International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Secured Mobile Communication using Audio Steganography by Mel-Frequency Cepstrum Analysis

Aswin.V[1], Narmadha.V[2]

[1, 2]SSN College of Engineering, Chennai, India

Abstract: ***The voluminous amounts of data that are transmitted through wireless network are majorly constituted by speech signal. In the prevailing environment where in speech data that are critical by its nature in terms of privacy must be protected. This paper would focus on suggesting a novel algorithm based on audio steganography to secure voice and speech related data during transmission in mobile devices. Audio steganography is a method of hiding the secret data in the cover medium by which only the sender and the intended receiver are only able to realize the existence of a secret data. The process of steganography is carried out in cepstral domain and the key is constructed using the Mel-frequency cepstral coefficients. The imperceptibility in hearing is exploited in a way where the data are embedded in low power levels to make the detection more complicated. The generic auto selection of the key members helps in making the critical data's existence highly non vulnerable in the cover medium. Thus the algorithm stands at high tolerance on interception.***

Keywords: Mobile Communication, Audio Steganography, Cepstral Analysis, Mel-Frequency Cepstral Coefficients

1. Introduction

Advancements in technology has allowed mobile communication to provide global connectivity to the people at a lower cost. Ever since the evolution of wireless networks radio communication has been used extensively. The transmitter power, type of antenna, frequency determine the range of mobility. The concept of cellular communication was evolved to accommodate increasing number of users. Time division multiple access (TDMA), Code division multiple access(CDMA), Frequency division multiple access(FDMA) are used for this purpose. In mobile communication a mobile station(MS) communicates to a Base station(BS) which is fixed. This in turn communicates to the end user. During transmission signal can take many different paths between the sender and receiver due to reflection, scattering, diffraction .Data transmitted through wireless networks in mobile communication is not secure due to interception. Audio steganography is an efficient technique used for reliable data transfer. Audio steganography is the act of hiding the critical information by concealing it under an audio file. The sender hides the data within a cover file. The exact position of the data hidden depends on the key shared between the sender and receiver. The recipient receives the stego data and performs extraction with the stego data and key as parameters. The main objective of audio steganography is to communicate in a secure and undetectable fashion. In addition, the suspicion that hidden data has been transmitted should be avoided. The audio steganographic model consists of a carrier audio file, the message to be hidden and the password. The stego file is identical to the cover file. This technique exploits the quality of human perception. Human senses are not trained to look for hidden information. There are various audio steganographic techniques for hiding information in such a manner that the hidden message cannot be recognized. In *LSB algorithm* the hidden data replaces the least significant bit of some bytes of the cover file. This technique is effective in cases where the substitution does not cause evidential quality degradation. *Parity coding* is a robust audio steganographic technique which segments a signal into separate samples and embeds each bit of the secret message from a parity bit. In *Phase coding* the phase of an initial audio segment is replaced by a reference phase that represents the secret information.

2. Related Works

2.1 Hiding in Temporal Domain

2.1.1 Low bit encoding(Least Significant Bit)

It embeds each bit from the message to be hidden in the LSB of the carrier audio. This method achieves imperceptibility at high embedding rate, It can be implemented easily and can be combined with other techniques .Nevertheless, it is less robust to the addition of noise making it insecure. Filtration, amplification, noise addition, lossy compression of the stego-audio are capable of destroying the data. Moreover, the message can be easily uncovered by removing the LSB plane completely. The depth of the embedding layer has been increased to improve the robustness but the hiding capacity decreases.

2.1.2 Echo Hiding

In this method an echo is introduced to the host signal following which the data is embedded. The echo eliminates the problem of HAS(Human Auditory System) sensitivity to the external noise. Despite the addition of echo the characteristics of the stego signal remain unchanged. Parameters of the echo signal namely decay rate, initial amplitude and delay are manipulated to achieve data hiding. However, this method suffers a disadvantage with respect to the induced echo signal size.

2.1.3 Hiding in silence intervals

Thee silence intervals of the speech along with the number of samples in the interval are ascertained. The values are decreased by a value x where $0<x<2^n$ and x is calculated as mod(newlength,2^n). Here n is the number of bits required to represent a value from the data to be hidden. The speech interval samples should be amplified and the silent interval samples should be reduced to avoid extraction of wrong data.

