❒ 665

Improved voice quality with the combination of transport layer & audio codec for wireless devices

Othman O. Khalifa, Raihan Jannati Binti Roslin, Sharif Shah Newaj Bhuiyan
Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

Article Info

ABSTRACT

Improving voice quality over wireless communication becomes a demanding feature for social media apps like facebook, whatsapp and other communication channels. Voice-over-internet protocol (VoIP) helps us to make quick telephone calls over the internet. It includes various mechanism which are signaling, controlling and transport layer. Over wireless links, packet loss and high transmission delay damage voice quality. Here VoIP quality will be measured by three main elements which are signaling protocol, audio codec and transport layer. To improve the overall voice quality, we need to combine these three elements properly to get the best score. Otherwise perceptual speech quality will not be the right tool to measure the voice quality. Here we will use Mean Opinion Score (MOS) for calculated jitter values and end to end delay. At the end, best combination of audio codec & signaling protocol produced the quality speech.

Corresponding Author:

Othman O. Khalifa,
Department of Electrical and Computer Engineering,
International Islamic University Malaysia,
Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
Email: khalifa@iium.edu.my

1. INTRODUCTION

Communication technology is one of the technology that keep on advancing throughout the years. Traditionally, people used Public Switched Telephone Network (PSTN) to make a voice call using their telephone. This traditional voice call usually is to pay as you use service provided by the Telco used by the user. As the technology advancing, the internet has been introduced to the world which enables data to be transmitted through the internet with the associated protocol [1-4]. Voice over internet protocol is used to enable audio such as speech or voices to be sent from the sender to the receiver as data through the internet. The example of VoIP's applications is Whatsapp Call, Facebook Messenger Call, Viber, and Skype. All of these well-known and used by almost everybody nowadays are using VoIP to allow the user to make internet voice call. A lot of research has been done to improve the quality of voice-over-internet protocol. Many methods were proposed in order to provide better VoIP experience and service quality such as using path-switching packet forwarding mechanism, using various call scheduling policies and by lowering the bandwidth consumption [5]. The method of combining the codec scheme and signaling protocol is proposed in this paper to improve the VoIP speech quality.

Three of the main reason that makes people favor to use VoIP is first, VoIP allows people to make internet voice call from and to any place in this world regardless the distance between the caller and the call receiver as long as they are connected to the internet. Second is internet voice call allow them to make an international call for free without any extra charge that a traditional voice call just like what is used to be if using the PSTN voice call. Lastly [6, 7], VoIP allows users to make voice and video conference which allow more than two callers to converse at the same time. These features that VoIP has allows users to maximize the usage of a voice call. To allow the audio to be transmitted from the sender to the receiver, three main

elements of VoIP are needed which are the audio codec scheme, signaling protocol, and the network layer of the internet. The signaling protocol is used to initialize, establish, sustain, and cut off the connection between the sender and receiver to allow and stop the transmission of the data. There are numbers of signaling protocol available, as for an example, Session Initiation Protocol (SIP) and H.323. Both of these signaling protocols have different architecture and features. Once the connection between the sender and the receiver is established, the voice audio from the sender then needs to be encoded from an analog voice audio into digitized voice signal as it is to be transmitted through the internet according to the agreed protocol. At the endpoint of the connection, the example of the audio codec that is commonly used in VoIP is G.726, G111, G.729, and Speex. These mentioned audio codecs have different algorithm to encode and decode the voice to data and vice versa so that the data can transmit through the internet from the sender to the receiver [8, 9]. The last element that is important in VoIP is the transport layer in the internet protocol. Real-time Protocol (RTP) and Real-time Control Protocol are the protocols that are used to move the voice audio that has been encoded into voice packets from the sender to the receiver over the internet [10].

