Avrupa Bilim ve Teknoloji Dergisi
Sayı 16, S. 638-645, Ağustos 2019

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 16, pp. 638-645, August 2019

Research Article

User-level Performance Evaluation of VoIP under Different Background TCP Traffic Conditions in ns-2

Fatih Abut[1*]

[1]Adana Alparslan Türkeş Science and Technology University, Faculty of Engineering, Department of Computer Engineering, Adana, Turkey
(ORCID: 0000-0001-5876-4116)

(First received 2 July 2019 and in final form 17 July 2019)

(**DOI:** 10.31590/ejosat.585736)

REFERENCE: Abut, F. (2019). User-level Performance Evaluation of VoIP under Different Background TCP Traffic Conditions in ns-2. *European Journal of Science and Technology*, (16), 638-645.

Abstract

Voice over IP (VoIP) is gaining more and more importance and displaces the traditional telephony. For example, more than 300 million monthly active users worldwide use the popular VoIP application "Skype". However, a big problem in the VoIP environment is the voice quality. The purpose of this study is to investigate the effects of background TCP traffic on perceived voice quality of a VoIP conversation at the user-level using the G.711 codec. Two different playout buffering policies including static buffering and optimal buffering have been applied by the VoIP server. For comparison purposes, the same experiments have also been repeated when no playout buffering policy has been used by the VoIP server. A three-hop network topology consisting of a source, a transit, and a destination subnetwork was simulated whereas the end-to-end capacity of the entire network was limited by a 1.5 Mbps Asymmetric Digital Subscriber Line (ADSL) link. Multiple simultaneous TCP connections with different segment sizes were established to simulate the various conditions of background traffic. By using the ns2voip framework, an enhancement to the popular Network Simulator 2 (ns-2), extensive simulation experiments for analyzing the VoIP user-level performance have been carried out and the voice quality has been evaluated by calculating the Mean Opinion Score (MOS). The results show that the voice quality is strongly negatively affected by background TCP traffic, even in the presence of a single TCP flow with 1500 Byte segments. Also, the size of background TCP segments significantly influences the achievable MOSs of VoIP conversations. However, it has also been observed that aggregating multiple speech frames into a single IP packet can increase the MOS. Particularly, depending on the number and segment size of background TCP flows, aggregation of the optimal number of speech frames into the same IP packet improves the MOSs up to 14.61% over a 1.5 Mbps ADSL link.

Keywords: Simulation, ns-2, VoIP, TCP, E-model, Mean Opinion Score.

Ns-2'de Farklı Arka Plan TCP Trafik Koşulları Altında VoIP'nin Kullanıcı Düzeyinde Performans Değerlendirmesi

Öz

IP üzerinden ses (Voice over IP; VoIP), giderek daha fazla önem kazanmakta ve geleneksel telefonun yerini almaktadır. Örneğin, dünya genelinde aylık 300 milyondan fazla aktif kullanıcı, popüler VoIP uygulaması Skype'ı kullanmaktadır. Ancak, VoIP ortamında ses kalitesi büyük bir sorun teşkil etmektedir. Bu çalışmanın amacı, arka plan TCP trafiğinin kullanıcı düzeyinde bir VoIP konuşmasının algılanan ses kalitesi üzerindeki etkilerini G.711 kodeğini kullanarak incelemektir. VoIP sunucusu tarafından statik ve optimal olmak üzere iki farklı tamponlama politikası uygulanmıştır. Karşılaştırma amacıyla, VoIP sunucusu tarafından herhangi bir tamponlama politikası uygulanmadığında da aynı deneyler tekrarlanmıştır. Kaynak, transit ve hedef alt ağlarından oluşan üç sekmeli bir ağ topolojisi

[*] Corresponding Author: Adana Alparslan Türkeş Science and Technology University, Faculty of Engineering, Department of Computer Engineering, Adana, Turkey, ORCID: 0000-0001-5876-4116, fabut@atu.edu.tr

simüle edilirken, tüm ağın uçtan uca kapasitesi 1.5 Mbps Asimetrik Dijital Abone Hattı (ADSL) bağlantısı ile sınırlandırılmıştır. Arka plan trafiğinin değişken koşullarını simüle etmek için farklı segment boyutlarında birden fazla eşzamanlı TCP bağlantısı kurulmuştur. Popüler Network Simulator 2 (ns-2) programına eklenen bir geliştirme olan ns2voip framework'ü kullanılarak VoIP kullanıcı seviyesi performansını analiz etmek için kapsamlı simülasyon deneyleri yapılmıştır ve ses kalitesi Ortalama Görüş Puanı (OGP) hesaplanarak değerlendirilmiştir. Sonuçlar, ses kalitesinin 1500 Bayt segmentli tek bir TCP akışında bile arka plan TCP trafiğinden güçlü bir şekilde olumsuz etkilendiğini göstermektedir. Ayrıca, arka plan TCP trafiğinin boyutu, VoIP konuşmalarının ulaşılabilir OGS'lerini önemli ölçüde etkilemektedir. Bununla birlikte, birden fazla konuşma çerçevelerinin tek bir IP paketi içerisine toplanmasının OGS'yi artırabileceği de gözlenmiştir. Özellikle, arka plan TCP akışlarının sayısına ve segment boyutuna bağlı olarak, en uygun sayıdaki ses çerçevelerinin aynı IP paketinde toplanması, OGS'leri 1.5 Mbps ADSL bağlantısı üzerinden %14.61'e kadar iyileştirdiği görülmüştür.