2.2 Hiding in Transform domain

The techniques listed under transform domain utilize the frequency masking effect of the HAS. Either the audio signal samples are changed or the masked regions are altered.

2.2.1 Spread Spectrum(SS)

The hidden data is spread across the frequency spectrum. Before the data could be hidden it is multiplied by an M-sequence code. This code is known to both the sender and the receiver. This enables us to recover the message in the event of any values getting damaged. This technique is used in the sub-band domain for an increased hiding rate. To increase the robustness SS is combined to phase shifting.

2.2.2 Discrete wavelet transform

Message is hidden in the LSB of the wavelet coefficient of the cover signal. While embedding the data a threshold is applied and data hiding in the silent parts of the signal is avoided to amend the imperceptibility of hidden data.

2.2.3 Tone insertion

While high power tones are present lower power tones will go unheard. Tone insertion technique is based on this. The tones are inserted at known frequencies and at a low power level. This method is tolerant to attacks such as bit truncation and low-pass filtering but it has low embedding capacity and the data can be easily extracted.

2.2.4 Phase coding

HAS is unreactive to relative phase of different spectral components. Phase coding makes use of this insensitivity. The selected phase components from the original audio signal spectrum is replaced with the data to be hidden. This method is resistant to signal distortion..

2.2.5 Amplitude coding

The HAS characteristics are dependent on the frequency values. Based on this principle, high capacity data is embedded in the magnitude speech spectrum. Distortion of the cover signal is controlled and hidden data security is assured. High hiding capacity is achieved with considerable speech quality.

2.2.6 Cepstral domain(log spectral domain)

In this method data is embedded in the cepstrum coefficients. Prior to this the carrier signal is transformed to Cepstral domain. The coefficients are tolerant to signal processing attacks and robustness is ensured.

2.2.7 Allpass Digital Filters(APF):

Using distinct patterns of APF data is embedded in selected sub bands. This method is resistant to noise addition, random chopping, re-sampling, requantization. Robustness can be increased further.

2.3 Coded domain

Voice encoders along with their corresponding encoding rates are used to hide data for real time communications.

2.3.1 In-Encoder Techniques:

Data is hidden into speech and other audio signals using sub band amplitude modulation. The LSB technique embeds data in the LSB of the Fourier transform in the prediction residual of the host audio signal. Data embedding in the LPC vocoder was proposed. The hidden data can be decoded by linear prediction analysis of the received signal. This method offers a reliable hiding rate.

2.3.2 Post-Encoder Techniques

Data is embedded in the bitstream of an ACELP codec. In this method data is hidden along with the analysis-by-synthesis codebook search. A lossless steganographic technique was proposed. One bit is embedded in 8-bits sample with zero absolute amplitude. A semi-lossless method was introduced to increase the hiding capacity.

3. Voice Recognition in Wireless Environments

Recent developments in voice recognition technology and the use of wireless networks and mobile communication has made voice recognition to become a common feature in mobile networks. Voice recognition in mobile devices is based on the location where the recognition takes place. It can happen in the central server, terminal device or in a distributed environment. Voice recognition in mobile communication is more vulnerable to performance degradation due to various constraints. Speech coding can be an evidential source of error during voice recognition, particularly when the acoustic models are mismatched. In case of mobile network speech recognition the recognizer is positioned in a location remote to the user. So the speech signal should be transmitted from the user's terminal to the recognition server through a wireless link. In case of mobile terminal speech recognition, the recognition is performed in the user's device. Here, the speech signal does not travel through a wireless link. In distributed speech recognition the automatic speech recognition application processing and computation functions are distributed between the user terminal and the central server. Distributed speech recognition represents the client- server architecture. Here, one part of the speech recognition system resides on the client, ASR search is performed on the remote server. This requires low data rates, higher sampling rates are possible, Better error handling methods can be developed, In network speech recognition both ASR font end and back end are shifted from the terminal to the remote server. The network speech recognition back end should efficiently serve hundreds of clients simultaneously. In embedded speech recognition systems the entire process of speech recognition is performed on the terminal device. The main advantage of

this method is that no communication between the client and the server is required but the limited availability of system resources is a major disadvantage.

