Speech Coding Combining Chaos Encryption and Error Recovery for G.722.2 Codec

Messaouda Boumaraf and Fatiha Merazka
LISIC Laboratory, Telecommunications Department
USTHB University, Algiers, Algeria
boumaraf.messa@gmail.com, fmerazka@usthb.dz

Abstract

With the evolution of network communication technology and advances in multimedia application, speech or data networks over an IP connection are vulnerable to threats. Therefore, the need to protect data attracts many researches on safe communications, especially speech secure communication. Additionally, with the large volume of unprotected speech data transmitted over the internet, Voice over Internet Protocol (VoIP) packets could be lost, and they cannot be recovered back, which would result in a degradation of speech quality. In this paper, we propose a secure speech communication approach based on chaotic cryptography combined with G.722.2 error recovery technique performed by interleaving. On the one hand, this approach uses the interleaving technique on inter-frames of G.722.2 speech in order to make a continuous packet loss becoming an isolated packets loss. On the other hand, speech will be encrypted using chaotic Lorenz system which achieves high encryption efficiency. To evaluate performance, the proposed design was evaluated through Enhanced Modified Bark Spectral Distortion (EMBSD) and Mean Opinion Score (MOS) with different packet loss rates to confirm the efficiency of our proposed scheme.

1 Introduction

Recently, with the development of network communication technology and signal processing techniques, it has become realistic to transmit speech, just like computer data, over the Internet (VoIP: Voice over Internet Protocol). However, the emergence of Internet use became very apparent; and the huge mass of data overloads the network (Mata-Díaz et al., 2014 - Labyd et al., 2014).

Networks must provide predictable, secure, measurable, and sometimes guaranteed services. Realizing the required Quality of Service (QoS) by managing the delay, delay variation (jitter), bandwidth, and packet loss parameters on a network become the secret to a successful end-to-end business solution. In real-time transmissions, IP networks are unpredictable and offer a best-effort transfer service with no QoS securities. Therefore, packets could be lost, causing an interruption in the conversation and a feeling of hatching of speech that is very annoying for the listeners. Therefore, it is fundamental to put a mechanism for concealing packet loss such as interleaving method, Forward Error Correction (FEC) (Nagano and Ito, 2013 - Shetty and Gibson, 2007).

In addition, speech data is vulnerable to corrupted or stolen by the hacker on the internet. For secure communication, it is necessary to protect data using encryption methods (Alvarez and Li, 2006).

Recently, research on chaotic cryptography increased expeditiously in order to improve chaos-based cryptosystems. In 1963, Edward Lorenz founded chaos theory, followed by the discovery of the Rössler attractor in 1976, since several chaotic systems are established (Jiang and Fu, 2008 - Kaur and Kumar, 2018). A chaotic system

is a non-linear, deterministic presenting good properties such as aperiodicity, pseudo-randomness and sensitivity to changes in initial conditions, which makes it unpredictable. Because of its characteristics, the chaos was used in the encryption system (Zhang and Cao, 2011 - Moon et al. 2017).

In (Afrizal, 2018), the authors' study focus on examining a few speech codec that usually used in connectionless communication such as G.711, G.722, G.729, AMR-NB, and AMR-WB for voice over LTE application and the impact of random and burst packet loss on voice communication against the codec using Evalid and NS-3 simulator. in (Li et al, 2015) the paper describes a method of digital encryption based on Lorenz continuous chaotic system, combined with chaotic dynamics, continuous sequence of numbers generated by the Lorenz chaotic system. Discrete the continuous data through the Euler method. Image encryption as an example, verify the Lorenz chaotic system digital encryption features. In (Guo et al., 2002) authors propose a VoIP technique combining the speech data encryption and G.729 error recovery. This technique uses the chaotic data interleaving on inter-frames of voice to make situation of continuous packet loss becoming an isolated packet loss situation. Then, they propose a Periodical Parameter Re-initialization (PPR) recovery approach to reduce the signal quality degradation in the G.729 decoder due to the lost of state synchronization to the G.729 encoder. Beside the proposed VoIP technique, also uses the idea of chaotic data encryption on intra-frames of speech to scramble the data sequence within a speech frame.