Anahtar Kelimeler: Simülasyon, ns-2, VoIP, TCP, E-model, Ortalama Görüş Puanı.

1. Introduction

VoIP (Voice over IP) is a modern service with high growth rates. For example, more than 300 million monthly active users worldwide use the popular VoIP application "Skype" (Microsoft, 2016). The basis of a VoIP network is optimal voice quality, low delays, and high reliability. Ensuring these key points requires a thorough analysis and performance test of the underlying network. It is important to use a simulation of network quality assessment before installing VoIP. This pre-measurement may reveal whether an existing network is VoIP-capable. In addition, the VoIP simulation will shed light on which factors influence the voice quality and how it can be optimized.

While in classical network applications the data related parameters such as throughput, packet delays, and losses are evaluated; in the VoIP environment, the voice quality subjectively perceived by a participant is of significant interest. Therefore, a human-judged assessment is needed that indicates how well a person feels the quality of a spoken language. For this reason, the ITU-T developed the so-called Mean Opinion Score (MOS), which determines the voice quality subjectively perceived by the communication participants. MOS can be in a range of 1 (i.e. poor) to 5 (i.e. excellent).

When determining the MOS of VoIP conversations, test persons are given speech samples who evaluate them subjectively. The average of the opinion of all subjects gives the MOS. For a meaningful and reliable measurement of the MOS, the entire acoustic environmental impact, such as the noise and echoes should be included. For such VoIP analyzes special measurement equipment and speech generators, as well as highly sensitive noise sensors on the transmitting and receiving sides are needed. Since such dedicated MOS measurements are quite expensive, the MOS is usually calculated using the so-called E-Model. In contrast to MOS, the speech quality in the E-model is calculated objectively considering all factors affecting the transmission quality, such as delays, packet losses, and jitter. The objective result calculated by the E-model, called the R-factor, can then be mapped to the MOS scale.

Several studies on evaluating the VoIP performance have been conducted in the related literature (Ahmed, 2017; Alharbi, Bahnasse, & Talea, 2017; Audah, Kamal, Abdullah, Hamzah, & Razak, 2015; Cao & Gregory, 2008; Chaudhary & Singh, 2014; Birke, Mellia, Petracca, & Rossi, 2007; Brak, Bouhorma, El Brak, & Bohdhir, 2013; Gurrapu, Mehta, & Panbude, 2016; Li, Chiang, Calderbank, & Diggavi, 2007; Meeran, Annus, & Le Moullec, 2017; Balan, Eggert, Niccolini, & Brunner, 2007; Haibeh, Hakem, & Safia, 2017; Perwej & Parwej, 2012; Tariq, Azad, Beuran, & Shinoda, 2013). Particularly, the studies in (Ahmed, 2017; Alharbi et al., 2017; Audah et al., 2015; Cao & Gregory, 2008; Chaudhary & Singh, 2014) focused on evaluating the performance of VoIP at the IP level, assessing objective parameters such as delays, jitter and packet losses, rather than evaluating the performance of VoIP at user-level which takes into account the user perception, i.e. the MOS. On the other hand, the studies in (Birke et al., 2007; Brak et al., 2013; Gurrapu et al., 2016; Li et al., 2007; Meeran et al., 2017) disregarded receiver's playout buffers which come as part of a VoIP application, and also plays a crucial role. Neglecting the playout buffer is an unrealistic assumption that can cause the voice quality to be significantly overestimated. This is caused by the fact that packets that are successfully delivered within a given deadline at the IP level can also be delayed or dropped at the playout buffer (Bacioccola, Cicconetti, & Stea, 2007). Finally, the studies in (Balan et al., 2007; Haibeh et al., 2017; Perwej & Parwej, 2012; Tariq et al., 2013) take playout buffers into account and use MOS as a metric, but they assess the performance of various playout buffers and codecs, and do not consider the effect of background TCP traffic on voice quality. TCP introduces several factors such as bi-directional data exchange, congestion control and retransmission mechanisms that can strongly interfere with VoIP's performance. Furthermore, all mentioned studies evaluated the traditional case of sending a single speech frame per IP packet, disregarding the aggregation of multiple speech frames into the same IP packet. Thus, it is also to be investigated to which extent an improvement in the MOS for the considered simulation scenario can be achieved by aggregating multiple speech frames into the same IP packet.