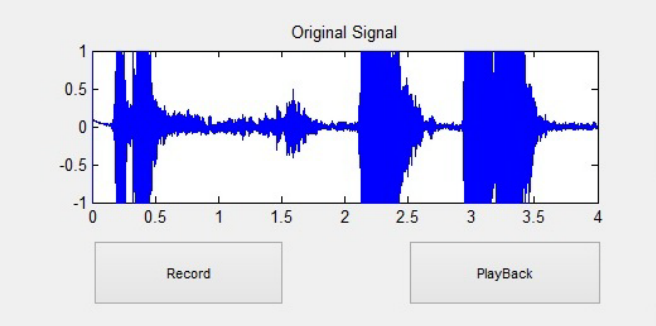


Figure 1: Voice Recognition in wireless networks

4. Noise and Additive Signal Removal

The process of voice recognition involves not only identifying the speaker voice but also the additive noise that is combined with the actual voice. When the combinational speech waveform is extended to a spectrogram the additive noise results in disturbance to the parameters of the original voice. In order to avoid this scenario the background additive noise is removed from the actual voice. The actual voice intensity is higher than that of the negligible quotient of additive noise hence the absolute values are alone considered in the estimation of the critical data such that the critical data is formed based on the absolute peaks of the actual voice. According to the equation stated below .

$$PE = Os * abs(Os) / max(abs(Os)) \quad \ldots(1)$$

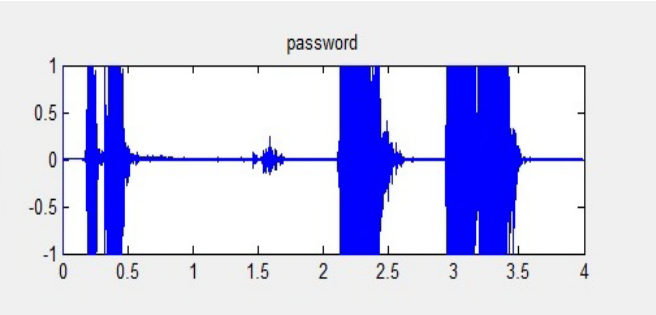


Figure 2: Noise Removal …(1)

5. Sampling the Critical Data

The sampling is the process of reducing the continuous signals into discrete multiple sets. In our study it is essential that the critical data is split into multiple samples in order to perform efficient embedding in the cover medium. The ratio of samples in cover medium to that of the critical medium is always a higher value to ensure better steganography. The sampler function is designed according to the duration of speech signal. The ideal sampler is constructed based on the instantaneous points that are located in the low power density regions. The average number of samples obtained per second is the sampling frequency. The sampling process is carried out as a function of time. This step forms the crucial step in the process because the cepstral analysis is individually performed and key are simultaneously generated as per the critical data. The below figure shows the sampling of the recognized speech

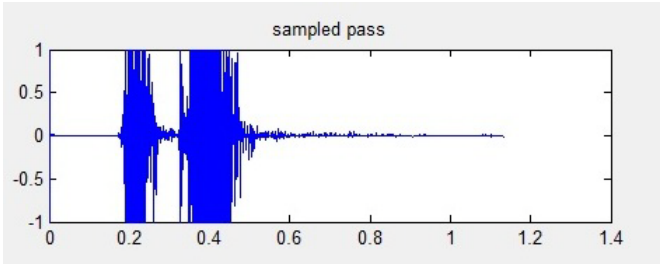


Figure 3: Sampling

6. Cepstral Analysis

Speech is composed of excitation source and vocal tract system components. These two components have to be separated from speech in order to analyse and model the excitation and system component independently and also use it in various speech processing applications. Without any prior knowledge about the source and system speech can be separated into its source and system components by cepstral analysis. Speech is the convolution of the excitation sequence and vocal tract filter characteristics.

