DOI: 10.24996/ijs.2021.SI.1.34

ISSN: 0067-2904

VoIP Speech Encryption System Using Stream Cipher with Chaotic Key Generator

Mahmood Khalel Ibrahem*, Hussein Ali Kassim
College of Information Engineering Al-Nahrain University, Baghdad, Iraq.

Abstract

Recently, with the development multimedia technologies and wireless telecommunication, Voice over Internet Protocol, becomes widely used in communication between connecting people, VoIP allows people that are connected to the local network or the Internet to make voice calls using digital connection instead of based on the analog traditional telephone network. The technologies of Internet doesn't give any security mechanism and there is no way to guarntee that the voice streams will be transmitted over Internet or network have not been intercepted in between. In this paper, VoIP is developed using stream cipher algorithm and the chaotic cryptography for key generator. It is based on the chaotic maps for generating a one-time random key used to encrypt each voice data in the RTP packet. Chaotic maps have been used successfully for encryption bulky data such as voice, image, and video, chaotic cryptography has good properties such as long periodicity, pseudo-randomness, and sensitivity to initial conditions and change in system parameters. A VoIP system was successfully implemented based on the on ITU-T G.729 for voice codec, as a multimedia encoding of Real-time Transport Protocol payload data, then, apply a proposed method to generate three-mixed logistic chaotic maps [1] and then analysis the encryption/ decryption quality measures for speech signal based this method. The experimental work demonstrates that the proposed scheme can provide confidentiality to voice data with voice over IP performance quality, minimum lost in transmitted packet, minimum average delay, and minimum jitter.

Keywords: Logistic Chaotic Map, G.729, QoS, Speech Encryption, VoIP.

Introduction:

In recent years, the large-scale in communications technology have been made people in more need to communicate. Voice over Internet Protocol (VoIP) connects people to put their voice call without any geographical limitation. A voice data is first captured from the sound device and then digitalized, encoded, compressed, and build RTP packet and then transmitted over the Internet or any other public network as a stream of packets. The interconnected router will route these sequence of the packet to correct destination depending on the

endpoint address of the packet. VoIP used the packet-switched mechanism of packet routing that different from fix connection in the circuit-switched network, the sequence packets of voice data of a session establish between two users will route to a network based on the packet-switched mechanism with a different duration time of the call. In a VoIP system, both digital (e.g., PC, PDA) and analog (e.g., telephone) devices coexist since the digital signal can be converted to analog signal by using the phone adapters. VoIP have been reducing the cost of maintenance and design and operating because it does not depend on the network properties and interoperability. Generally speaking, VoIP used Real-time Transport Protocol RTP, and it's not reserved for voice communication, as well, can be extended

* Email: mahmoodkhalel@coie-nahrain.edu.iq

to support video communications. The network that used the circuit-switched mechanism (e.g., Public Switched Telephone Network (PSTN)) only supports minimum bandwidth with 64kbps and it doesn't support video transmission, it can carry the voice call and other texts contain. Therefore, with the development multimedia technologies and new technologies support high-speed connection with bandwidth more than 1Mbps is necessary to use VoIP system with high quality, bandwidth with more than 1Mpbs can utilize transmission capability for video, voice, and combination video and voice data. Nowadays, the evaluation of broadband networks and enterprise LANs make VoIP grow even faster and gain popularity. The major problem to people that have been used VoIP is security threat when used a public network such as Internet, the conversation across a public network may be intercepted by malicious attackers or someone eavesdropped to it. To make a public network secure environment and protect conversation against eavesdropping, cryptographic techniques should be used to provide confidentiality and security enhancement[1].