The purpose of this study is to investigate the effects of background TCP traffic on perceived voice quality of a simulated VoIP conversation at the user-level using the G.711 codec. Two different playout buffering policies including static buffering and optimal buffering have been applied by the VoIP server. For comparison purposes, the same experiments have also been repeated when no playout buffering policy has been applied by the VoIP server. Furthermore, the effect of aggregating multiple speech frames into a single IP payload on perceived voice quality has also been investigated. A three-hop network topology consisting of a source, a transit, and a destination subnetwork was simulated in Network Simulator 2 (ns-2). The end-to-end capacity of the entire network was configured with 1.5 Mbps, which is the lowest capacity provided by Asymmetric Digital Subscriber Line (ADSL) links. A theoretical model for typical VoIP traffic generated by G.711 was introduced to simulate the conversation behavior of a VoIP user as exactly as possible. By using the ns2voip framework, an enhancement to the popular ns-2, extensive simulation experiments for analyzing the VoIP user-level performance have been carried out. The MOS has been used to evaluate the voice quality of a VoIP conversation.

The rest of the paper is organized as follows. Section 2 introduces the utilized model for typical VoIP traffic generated by G.711 over the TCP/IP stack. Section 3 presents the simulation scenarios and evaluation methodology. Section 4 gives the results and discussion. Finally, Section 5 concludes this paper.

2. Theoretical Model for Generating Typical VoIP Traffic

The activity of a VoIP client is typically modeled by a series of talkspurt and silence periods. The talkspurt period means that the participant speaks and thus generates speech frames. The silence period, on the other hand, indicates a pause in the conversation. To simulate the duration of ON/OFF state transitions for the talkspurt and silence periods, various distributions such as Pareto, Weibull, constant, uniform and exponential can be used.

During the ON states of a VoIP conversation, a realistic workload generator should be designed. For a realistic simulation as well as for its optimal parameterization, one must understand the exact functionality of VoIP to determine how much traffic a VoIP user typically generates. For example, the typical Internet traffic generated by a browser usually has a relatively unpredictable course, and therefore cannot be simulated exactly as in reality. The voice traffic, however, is a continuous data stream as the sampler and the quantizer always provide a speech frame at certain intervals.

To design a theoretical model for VoIP workload generator, it is necessary to understand what steps are taken to convert an analog signal to a digital one. The objective is that a VoIP terminal ultimately generates a digital signal from an analog sound signal. As input, the VoIP terminal receives a sound signal s(t), which is of an analog nature, and provides at the output a signal g(t), which is in binary form.

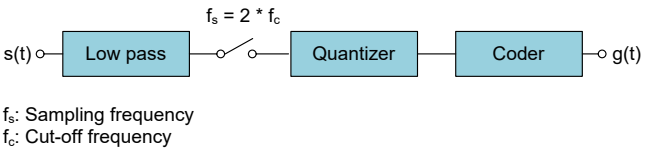

Figure 1. Conversion of an analog signal into a digital signal.

If an analog signal s(t) arrives, first a band limitation takes place. Particularly, the analog signal is narrowed in its frequency spectrum, provided that it can be reconstructed as clearly as possible afterward. This band limitation is caused by an element called a low pass. This maximum band-limited signal, also referred to as the cut-off frequency f_c, is determined by speech intelligibility. The conventional telephony has a cutoff frequency of 4 kHZ.

The steps to be taken to convert this band-limited signal s(t) into a binary stream g(t), as illustrated in Figure 1, can be outlined as follows:

i. **Sampling:** The first step is to convert the signal from a time- and value-continuous form into a time-discrete form. For this purpose, the analog signal is sampled at twice the cutoff frequency f_c.

ii. **Quantization:** Thereafter, the sampled signal must be quantized to convert it from the discrete-time, but value-continuous form into a time- and value-discrete form.

iii. **Coding and compression:** The time- and value-discrete signal must finally be coded and, if necessary, compressed.

First, the sampling rate must be determined. As already mentioned, the cut-off frequency of conventional telephony is 4 kHZ. This results in the sampling rate of 8 kHZ. Also, a certain form of coding should be parametrized. There are quite a few coding schemes to choose from, including the A-law encoding. The A-law encoding used by G.711 generates 8 bits per sample. In the simulation, therefore, both the sampling rate and the coding rate of the data stream to be generated must be parameterized.