In this paper, we propose a secure speech communication approach based on chaotic cryptography combined with G.722.2 error recovery technique performed by interleaving, and it is organized as follows. In Section 2, an overview of the AMR-WB G.722.2 is introduced. Section 3 gives a very brief description of the proposed technique, which has a direct relation to our contribution. Simulations and interpretation are presented in Section 4. Finally, the conclusion is provided in section 5.

2 Overview of the AMR-WB G.722.2

The adaptive Multi-Rate Wideband (AMR-WB) speech codec is based on Adaptive Multi-Rate encoding, using similar methodology as algebraic code excited linear prediction (ACELP). AMR-WB is codified as G.722.2, an ITU-T standard speech codec, then was improved by Nokia and VoiceAge and it was first defined by 3GPP.AMR-WB offers enhanced speech quality due to a larger speech bandwidth of 50–7000 Hz compared to narrowband speech coders. G.722 sample audio data at a rate of 16 kHz, it contains nine bit rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.6 kbps, these ones are presented by modes 8, 7, 6, 5, 4, 3, 2, 1 and 0 respectively. To reduce average bit rate, this codec supports the discontinuous transmission (DTX), using Voice Activity Detection (VAD) and Comfort Noise Generation (CNG) algorithms (ITU-T Standard G.722.2, 2003).

The coder works with a frame size of 20-ms and the algorithmic delay for the coder is 25-ms. The AMR-WB G722.2 uses six parameters (VAD-flag, ISP, pitch delay, LTP-filtering, algebraic code, and gain) to represent the speech and these are shown in Figure 1 for bit rate 6,60 kbps.

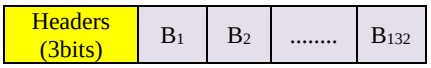


Figure 1: The bitstream of the coder parameters (coder output / decoder input) for the 20-ms frame in mode 0

where B_1, B_2, ..., B_{133} represent the bit 0 (BIT-0: FF81) or the bit 1 (BIT-1: 007F) of the coder parameters which is codified on 16 bits (WORD16).

3 The proposed technique

In this study, two techniques are combined employing interleaving and encryption processes. The encoded bitstream will be reordered using the interleaving, then transmitted over lossy IP channel after encryption, channel encoding and modulation. All these steps will be reversed at receiver as depicted in Figure 2.

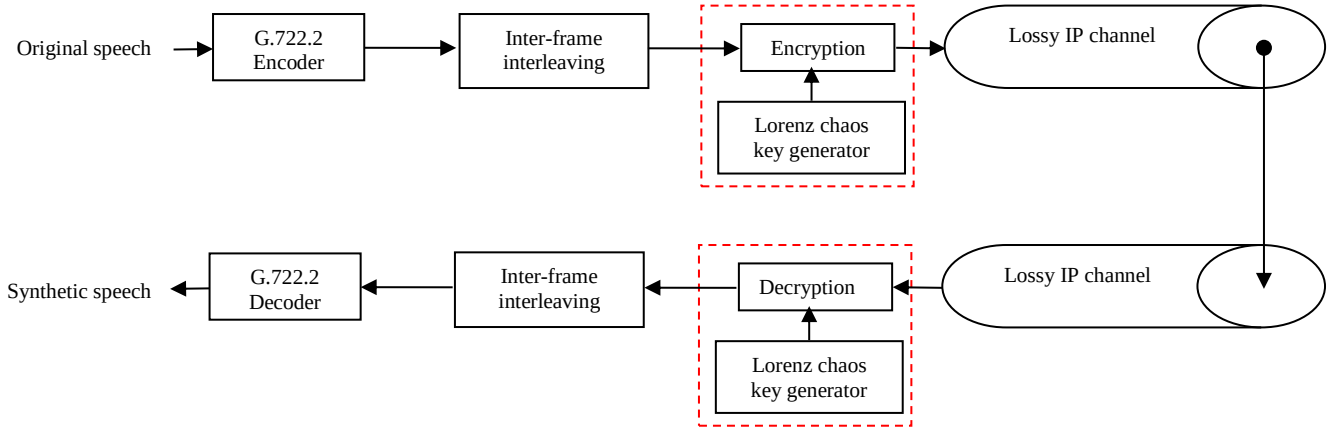


Figure 2: Proposed scheme of combined speech encryption with error recovery based on interleaving

