Implementation and Performance Analysis of Reliable and Secure End to End Voice Encryption over Public Mobile Network Based on Frequency Domain Using Dual Processor in FPGA Platform

Yohan Suryanto¹, Kalamullah Ramli²

Department Electrical Engineering Faculty of Engineering, University of Indonesia
Depok 16424, INDONESIA
¹yohan.suryanto@ui.ac.id, ²k.ramli@ee.ui.ac.id

Abstract—This research aims in designing a method for implementing end-to-end voice encryption over mobile network such as GSM that independent to phone devices, operating system and network devices. The proposing method is an encryption system in domain frequency to keep the encryption signal remain within the range of sound frequencies that can be transmitted through GSM/2G/3G. This method differ compare to the method used in the previous studies, in which they use the time-based encryption method to get a low rate encrypted data so the results can be communicated via GSM network. We simulated the method using Matlab. The simulation results showed that the voice encryption on the frequency domain characteristics is very different from the original signal and can be passed completely through the GSM network. This method was implemented in the Xilinx FPGA using microblaze dual core processor on the Atlys board. The results of the signal reconstruction using FFT/IFFT module in FPGA Xilinx varied in a relatively small deviation, maximum -10.42 dB, compared to the original signal. Performance analysis of the prototype showed the sender speech can be reconstructed real time in the mobile handset of the receiver side, even in the unsynchronized condition. However, in a further prototype development, the precision level of the FFT/IFFT module needs to be improved, meanwhile the synchronization module and echo canceller need to be added.

Keywords—Voice encryption, End-to-end mobile network, GSM/2G/3G, Dual core microblaze, Xilinx FPGA, FFT-IFFT, Domain frequency.

I. INTRODUCTION

GSM/2G/3G cellular is a communications network that is widely available and easy to use, thus becoming an attractive alternative for use as a communication path of high-value messages. Unfortunately, GSM/2G/3G cellular networks do not offer the end-to-end security, resulting in the voice being transmitted through the GSM network can be tapped by someone who intercepted the communication lines. To that required end-to-end encryption of voice over cellular networks.

Framework for end-to-end encryption of voice over GSM networks can be classified into five groups:

a. Voice is converted into digital form, which is then compressed in a low bit rate format, then the encrypted data sent through a special modem [1][2][8].

b. Voice is converted into digital form, which is then compressed in a low bit rate format, and then the encrypt data sent over the circuit switched service [5][6].

c. Voice is converted into digital form, which is then compressed into a low bit rate format, and then the encrypted data sent through the packet switch data service such as GPRS [1].

d. Encryption in the time domain by choosing an encryption algorithm that results as a normal conversation [3][4].


Challenges in encryption voice communication over mobile networks is how to create a chipper code pass the GSM/2G/3G network bandwidth limitations, meanwhile many kinds of devices being used, the number of service providers involved, and the number of operating systems that are used by mobile phone or smart phone should be considered. In terms of bandwidth limitations that can be passed, the GSM network can deliver only a range of voice frequencies less than 3400 Hz or data rate of less than 6.5 kbps for half rate mode [2].

II. SUPPORTING THEORY

A. Encryption

Encryption is a method to disguise or scramble the original signal into a difference signal, which is referred to as the encrypted code, which can not be read by others who do not know the mechanism and the decryption key. In terms of the method of disguise original code, encryption can be divided into two main groups, namely:

- Cryptography: Cryptography is the technique of randomization original code into a form that can not be understood by others. Randomization techniques can be as simple as a character mapping or as complicated as AES and application of dummy files. Judging from the use of a key to encrypt and decrypt, cryptographic techniques are divided into two main groups that are symmetric key and asymmetric key. In symmetric key, decryption and encryption using the same key. While in the asymmetric key, the decryption use a key pairs used by encryption.

- Steganography: An undercover techniques to hide original code into a specific media. With the
steganography techniques, a person didn’t aware that the commonplace media is containing certain information. Example of the steganography techniques is hiding the text message into the LSB bit of image pixel.

Cryptographic methods that will be used in this research is the code mapping using bitwise xor or frequency bean substitution, linked to the key id using the sliding window technique. Substitution or bitwise xor of the frequency bean according to the id password can be depicted in Figure 2.1 below.