This results in a generated continuous data stream of net 64 Kbps. The data transfer, however, takes place packet-oriented. Therefore, the next step is to break this continuous data stream into individual packets and pass them to a specific protocol. For this process, a packetization time is to be considered. The question is under which aspects the packetization time is to be determined. On the one hand, it may be beneficial to maximize the size of a speech frame to make the most efficient use of the underlying network. On the other hand, an end-to-end delay of 150 to 300 ms must be observed. Otherwise, the voice quality will become unacceptable. In the same way, one cannot transmit a speech frame every millisecond, as this would cause considerable overhead. Typical packetization time used by G.711 is 20 ms.

Given the TCP/IP reference model, the overhead caused by the protocols should be considered, too. Particularly, 64 Kbps of raw data is delivered into the TCP/IP protocol stack. A packetization time of 20 ms at a rate of 64 Kbps means an effective packet size of 160 Byte. This packet size runs through the individual layers of the TCP/IP reference model, as shown in Figure 2. Typically, the RTP, UDP and IP protocols are used at the application, transport and network layers, respectively. The next question is how much overhead each of these protocols generates. In the case of an IP packet, it is known that the header without options is always 20 Byte. Similarly, by default, UDP has a header size of 8 Byte and the size of the RTP header is typically 12 Byte. This results in a total header size of 40 bytes. The generator, therefore, must generate a packet of size of 160 Byte + 40 Byte = 200 Byte every 20 ms, simulating the call behavior of a VoIP participant.

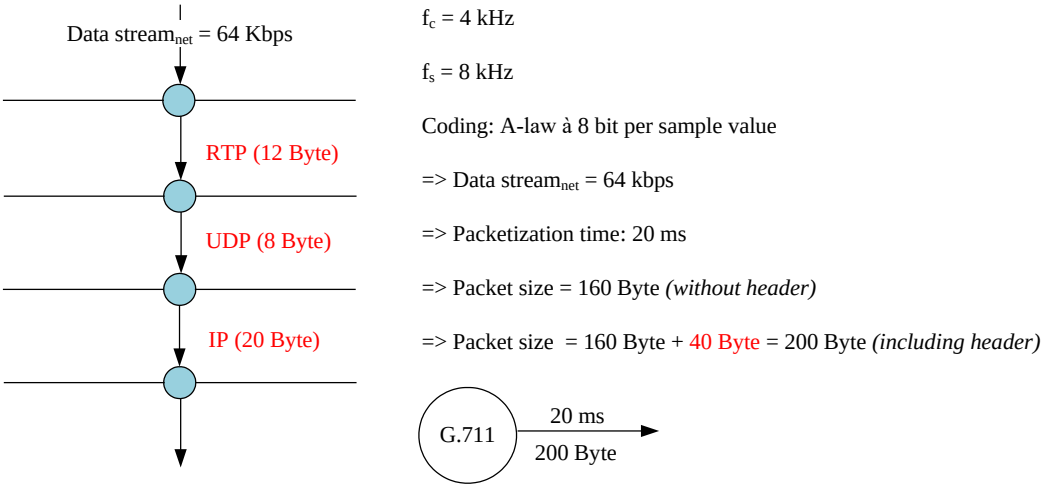

Figure 2. Additional overhead caused by TCP/IP protocol stack regarding VoIP

It is noteworthy that in contrast to other internet traffic generators, the VoIP model has the significant advantage that it is very close to the details of the real telephony and does not make any significant abstraction from reality to achieve a useful workload generator. Furthermore, the overhead of 40 Byte in relation to 160 Byte, which one would only have to transfer in the classical telephony, shows the immense overhead caused by VoIP.

3. Evaluation Methodology

Figure 3 illustrates the simulation scenario of a VoIP conversation between two participants. Particularly, the simulation setup consists of one VoIP client and one VoIP server connected through two routers, the capacities of which are set to 1.5 Mbps forming the bottleneck link of the end-to-end path. The routers were configured with DropTail queue (i.e. First-In First-Out) and the queue size was set to 20 packets. The links of the VoIP client and VoIP server were configured with 100 Mbps to avoid any queuing delays on these links. The latency of all links is set to 2 ms.

The ns2voip framework (Bacioccola et al., 2007) was used to simulate and perform user-level performance analysis of the VoIP conversations. Particularly, by using ns2voip, the client was configured to use the G.711 codec which generates 64 Kbps (160 Byte packet size) data stream according to the generation model described in Section 2. The frame aggregation and compression features of the framework are disabled, if not specified otherwise. For the simulation of the VoIP call, the "one-to-one" model is used, i.e. there is a one-to-one conversation between the client and the server. Only the VoIP traffic generated by the client is simulated, i.e. the generated VoIP traffic always flows from the client to the server and never vice versa.