The aim of this paper is to proof that with the appropriate pairing of the audio codec together with signaling protocol can help in producing a better quality of voice over internet protocol. The performance of each paired audio codec and signaling protocol are measured using the parameters of Quality of Service (QoS) of the decoded audio transmitted through the internet from the sender. However, Perceived QoS is measured from 'mouth to ear', i.e. end-to-end and depends on the performance of IP network and terminal/gateway. Figure 1 shows the VoIP network and perceived QoS.
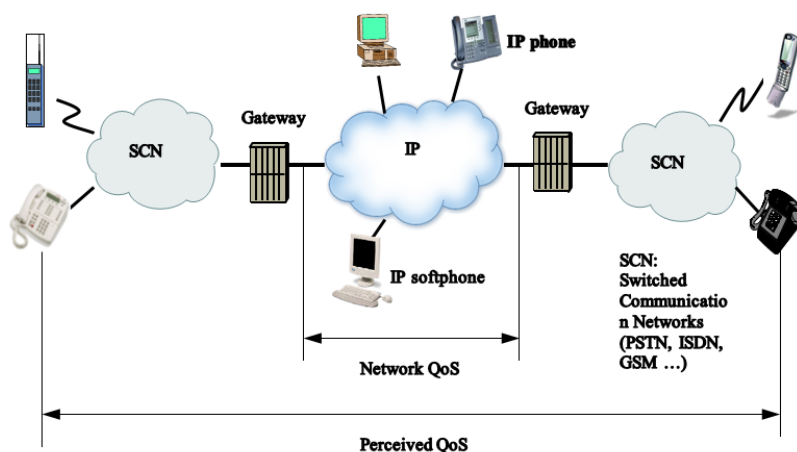

Figure 1. VoIP network and perceived QoS

There are many factors affect voice quality such as; coding distortion, codec delay in the sender side and packet loss network delay and jitter in IP networks. The receiver side also effected by buffer-delay buffer-loss as wel as codec impairment delay in addition to other factors such as; language, gender, FEC, packet loss concealment. Figure 2 shows factors affect voice quality.

2.   RELATED WORK

There are many methods have been proposed by the other researcher to improvise the voice-over-internet protocol's speech quality, especially for the wireless device. However, the performance voice transmission (Voice over IP) over wireless links is still unreliable in terms of throughput and perceptual speech quality.

Tao et al. demonstrated [7] the improved path switching technique to improve the quality of VoIP through their proposed scheme's prototype. An algorithm was used to determine the most efficient path at the gateway of VoIP. The feature of the proposed scheme is it has the capability to predict the path quality by referring to past path performance. This is the easiest way for path quality prediction. Other than that, this algorithm also can estimate the benefits of path switching of all paths available during the transmission.
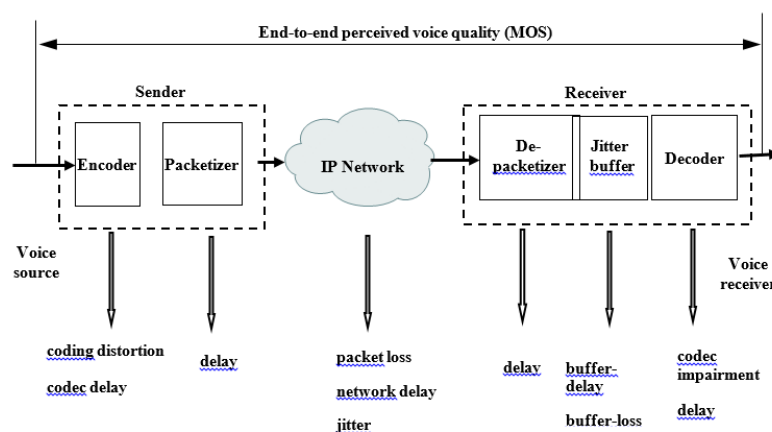
Figure 2. Factors affect voice quality

In 2006, a method using routing diversity was proposed by Jyoti et al. in [8] to improve the quality of VoIP. It was mentioned in this paper that routing diversity can reduce packet loss through explicitly using different path on sending the subsets of a packet stream. The method suggested was implemented at the postprocess of audio encoding where RTCP send the reports about the receiver to the sender, and the information like delay, packet-loss and jitter were used to indicate the congestion and length of the links, at the same time, it was also used for routing diversion via relays. The packets then buffered and arranged in the right sequence before the payload is played. Routing diversity is one the good ways to increase the packet-switched networks reliability for multimedia communication. However, this test result assessment of path diversity technique is inexact because delay overhead was added to the voice packets due to sending a stream via a series of known nodes.