![Figure 2.1 Frequency bean substitutions using sliding window in encryption](image)

Bitwise xor method or substitution code is a simple encryption technique based on a map that shifted towards id password, is used to indicate that the encryption in the frequency domain gives satisfactory level of results in disguising original voice. Decryption process is done through the reverse process of the encryption process as shown in Figure 2.2 below.

![Figure 2.2 Frequency bean substitutions using sliding window in encryption](image)

### B. Time and Frequency Domain Transformation

DFT is one of the Fourier transform family that operates on a limited number of data (N) of the digitized signal samples. When the number of samples (N) is repeated and placed end to end, they appear as a periodic signal to the Fourier transform. The correlation between samples in the frequency domain $X_k$ and samples in the time domain $x_n$ can be seen in Figure 2.3 below.

![Figure 2.3 DFT and IDFT relationship](image)

Transformation from the time domain samples $x_n$ into frequency domain $X_k$ can be done using equation 1:

$$X_k = \sum_{n=0}^{N-1} x_n(\cos\left(\frac{2\pi nk}{N}\right) - j\sin(2\pi nk/N))$$ (1)

Transformation from the frequency domain $X_k$ into the time domain $x_n$ known as the Inverse DFT (IDFT), can be written in the following equation 2:

$$x_n = \sum_{k=0}^{N-1} X_k(\cos\left(\frac{2\pi nk}{N}\right) + j\sin(2\pi nk/N))$$ (2)

FFT is an algorithm to accelerate DFT calculations by reducing the number of multiplication and summation involved in the DFT. It was popularized by J. W. Cooley and J. W. Tukey in 1960 and is based on reinvention of the idea of Runge (1903), Danielson and Lanczos (1942). FFT exploit symmetry properties and repetition of the sinusoidal function is called as a twiddle factor. If the DFT formula (1) and IDFT (2) rearranged, then the result is the FFT/IFFT operation that only requires a number of complex multiplication as $\frac{N}{2}\log_2(N)$. FFT will speed up the counting process, because the original DFT formula requires as much as $N^2$ complex multiplications.

### III. DESIGNING PROTOTYPE END TO END VOICE OVER GSM/2G/3G NETWORK

The system is designed to be able to cope with the condition of voice communication over the cellular network and should be used in all mobile devices, with the following criteria:

a. Cellular networks only pass voice signals in the frequency range of less than 3400 kHz.
b. The maximum rate if the encrypted digital data to be transmitted over the GSM network is 6.5 kbps.
c. The system is designed to be reliable and allows for secure encryption.
d. End to end delay should no more than 150 ms in accordance with the standards of voice communication QoS [21].
e. Encryption or decryption total is less than 30 ms

In order to make the designed system is independent to the type of phone and the cellular network are used, we proposed encryption before the voice goes into the phone and decryption after it exit from mobile phone receiver as depicted in Figure 3.1 below:

![Figure 3.1 End to End Encryption System voice](image)
Advantages of the offered system are:
- Can be implemented for various types of mobile phones
- Independent to the operating system
- Independent to the cellular networks

Nevertheless, the considered system is consist of 3 separate digital networks, namely the decryption system, cellular network system, and the encryption system, each bounded by a separate analogue system. System in this state has some of the challenges that must be solved in order to be implemented, namely:
- Time synchronisation between encryption and decryption.
- Normalization of data between encryption and decryption.
- Limitations of GSM/2G/3G network bandwidth for voice communication as a bottleneck.
- The time of encryption and decryption process
- Jitter and end-to-end delay.

A. Design of Encryption in Frequency Domain Using MatLab

One of the challenges that must be solved in voice encryption that can be transmitted by cellular network is chipper code must be maintained in order to always be in the voice frequency range. This is must be done because the mobile network only forwards the signal in that range. The proposed method is encryption or decryption in the frequency domain, using the techniques:

☑ FFT - encryption - IFFT
☑ FFT – decryption - IFFT

In the design of the encryption and decryption systems using MatLab, the encryption method chosen is the frequency beam mapping technique. It is chosen because this technique is simple and sufficient to show that the encryption voice in the frequency domain can be effectively use in order to maintain the encryption results are within the range of sound frequencies that can be transmitted by the GSM network. Design scenarios to be conducted are:

a. The process of encryption and decryption for the entire duration of the speech signal in a given time.
b. The process of encryption and decryption will be simulated per frame.
c. Decryption process is not synchronized to the encryption process as many as 32 arrays.

