The Design of Secure File Transfer with Speech Recognition

Vinita Malik
Information Scientist
Central Library
Central University of Haryana
Mahendergarh, Haryana, India

Anu Saini
Assistant Professor
Department of Computer Science
G.B.Pant.Engineering College
New Delhi, India

Sukhdip Sangwan
Assistant Professor
Department of Computer Science
D.C.R.U.S.T, Murthal
Sonipat, Haryana, India

Abstract: The design of a secure file transfer system using Speech Recognition technique is investigated. Though cryptography enforces protection specifications, complements various other protection mechanisms. Different file types are enciphered by different cryptographic keys, supplied at access time by the users, who are being authorized using speech recognition technique. Several protection implementation schemes are analyzed according to the criteria such as: security, efficiency, and user convenience.

Keywords: File Transfer, Speech Recognition, Cryptography

I. INTRODUCTION

"Secure File Transfer" is an application written in JAVA programming language in which Java Files are transferred between Client and Server over a TCP based network [1], such as the Internet. Files sent by Client are Digitally Encrypted and Digitally Signed as well and Server checks the HASH CODE and DIGITAL SIGNATURE of the file sent by Client for Security purposes [2]. Various methods in cryptography try to achieve secure the required information to guarantee that only authenticated users can access the data. It can make signal difficult to understand by the unauthorized users [3].We have two classes for speaker recognition and these are broadly divided into speaker identification and then speaker recognition[4]. These methods are used to identify speaker by collecting individual information and then by integrating them into speech waves. Speaker recognition techniques are applied in speaker's voice for verification of their identification and then controlling access to various services like telephone banking, accessing databases, telephone shopping, information providing services and to secure confidential information.

Other applications in speaker recognition includes forensic reports development .This technology has a bright future in terms of that its going to resolve a number of unsolved issues[5].Speech recognition can also be done by coherence-based post filter which gives really a good performance. Coherent-to-diffuse power ratio can be used to find out signal to noise ratio[6].Speaker recognition is again fragmented in two methods and these can be text- dependent as well as text independent methods both[7] . In the first type of method, the speaker will tell key words or he will talk in sentences and key words must have the same text for both recognition as well as training trials. Whereas in the second method speaker is not dependent on a specific text being spoken.

Our project main aim is first to identify the speaker by comparing speech signal from speaker who is not known to our database where known speakers are stored. In the process of identification of speaker identification we first determine which registered speaker are going to provide a given speech. It's being used to authenticate the users to securely transfer the files.

II. RELATED WORK

Since years, efforts are being made to make file transfer system as easy, compliant and secure. Firstly, the design of a user-controlled cryptographic secure file transfer system was investigated. Later on various algorithms, techniques and security extensions are established in time to provide secure file transfer over the expanding networks as timely, consistent or efficient service.

Recently, systems have been developed to securely transfer the files by encrypting and decrypting the files using secure algorithms like RSA and many more [8]. These days we have email as the most important file-transfer solution in most of organizations. Due to implementation of telework or work-from-home policies in company, for collaborating on software projects the employees are more dependent on sharing of documents by help of e-mail. Although the email system was never designed for file transferring only in organizations and it got suffered in many ways : (1) Email management systems overflowed with lot of contents and it got very difficult to manage them by IT staff. (2) If various workgroups wanted to have access to the shared data then they must be member of the sender's list of recipients [9].

The Administrator speech can be authenticated by conversion of speech waveform to parametric forms for its better recognition .One possibility can be for this is Mel-Frequency Cepstrum Coefficients (MFCC) [10]. Recognition of speech can be based on audio visual classifiers also [11].

Speech Systems use several pattern classifiers i.e. for classification of emotions, recognition of speech, verification of speaker. Mostly classifiers are divided in two forms Linear and Non-Linear [12].

As we are aware that articulatory information can help in modelling of speech variability .Neural networks of convolutional nature can help in collecting articulatory information in a better way than deep learning neural networks [13].

As various types of files need to be exchanged between various organizations then no intermediate documents should be made while transfer of files and transforming process which can again cause further security problems [14]. The file transfer systems that can be used for transferring large files again suffers from several issues like they may lack security because various users share their log in credentials for transferring file among them and again the content of file sharing systems is very unmanaged, So we require a secure file transfer solution .

III. PROPOSED WORK

As per need of the hour, what we need to focus is high security considerations for secure file transfer. Thus we propose securing file transfer scheme, by authenticating the users by using speech recognition technique. It has been well established as a small scale application.

The project includes the following secure features:

A. Encryption: Its basically tells how can we convert data into cipher text which unauthorized users can not easily understand. To get the encrypted data into original state for understanding purpose is called as decryption.

B. Digital Signature: It is a mathematical system for demonstration to tell how authentic a digital message is. Validity of a digital is inherent in giving belief to a recipient that message was from a right sender and was not altered in between.

C. Secure System: The requirements include:

   •The system username and password will be authenticated so that the information is not accessible to any unauthorized person.

