, March-2014, pp: (8-11), Impact Factor: 1.252, Available online at: www.erpublications.com

Voice Compression and Speech Coding Techniques for faster & secured communication (A Review)

Karan Yadav[1], Sukhwinder Singh[2]
[1]Student, [2]Mentor
[12]Dept. of Electronics and Communication Engineering, PEC University of Technology, Chandigarh, India

Abstract: Ever since time immemorial living things to include humans and animals, have been using sound to convey information to one another. The human beings evolved leaving other animal species far behind and discovered their own special language in which they are able to communicate with one another. Even in the past centuries flag signals, hand signals, smoke, drums and pigeons were used for communication. With the fast pace of modernization and advancement of technology the simple phenomenon of speech which we now call voice has become the lifeline of mankind. This has shaped in various keywords like analog, digital, voice compression, encryption, decryption to name a few. This paper gives an overview on the various modes of communication techniques, voice compression and encryption.

Keywords: Decryption, Digital Telecommunication, Encryption, Speech Coding, Voice Compression.

Introduction

In the news recently we are aware how USA used modern communication techniques to eavesdrop on voice and cyberspace of almost all countries to which there was an upsurge of anger all over the world. The world has now realized the importance of encryption of data to secure oneself from unwanted agencies and also to have capacity to handle large amount of data. As we are aware initially analog communication was used which had basic limitations on the number of voice channels that could be used and also limited encryption techniques. With the advent of digital communication voice can now be made available on a large scale to fulfill the requirement of the ever growing population and latest state of the art gadgets. Digital communication has completely revolutionized the encryption process being used now days. The process of compression, encryption and decryption can be explained as follows:

The original speech is given as the input to the speaker identification for authentication. Using the speech recognition techniques the speaker is identified. The authorized speakers' names are recorded in a database and a copy of the database will be loaded in the destination. For all speakers the index value is obtained from the data store. Names of all the speakers are indexed so that by using the index number the speaker is retrieved from the database. In order to find the message digest, secure hashing algorithm is applied to the speaker index. Next the original speech is send to the encoder to perform compression. If the data is too long then it will take more time for transmission so in order to avoid that the data is compressed and it is sent to the next conversion. By combining secure hashing algorithm bits with encoded speech the message digest is recovered, then the next process encryption is done. It is the process of converting the speech data into another form that is not understood by unauthorized people. After receiving the data, the receiver decrypts the message to convert it into the original form and it is send to the decoder to decompress the data. At the receiving end new speaker Id is calculated using the synthetic speech and the index value is obtained from the local database. A message digest is obtained by applying hashing algorithm on the index value. Both the received message digest and the new message digest are compared, if there is any mismatch the user at the receiving end is alerted [1].

This paper is organized as follows: Section-II describes speaker identification, section-III describes voice compression and section-IV focuses on speech encryption and decryption. Finally, the conclusion is summarized in section V with future work.

Speaker Identification

Speech signal is used to convey the linguistic information as well as speaker information such as emotional, regional and physiological characteristics. In today's world of ever increasing competition in intelligence, surveillance and military technology, secrecy and data protection from unauthorized access is a must. Such unauthorized access to information can

be minimized to a great deal by a technology known as speaker recognition which is used to find out "who is speaking". The identified speaker is authenticated against such attacks.

By combining secure hashing algorithm bits with encoded speech the message digest is recovered. There are many cryptographic hash functions available and it includes some of the important properties such as [2]

- Hash value computation for any given message is very easy.
- It is difficult to reproduce a message that has a given hash.
- Without changing the hash it is infeasible to modify a message.
- Two different messages with the same hash are very difficult to find.

Some of the hashing functions are Message Digest, Secure Hashing Algorithms (SHA), RIPEMD, GOST and HAVAL. Although there is a long list of hash functions many of them are found to be vulnerable. For example, in August 2004 the popular algorithm such as SHA-0, RIPEMD were found to be weak and long term security algorithm was derived such as SHA1, RIPEMED 8 and 160. In 2009 the most commonly used algorithm are MD5 and SHA1. The SHA-0 and SHA-1 were developed by National Security Agency. Still there is a competition for replacement of SHA-2 [3], and also to ensure the long term toughness of applications that use hash functions.