In practice, many cryptographic methods are existing to improve the security services on VoIP. For Example, Secure Real-time Transport Protocol (SRTP) that attempts to provide the protection to speech through traditional encryption schemes, public-key algorithms such as RSA to provide authentication for RTP traffic and management distributed a session key, and AES to provide confidentiality to the RTP payload data packet, but still SRTP suffer from some weakness, the weakness due to SRTP depend on block cipher AES algorithm for encrypting the packet data contents" plaintext data"[2] AES algorithm and any other block cipher cryptosystems require block size of input data should be multiple of 128 bits. If the size of the data is not fitting to multiple of 128 bits the algorithms will increase "padding" to make the size of data multiple of 128 bits, and then encrypts data (original plus padding data). In this situation, there are several threats in using SRTP protocol. The attacker may use the encrypted padding part of message and apply the brute-force attack to conclude the encryption Key, and in general traditional block cipher cryptosystem such as AES, 3DES, and DES are not efficient scheme for multimedia data, speech and video, due to the large data size, high correlation and redundancy among data. Therefore, researchers have been begun to develop a novel technique aims to make the original speech corrupt and reduce the residual intelligibility of speech data, and the output of this technique in general is noisy data.

Chaos-based cryptography has been effectively used for encrypting large-scale data such as image, audio, and video data, because Chaos-based have a good characteristic like generating a key with long periodicity, pseudo-randomness, and sensitivity to change in initial conditions and system parameters.
In this paper, One-Dimensional logistic chaotic maps has been used to generate a one-time key that successfully used to encrypt payload data of RTP protocol. Three-mixed One-Dimensional maps are combined in a novel algorithmic operation for key generation with a block size of 32 random bits is produced at each iteration. The binary floating-point 64-bit format is used from the IEEE 754-2008 standard for floating point arithmetic.

RELATED WORK

Several chaotic maps have been proposed for speech encryption schemes. R. Gnanajeyaraman [3] present a new speech encryption system based on generating a look-up table using eight dimensions chaotic cat map, this high dimension called 8D chaos-based cat map and then used for encryption, decryption system with high dimension algorithms enhance the security level of the algorithm and the key space, and that make the audio sample distributed uniformity. Sheu [4] presented a Two-channel chaos-based Speech Encryption using fractional Lorenz system for speech communication (henceforth called TCSE). The TCSE can achieve high key sensitivity, large key space, and increase ability to resist chosen plaintext\cipher-text attack. Prabu [5] proposed a novel speech encryption techniques, it depends on a one-dimensional logistic chaotic map to generate random sequence of values to encrypt the sequence of character, and speech data stream that transferred between two individual phones have been encrypted by the generated these sequence random values. The authors suggested using a logistic map to mask every character by one value of the sequence that obtained by the logistic map. The proposed method was successfully implemented in the real-time application on mobile phones. Ashtiyani et al [6], presented speech encryption based on symmetric cryptography via the Cat Map. The speech signal was encrypted based on a combination of two operations of scrambling speech samples and then confusion. The Cat map was used for scrambling the speech samples, chaos cryptosystem also used in improvement of the simple form of the Advance Encryption Standard (AES), the improvement occur on the contain of original S-box, the cat map used to generate random 16 bits that substituted into the AES S-box, rather than based on fixed values. Su et al [7], suggest encryption scheme for G.729 standard speech based on two selective encryption methods. The research used chaos cryptosystems to minimize the computational complexity and provide full encryption to G.729 speech data. The algorithm started by portioning the G.729 speech data into two parts according to the sensitivity bits, the sensitivity part was encrypted by using a strong cipher used a companion the logistic and Cat maps, and the remaining part that has less sensitive bits was encrypted using a lightweight cipher based on one of the logistic or cat maps.

Chaotic System

Background Chaos system have been used to implement nonlinear dynamical systems that have mathematical models. Chaotic maps classify into two categories, according to the time range that described by equation of system, continuous systems that have differential equations, or discrete systems that have difference equations. Logistic map and Henan map are example of the discrete systems. The Lorenz system and Rossler system are example of the continuous systems. In the 1960's, the chaotic systems behavior were first studied and show have numerous random good properties. The iterative random numbers that generated from one of chaotic maps have completely random, and sensitiveness to change in the initial conditions, this means that the change in the initial condition (initial values of a chaotic map) generating different trajectory, and the similar initial conditions (initial values of a chaotic map) can generate the same trajectory. The property of “sensitivity to change in the initial” makes chaotic systems used effectively in the diffusion function for cryptography application [8].