Lambrinos and Djouvas mentioned that in a real-time application like VoIP, data streamed over multiple links at the same time is common where it may result in the links utilization decrement. One of the ways to solve this problem is through applying scheduling policies in Call Admission Control (CAC) as proposed in [9]. The scheduling policies defined in this paper are 1) "Call back first call" where call requests were favored based on their receiving time, 2) "Call back first exit" where the first call request that estimated to exit the coverage area is favored and the last one, 3) "No call back" reject the calls when the lines are unavailable. This scheduling scheme was implemented based on SIP [9]. This method can improve the Quality of Experience of VoIP, especially in wireless mesh network and multi-hop user connectivity. However, the test of the proposed scheduling policies is assuming that users bit rates are steady which is not realistic in a real-world situation and may result in imprecise results and analysis.

Bikes Agayev, in [10] investigated the effects of parameters of VoIP networks on the quality of the transmitter speech. The test had been done to G.711 and G.729 codec by analyzing the rank in R units (R-quality) which is in hundred-point scale and in MOS units (MOS-quality) based on a five-point scale for each codec. Five main causes of speech signal distortion in the network; packet loss in the internet, packet jitter, codec time and delivery delay of the speech signal to the receiver. Form the test conducted, G.711 scores higher than G.729 codec in both R units and MOS units. This test shows that changes in codec can give a huge impact on the quality of the transmitted speech. However, this test only considers two audio codecs while there is numerous audio codec available for VoIP.

Dantas et al. proposed a method that aimed to reduce the bandwidth usage but maintaining the quality of the call quality in [11]. The basic work of the proposed method is the central node is contacted prior finding for the other user within the network. Then a VoIP server is assigned by the central node for call handling task and both participating sides are connected to the VoIP server. All the call relay between two sides is controlled and handled by the VoIP server. The improvisation introduced by this method is the header elimination which requires both communicating sides to acknowledge that the particular header section is redundant and ready to be removed. The other feature introduced is the header-to-header payload ratio reduction using Neagle's algorithm which can be implemented in the network protocol. Nonetheless, as the scope is not including the network protocol, this situation cannot be applied here. Silence detection is a process using silence algorithm which enables the normal packet to be replaced with the silence packet that has less bit rate but still preserving the connection. However, this method tested using an application; The Horizon Global Exchange but the performance was compared to other application like Whatsapp, Viber, and Facebook. Besides, the payload-to-header ratio, payload size, and frame size still can't compete with the

Viber application although The Horizon was claimed to use less than half bandwidth compared to the second rank application, Whatsapp.

In a paper by Gueham & Merazka, they found an enhanced insertion packet loss concealment method in [12] to solve one of the major problems that always occur in VoIP, a real-time application is packet loss. In this paper, they proved that the commonly used method for packet lost concealment (PLC), last received frame repetition is not efficient. PLC is a way to prevent VoIP quality degradation due to loss of voice data packet. Thus, they proposed an algorithm to determine the best packet to repeat. This method can solve the redundancy problem that occurs in the formerly used PLC method by adding extra information about the Nth frame at the end of its frame. This information will help decoder at the receiver sider to produce a coherent signal to cover the Nth frame lost. Thus, the listener will remain unaware of the frame lost and at the same point giving a better MOS score. However, an additional delay max: 25ms was introduced as the decoder need a recovery phase of the Nth packet where it has to wait for the information extraction by the Nth+1 packet. Besides, this test only was done on G.722.2 codec while there are many other commonly used VoIP codecs not being tested through this method and may produce a different result.

Last year, Olariu et al. in [13] suggested to use delay-based priority queueing for VoIP where the priority is divided into five level; 1) packets with less than 5ms delay, 2) packets with delay between 5 and 10ms, 3) packets that have 10 to 15ms delay, 4) packets with delay between 15ms to 20ms and lastly 5) packets that have more than 5ms delay. The priority increased as the level increased which means 1 has the lowest priority and 5 has the highest priority. This method is suitable for a network with multi-queue as single queue will give a longer period of overall delay. Opus codec used during the test held for this method. MOS of VoIP can be improved by applying this method as it reduces the overall delay even in a crowd network situation. However, this method was developed for software defined networks (SDN) which mean the efficiency of this method on non-SDN is uncertain as the network architecture might be different.