Voice Compression

Communication in the past suffered from inherent limitations of limited bandwidth due to limited transmission lines and end equipments. The need was felt to provide the basic necessity of voice communication to as large base of subscribers as possible. This led to the advent of voice compression techniques where in large amount of data can be passed on lesser number of channels. Example 2 Mbps, 8 Mbps, 34 Mbps, 155 Mbps, SDH etc. Voice compression is achieved by various coding and decoding techniques. The less noisy the voice after compression the better the coding technique. The tandem pairs of coder and decoder is known as codec.

When an analog voice signal is sampled 8000 times per second, and each sample is stored in an 8-bit byte, the bandwidth required is 64,000 bits per seconds (64 Kbps). But carrying 64kbps per channel greatly reduces the number of subscribers because of the limited capacity of T-1(24 channels) and E-1 carriers(30 channels). In order to increase the number of subscribers or to send a larger amount of data on the same carrier, voice compression techniques are used. Only 256 possible amplitude measurements can be represented with 8 bits. 256 digital values are not enough to represent the entire amplitude range of the human voice at a usable quality level. However, most of the characteristics of a voice signal that make it understandable to the human ear exist at the lower end of the amplitude range. Therefore, the values are assigned to amplitude values non-linearly, with many values available to represent various amplitudes in the low end of the range, and few values to measure the high end. This compression method is called companding. Different companding algorithms are used in different geographic regions. A companding method called mu-law is used in the US, Canada, and Japan. Another method, called A-law, is used in the rest of the world.[4]

Voice compression is handled by coder/decoders, or codecs, which are implemented in either hardware or software. A hardware implementation in a silicon chip is preferable because codecs must work quickly so that no delay is perceptible. In recent years, however, typical personal computers have become so fast that software codecs are now effective. International standards organizations have developed three types of speech compression algorithms: waveform codecs, source codecs, and hybrid codecs. Hybrid codecs use a mix of waveform and source methods [5].

Types of Voice Coding Techniques:

- Waveform encoding is used to encode the waveform itself in an efficient way. The waveform coding in time domain is the traditional speech coding which attempts to code the exact shape of the speech signal, without considering the nature of human speech production with speech perception and these coders are high bit rate coders. The PCM, DPCM and ADPCM are used to directly code the received audio signal. The Pulse Code Modulation (PCM) is used to digitize the signals through signal conversion by applying Sampling Theorem (sampling frequency =8000Hz). Differential Pulse Code Modulation (DPCM) uses the baseline of PCM but compresses the voice by reducing the quantization bits by predicting the future signal values. The Adaptive Differential Pulse Code Modulation (ADPCM) which is then used to provide even more compression uses a functional model of human speaking mechanism at the receiver end [6].

- The Sub band coders are not widely used and it is used to parameterize the speech signal in terms of spectral properties in different frequency bands, it can be used for high quality audio coding. In the transform coding the signal is transformed to its representation in another domain in which it can be compressed well than its original form. Here if the signal is decompressed, an inverse transformation is applied to restore an approximation of the original signal. LPC (Linear Predictive Coding) techniques are the most popular techniques in the synthesis and coding of speech which is used to provide low bit rate speech data [8].

- Hybrid coders are used to encode speech, and the bandwidth requirement lies between 4.8 and 16kbps. The hybrid coders include CELP, MPE and RPE coders. Multiple Pulse Excited coding (MPE) and Regular Pulse Excited coding (RPE) techniques try to improve the speech quality by giving a better representation of the excitation signal. The MPE method produces high quality speech at rates around 9.6 kbps. Codebook Excited Linear Prediction (CELP) technique can also be called as analysis by synthesis technique and it allows bit rates of even 4.8 kbps [9].

Properties of a speech coder are:

- Low bit rate: If the bit rate is lower for encoded bit- stream, then the bandwidth will also be lower for transmission which leads to a more efficient system [7].
- High speech quality: The decoded speech should have a high quality so that it will be understandable and acceptable by the target application.
- Robustness in the presence of channel errors: It should be error tolerant to work under the harsh acoustic environments.
- Low coding delay: The coding algorithm should be flexible and faster to convert the input speech of the encoder with respect to the output speech of the decoder.