The logistic map

Frequently used in the application that based on chaos theory, as well as in chaotic cryptosystems, the simple mathematical equation of the logistic map is described as: [8]

$$F(x)=\beta\, x\_(n\ \ )\ (1\text{-}x\_n) \qquad (1)$$

Where xn is a number between zero and one, x0 represents the initial population, and β is a positive number between zero and four. To get chaotic behavior the value of β should be between [3.57, 4.0]. Logistic map has been successfully used for generating pseudo-random numbers. To avoid a non-chaotic behavior of logistic chaotic map, adjust the value of β necessary be near 4.0, that leads to generate a highly random behavior. The logistic map is used under the iterative form:

$$x\_(n+1)=\beta\, x\_(n\ \ )\ (1\text{-}x\_(n\ \ )\ )\quad \forall\geq 0 \qquad (2)$$

Where the 〖X 〗 _0 is a real number between zero and one All the computed elements Xn are also real numbers between zero and one.

Generation a One Time key using chaotic logistic map

The algorithm used for generating pseudo random numbers in this paper based on the chaotic logistic map given by Eq.1. In this paper, to get a highly chaotic behavior β assigned a constant value equal to 3.9999. Indeed, the chaotic behavior of the logistic function measures using the Lyapunov exponent [9] and when make the value of β of the logistic map equal to (3.9999) and the Lyapunov exponent equal to 0.69 that is very close to its maximum value which is 0.59. The equation of logistic map is written with fixedβ.

$$x\_(n+1)=3.9999\ x\_(n\ \ )\ (1\text{-}x\_(n\ \ )\ )\quad \forall\geq 0 \qquad (3)$$

$$y\_(n+1)=3.9999\ y\_(n\ \ )\ (1\text{-}y\_(n\ \ )\ )\quad \forall\geq 0 \qquad (4)$$

$$z\_(n+1)=3.9999\ z\_(n\ \ )\ (1\text{-}z\_(n\ \ )\ )\quad \forall\geq 0 \qquad (5)$$

A binary64 floating-point as shown in Fig. 1, used to represent each computed values of (Xn; Yn and Z_n).In each iteration, apply an xor operations to the last part 32 bits of binary64 floating-point that called mantissa1 of the three output valuesXn; Yn and Zn, at each iteration, the proposed algorithm allows product producing 32 random sequences of bits that increasing the throughput of key generation, Figure-2 shows the proposed algorithm to generate one time key.

The Real-Time Transport Protocol (RTP)

RTP is an Internet protocol that is designed essentially for transmitting a real-time multimedia packet, and it is standard for a real-time multimedia application that require timely delivery and synchronization traffic of packets stream to compensate the lost in the packets that may cause de-sequencing and delay variations(jitter)[10].

The RTP packet format:

The RTP packet consists of two parts, packet header with a smallest size of 12 bytes, followed by payload data with different size. The RTP packet as follows:

Version (V): (2 bits) shows the version of the RTP protocol. We used version 2.

Padding (P): (1-bit) padding indicator of the payload data size that result from the block encryption. In our case, padding is zero value which meaning no padding.

Extension(X): (1 bit) used to indicate existing of an “Extension header “between RTP packet standard header and RTP packet payload data.

CSRC count (CC): indicates the count of Contributing source IDs.

Marker (M): (1-bit), used to indicate the current data that used in specific application and it defined by profile.

Payload type (PT): (7 bits) is important filed in RTP packet that indicate the format of the payload data, in the proposed algorithm we used the G.729 standard to represent speech data.

Sequence number: (16 bits) shows the sequence number of the current packet and incremented after each packet sent. It is very important to measure the VoIP quality to determine the packet lost.

Timestamp: (32 bits) contain the time of the RTP packet data that have been sent and used for synchronization between the sender and receiver and determine the delay variations.

Synchronization source identifier SSRC: (32 bits) random number used to solve stream conflict that used to distinguish between RTP streams, when destination received RTP packet from different source on the same session.

CSRC: (32 bits) allows destination to contributing sources of a stream packets when a stream has been generated by different sources.