Multiple simultaneous TCP connections were established to simulate the various conditions of the background TCP traffic. Particularly, four separate simulation scenarios were created. The first scenario investigates the effects of a single TCP transfer during the VoIP conversation. The other three simulation scenarios consider the concurrent establishments of 2, 3 and 4 TCP connections, respectively. Each simulation scenario was configured to produce TCP traffic with fixed segment sizes. Particularly, three segment sizes including 64 Byte, 800 Byte, and 1500 Byte were adapted from the previous VoIP work (Triyason, Kanthamanon, Warasup, Yamsaengsung, & Supattatham, 2010) which represent low, medium and high level segment sizes of TCP, respectively. All TCP nodes are operating an FTP application, and bidirectionally exchange data. With all these scenarios, it is intended to investigate the effects of the network congestion caused by concurrent TCP connections and the resulting queuing over the 1.5 Mbps link, which is the lowest capacity provided by the ADSL technology.

The VoIP server applies two different playout buffering policies, namely static and optimal playout buffering. For comparison purposes, the same experiments have also been repeated when no playout buffering policy has been applied by the VoIP server. Particularly, the no buffering option, as the name suggests, does not use any playout buffering scheme and is merely used to show to which extent the MOSs will be overestimated compared to the case where a more realistic static playout buffering policy is applied. Static buffering delays arriving frames of a selectable but fixed amount of time before passing them to the decoder. The size and the initial delay of the static buffer are set to 20 speech frames and 80 ms, respectively. Finally, the optimal playout buffer accumulates the whole set of arriving frames of a given talkspurt, and then it selects the playback delay so that the best possible voice quality is obtained. Each simulation scenario lasts a total of 600 seconds. In the first 30 seconds (i.e. during the warm-up phase) no measurement results are collected. Table 1 gives an overview of values of the utilized parameters for the simulated VoIP conversation.

The performance of simulation scenarios has been evaluated by calculating the R scores which are finally mapped to MOS by the ns2voip framework. In contrast to the subjectively determined MOS, the E-model is the objective evaluation of the voice quality. The result of the E-model is the R-factor (R: Rating). The calculation of this R-factor requires a whole series of parameters which influence

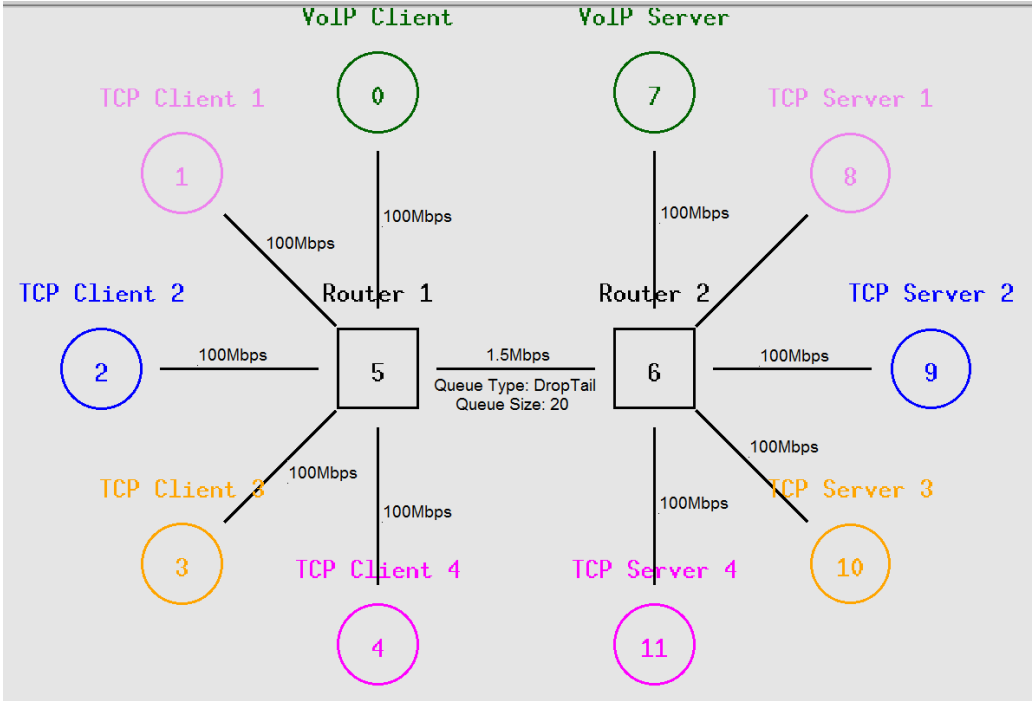

Figure 3. Ns-2 representation of the network topology and the simulated VoIP conversation between two participants

Table 1. List of values of the utilized parameters for the simulated VoIP conversation