$$s(n) = e(n) * h(n) \quad \ldots(2)$$

where s(n) is the speech sequence
e(n) is the excitation sequence
h(n) is the vocal tract filter sequence

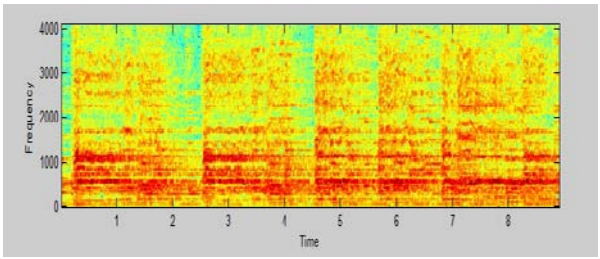


Figure 4: Spectogram of speech signal

In the frequency domain this can be represented as

$$S(\omega) = E(\omega) \cdot H(\omega) \quad \ldots(3)$$

For the speech sequence to be deconvolved into the vocal tract and excitation components in the time domain multiplication of the two components in the frequency domain has to be converted to a linear combination of the components. Cepstral analysis transforms the multiplied source and system components in the frequency domain to a linear combination of the components in the cepstral domain. Logarithm representation is used to linearly combine E(ω) and H(ω) in the frequency domain. In speech processing we generally use real cepstrum, which is obtained by applying an inverse fourier transform of the log spectrum of the signal. However, this will not enable the reconstruction of the sequence from the cepstrum. Complex cepstrum is used for reconstruction of sequence from the cepstrum. The inverse fourier transform of the logarithm of complex spectrum is used to compute complex cepstrum. The phase is preserved

in the complex cepstral sequence as the logarithm of all the spectral values are used. This reconstructs the sequence back. Homomorphic filtering is used in speech processing to separate the filter from the excitation and cepstrum is one such homomorphic transformation that performs such a separation. An appropriate windowing technique has to be used with the cepstrum to achieve deconvolution. The power cepstrum is often used as a feature detector for representing the human voice and musical signals. The mel scale first transforms the spectrum for such applications resulting in mel-frequency cepstrum. It is used to detect pitch, identify voice and so on. Cepstral analysis can be used for pitch extraction and formant tracking. The cepstral coefficients computed at a fixed frame rate represent the speech signal in many speech recognition systems.

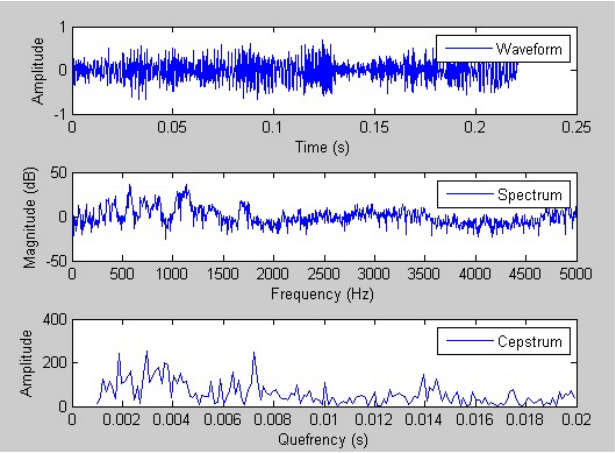


Figure 5: Cepstral analysis of speech signal

7. Power Cepstrum

The power cepstrum of the signal is the squared magnitude of the fourier transform of the logarithm of the squared magnitude of the fourier transform of a signal. It is given by