3.1 Interleaving process

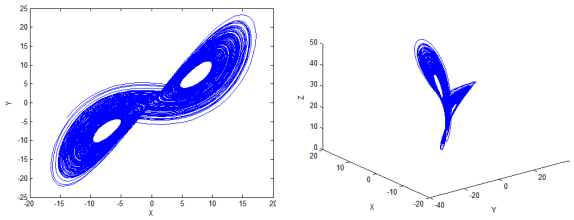
Interleaving technique is very useful when the packets contain multiple frames and the end-to-end delay is not important. Before transmission of the bitstream, the frames are re-arranged in such a way that the initially adjacent ones are separated in the transmitted bitstream and then put back in their original order at receiver level. As a result, the packet erase effects are scattered and produce situation of continuous packet loss becoming an isolated packet loss situation (Okamoto, et al., 2014).

3.2 Encryption process

Some important properties of chaos, such as the ergodicity, high sensitivity to the changes of control parameters, initial conditions and unpredictable behavior can be used in the generation of random numbers. So we use Lorenz model, the first well known dynamical system, governed by the differential equations (Lorenz, 1963):

$$\begin{cases} \dot{x} = a(y - x) & (a) \\ \dot{y} = cx - y - xz & (b) \\ \dot{z} = xy - bz & (c) \end{cases} \quad (1)$$

where x, y, z are state variables and a, b, c are real constant parameters of the system. With $a = 10$, $b = \frac{8}{3}$, $c = 28$, Lorenz system generates a chaotic behavior and its attractor is depicted in Figure 3:



phase portraits: x−y plane Attractors in (x, y, z)

Figure 3: Phase portraits and chaotic attractors of Lorenz model

The speech encryption algorithm is done in two stages: confusion and diffusion.

Step1: In the confusion stage, the parameters of the frame are permuted by using the keys x_n of Lorenz formula (1-a). So, the values are sorted in decreasing order while safeguarding the position or the index of each key values. then, the position of data speech is changed according to indexes' keys.

Step2: In the diffusion stage, the permuted parameters of frames are substituted formula (1-b) of Lorenz equation. The obtained keys are calculated as follows:

$$\text{key(i)} = \left[\text{y(i)} - \text{floor}\left(\text{y(i)}\right)\right] * 32767$$

So, the diffusion is performed using XOR operation between data and the key.

4 Simulation and discussion

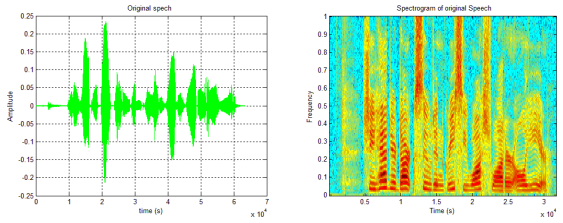
In this section, we study the performance in terms of security and recovery quality of lost packets. Several experiments are carried out to test the interleaving and encryption efficiency of the presented wideband speech cryptosystem. The quality of the encrypted interleaved speech and the reconstructed signals is assessed for the standard AMR-WB G.722.2. Thus, the speech file was

encoded using AMR-WB G.722.2 CS-ACELP. The resulting bit streams were rearranged employing interleaving technique and encrypted using Lorenz model. In the experiments, signal assessment in both the time and frequency domains is done to evaluate the distortion degree between the original and reconstructed speech. Therefore, the speech signal is displayed in two representations: waveform and spectrogram.