Simulation Parameter	Value
Codec used by VoIP client	G.711
VoIP source model	One-to-one
Duration of talkspurt and silent state transitions	Random variable with constant distribution and a mean of 10 ms
Size of static buffer at VoIP server	20 speech frames
Initial delay of static buffer at VoIP server	80 ms
Simulation duration	600 s
Warm-up phase	30 s

the transmission and voice quality. These parameters include, among other things, the signal-to-noise ratio, delays, jitter, echoes and packet losses. The R-factor is calculated using the R-formula in the E-model as

$$R = R_0 - I_s - I_d - I_{e,eff} + A \ , \tag{1}$$

where R_0 is the signal-to-noise ratio, I_s is the interference occurring to the speech signal, I_d indicates the sum of the delays, $I_{e,eff}$ is the noise components due to encoders and packet losses. In addition, the A factor is a purely abstract correction value. It symbolizes the influences of the used communication infrastructure on the R-factor. The effective $I_{e,eff}$ factor is calculated as follows:

$$I_{e,eff} = I_e + (95 + I_e) * \frac{P_{pl}}{\frac{P_{pl}}{P_{pl} + B_{pl}}} \tag{2}$$

In Eq. (2), I_e indicates the influence of the components used in the communication on the quality of the connection. The effective $I_{e,eff}$ value is additionally weighted by the empirical factor $(95 + I_e)$. P_{pl} is the packet loss probability and B_{pl} expresses a value indicating the robustness of the used codec to packet loss (Bacioccola et al., 2007). The default values of the above parameters for the most commonly used codecs can be found in (ITU-T Recommendation, 2001).

The objective result called the R-factor, calculated by the E-model using Eq. (1) can then be mapped to the MOS scale. Table 2 shows the relation between R-factor and MOS in coarse values. A more exact statement about the relationship requires a calculation.

Table 2. Relation between R-Factor and MOS

R-Factor	MOS	Subjective Feeling
100	5.0	Excellent
90	4.3	Very satisfied
80	4.0	Satisfied
70	3.6	Few are dissatisfied
60	3.1	Many are dissatisfied
50	2.6	Almost all are dissatisfied
0	1.0	All are dissatisfied

4. Results and Discussion

The overall simulation experiments have been conducted in four parts. In the first part of experiments, plausibility tests were conducted. To this end, no background TCP traffic was generated on the path during the VoIP conversation to simulate the ideal condition. As expected, due to the lack of any other cross-traffic, the end-to-end delay only included propagation and transmission delays, and no queueing delays. Propagation delays are 2 ms for three links (i.e. 6 ms total) and the transmission delay ranges from 0.34 ms to 8 ms which can be calculated by dividing packet length (i.e. 64 Byte, 160 Byte, 1500 Byte) to link capacity (i.e. 1.5 Mbps). Under these ideal conditions, the achieved MOSs using no, static and optimal playout buffering policies are 4.46, 4.48 and 4.48, respectively. It is noteworthy that the theoretical MOS of 5.0 cannot be achieved in practice. The lossy conversion of the analog voice signals into digital signals and vice versa reduces the maximum achievable MOS to around 4.5.

In the second part of experiments, the performance of the three playout buffering policies has been compared in terms of achieved MOS. Table 3 through Table 5 show the MOSs for different TCP-based cross-traffic scenarios using no, static and optimal playout buffering policies applied by the VoIP server. The no playout buffering policy is merely used to demonstrate to which extent the MOSs are overestimated in comparison to the case where the static playout buffering policy is applied. Similarly, the optimal playout buffering policy acts as a reliable method for determining the reference values (i.e. highest MOSs) obtainable using an ideal playout buffering policy. The results show that compared to static and optimal playout buffering policies, the overestimation errors of no playout buffering policy range from 0.91% to 17.42%, and from 0.23% and 3.37%, respectively. This result, in turn, signifies the need for the consideration of playout buffering policies in simulative performance evaluations of VoIP conversations.

Table 3. Resulted MOS values when establishing 1, 2, 3 and 4 TCP connections during VoIP conversation using no playout buffering policy

Segment Size (Byte)	Number of TCP Flows			
	1	*2*	*3*	*4*
64	4.43	4.27	4.18	4.05
800	4.28	3.96	3.54	3.15
1500	3.29	2.76	2.41	2.09

Table 4. Resulted MOS values when establishing 1, 2, 3 and 4 TCP connections during VoIP conversation using static playout buffering policy

Segment Size (Byte)	Number of TCP Flows			
	1	*2*	*3*	*4*
64	4.39	4.20	4.11	4.00
800	4.22	3.85	3.42	3.02
1500	3.13	2.49	2.10	1.78

Table 5. Resulted MOS values when establishing 1, 2, 3 and 4 TCP connections during VoIP conversation using optimal playout buffering policy

Segment Size (Byte)	Number of TCP Flows			
	1	*2*	*3*	*4*
64	4.42	4.26	4.16	4.08
800	4.26	3.93	3.50	3.12
1500	3.23	2.69	2.34	2.03