Mekki and Mohammed mentioned in [14] that one of the ways to improve the quality of VoIP is by coupling the signaling protocols and codecs scheme. In this paper, the test was handled by several steps. First, the signaling protocol was chosen between H.323 and SIP. Second, one codec between G.711 and G729codec was selected. Next, the number of frame per packet was defined and one of the transport protocol was picked between TCP and UDP. It is important to test and determine the best combination of these components as they give a huge impact on the Quality of Service (QoS) of VoIP. Finally, run the simulation and the result was analyzed by comparing the values of throughput, packet loss, end-to-end delay, and jitter. To determine the best combination of the components, MOS and R-factor were calculated. Based on the test, the highest MOS and R-factor is the combination of UDP transport protocol and G.711 codec with one frame size. This method is good at determining the best combination of the signaling protocol, transport protocol, and the frame size. However, the test only made for two type of audio codec only while there are many other audio codecs available for VoIP and it is good to further this study as it can determine the best pair of audio codec and signaling protocol combination as a guideline for the future developer to choose the best pairing of this two element for their usage.

There are many types of research and founds to in improving the quality of voice-over internet protocol, but there is still many ways enhancement that can be made to obtain the best quality of VoIP especially in speech quality experienced by VoIP user [15, 16]. Therefore, an improvised VoIP speech quality using coupling signaling protocol and codecs scheme is proposed in this study.

3. PROPOSING SCHEME FOR VoIP

The proposed method for this paper is using signaling protocols and codecs scheme coupling to improve the quality of VoIP. By using this technique, there are steps that will be used throughout the project in obtaining the result which is the best match for codecs and signaling protocol. Riverbed OPNET Modeler was used to simulate the behavior of voice over internet protocol. Technically, the steps proposed for this project are as shown in Figure 3.
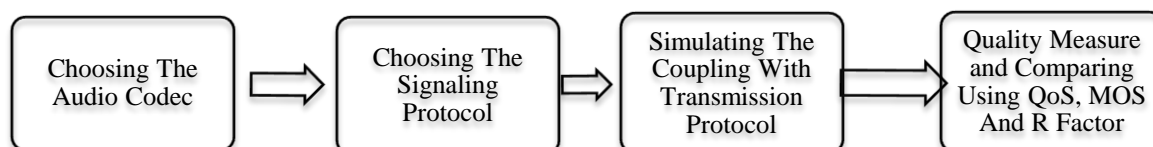

Figure 3. Method proposed to improvise the VoIP speech quality

3.1. Choosing the audio codec

The audio codec is one of the important element in VoIP. There are many available audio codecs that can be used in VoIP which every one of them has different features and characteristics. For this project, three commonly used audio codecs have been chosen which are G.711, G.726, and G7.29. This audio codec later will be paired with the signaling protocol and the VoIP speech quality produced will be analyzed and ranked based on their performance. Table 1 shows comparison for the G.711, G7.26, and G.7.29.

Table 1. Properties and features table of g.711, g.726and g.729 audio codec

Feature/Properties	G.711	G.726	G.729
Created By	ITU-T	ITU-T	ITU-T
Release Year	1972	1990	March 1996
Formal Name	Pulse Code Modulation (PCM)	Adaptive Differential Pulse Code Modulation (ADPCM)	Coding of speech at 8 kbit/s using code-excited linear prediction speech coding (CS-ACELP)
Cost	Free	Free	Free
Lossless audio compression	No. Lossy compression	No. Lossy compression	No. Lossy compression
Bit rate	64 kb/s	16 kb/s, 24 kb/s, 32 kb/s, 40kb/s.	8kb/s
Channels (Mono/ Stereo)	Mono	Mono	Multi-channel capable.
Audio Bandwidth	300Hz-3.4kHz (narrow band)	300Hz-3.4kHz (narrow band)	300Hz-3.4kHz (narrow band)
Frame Size/ Duration	5ms	10ms	10ms
Packet Loss Concealment	Yes	Yes	Yes
Forward Error Correction	No	No	Yes
Discontinuous Transmission	Yes	Yes	No
Voice Activity Detection	Yes. Comfort Noise Generation (CNG) is also used on silence moments through bandwidth usage reduction.	Yes. Also, use Comfort Noise Generation (CNG).	No

Using the Mean Opinion Score value, theses audio codecs are rated according Table 2. As the proposed method used for this paper was by coupling the audio codec with the signaling protocol, the features and the characteristics of the signaling protocol has been studied and the table of comparison below was obtained.