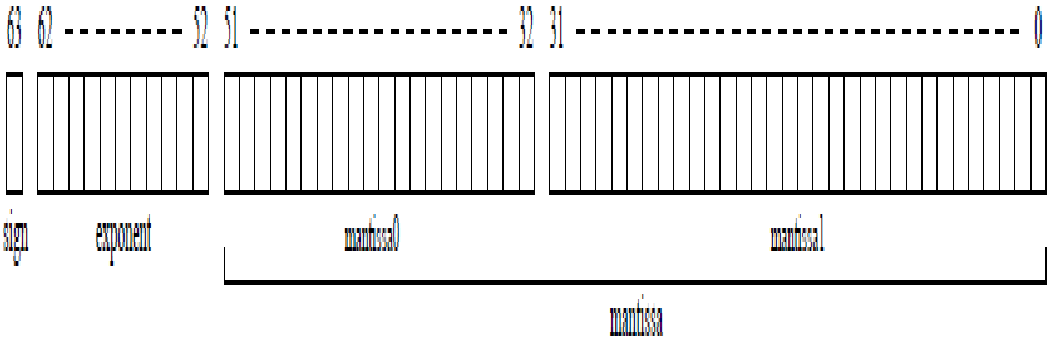

Figure 1-Floating-point representation in binary64 format [8].

Proposed Integrated VoIP protocol

The proposed protocol provided secure environment to start speech communication between two individual users. It uses public key scheme for user authentication to server, secure key exchange, and symmetric stream cipher encryption scheme for speech data that encoded and compressed by ITU G.729 standard. The protocol comprises of four stages; User Connection Stage, User Authentication Stage, Distributed chaotic initial parameters Stage and Communication Stage. Stages are described as follows:

Stage I: User Connection Stage: User send information to server includes (User ID and network information) Server accept request and validate the information with server Database, then send public key (KUs) to user .Stage I consist of steps of (1_3) of protocol.

Stage II: User Authentication Stage: In this stage, User try to establish secure communication with server based on Real Time Control Transport Protocol (RTCP) for network management and packet format and used the public key encryption algorithms (RSA) to encrypt encryption data packet. After user have received server public key, generate private, share public key, and send public key to server encrypted with server public key. Server receives public key and update online user list to all others user, after this stage user will be ready to start secure communication with any online user.

Stage III: Distribution of chaotic initial parameters Stage: User A request start voice communication with destination user B (B must be online). Server generate chaotic parameters and implements message encryption by user A public key.

Stage IV: Communication Stage: User B build lockup table, as shown in Table 2, depend on the chaotic key generated as separate process to store one time key generated by chaotic logistic map and each packet will be encrypted with one time key of chaotic logistic map and a reception used the packet index to select the correct key from the lockup table that packet encrypt for successive packet decryption.

Table 2-lockup table

Seq.	**From/To**	**Activity**
Stage I: User Connection Stage		
1	U → S	Send User Information (ID*u*, NI*u*).
2	S	Check User validate using sever database, if not terminate connection (M1).
3	S → U	Send PUS to User
Stage II: User Authentication Stage		
4	U	Generate Private, Public Keys of RSA algorithms.(PR*u*,PU*u*)
5	U → S →	Send RSA(PUs,N||ID*u*||PU*u*) to server
6	S all users	Server update online user list and store PU*u* in DBs
Stage III: Distributed chaotic initial parameters		
7	U*a* → S	User *a* request start session with user B. RSA(PR*a*, N1||ID*a*||ID*b*||T1)
8	S	Generate logistic chaotic values (ICV)
9	S → U*a*	Send RSA(PU*a*,N1||ICV||ID*b*||T2||RSA(PU*b*,N1||ICV||ID*a*||T2)
10	U*a*	Decrypt message using PRA
11	U*a* → U*b*	User a send RSA(PU*b*,N1||ICV||ID*a*||T2) message part to user b
12	U*b* → U*a*	Send RSA(PU*a*,N1||OK||T3) and Start Session
Stage IV: Communication Stage		
13	a,b	Build lockup table ,IUT(I) Chos(Xi)
15	A	Encoding and **Encryption voice Data**,RTP payload =RC4(Chos(Xi),VD)
16	a → b	Send RTP packet over UDP protocol
17	B	Received Packet and decrypt RTP payload,VD =RC4(IUT(I), Payload).