B. Designing Prototype Using Xilinx FPGA

After going through the stages of designing an encryption method based on frequency domain using MatLab, then the next step is a prototype implementation on Xilinx FPGA on the Atlys board. On the encryption side, the user’s voice will enter into prototype through the microphone, and then go into the ADC (Analogue to Digital Converter). This data will then be entered into the FFT module, the domain changed from the time domain into the frequency domain. Signal in the frequency domain will then be encrypted using frequency beam mapping technique. The encrypted output signal will be processed by IFFT module to restore process into the time domain. Signal in the time domain will then enter the module DAC (Digital to Analogue Converter) which will subsequently enter the smart phone microphone port to be forwarded over the cellular network to the receiving party. Encryption system block diagram can be seen in Figure 3.2 below.

In decryption side, the received voice from a cell phone will enter through line-in then entered into the ADC (Analogue to Digital Converter). This data will then be entered into the FFT module, the domain changed from the time domain into the frequency domain. Signal in the frequency domain will then be decrypted using frequency beam mapping technique. Decrypted output signal will be processed in IFFT module to restore the domain into the time domain. Signal in the time domain will then enter the module DAC (Digital to Analogue Converter) which in turn will enter the smart phone headset port and can be heard by the user. Decryption system block diagram can be seen in Figure 3.3 below.

Noting the need for end-to-end delay for voice communication is less than 150 ms and the end to end delay on the existing GSM network is approximately 90 milliseconds, and then the total number of encryption/decryption in accordance block diagram in Figure 3.1 and Figure 3.2 are expected to total less than 60 millisecond. The maximum time taken at the encryption side is 30 ms and decryption side is also 30 ms. The total time of this process included an audio frame buffer with 8 kHz sampling rate.

Design system are selected either using the FFT-IFFT in microprocessor or using dedicated FPGA FFT/IFFT module whichever the processing time had a total response less than 5 ms. The first option is preferred, as it will simplify the process of encryption and decryption including easier to customize the...
data format used. We can use dedicated FFT/IFFT module, nevertheless module FFT/IFFT still controlled by the microblaze processor and the encryption and decryption process is still being done in the processor.

IV. PERFORMANCE ANALYSIS AND PROTOTYPE TESTING

A. Analysis Simulation in Matlab

The simulation process of FFT/IFFT in overall signal is intended to look for encryption patterns in the frequency domain in non-real time mode. Incoming voice samples will be processed FFT to transform the signal from the time domain into the frequency domain. Voice samples "ABC" with the signal shape as illustrated in Figure 4.1 below is used for simulation.

![Figure 4.1 Original sample signal](image)

When the signal was encrypted in the frequency domain with frequency bean exchange method, it will produce a signal shape as shown in Figure 4.2 below.

![Figure 4.2 ABC encrypted signal samples](image)

Although the performed encryption process to the signal "ABC" in the frequency domain is considered as a simple method, because it involves only exchange position of the frequency bean, but the sound produced by the encryption process is very different from the original sample signal and can not be identified without using the appropriate decryption process.

The encrypted signal is then sent through the GSM network, using two smart phones that are connected to 2 different GSM network operators. As we expected, the encrypted signal can be received by the receiver without obstacles. Comparison between the original voice encryption the encryption result after passing the GSM network, can be seen in Figure 4.3 below.

![Figure 4.3 Comparison of original encrypted signal and after passing through the GSM network](image)

It appears from Figure 4.3, the encrypted signal is fully transmitted by the GSM network. The deviation value is very small when compared with the original encrypted signal, which is a maximum of -79.9 dB, and can be ignored.

If you see a comparison between the frequency plot of the original encrypted signal and the encrypted signal after passing through the GSM network, as shown in Figure 4.4 below, it can be concluded that the signal is identical. Deviation between the original encrypted frequency plot and the passed encrypted frequency plot through GSM is very small (maximum -79.9 dB) and can be ignored.

![Figure 4.4 Comparison of frequency plot between the original encryption and after passing through the GSM network](image)

The frequency plot of the original signal and the decrypted signal has the same pattern. While the frequency plot of encryption signal has different patterns compare to the original signal according to the pattern of frequency bean substitution performed onsite as can be seen in Figure 4.5 below.
Analysing the signal in the time domain, showed that the decryption successfully restore the shape of the original signal with deviation maximum -118.58 dB as shown in Figure 4.6 below. In addition, the performed of frequency bean exchanged, resulting in the varied difference between the original signal and the encryption signal.