$$|F\{\log(|F\{f(t)\}|^2\}|^2$$

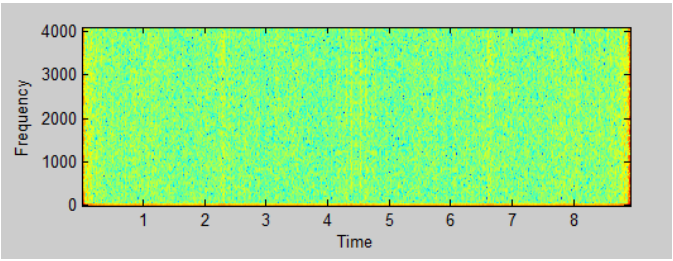


Figure 6: Power Cepstrum

The power cepstrum is obtained if the phase is discarded. For the power cepstrum to be uses in analysis of human voice it has to be transformed using the mel scale The mel frequency cepstral coefficients can be computed directly on the power spectrum. This yields a good recognition performance without a filter bank reducing the number of parameters that need to be optimized.. The cepstrum is useful in these applications because the low-frequency periodic excitation from the vocal cords and the formant filtering of the vocal tract, which convolve in the time and multiply in the frequency domain, are additive and in different regions in the quefreqency domain. The low density power level points are identified in the cover medium so as to embed the critical data in these layers. Thus making the existence of critical data highly non vulnerable.

8. MFCC Extraction

The components of speech signal which are necessary to decipher the linguistic content are to be identified and the redundant information such as background noise and emotions should be cast aside. Feature extraction is thus a capital step in any automatic speech recognition system(ASR). The shape of the vocal tract provides a precise representation of the sound that is generated. The shape of the tract demonstrates itself in the envelope of the short time power spectrum. Mel frequency cepstral coefficients accurately represent the envelope. MFCC can be deduced directly from the power spectrum of a speech signal. Using this approach consequent signal analysis steps can be fused into the cepstrum transformation avoiding interpolation and discretization errors. Mel frequency cepstral coefficients form the basis for most of the speech recognition systems and are found to be robust and efficient. The signal analysis front end of an automatic speech recognition system comprises of a series of steps. The sampled speech waveform is differentiated and brought down into a number of overlapping segments. A Hamming window is used with an intent to calculate the Fast Fourier Transform(FFT) for each frame. The power spectrum is bent out of shape according to the mel-scale. This is then divided into a number of critical bands by a filter bank which consists of overlapping triangular filters. The raw MFCC is obtained by applying Discrete Cosine Transform(DCT) to the logarithm of the filter bank outputs. The highest cepstral coefficients are excluded. The variance of each cepstral coefficient is normalized and the mean is subtracted. Both the traditional and the integrated approach can be used for the computation of cepstral coefficients from the speech spectrum. In the traditional approach the mel warped spectrum can be computed by interpolation from the original discrete frequency power spectrum. In this case the triangular filters have the same size. Another way is to place the filters non-uniformly at the unwarped spectrum thus incorporating mel-frequency scaling. Nonetheless discretization errors may occur. It is also not clear as to how many filters are required and which filter shape fits the best. The logarithm of the output is cosine transformed to obtain the coefficients. MFCC can also be computed directly on the power spectrum thus avoiding problems encountered in the standard approach

8.1 Steps involved in MFCC extraction

Step 1: The passing of signal through a filter which emphasizes higher frequencies is produced. The energy of signal at high frequencies is increased.

$$Y[n]=X[n] - 0.95\ X[n] \quad \ldots(4)$$

Step 2: The speech samples obtained from ADC is segmented into a small frame.

Step 3: Hamming window is used to integrate all the closest frequency lines

N= Number of samples in a frame
Y[n]= Output signal
X(n)= Input signal
W(n)= Hamming window

$$Y(n)=X(n)\times W(n) \quad ...(5)$$
$$W(n)= 0.54-0.46 (\cos 2\pi/N-1)\ 0\leq n \leq N-1 \quad ...(6)$$

Step 4: Fast Fourier Transform(FFT) is to convert each frame of N samples from time domain into frequency domain

$$Y(\omega)= FFT [h(t) * x(t)]= H(\omega)*X(\omega) \quad ...(7)$$

X(ω), H(ω) and Y(ω) are the Fourier transforms of x(t), h(t) and y(t) respectively.

Step 5: Mel filter bank processing is done to narrow down the frequencies range in FFT spectrum. The equation below is used to compute the mel for the given frequency f in hz.

$$F(Mel)= [2595*\log 10[1+f]\ 700] \quad ...(8)$$

Step 6: Discrete Cosine Transform(DCT) converts the log mel spectrum into time domain. This results in Mel Frequency Cepstral Coefficients.