Encryption Quality and VoIP Performance

Encryption Quality

In this section, we demonstrate the quality of proposed encryption scheme that based on the stream cipher RC4 algorithms to encryption voice data and three-mixed logistic map to generate sequence of random keys. The results have been implemented using Visual C# 2012 on the laptop Windows 7, Intel® Core™ i3 with speed 2.3 GHz, and Ram 8GB. For encryption experimentations, we used 5-wave sound files with different size. Fig. 3 shows a waveform of an original, encryption and decryption speech signal. From the Fig. 3, as illustrated in the Fig. 4, the encrypted speech has been distributed uniformly and unintelligible. It is different from the waveform of original speech. In addition, the waveform of the decryption speech is identical to the original speech waveform.

Mean Square Error (MSE): MSE is a frequently calculate the difference between two samples speech and it indicates the measurement of the error with estimate to the center of the mean of the value of speech samples, MSE is describe in equation (6). At most, it has already been used to estimate the error that has occurred due to the encryption and decryption process in the recovered speech data.

$$MSR = \frac{1}{N}\sum_{i=0}^{N-1}(i\hat{\mathrm{O}} - \mathrm{O}\, i)^2 \qquad (6)$$

Where O represent the samples of audio file, Ô represent the samples of encrypted or decrypted audio samples, and N represent the length of audio samples. When MSE equal to zero or near to zero it indicates that decryption process have a perfect recovery operation to return the original audio samples, otherwise, when MSE of is greater value that indicate the distortion between two sequence of samples is hugely. .Table 1 shows MSE measures of the tested speech files

Peak Signal to Noise Ratio (PSNR): PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that generated from the encryption/decryption process. It is easily calculated by MSE, PSNR is describe in equation (7).

$$PSNR_{dB} = 10\log\left(\frac{MAX^{2}}{MSE}\right) \qquad (7)$$

MSE refers to Mean Square Error and MAX to maximum possible value of audio sample, which is equal to 65,535. A lower PSNR indicates that, decrees residual intelligibility of speech signal, and a higher PSNR indicates to increase recovered speech signal. In the specific case, when the MSE is zero that mean the original and recovered speech signal is equal, In this case the PSNR becomes infinite. Table-2 shows PSNR measures of the tested speech files.

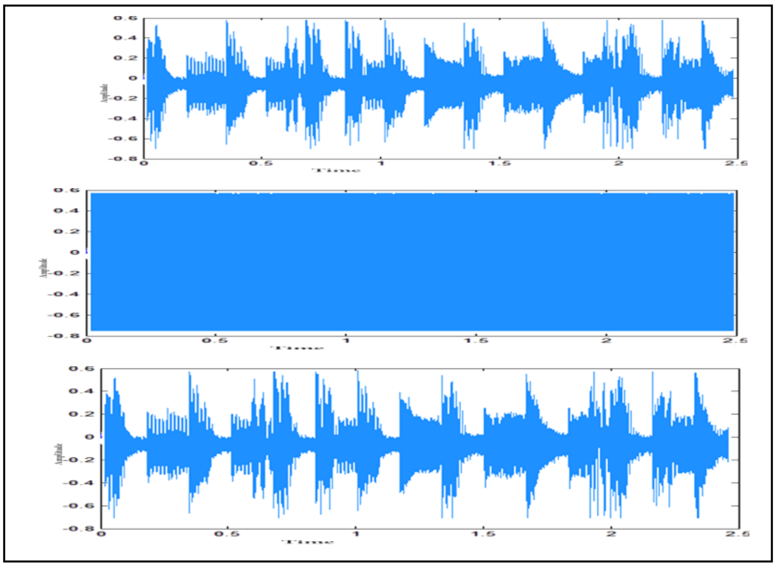

Figure 3-Waveform of (a) original, (b) encrypted, (c) decrypted speech.