The evaluation of speech quality includes two measures: objective and subjective, Enhanced Modified Bark Spectral Distortion (EMBSD) (Yang, 1999) and Mean Opinion Score (MOS) (ITU-T, 2006) respectively. Note: for the MOS assessment, scores on the scale range from 1 to 5 (1: Unsatisfactory, 2: Poor, 3: Fair, 4: Good 5: Excellent). To demonstrate the efficiency of our proposed VoIP scheme combining G.722.2 frames with interleaving and chaos encryption, we have performed individually simulations for AMR-WBG, followed by the interleaving technique, then,the chaos encryption and finally, we combine them.

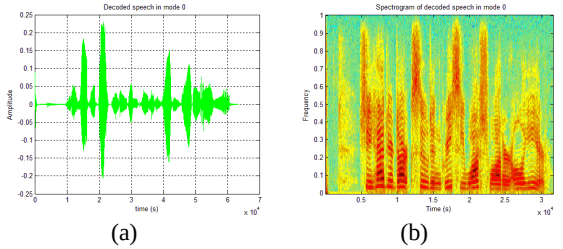
4.1 Performance of AMR-WB

In our test, a speech file with 198 frames is used which is represented in Figure 4. Recall that encryption uses 9 modes, of which we opt, in our experiments, for mode 0 (6.6 kbps) and mode 7 (23.05 kbps).



(a) (b)

Figure 4: (a) Original speech, (b) its spectrogram



(a) (b)

Figure 5: **(**a)Decoded speech in mode 0 (b) its Spectrogram

Figure 5 shows the speech decoded in mode 0.

We can see that the original and the decoded speech seem identical in waveforms (Figure 4-a and Figure 5-a) and spectrograms (Figure 4-b and Figure 5-b) representations.

The EMBSD and MOS assessments of speech quality are given in Figure 6. The values given by the two metrics show that the speech encoded in mode 7 is better than the one encoded in mode 0, while noticing that the original speech (no coding) is the best. A small difference between the original and the encoded speech is because we have a lossy codec. But generally, the encoded speech in both modes is classified good.

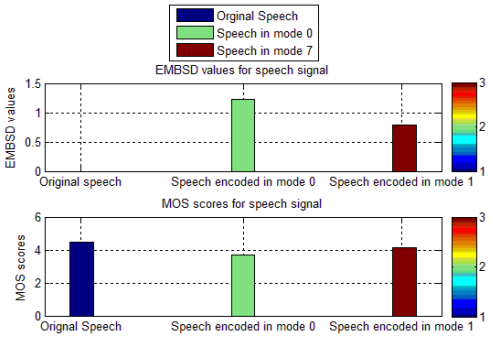


Figure 6: EMBSD and MOS scores

4.2 Interleaving tests

The encoded speech data will be scrambled using interleaving method. To simulate VoIP network losses, we use two-state Gilbert model.

Rate (%)	P	q
00	00	00
5	0,05	0,15
10	0,09	0,15
20	0,22	0,20
30	0,31	0,23
40	0.39	0.38

Table 1: shows the loss rates.

We use interleaving method to recover the lost packets during network congestion or degradation. Figures 7 and 8 give the obtained results from tests with EMBSD and MOS objective and subjective measurement tool respectively. We can see that the proposed method in both modes performs well than the original for the two losses rates 5% and 10%, contrariwise for the higher i.e more than 10%.