In real time systems, the entire signal processing is not possible, because the whole process is only possible if the signal has been collected previously. The process of collecting the signal takes delay time associated with the end-to-end communication. In voice communications, to maintain good QoS for interactive communication, the maximum delay allowance is equal to or less than 150 ms [21]. To that, we also simulate encryption/decryption process in the frequency domain with a size of 128 array per frames.

Plot frequency of the decrypted signal per frame 128 array has the same pattern as the original signal as shown in Figure 4.7 below. Meanwhile, the plot of the frequency of the signal encryption has a different pattern than the original signal according to the frequency bean exchange pattern.

Signals sent by the encryption module and received by decryption module sometimes out of sync, this signal can be shifted a few points due to the lack of initial synchronization point. To simulate whether the signal received by decryption module is still be reconstructed in the event of unsynchronized, we simulate the decryption process shifting at 32 points. The form of the decryption signal with unsynchronized 32 arrays has the similar form to the original signal as shown in the following Figure 4.9. Received voice can also be understood by the recipient, yet feels vibration nuisance.

The vibration noise occurred because when viewed in a more detailed scale, in the time domain, there is a significant difference between the original signal and the decrypted signal with unsynchronized 32 arrays as shown in Figure 4.10 below.
B. Speed Transformation process for several scenarios

Processing speed of the prototype encryption consists of:
- 128 samples collection arrays
- The process of reading and writing to the buffer
- The process of encryption and encryption
- The process of transformation from the time domain to the frequency domain and vice versa

Read and write speed of the microblaze processor to the FIFO can be seen in the following Figure 4.11. Data being written or read for audio FIFO is real data, while complex data is written or read to the FFT/IFFT FIFO.

![Figure 4.11 Speed of the microblaze writing and reading to FIFO](image)

It appears that the process of writing and reading complex array slightly longer than the process of writing and reading a real array. For example, to write a 128-point real, it takes 31 microseconds while to write a 128-point complex takes 37 microseconds. To read a 128-point real, it takes 32 microseconds while to write a 128-point complex takes 39 microseconds. The process of writing and reading by 128 points for the microblaze processor is much faster than the process of collecting the audio frame buffer consisting of 128 points, which takes over 16 ms. This means that the use of FIFO in the design qualified to run as a real-time system.

Performance implementation of DFT using microblaze MCS with a clock of 100 MHz and using an external module in FPGA Xilinx FFT, can be seen in Figure 4.12 below.

![Figure 4.12 Comparison of DFT processing time using some scenarios](image)

It appears from the picture above, the function calls of the sine and cosines in the math.h header causes the DFT program with this method is running very slow. For the 128 point number array, DFT processing time sine function takes 7.5 seconds, much longer than the allowed maximum time required for processing single frame 16 ms.

DFT process using a lookup table in microblaze has a speed 35 times faster than the sine function. Nevertheless, the DFT program using a lookup table still does not meet the need for real-time process. For the frame with 128 points, DFT with lookup tables takes 214.4 ms, much slower than the time required for the collection of single frame 16 ms.

Processing time of the FFT with table lookup process is faster than the DFT program using lookup tables. For a frame with 128 points, FFT program performed 10.3 times faster than the DFT program using lookup tables. Nevertheless, the FFT program in the microblaze still not meet the minimum requirements for real time process because it is still longer than the maximum time allowed for processing one frame. For example, for a 128 point FFT, processing time is 20.7 ms, it is much longer than the allowed maximum time required for processing single frame 16 ms.

FFT process using external modules qualify for real time voice encryption/decryption. FFT process using external modules including to write and read from the FIFO with the number of complex points 128 array, it only takes as long as 110 microseconds, or 145.15 times faster than the maximum time required for the processing of one audio frame 128 points.
C. Testing Prototype Implementation in FPGA

On the side of the FFT module, we verified prototype by comparing the original data with the data results of the FFT and IFFT reconstruction process is performed by the FFT/IFFT external module. For data blocks containing simple data like all worth 255, all worth -1, or signal box -32 676 and 32767, prototype can reconstructed the signal almost perfect with very little deviation as shown in Figure 4.13 below.