Table 2. The rate of G.711, G.726, and G.729 according to MOS value

Audio Codec	Mean Opinion Score (MOS) value
G.711	4.1
G.726	3.85
G7.29	3.92

3.2. Choosing the signaling protocol

The other VoIP element which to is paired with the audio codec is the signaling protocol. Just like audio codec, there is the various choice of signaling protocol available but the mostly used for VoIP are SIP and H.323. Each SIP and H.323 have different network architecture which has been summarized in Table 3.

Table 3. Table of comparison between SIP and H.323 signaling protocol

Features characterestic	H.323	SIP
Owner	ITU	IETF
Year	1996	1999
Elements	Terminals Gatekeepr Gateways Multiconference Unit (MCU)	User Agent Proxy Servers
Real Time Data Transmission	RTP/RTCP	RTP / RTCP
Signaling Procedure	Basic Call Setuo or Fast Connect	SIP-INVIT Transaction

Improved voice quality with the combination of transport layer & audio codec for… (Othman O. Khalifa)

Based on the study of these two signaling protocol, it can be deduced that H.323 offer more functionalities and features compared to SIP, but in the other hand, SIP has a lot simpler and flexible protocol. Despite the functionalities and sizes that they have, both of the signaling protocol can provide the same high level requirement of setting up a voice call which has been the reason for their succeed over the other.

3.3. Simulating the paired signaling protocol and audio codec

Using the Riverbed OPNET Modeler, the signaling protocol and audio codec coupling have been simulated. The H.323 and SIP network scenario have been setup to test the G.711, G.726 and G729 with the TCP an UDP transport layer on them. Figure 4 & 5 show the scenarios of a network architecture for H.323 and SIP respectively.
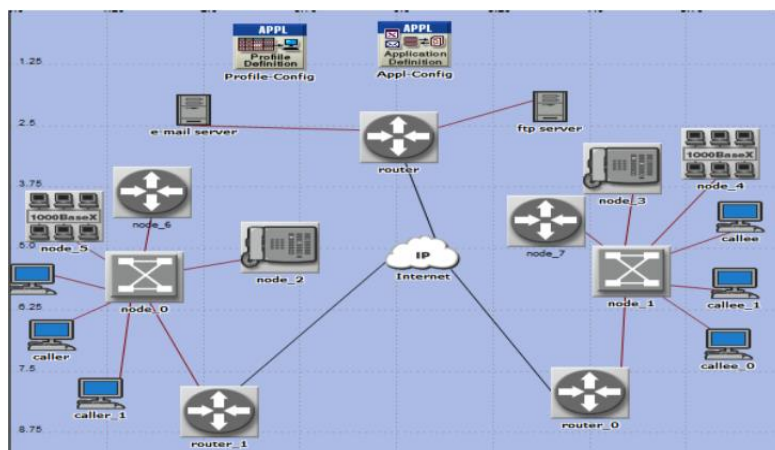

Figure 4. Configuration of H.323 architecture in OPNET scenario
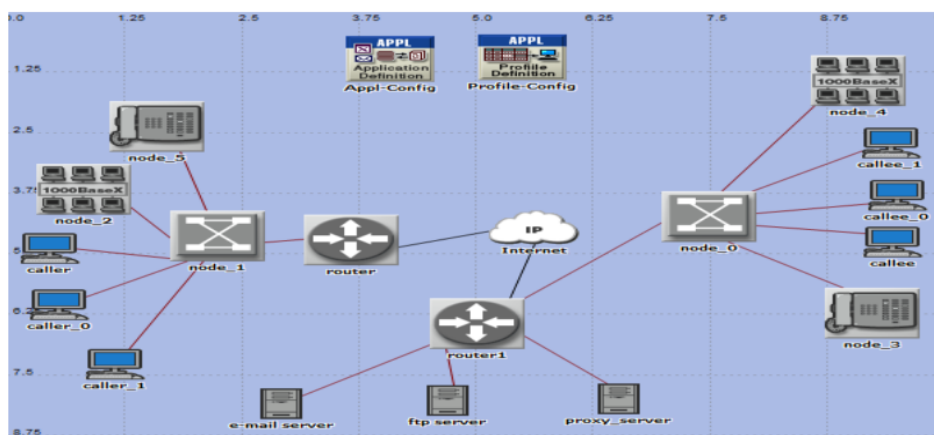

Figure 5. Configuration of SIP architecture in OPNET scenario

Then the result obtained from the simulation was then used to determine the best quality of the speech produced by each pair. To determine the quality of performance and speech of VoIP, Mean Opinion Score (MOS) and R-Factor were used [14]. Table 4 shows the score of R-Factor and MOS together with the score's description.