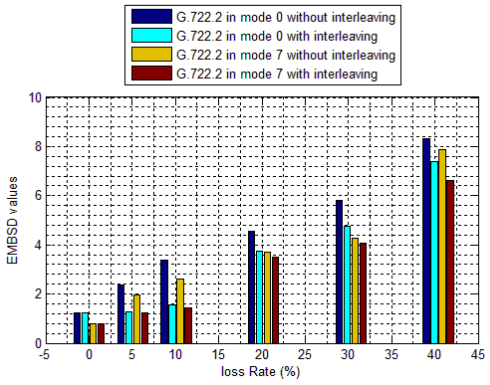


Figure 7: EMBSD values for interleaving

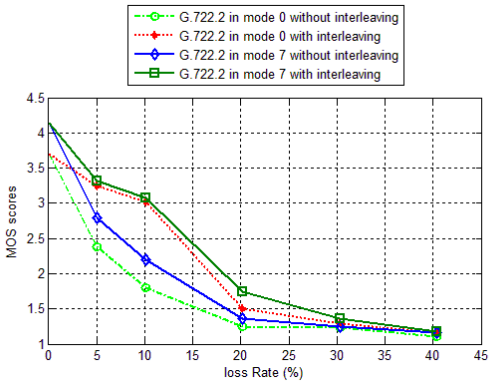


Figure 8: MOS scores for interleaving

we can confirm that by analyzing the audiograms speech slices in Figure 9.

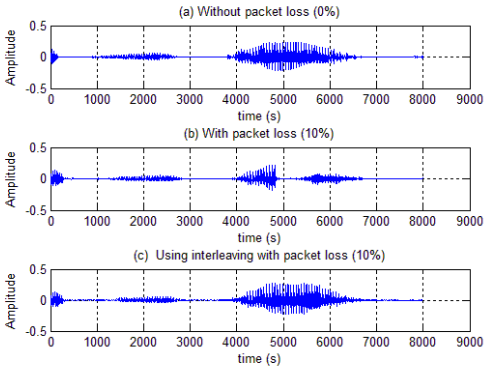


Figure 9: Portion of G.722.2 speech in mode 0: (a) original speech (b) Original speech with packet loss (10%)(c) Using interleaving with packet loss (10%)

4.3 Encryption tests

The speech file was encoded using AMR-WB G.722.2 CS-ACELP. The resulting bitstreams were encrypted using chaotic full encryption performed by both confusion & diffusion processes. Figure 10 depicts the signal inspection in both the time and frequency domains.

We can see from Figures 10-a and 10-b that the encrypted speech signals are similar to the white noise, which indicates that no residual intelligibility can be useful for eavesdroppers at the communication channel. However, the reconstructed speech signals (Figures 10-c and 10-d) using the right keys are the same as the original.

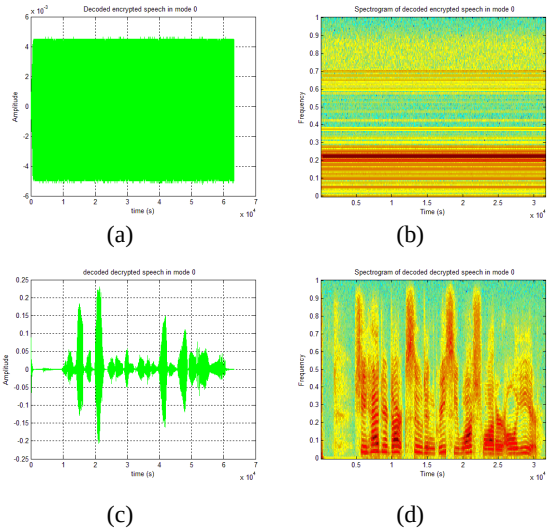


Figure 10: Full encryption using mode0 of WB-G722.2: (a) Decoded encrypted speech (b) its spectrogram (c) decoded decrypted speech (d) its spectrogram

To evaluate the efficiency of the encryption schemes, we have used the EMBSD and MOS tools. We can see that the EMBSD (Figure 11) values for the original speech coded in the modes 0 and 7 are near zero which indicates its good quality.

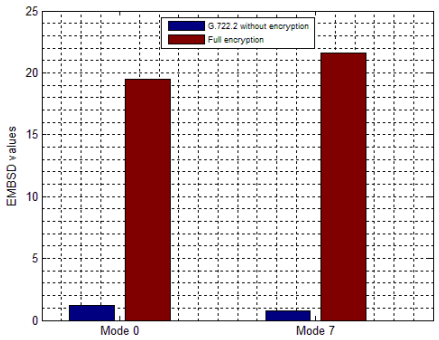


Figure 11: EMBSD values for full encryption

In return, significantly greater values increase for encrypted speech data which indicates its worse quality.