Table 2- Encryption Quality (MSE and PSNR) Measures

File	MSE of Encrypted speech	MSE of Decrypted speech	PSNR of Encrypted speech	PSNR of Decrypted speech
Example1.wav	265823847.9	0	12.0835266723259	Infinity
Example2.wav	191737616.8	0	13.5023928230192	Infinity
Example3.wav	170480586.7	0	14.012716759745	Infinity
Example4.wav	175358474.3	0	13.8901984918816	Infinity
Example5.wav	212262913.7	0	13.0607248605054	Infinity

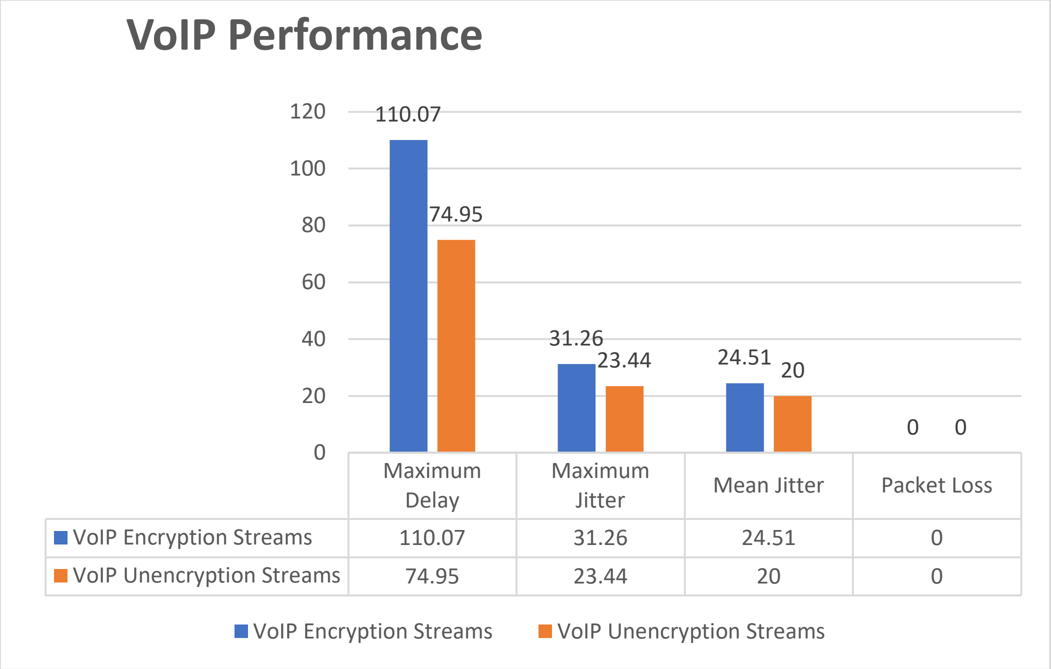

Figure 4- VoIP Performance metrics

The quality of the VoIP performance is represented in three important parameters packet loss, delay, and jitter [1].

Packet loss: loss in the transmitted packet defines the percentage of the packet that transmitted via sender and never reaches the correct destination, or the destination has dropped these packets deliberately due to an error in the packet header (e.g. TTL=0), or a transmitted packets discarded by intermediate links. The VoIP system should be implemented with packet loss less than 1.5%.The packet loss through reasonably reliable estimates to define the grade of performance, Good between 0% and 0.5%, Acceptable between 0.5%- and 1.5%, Poor greater than 1.5%.

Delay: is defined as the time that RTP packet takes to successful transmitting from the sender to destination. Delay defined as three categories: Good between 0ms and 150ms, Acceptable between

150ms and 300ms, and poor greater than 300ms.

Jitter: is defined as the variation in the time delay from one end to another. If the time delay of transmitting packet has widely variation in a VoIP call, it causes greatly degraded in voice call quality. In VOIP network, at each endpoints having jitter buffers to store received packets, it is designed to deliver traffic to end user at the constant rate. When sender have been generated packets with different rates, this variation that called jitter result from delay in time for construction RTP packets, voice data compression, and data encryption. VoIP network that has higher jitter values, it causes loss in packet and degradation in voice quality. The grade of Jitter classify into three categories: Good between 0ms and 20ms, Acceptable between 20ms and 50ms, and poor greater than 50ms.

RESULTS: The main goal here to analysis show the effect of proposed encryption algorithm on the VoIP quality, two types of VoIP streams were analyzed, stream without encryption (Plain-text stream), and stream based on stream cipher and logistic key generator. As mentioned earlier, four parameters were calculated using Wireshark program to analyze 50,000 RTP packets per duration for each stream. These four parameters were maximum delay, maximum and mean jitter, and packet loss. These results have been shown in Figure-4.