Table 4. F-factor and MOS mapped to the user satisfaction level

R-Factor	User Satisfaction Level	MOS
90–100	Very satisfied	4.3–5.0
80–90	Satisfied	4.0–4.3
70–80	Some users satisfied	3.6–4.0
60–70	Many users dissatisfied	3.1–3.6
50–60	Nearly all users dissatisfied	2.6–3.1
Below 50	Not recommended	1.0–2.6

4.   RESULT AND ANALYSIS

Through the simulation of the paired audio codec and the signaling protocol using the Riverbed OPNET Modeler, the result below were obtained and compared. As shown in Table 5 & 6, the End-to-End delay of the three codec (G.711, G.729 and G.726) were compared using H.323 and SIP protocols. It shows that that G.711 has the smallest end-to-end delay.

Table 5. Result comparison end to end delay (H.323 protocol)

Codec Scheme	Transport Layer Protocol	One Frame (sec)	Three Frames(sec)
G.711	TCP	0.22	0.17
	UDP	0.18	0.20
G.729	TCP	0.24	0.21
	UDP	0.22	0.22
G.726	TCP	0.26	0.36
	UDP	0.25	0.34

Using the data of end-to-end delay and jitter obtained for both transport layer protocol as in the Table 5, the values were used to determine the MOS values for each pairing. The MOS value also can be obtained directly from the Result Browser in the Riverbed OPNET Modeler software.

Table 6. Result comparison end to end delay (SIP protocol)

Codec Scheme	Transport Layer Protocol	One Frame (sec)	Three Frames(sec)
G.711	TCP	0.24	0.24
	UDP	0.16	0.22
G.729	TCP	0.17	0.20
	UDP	0.16	0.18
G.726	TCP	0.24	0.35
	UDP	0.26	0.29

The second experiment is compare the jitter buffer for the three codec (G.711, G.729 and G.726) using two both transport layer protocols; H.323 and SIP. It observed that The G.711 has the lower Jitter buffer as shown in Tables 7 and 8. Tables 9 and 10 show the R-Factor and MOS for both transport layer protocols.

Table 7. Result comparison for jitter (H.323)

Codec Scheme	Transpor Layer Protocol	One Frame (sec)	Three Frames(sec)
G.711	TCP	0.000090	0.000002
	UDP	0.000061	0.00017
G.729	TCP	0.00010	0.000070
	UDP	0.000035	0.000061
G.726	TCP	0.00007	0.00007
	UDP	0.00008	0.00008

Table 8. Result comparison for jitter (SIP)

Codec Scheme	Transpor tLayer Protocol	One Frame (sec)	Three Frames(sec)
G.711	TCP	0.00015	0.00028
	UDP	0.000005	0.00020
G.729	TCP	0.000044	0.00013
	UDP	0.000045	0.00000
G.726	TCP	0.00036	0.00020
	UDP	0.00040	0.00012

Table 9. Result comparison for R-factor and MOS (H.323)

Transport Layer Protocol	Codec	Average R-Factors	Average MOS
TCP	G.711	85.95	4.22
	G.729	71.5	3.67
	G.726	57.4	2.92
UDP	G.711	88.65	4.31
	G.729	73.55	3.76
	G.726	59.05	3.05

Table 10. Result comparisons for R factor and MOS (SIP protocol)

Transport Layer Protocol	Codec	Average R-Factors	Average MOS
TCP	G.711	80.6	4.04
	G.729	76.45	3.88
	G.726	58.1	2.995
UDP	G.711	86.2	4.23
	G.729	77.25	3.915
	G.726	63.6	3.28

In summary, the end-to-end delay and jitter values in the tables of results above show that G.711 has the smallest end-to-end delay when paired with H.323 while G. 729 has the smallest end-to-end delay when paired with SIP. In the other hand, for the voice jitter, G.726 has a higher value than G.711 and G.729 gives the best result by having the least jitter value. The high value of voice jitter will result in the difficulty to understand the received voice due to the packets arrived at different time. This will contribute to a lower value of MOS for G.726, vice versa to G.729 that has the least voice jitter. From the above founding, firstly, the properties and features of G.711, G726, and G.729 audio codecs together with the signaling protocol SIP and H.323 was extracted which was very well used in this study. Next, the network architecture of H.323 and SIP was configured in the Riverbed OPNET Modeler as the software that was used for the simulation.