Step 7: Features related to change in cepstral features over time should be added. The energy in a frame for a signal x in a window from time sample t1 to time sample t2 is given by the equation,

$$Energy= \Sigma X^2 [t] \quad ...(9)$$
$$d(t)= ½* [c(t+1) – c(t-1)] \quad ...(10)$$

9. MFCC Key Generation

The coefficients that are segregated in the previous step are use to form the key. This secret key plays a vital role in the steganography as it is the only connecting medium between the sender and receiver algorithms. The mel frequency cepstral coefficients are feature values that contains the magnitude and the direction when plotted over the two dimensional graphs. The key construction involves initial displacement that locates the first occurrence of the critical data in the cover medium and also the mfcc. The cumulative value of these two parameters is used in locating the next iteration of critical data in the cover medium. The number of coefficients considered at that point is also kept as a dynamic value considered according to the time duration of the critical data sample. The equation for the key is given below :

$$K(n) = [d * N \{ M(c_1) + M(c_2) + ...M(c_n)\}] – Tc \quad ...(11)$$

where,

K(n) = key for the n^{th} sample
d = initial displacement
N = Number of Mfscc coefficients
Tc = constant to time limit.

10. Critical Data Embedding in Cover Medium

The key generated is used to identify the points where critical data is embedded over the cover medium as shown in the figure below. The location of initial critical data is by then auto generated based on the sampling results. The number of key elements generated are equal to the number of samples created in the process of sampling. The time limit constant is essential when the value of the coefficients exceeds its threshold there by pushing the critical data position beyond the cover medium. This combined medium is now transmitted from the base station to the receiver end. Any foreign agent who is able to acquire the combined wave will still not be able to listen to the actual voice as the voice is hidden in the low density power level. Thus the ratio of cover samples to speech data is maintained high keeping the overall time and space complexity of the mobile environment.

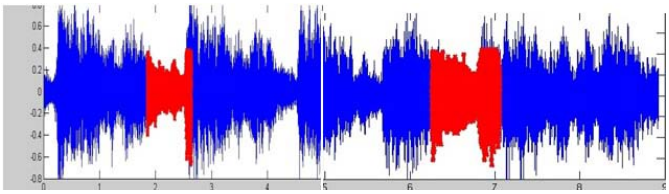
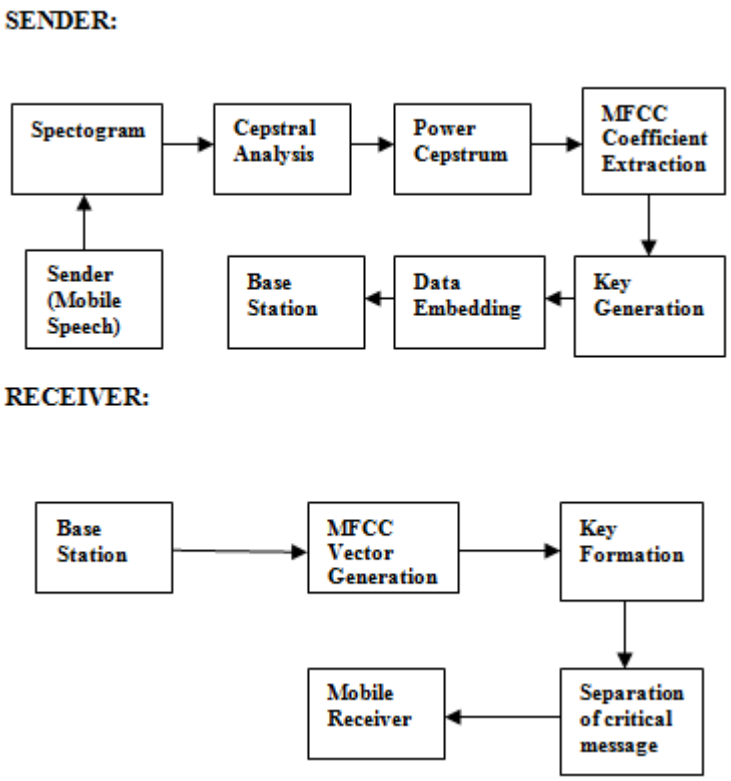


Figure 7: Critical data in cover medium

The Basic threats that are possible to happen in the process of steganography are handled in a proper way by prompt anticipation of the algorithmic steps. The threats like eavesdropping between the medium is not expected to happen since the voice data is kept in the low power density region. The data loss or compression over the cover medium will not have greater impact over the voice medium due to the low existence of the voice data. The flow of algorithm is designed in such a way that the sender medium will involve speech analysis process and the other end uses the key to retrieve the actual data. Thus making the algorithm a light weight process that fits the mobile operating system.