   •New user can be created only by making an account.

Secondly, the user can be authenticated by doing speaker identification which we get by comparison of speech signal of speaker who is unknown in nature to our database that stores all known speakers as shown in the Figure 1 below. Identification of speaker is basically the process of recognition of registered users   .Our system have limitation of that if we play back the registered users recorded voice then the system can be deceived sometimes.

For speech recognition the extracted features of speakers voice should meet some criteria such as:

- It must not be mimicry susceptible.
- Whenever we change one speaking environment to another then it should be very little fluctuated .
- It must show stability over time.
- Speech should be frequent and natural in form.

We have used   Mel Frequency Cepstral Coefficients (MFCC) method to accomplish this . From speech signal to extract the required features and for comparison of the speaker who is unknown in nature with the already stored speakers in the database. Thus a referential model (as shown below) has been designed to get a view how speech

recognition module works. In the model, a speech signal is input; its features like pitch, frequency, etc. are extracted and compared with a known database based on some criteria, as explained further. The identification result is given by providing Speakers' ID.
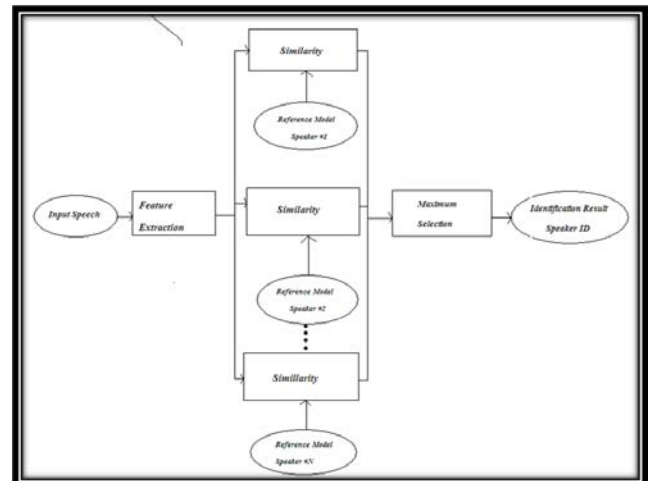

Figure 1 : Proposed system for speech recognition

IV. IMPLEMENTATION

Here in our project   we   use Mel Frequency Cepstral Coefficients (MFCC) technique   to do the   speech recognition by extraction of the features from the speech signal .

Our technique is using one type of Fourier transform named as Cepstrum . Cepstrum is basically   the   Fourier Transformer of the log with unwrapped dphase. How speech signal gets pipelined to cepstrum is shown below in the Figure 2 given as below .

In terms of Mathematics, Cepstrum of signal is FT $(\log(FT(\text{the signal})) + j2\pi m)$.
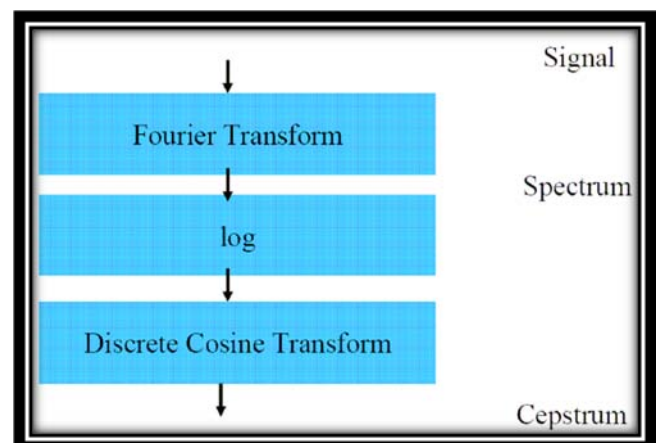

Figure 2: Speech signal gets pipelined to Cepstrum

We use Mel-frequency cepstral coefficients (MFCCs) rather than filter bank coefficients for reliable recognition of speech is used to represent audio signals based on perception [16]. This coefficient is very important role in recognition of speaker. Here all frequency bands are mapped and placed logarithmically and used for approximation of response system more accurately and better than other recognition systems but in the case of Fourier Transform all the frequency bands may not be positioned in logarithmic

manner . Data is processed better by using these coefficients.

Evaluation is done by Mel Frequency Cepstral Coefficients differs from real Cepstrum in that the frequency scale of Mel Cepstrum's is mapped to the Mel scale which is based on observation of speech frequency of human beings. The scale in this technique gets divided into various units or mel. How do we map normal frequency to the Mel frequency is shown in the Figure 3 given below.
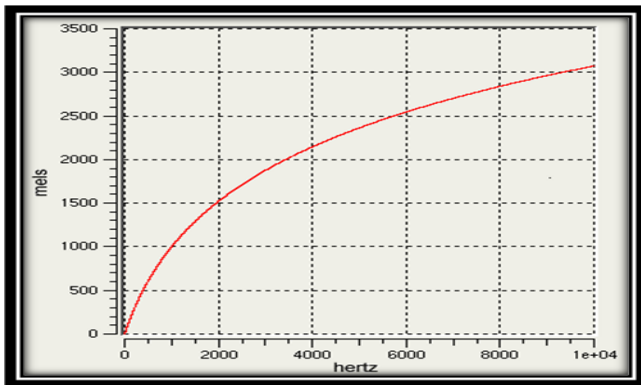

 Figure 3: How to map Normal frequency to Mel Frequency
First we map the normal frequency into the Mel frequency and then we get back to normal frequency by doing the same process in reverse .