5.   CONCLUSION

VoIP over wireless is a promising and a challenging service that started to flourish between mobile and wireless users. Cost effectiveness and also the additional available services that VoIP can deliver are two driving forces that greatly motivate the Internet service providers and solution developers to develop new software and systems that are capable of delivering a competitive VoIP quality over the mobile network. In this paper, an improvised speech quality of voice-over-internet protocol by coupling signaling protocols and codec schemes is proposed to improve the quality of VoIP is presented. The SIP and H.323 architectures were used to evaluate the QoS parameters and calculate the number of frame per packet. With regards to criteria parameters, that TCP outperforms UDP/RTP. Packet losses were not observed for TCP unless tqhe background traffic was at maximum load.

REFERENCES

[1] Hur, J. The History of VoIP–From The Beginning Til Present Time, [Online]. Available: https://bebusinessed.com/history/voip-history/. [Accessed: 23-August-2018].
[2] Mehta P. and Udani, S., "Voice over IP," *IEEE Potentials*, vol. 20, no. 4, pp. 36–40, 200.
[3] "What is SIP? | Network World,". [Online]. Available: https://www.networkworld.com/article/2332980/lan-wan/lan-wan-what-is-sip.html. [Accessed: 13-Feb-2019].
[4] Khiat, A., El Khaili, M., Bakkoury, J. and Bahnasse, A., "Study and evaluation of voice over IP signaling protocols performances on MIPv6 protocol in mobile 802.11 network: SIP and H.323," in *2017 International Symposium on Networks, Computers and Communications (ISNCC)*, 2017, pp. 1–8.
[5] Mohammad Behdadfar, Ehsan Faghihi, Mohammad Ebrahim Sadeghi, "QoS parameters analysis in VoIP network using adaptive quality improvement", *Signal Processing and Intelligent Systems Conference (SPIS)* 2015, pp. 73-77, 2015.
[6] S. Tao *et al.*, "Improving VoIP quality through path switching," *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, Miami, FL, 2005, pp. 2268-2278 vol. 4.
[7] T.S. Gunawan, M. Kartiwi, "Performance evaluation of multichannel audio compression," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, 2018, 10(1). pp. 146-153.
[8] Jyoti, J. El-Tawab, S., El-Derini, M. N., Aboelela, E. and Aly, H., "Improving Quality of Service for Voice-Over-IP Using Routing Diversity," in *23rd Biennial Symposium on Communications,* 2006, pp. 364–367.

[9]   L. Lambrinos and C. Djouvas, "Applying scheduling policies to improve QoE in wireless Voice-over-IP," *2009 IEEE 28th International Performance Computing and Communications Conference*, Scottsdale, AZ, 2009, pp. 366-370.

[10] Agayev, B. "Investigation of Affects of Some Parameters of VoIP Networks on the Quality of Transmitted Speech", *2012 IV International Conference "Problems of Cybernetics and Informatics" (PCI)*, September 2012.

[11] Dantas, C. Exton, and A. Le Gear, Improving Mobile VoIP Quality Through Bandwidth Optimisation, [Online]. Available: https://bebusinessed.com/history/voip-history/. [Accessed: 23-August-2018].

[12] Gueham, T.and Merazka, F., "An enhanced insertion packet loss concealment method for voice over IP network services," in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, 2017, pp. 377–382.

[13] C. Olariu, M. Zuber and C. Thorpe, "Delay-based priority queueing for VoIP over Software Defined Networks," *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, Lisbon, 2017, pp. 652-655.

[14] H. A. S. Mekki and Y. A. Mohammed, "The Impact of Coupling Signaling Protocols and Codecs Scheme in Achieving VoIP Quality". *American Scientific Research Journal for Engineering, Technology, and Sciences*. 2017; 32(1): 192-199.

[15] R. J. B. Roslin, O. O. Khalifa and S. S. N. Bhuiyan, "Improved Voice Over Internet Protocol for Wireless Devices," *2018 7th International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, 2018, pp. 498-503.

[16] Hery Dian Septama, Ardian Ulvan, Jiri Hlavacek, Robert Bestak, "Voice Quality of VoIP in High Availability Environment", *Proceeding of the Electrical Engineering, Computer Science and Informatics*, 2014.