Also, the MOS evaluation in Figure 12 confirms and gives scores "Good" for the original speech and "unsatisfactory" for the encrypted one. We can also notice that the quality of the decrypted speech employing the same keys than the encrypted one give a signal quality identical to the original speech.

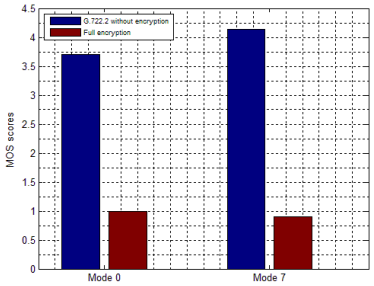
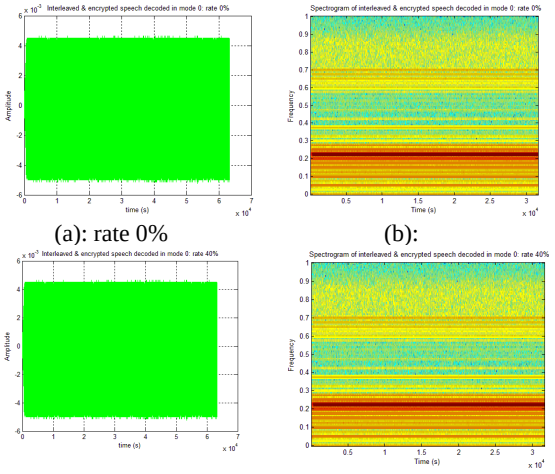


Figure 12: MOS scores for full encryption

4.4 Combined tests

The speech file will be encoded then scrambled using the interleaving process, in order to make the continuous multiple-packet loss situation to isolated packet loss situation. Next, it is encrypted by chaotic Lorenz mode. Figure 13 shows the combination of interleaving and encryption processes. We can see that, for the two losses rates, the speech data appears as a white noise.

Note: The EMBSD values and MOS scores for the interleaved and encrypted file in mode 0 or mode 7 give the same value than the only encrypted speech which indicate the efficiency of the full encryption.



(a): rate 0% (b): (c): rate 40% d):

Figure 13: Interleaved & encrypted speech decoded in mode 0

Figure 14 shows the speech audiograms of the proposed schema.

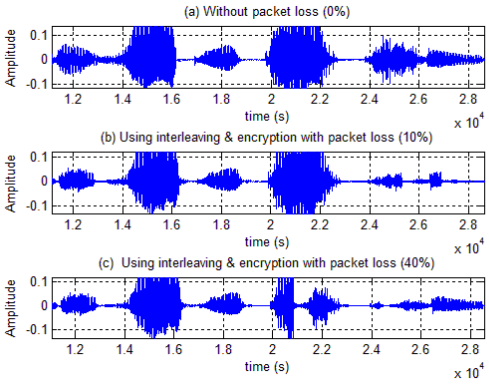


Figure 14: Portion of G.722.2 speech in mode 0: (a) original speech (b) Using interleaving & encryption with packet loss (10%) (c) Using interleaving & encryption with packet loss (10%)

5 Conclusion

In this paper, we have presented our proposed method which combines chaos encryption using the Lorenz system and error recovery based on interleaving techniques for the standard ITU-T AMR-WB G.722.2 codec. The purpose of interleaving is to improve speech quality degradation caused by packet losses. In addition, the experimental results and analysis show that the cryptosystem is efficient in terms of security which is suitable for transmission over public transmission channels.

References

Mata-Díaz, J., Alins, J., Muñoz, J. L., and Esparza, O. 2014. A simple closed-form approximation for the packet loss rate of a TCP connection over wireless links. *IEEE Communications Letters*, 18(9), 1595-1598.

Labyad, Y., MOUGHIT, M., Marzouk, A., and HAQIQ, A. 2014. Impact of Using G. 729 on the Voice over LTE Performance. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(10), 5974-5981.

Labyd, Y., Moughit, M., Marzouk, A., and Haqiq, A. 2014. Performance Evaluation for Voice over LTE by using G. 711 as a Codec. *International Journal of Engineering Research and Technology*, 3(10), 758-763.