Encryption and Decryption

Encryption and decryption is a technique in which the intelligent signal is converted into an incomprehensible form for transmission over unsecure line. In encryption the data is converted into an incomprehensible form and after transmission from point A to point B at the receiving end it is converted back into the intelligent signal by using decryption techniques. This saves the valuable information from being monitored by unwanted agencies. Since voice is transmitted by various media like satellite, underground, submarine cable hence it is susceptible to monitoring.

Techniques of encryption

The two basic techniques of encryption are symmetric and asymmetric encryption.

➢ Symmetric Encryption

Symmetric encryption also called single key encryption, one key encryption or private key encryption. It uses the same secret key to encrypt and decrypt information. It is essential that the sender and the receiver should know the secret key, which will be helpful to encrypt and decrypt the whole information. Some of the commonly used encryption techniques are listed in “table1” [9].

➢ Asymmetric Encryption

This method uses different keys for encryption and decryption. Here the public key is made available to everyone so that they can send messages but the private key is made available only to the person it belongs to. However, it has two major disadvantages such as it is based on non trivial mathematical computations, and it is very slower than the symmetric ones. Some of the popular examples of asymmetric encryption algorithm include RSA, DGA and PGP [11]. The RSA algorithm is the best known public key algorithm. The key used for encryption is a public key and the key used for decryption is a private key.

Table 1: Commonly used Symmetric Encryption

Symmetric Encryption Algorithm	*Developer*	*Block size*	*Cryptanalysis resistance*	*Security*
Advanced Encryption Standard	Vincent Rijmen and Joan Daemen in 2000	128-, 192- or 256-bit	It is very Strong against truncated differential, linear, interpolation and square attacks	More secure
Data Encryption Standard	IBM in 1977	64bit block	Vulnerable to differential and linear crypt analysis	Proven inadequate
Triple Data Encryption Standard	1978	64bit block	Vulnerable to differential, Brute force attacker could be analyze plain text using differential crypt analysis	One only weak which is exit in DES
CAST	Carlisle Adams and Stafford Tavares in1996	64bit block		Very fast and efficient [10]

Conclusion and Future Work

In this paper the importance of voice compression and encryption and decryption techniques was discovered to provide swift and secure communication to a large subscriber base. So that communication is possible in real time and with utmost secrecy without loss of information unscrupulous elements. The future work is to develop more advanced voice compression techniques and to provide foolproof encryption technology to meet the ever growing requirements of this modern world.

References

[1]. D.Ambika, "Secured Speech Communication - A review", International Journal of Engineering Research and Applications, Vol 2, Issue 5, September-October, 2012.
[2]. For speaker identification: http://en.wikipedia.org/wiki/Cryptographic_hash_functions.
[3]. http://csrc.nist.gov/groups/ST/hash/sha-3/index.html.
[4]. T1 & E1 Trunk Channels, NMS Communications, 1 Feb, 2001.
[5]. Priscilla Oppenheimer, "Digitizing Human Vocal Communication", Article, [www.priscilla.com/troubleshootingnetworks/html].
[6]. Mark Johnson, University of Illinois, "Speech Coding: Fundamentals and Applications", Paper, [www.ee.ucla.edu/~spapl/paper/mark_eot156.pdf].Pg 1-5.
[7]. TE-4107 Digital Telephony, "Voice Digitization-I"[pdf].
[8]. Prof Murat Torlak, EE 4367 Telecom Switching and Transmission, "Voice Transmission", Lecture, University of Texas, Dallas, USA, March 27, 2007.
[9]. D.Ambika, "Speech coding and Identification Techniques for Net-Centric Communication", International Journal of Emerging Trends in Engineering Development, Vol 7, Issue 2, November, 2012. Pg 433-435.
[10]. For encryption and decryption: http://omnisecu.com/security/public-key-infrastructure/symmetric-encryption-algorithms.html
[11]. http://zybersene.blogspot.in/2012/06/symmetric-encryption-vs-asymmetric.html