Conclusion

In this paper, a novel VoIP speech-encrypting algorithm is proposed based on stream cipher RC4 method with three-mixed 1- D logistic chaotic maps for one time pad key generation. The VoIP system based chaotic encryption scheme and quality of encryption are implemented using C#.net and the VoIP performance analysis using Wireshark. The experimental results demonstrate that the proposed method satisfies the speech encryption requirements. The encrypted voice is unintelligible, and the payload of RTP packet has the same size without any padding .The recovered speech has perfect quality with MSR equal to zero and PSNR equal to infinity. The VoIP system is implemented using Microsoft windows environment with Framework version 4.

The future work may envisage the implementing using the Android environments, AICS or a reconfigurable hardware, e.g., FPGA [11-16].

References

1. Alani MM. **2010**. Measuring the effect of AES encryption on VoWLAN QoS. In Soft COM 2010, 18th International Conference on Software, Telecommunications and Computer Networks 2010 Sep 23 (pp. 51-54)IEEE.
2. Man KP., Wong KW. and Man KF. **2006**. Security Enhancement on VoIP using Chaotic Cryptography. Computer Science IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics, pp: 3706-3708,IEEE .
3. Gnanajeyaraman R. and Prasadh K. and Ramar Dr. **2009**. Audio encryption using higher dimensional chaotic map. *International Journal of Recent Trends in Engineering*, **1**(2): 103-107.
4. Sheu LJ. **2011**. A speech encryption using fractional chaotic systems. *Nonlinear Dynamics* , **65**(1-2): 103-108.
5. Ibrahem MK, Kassim HA. **2018**. VoIP Speech Encryption System Using Stream Cipher with Chaotic Key Generator. *Journal of Fundamental and Applied Sciences*. **10**(6S): 204-210.
6. Ashtiyani, M., P.M. Birgani and Madahi, S.K. **2012**. Speech signal encryption using chaotic symmetric cryptography. *J. Basic Appl. Sci. Res*., **2**: 1668-1674.
7. Su ZP, Jiang JG, Lian SG, Zhang GF, Hu DH. **2010**. Hierarchical selective encryption for G. 729 speech based on bit sensitivity. *JIT*. 2010 Sep 1; **11**(5):599-607.
8. Wolf A., Swift JB., Swinney HL. **1985**. Vastano JA. Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*. 1985 Jul 1; **16**(3): 285-317.
9. Mohammed MT., Rohiem AE., El-moghazy A. **2012**. Confidentiality enhancement of secure real time transport protocol. In2012 8th International Computer Engineering Conference (ICENCO) 2012 Dec 29 (pp. 43-48). IEEE.
10. Hasan S., Boussakta S. and Yakovlev A. **2010**. Improved parameterized efficient FPGA implementations of parallel 1-D filtering algorithms using Xilinx System Generator. InThe 10th IEEE International Symposium on Signal Processing and Information Technology 2010 Dec 15 (pp. 382-387). IEEE.

11. Hasan S., Boussakta S. and Yakovlev A. **2011**. Parameterized FPGA-based architecture for parallel 1-D filtering algorithms. InInternational Workshop on Systems, Signal Processing and their Applications, WOSSPA 2011 May 9 (pp. 171-174). IEEE.

12. Hasan S. **2016**. Performance-vetted 3-D MAC processors for parallel volumetric convolution algorithm: A 256× 256× 20 MRI filtering case study. In2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA) 2016 May 9 (pp. 1-6). IEEE.

13. Humaidi AJ., Hassan S. and Fadhel MA. **2016**. Rapidly-fabricated nightly-detected lane system: An FPGA implemented architecture. *The Asian International Journal of Life Sciences*. 2018; **16**(1): 343-355.

14. Humaidi AJ., Hassan S. and Fadhel MA. **2018**. FPGA-based lane-detection architecture for autonomous vehicles: A real-time design and development. *The Asian International Journal of Life Sciences*.2018; **16**(1): 223-237.

15. Humaidi AJ., Hasan S. and Al-Jodah AA. **2018**. Design of Second Order Sliding Mode for Glucose Regulation Systems with Disturbance. *International Journal of Engineering & Technology*. 2018; **7**(2.28):243-7.