The same technique is implemented by coding a MATLAB project using some in-built functions. This MATLAB project mainly features to either add a new speech signal in the database or to recognize a speech sample from the stored database. Thus, in this project module during the recognition of speaker:

In first step, representation of unknown speaker's voice is done by a feature vector sequence taken as {a1, a2 ….ai} and then in second step this sequence is compared with the codebooks which is taken from the database. For the identification of an unknown speaker we find out the distortion distance of two vector sets which is measured by reduction of the Euclidean distance. The Euclidean distance is found out between two points X = (x1, x2…xn) and Y= (y1, y2...yn). The speaker who have the lowest distortion distance is taken as unknown person.

This speech recognition module was created in MATLAB then has been integrated with a file transferring module created in NETBEANS (7.2.1) by importing .jar files from MATLAB project to the NETBEANS source files of the project. This accomplishes the task to securely transfer the files between a client and a server by using speakers' recognition to authenticate the users and then encrypting the transferring file from client to server which decrypts at the server on receive.

V. CONCLUSION

Various users in an organization can access all programs as if they reside on their local personal machines although the same reside on a single file server while transferring files over the network .In case of small workstations where disk space is to be used qualitatively and efficiently, this system can be a great benefit. Our proposed system is highly secure. It helps its users to Transfer files with .JAVA extension to be transferred over internet via secure means. It incorporates various security functions like HASHING and ENCRYPTION for maintaining the privacy of data and also includes DIGITAL SIGNATURES for integrity of the data and Client-Server communications.

VI. REFERENCES

[1] L.Brown , M. G. Jaatun , "Secure File Transfer Over TCP/IP" , IEEE, International Conference on "Technology Enabling Tomorrow : Computers, Communications and Automation towards the 21st Century", 1992

[2] E.Gudes, "The Design of a Cryptography Based Secure File System"; September 1980, IEEE Vol. SE-6

[3] S. Rekik, D. Guerchi , S. Selouani3, "Speech steganography using wavelet and Fourier transforms", EURASIP Journal on Audio, Speech, and Music Processing , 2012, p 1-14

[4] T.Nilsson ,"Speech Recognition Software and Vidispine"; Master's Thesis in Computing Science, UmeaUniversity Department of Computing Science, April 2,2013

[5] A. Srinivasan, "Real time speaker recognition of letter 'zha' in Tamil language" , 2013 Fourth International Conference on Computing Communications and Networking Technologies (ICCCNT), 2013

[6] H.Barfuss, C.Huemmer, A.Schwarz, "Robust coherence-based spectral enhancement for speech recognition in adverse real-world environments", Computer Speech & Language ,2017, p 1-13

[7] S. Sunny , "Performance of Different Classifiers In Speech Recognition"; IJESE, ISSN: 2319 – 1163, Volume: 2, Issue: 4,2013

[8] R.S.Jamgekar, G.S. Joshi "File Encryption and Decryption Using Secure RSA", (IJESE) ISSN: 2319–6378, Volume-1, Issue-4, February 2013

[9] "Overcoming Key File Transfer Problems"; An Osterman Research White Paper, September 2010

[10] H.N.M.Shah, M.Z.A.Rashid "Biometric Voice Recognition in Security System", Indian Journal of Science and Technology, Vol 7(2), 104–112, February 2014

[11] L. D. Terissi , G. D. Sad ,"Noisy Speech Recognition Based on Combined Audio-Visual Classifiers Lecture Notes in Computer Science, vol 8869. Springer, 2014

[12] S. S.Jarandel ,S. Waghmare , A Survey on Different Classifier in Speech Recognition Techniques, International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 3, March 2014

[13] V. Mitra, G. Sivaraman, "Hybrid Convolutional neural networks for articulatory and acoustic information based speech recognition " International Journal on Speech Communication, Volume 89, May 2017, Pages 103–112

[14] "Making File Transfer Easier, Compliant and More Secure";An Osterman Research White Paper, February 2012

[15] T.Nilsson ,"Speech Recognition Software and Vidispine"; Master's Thesis in Computing Science, UmeaUniversity Department of Computing Science, April 2,2013

[16] Z. Jun, S. Kwong, Using Mel-Frequency Cepstral Coefficients in Missing Data Technique, EURASIP Journal on Applied Signal Processing :3, 340–346 ,2004