11. Block Diagram



12. Algorithm

Sender

1. Define speech data (critical), cover medium.
2. Analyse the mobile speech data in spectrogram
3. Identify the key changes in the level
4. Extend the cover medium to cepstral domain
5. Perform Power cepstrum analysis on speech data.
6. Identify the low power density points.
7. Extract the MFCC from speech wave.
8. Auto generate key based on mfcc.
9. Embed the critical data over cover medium based on key
10. Transmit the combined signal.

Receiver

1. Retrieve the combined speech form.
2. Extend the wave to cepstral domain.
3. Identify the MFCC from the key.
4. Analyse the cover medium to isolate the critical data using key.
5. Form the critical data synchronously into speech form.

13. Conclusion

In the growing era where in usage of mobile communication is growing rapidly fast, the privacy and security seem to be the major concern. The interoperable mobile environment enables us to bring about an algorithm based on key architecture that can communicate between the sender and receiver.

The execution of this algorithm secures the mobile communication to greater extent as it can tolerate the basic threats and as well as the complex steg-analysis threats. The speech analysis is an efficient technique to bring about the security in terms of cost and energy. The future work would focus on dynamic auto generation of the key when the speech data is transmitted in focus to solve the space complexity issues.

References

[1] Sirko Molau, Michael Pitz, Ralf Schlüter, Hermann Ney: "Computation of Mel-Frequency Cepstral Coefficients on the Power Spectrum" , IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake city, UT. Vol 1, pp 73-76.
[2] Jayaram P, Ranganatha H R, Anupama H S: " Information hiding using Audio Steganography- A Survey", The International Journal od Multimedia and its applications(IJMA), Vol.3, No.3, August 2011.
[3] Fatiha Djebbar, Beghdad Ayad, Karim Abed Meraim, Habib Hamam: "Comparitive Study of Digital Audio Steganography Techniques", EURASIP Journal on Audio, Speech and Music processing, October 2012.
[4] Aaron Bere: "Toward Assigning the Impact of Mobile Security Issues in Pedagogical Delivery: A Mobile Learning Case Study", Science and Information Conference, London, UK, 2013.

[5] Honggang Wang, Shaoen Wu, Min Chan, Wei Wang : "Security Protection Between Users and the Mobile Media Cloud", IEEE Communications Magazine, March 2014.
[6] Marc Lacoste, Aurelien Wailly, Aymeric Tabourin, Jean-Philippe Wary, Loic Habermacher, Xavier Le Guillou:"Flying over Mobile Clouds with Security Planes: Select your class of SLA for End-to-End Security", IEEE/ACM 6th International Conference on Utility and Cloud Computing, 2013.
[7] Zhibin Zhou and Dijian Huang: "Efficient and Secure Data Storage Operations for Mobile Cloud Computing", 8th International Conference on Network and Service Management (CNSM 2012), 2012.
[8] Claudio A. Ardagna, Mauro Conti, Mario Leone, Julinda Stefa: "An Anonymous End-to-End Communication Protocol for Mobile Cloud Environments" IEEE Transactions on Services Computing, 2013.
[9] Dijiang Huang, Xinwen Zhang, Myong Kang, Jim Luo: "MobiCloud-Building Secure Cloud Framework for Mobile Computing and Communication" Fifth International Symposium on Service Oriented System Engineering, 2010.
[10] Lindasalwa Muda, Mumtaj Begam, I. Elamvazhuthi: "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient(FCC) and Dynamic Time Warping (DTW) Techniques", Journal of Comuting, Volume 2, Issue 3, March 2010.