In the third part of experiments, the effects of the number of background TCP flows and their different segment sizes on the achieved MOSs were investigated. According to the results shown in Table 3 through Table 5, it is seen that increasing the size of TCP's segments significantly leads to a parallel decrease in the MOSs. More specifically, as the background segment size gets bigger, speech frames

suffer from longer queuing delays due to the long waiting times at buffers. When TCP segments of size 64 Byte are generated, a satisfactory user-level performance of VoIP can be experienced at least up to 4 concurrently established TCP connections without major problems, independent of whether no, static or optimal playout buffering policy is applied by the VoIP server. In the case of applying static playout buffer policy, the MOSs range from 4.00 to 4.39, whereas in the case of applying optimal playout buffering policy, the MOSs change between 4.08 and 4.42. In contrast, when TCP segments of size 800 Byte are generated, a satisfactory user-level performance of VoIP can be experienced only up to two concurrently established TCP connections for static and optimal playout buffering policies. In these cases, the applications of static and optimal playout buffering policies yield MOSs ranging from 4.22 to 3.85, and from 4.26 and 3.93, respectively. Finally, when TCP segments of size 1500 Byte are generated, the voice quality is strongly influenced even in the presence of single TCP connection, lowering the MOSs up to 1.78 and 2.03 for static and optimal playout buffering policies, respectively. The negative effects of TCP connections on voice quality are caused by the fact that during the slow start and additive increase phases, TCP sends a continually increasing number of segments in batches, filling the router's queue, and causing unacceptable longer delays and even losses of speech frames.

Table 6. Resulted MOS values when establishing 1, 2, 3 and 4 TCP connections during VoIP conversation with frame aggregation

Segment Size (Byte)	**Aggregation Level**	**Number of TCP Flows**			
		1	*2*	*3*	*4*
64	3	4.39	4.42	4.37	4.31
800	4	4.28	3.97	3.73	3.34
1500	8	2.89	2.48	2.14	2.04

Finally, in the fourth part of experiments, the effect of aggregating multiple speech frames into a single IP payload has been investigated. After extensive simulation experiments, it is seen that aggregating speech frames up to a certain extent entails a performance gain in increasing the MOS, as also mentioned in (Bacioccola et al., 2007). Particularly, the number of speech frames per IP payload was varied from 1 to 10, and the number (i.e. the aggregation level) leading to the highest MOS for each segment size on the considered network topology has been determined. It is observed that depending on the segment size of the TCP cross-traffic, the aggregation level leading to the highest MOS also varies. In more detail, when during the VoIP conversation TCP segments of sizes 64 Byte, 800 Byte, and 1500 Byte are generated, the optimum aggregation levels have been revealed as 3, 4 and 8, respectively. Table 6 shows the MOS values for different TCP-based cross-traffic scenarios when the aggregation level is set to 3, 4 and 8 for 64 Byte, 800 Byte and 1500 Byte segment sizes, respectively. According to these results, it is seen that the percentage increase rates in MOS obtained by using the static playout buffering policy with frame aggregation compared to the ones obtained by using the static playout buffering policy without frame aggregation for different experiment scenarios range from 1.42% to 14.61%. This result can be interpreted as follows: On one side, the aggregation of the frames on the client causes an additional delay. On the other side, all frames aggregated into a payload suffer from the same network delay, i.e. the least possible delay variation. Also, the number of frame reorderings at receiver's playout buffer is minimized when multiple speech frames are aggregated. These two points bring a gain in terms of the MOS, which offsets the additional delay on the transmitter side.

5. Conclusion

The purpose of this study was to investigate the effects of TCP background traffic on perceived voice quality of a VoIP conversation at the user-level using the G.711 codec. Two different playout buffering policies including static buffering and optimal buffering have been applied by the VoIP server. For comparison purposes, the same experiments have also been repeated when no playout buffering policy has been used by the VoIP server. First, a theoretical model for typical VoIP traffic generated by the G.711 codec was introduced to simulate the conversation behavior of a VoIP user as exactly as possible. By using the ns2voip framework, the described G.711 traffic generation model was applied, and extensive simulation experiments related to the user-level performance analysis of the simulated VoIP conversation on a three-hop network topology were conducted.

The simulation results show that the voice quality is strongly negatively affected by background TCP traffic, even in the presence of a single TCP flow with 1500 Byte segments. In case of concurrently established TCP connections, the voice quality reaches unacceptable MOSs, even with smaller background segment sizes. Also, it is observed that the size of TCP segments strongly influences the MOSs during the VoIP conversation. Particularly, increasing the segment size of the background TCP traffic leads to a parallel decrease in the MOSs. However, it has also been observed that aggregating multiple speech frames into a single IP payload can significantly increase the MOS. Depending on the number and segment size of background TCP flows, aggregation of the optimal number of speech frames into the same IP packet improves the MOSs up to 14.61% over a 1.5 Mbps ADSL link.