Nagano, T., and Ito, A. 2013. A Packet Loss Recovery of G. 729 speech using discriminative model and N-gram. In *2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.*2013: 267-270.

MITTAG, Gabriel et MÖLLER, Sebastian. Single-ended packet loss rate estimation of transmitted speech signals. *In : 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. 2018: 226-230.

Shetty, N., and Gibson, J. D. 2007. Packet Loss Concealment for G. 722 using Side Information with Application to Voice over Wireless LANs. *Journal of Multimedia*, *2*(3).

Alvarez, G., and Li, S. 2006. Some basic cryptographic requirements for chaos-based cryptosystems. *International journal of bifurcation and chaos*, 16(08): 2129-2151.

Jiang, H. Y., and Fu, C. 2008. An image encryption scheme based on Lorenz chaos system. *Fourth International Conference on Natural Computation*. 4: 600-604.

Alshammari, A. S., Sobhy, M. I., and Lee, P. 2017. Secure digital communication based on Lorenz stream cipher. *30th IEEE International System-on-Chip Conference (SOCC)*.2017: 23-28.

Zhang, J. 2015. An image encryption scheme based on cat map and hyperchaotic lorenz system. *IEEE International Conference on Computational Intelligence & Communication Technology* . 78-82

Kaur, M., and Kumar, V. 2018. Efficient image encryption method based on improved Lorenz chaotic system. *Electronics Letters*, 54(9): 562-564.

Zhang, Z. X., and Cao, T. 2011. A chaos-based image encryption scheme with confusion-diffusion architecture. *In International Conference on Computer Science and Information Engineering*. *Springer, Berlin, Heidelberg*. 258-263.

Zhu, Z. L., Zhang, W., Wong, K. W., and Yu, H. 2011. A chaos-based symmetric image encryption scheme using a bit-level permutation. *Information Sciences*, 181(6): 1171-1186.

Wong, K. W., Kwok, B. S. H., and Law, W. S. 2008. A fast image encryption scheme based on chaotic standard map. *Physics Letters A*, 372(15): 2645-2652.

Wang, B., Xie, Y., Zhou, C., Zhou, S., and Zheng, X. 2016. Evaluating the permutation and diffusion operations used in image encryption based on chaotic maps. *Optik-International Journal for Light and Electron Optics*, 127(7): 3541-3545.

Moon, S., Han, B. S., Park, J., Seo, J. M., and Baik, J. J. 2017. Periodicity and chaos of high-order Lorenz systems. *International Journal of Bifurcation and Chaos*, 27(11): 1750176

Afrizal, G. 2018. Impact of Random and Burst Packet Loss on Voice Codec G. 711, G. 722, G. 729, AMR-NB, AMR-WB. *In 2018 4th International Conference on Wireless and Telematics (ICWT)*. 2018: 1-4.

Li, W., Zhang, Q., and Ding, Q. 2015. Digital encryption method based on lorenz continuous chaotic system. *Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*. 2015: 262-266.

Guo, J. I., Lin, C. C., Tsai, M. C., & Lin, S. W. 2002. An efficient voice over Internet protocol technique combining the speech data encryption and G. 729 error recovery. *In Proc. Int. Computer Symposium (ICS'2002)*. 2002

ITU-T Standard G.722.2, 2003.Wideband coding of speech at around 16 kbps using Adaptive Multi-RateWideband (AMR-WB).

Okamoto, M., Nose, T., Ito, A., and Nagano, T. 2014. Subjective evaluation of packet loss recovery techniques for voice over IP. *In 2014 International Conference on Audio, Language and Image Processing.* 2014: 711-714.

Lorenz, E. N. 1963. Deterministic non periodic flow. *Journal of the atmospheric sciences*, 20(2): 130-141.

Yang, W. 1999. Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure Based on Audible Distortion and Cognitive Model. *Temple University*.

ITU-T. 2006. Mean opinion score (MOS) terminology. *Recommendation P.800.1.*