![Figure 4.13 Deviation of the reconstruction signal using FFT/IFFT module with a simple signal](image)

When the data replaced using a dynamic voice signal, the reconstruction result of an external FFT/IFFT has a significant deviation compared with the original signal, a maximum of -10.42 dB. This can be seen in Figure 4.14 below. When the same data is processed using Matlab, the deviation between the original and reconstructed signal using FFT-IFFT process is maximum -118.58 dB or the signal can be reconstructed perfectly.

![Figure 4.14 Deviation of the reconstruction result using FFT/IFFT module against the original voice signal](image)

In the test system, this deviation generates disturbing noise when we make a conversation between two phones via a cellular network that previously connected to a prototype device.

Prototype testing was performed using 2 smart phone are connected to two different GSM operators. But because there is an echo that cause noise when the prototype is used for two-way communication, both at the mode of encryption or no encryption, then testing is done using one way half duplex communication. In order to makes the echo does not appear when the prototype is used for 2 way communication through the network with high delay such as GSM, adding echo canceller module needs to be done. However, the addition of an echo canceller modules in the encryption system is beyond the scope of the prototype defined in this research. Therefore, to represent two way testing, we test the prototype in back to back configuration.

The prototype test result for voice communications without encryption indicates that the sample signal "ABC" sent by one user can be understood by the user 2. Waveforms between the original signal and the reconstructed signal is identical, with a relatively small deviation, maximum of -6.02 dB as shown in Figure 4.15 below.

![Figure 4.15 Comparison of the original signal and the signal received by the second prototype without the encryption/decryption](image)
Figure 4.16 Comparison of the frequency plot between the original signal and the received signal without the encryption/decryption

Test results of the back to back voice communication shows that the signal sample encryption “ABC” sent by one user can be understood by the user 2. Noise that occurs in the process of causing a form of communication signal received by user 2 is somewhat different compared to the signal sent by user 1 as shown in Figure 4.17 below.

Plot of frequency noise due to differences occurring between the original signal and the received signal in the system back to back with the encryption/decryption is shown in Figure 4.18 below.

Figure 4.17 Comparison of the original signal and the signal received by the second prototype without the encryption / decryption

Test results with half duplex configuration showed that the encrypted signal can passes the GSM network and signal samples "ABC" is sent by the user 1 can be understood by the user 2. Noise that occurs in the process of causing a form of communication signal received by user 2 is somewhat different compared to the signal sent by user 1 as shown in Figure 4.19 below.

Figure 4.19 Comparison of the original signal and the signal decryption via GSM

Plot of frequency noise due to differences occurring between the original signal and the signal received by user 2 via GSM communication system using two prototype encryption/decryption with half duplex mode can be seen in Figure 4.20.
be further developed to produce a smaller model, over GSM/2G/3G networks based on frequency domain can implementing different encryption methods for each frame in the unsynchronized condition, but has a quality that is not clear yet because the module FFT/IFFT using single precision and lack of synchronization module between encryption and decryption. We hope that the prototype of the voice encryption/decryption over GSM/2G/3G networks based on frequency domain can be further developed to produce a smaller model, implementing different encryption methods for each frame in a super frame, the decrypted result clearer, and the higher accuracy level.

V. CONCLUSIONS

Based on the simulation results of the voice encryption/decryption in the frequency domain using Matlab and performance analysis of prototype of real time voice encryption using dual processors in FPGA Xilinx, concluded that:

1. The results of encryption are in the voice frequency range from 0 to 3400 Hz and can be transmitted through GSM network with a maximum deviation of -79.9 dB.
2. The level of precision of the sound signal reconstruction using Xilinx FFT module still needs to be improved to eliminate the noise generated during the process of encryption/decryption in the frequency domain. Deviation between the original voice signal and the reconstructed signal Xilinx FFT module maximum is -10.42 dB, while the deviation of the reconstruction results using Matlab is much smaller, maximum -118.58 dB.
3. The speed of the reconstruction process using FFT/IFFT Xilinx module is 110 microseconds. The total additional delay made by the prototype of voice encryption/decryption on voice communication using the network GSM/2G/3G is 18 ms, including a voice sample buffer to process as many as 128 samples for 16 ms.
4. The sound produced by prototype in the process of encryption and decryption can be understandable even in unsynchronized condition, but has a quality that is not clear yet because the module FFT/IFFT using single precision and lack of synchronization module between encryption and decryption.

REFERENCES

[12] Xilinx support team, Spartan-6 FPGA Data Sheet, DS162 (v3.0), 17 October 2011
[21] Digilent team, Atlys schematic, 6 November 2011