References

Ahmed, A. (2017). Performance Analysis of VoIP in WiFi Campus Network. *International Journal of Computer Applications*, *174*(3), 9–13. https://doi.org/10.5120/ijca2017915339

Alharbi, A., Bahnasse, A., & Talea, M. (2017). A Comparison of VoIP Performance Evaluation on different environments Over VPN Multipoint Network. *International Journal of Computer Science and Network Security*, *17*(4), 123–128.

Audah, L. M., Kamal, A. M., Abdullah, J., Hamzah, S. A., & Razak, M. H. S. A. (2015). Performance evaluation of voice over IP using multiple audio codec schemes. *ARPN Journal of Engineering and Applied Sciences*, *10*(19), 8912–8919.

Bacioccola, A., Cicconetti, C., & Stea, G. (2007). User-level performance evaluation of VoIP using ns-2. In *2nd International Conference on Performance Evaluation Methodologies and Tools*. Nantes, France. https://doi.org/10.4108/nstools.2007.2014

Balan, H. V., Eggert, L., Niccolini, S., & Brunner, M. (2007). An Experimental Evaluation of Voice Quality Over the Datagram Congestion Control Protocol. In *26th IEEE International Conference on Computer Communications* (pp. 2009–2017). IEEE. https://doi.org/10.1109/INFCOM.2007.233

Birke, R., Mellia, M., Petracca, M., & Rossi, D. (2007). Understanding VoIP from Backbone Measurements. In *IEEE International Conference on Computer Communications* (pp. 2027–2035). https://doi.org/10.1109/-INFCOM.2007.235

Brak, S. El, Bouhorma, M., El Brak, M., & Bohdhir, A. (2013). Speech Quality Evaluation Based Codec for VoIP Over 802.11P. *International Journal of Wireless & Mobile Networks*, *5*(2), 59–69. https://doi.org/10.5121/ijwmn.2013.5205

Cao, J., & Gregory, M. (2008). Performance Evaluation of VoIP Services using Different CODECs over a UMTS Network. In *2008 Australasian Telecommunication Networks and Applications Conference* (pp. 67–71). https://doi.org/10.1109/-ATNAC.2008.4783297

Chaudhary, D. A., & Singh, D. S. P. (2014). Performance Evaluation of VoIP in MPLS network using NS-2. *International Journal of Computers & Technology*, *13*(9), 4792–4798. https://doi.org/10.24297/ijct.v13i9.2355

Gurrapu, S., Mehta, S., & Panbude, S. (2016). Comparative Study for Performance Analysis of VOIP Codecs Over WLAN in Nonmobility Scenarios. *International Journal of Information Technology, Modeling and Computing*, *4*(3). https://doi.org/-10.2139/ssrn.3389769

Haibeh, L. A., Hakem, N., & Safia, O. A. (2017). Performance evaluation of VoIP calls over MANET for different voice codecs. In *IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 1–6). IEEE. https://doi.org/10.1109/-CCWC.2017.7868479

ITU-T Recommendation. (2001). G.113 Transmission impairments due to speech processing. Retrieved from https://www.itu.int/rec/T-REC-G.113

Li, Y., Chiang, M., Calderbank, A. R., & Diggavi, S. N. (2007). Optimal Rate-Reliability-Delay Tradeoff in Networks with Composite Links. In *26th IEEE International Conference on Computer Communications* (pp. 526–534). https://doi.org/10.1109/-INFCOM.2007.68

Meeran, M. T., Annus, P., & Le Moullec, Y. (2017). Approaches for improving VoIP QoS in WMNs. In *International Conference on Electrical Engineering and Computer Science (ICECOS)* (pp. 22–27). IEEE. https://doi.org/10.1109/ICECOS.2017.8167138

Microsoft. (2016). Skype has over 300 million monthly active users. Retrieved July 1, 2019, from https://windowsreport.com/skype-number-of-users/

Perwej, Y., & Parwej, F. (2012). Perceptual Evaluation Of Playout Buffer Algorithm For Enhancing Perceived Quality Of Voice Transmission Over Ip Network. *International Journal of Mobile Network Communications & Telematics*, *2*(2), 1–19. https://doi.org/10.5121/ijmnct.2012.2201

Tariq, M. I., Azad, M. A., Beuran, R., & Shinoda, Y. (2013). Performance Analysis of VoIP Codecs over BE WiMAX Network. *International Journal of Computer and Electrical Engineering*, 345–349. https://doi.org/10.7763/IJCEE.2013.V5.729

Triyason, T., Kanthamanon, P., Warasup, K., Yamsaengsung, S., & Supattatham, M. (2010). The Effect of Background Traffic Packet Size to VoIP Speech Quality (pp. 175–182). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-16699-0_19