Payload size-varying VoIP Quality of Service

Y.C. Jung and H.J. Kim
School of Information, Communications
and Electronic Engineering
The Catholic University of Korea
43-1 Yokkok Wonmi-gu, Puchon City,
Kyonggi-do 420-743, Korea
Telephone: +82-2-2164-4364
Fax: +82-2-2164-4777
Email: ycjung@catholic.ac.kr
and liasgard@naver.com

J. William Atwood
Department of Computer Science
and Software Engineering
Concordia University
1455 de Maisonneuve Blvd. West
Montreal, Quebec Canada H3G 1M8
Telephone: +1-514-848-2424 ext 3046
Fax: +1-514-848-2830
Email: bill@cse.concordia.ca

***Abstract*—We present measurements of the relationship between VoIP QoS and payload sizes. Network delays, variation in delay (jitter) and lateness loss are measured using a new one-way measurement method. Existing works show that in order to minimize latency in the voice packet, the payload size of the VoIP packet needs to be below 160 bytes. However, based on our experimental results, we propose that the size of VoIP packet needs to be as large as possible as long as its end-to-end delay and lateness do not degrade the perceived quality of voice. According to our measurements, 640 bytes of the payload size shows similar Quality of Service performance as 160 bytes of it can provide.**

***Index Terms*—Audio; Quality of Service; Measurement**

I. Introduction

Voice over Internet Protocol (VoIP) is a technology that allows voice communication using Internet connections. VoIP is one of fastest growing technologies in the world at the moment [1]. When a human voice is active, the resulting speech signal alternates between active periods, when sound energy is produced, and silent periods. Each active period is called a *talk spurt*. The speech signal is encoded, using one of many possible encodings, and generates a stream of packets.

Figure 1 shows a typical time sequence of packet generation and packet arrival, for fixed packet size. Because network congestion impedes the flow of packets, the packet arrival staircase is irregular, i.e., the delay between generation and arrival is not uniform. The two properties of the arrival sequence that are important are the one-way delay, and the delay variation (jitter) [2].

To compensate for the delay variation, artificial delay (buffering) is introduced in the receiver, and many algorithms have been proposed for different *playout schedules* [3], [4], [5], such as the one illustrated by the dotted line beginning at t_1 in Figure 1. The goal of these schedules is to try to balance the opposing goals of minimizing buffer delay and minimizing lateness loss (a packet that arrives after its scheduled playout time is discarded, i.e., it is treated as if it had never arrived).

Within each talk spurt, because humans react poorly when the round-trip delay between finishing a talk spurt and hearing the response is greater than about 300 ms., it is important to keep the overall one-way delay below 150 ms. This leads to the idea of a *delay budget*, with specific amounts of delay allocated to the encoding delay, the network delay, the jitter-removal buffer, and the decoding delay.

When routing capacities and transmission speeds were (relatively) small, it was important to minimize the encoding (and decoding) delay. This has led to many algorithms for encoding and for jitter buffering where minimal delay is the overriding goal [6], [7].

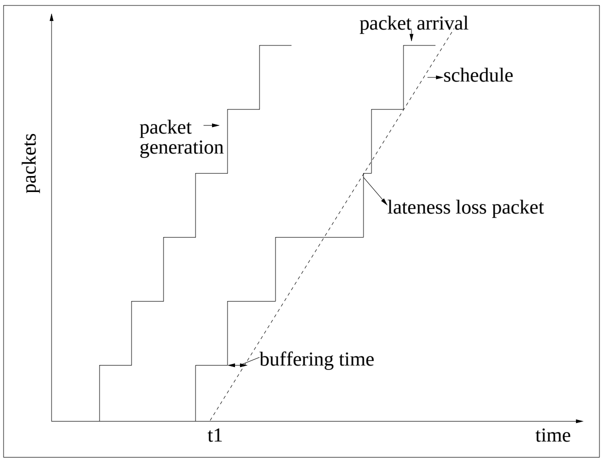

Fig. 1. Divergent goals of minimizing buffer time and minimizing late packet loss

The payload size of the voice packet is 160 bytes and 20 bytes in G.711 and G.729, respectively [8], [9]. These CODECs cover the constant bit rate (CBR) traffic streams. From the delay budget viewpoint, the voice CODEC's latency is assigned 20 ms. or 2.5 ms., respectively. This use of small payload size is based on the idea that delay for the CODEC encoding should be minimized to reduce the whole end-to-end delay. This allows more allocation from the delay budget for queuing delays in the intermediate routers. However, this approach ignores the fact that voice packets of smaller payload size can cause more network congestion. Moreover, the high bandwidth of home broadband and core networks provides more than enough bandwidth for a high-quality phone call.

In this paper, we measured the VoIP performance as a function of different payload sizes. Our goal in the measurements is to show that payload size in voice packets has a minimal effect on VoIP quality. This approach is exactly the opposite of existing works, where any latency in voice CODECs is considered an important reason to degrade VoIP quality. Our proposal is to use voice packets with a large payload size, thus decreasing network traffic while still maintaining similar VoIP quality to what can be achieved with VoIP packets with a small payload size.

In the packetized continuous voice stream, voice packets are continuously generated at the source and are generally injected at the constant intervals into the network. The packets then travel independently through the network, and each packet is delayed by a different amount of time. These statistical variations in delay are affected by network parameters. The total delay of the streaming voice packets consists of four elements. First, there is the overhead delay associated with the packetizing process at the transmitting end. The second element is associated with the nodal delays (the delay from reception of the packet at a certain node to passing the packet toward its next node), and the third element is the packet transmission time on an access line. The last element is propagation delay. Total packet delay is variable as shown in Figure 2. Also there is additional buffering delay. Buffering is necessary for the destination node to pass the packets to the destination process at the constant rate and to restore the constant intervals between the packets. Buffering effectively reduces the delay jitter. However, voice conversation cannot be delayed for more than few hundred milliseconds without having a disturbing effect, so there is a limit to how long the packets can be delayed [10]. Usually, buffering delay is considered at the destination for the playout algorithm.

Existing works focus on minimizing the total end-to-end delay. Their approaches result in the conclusion that the size of a voice packet must be minimized. On the other hand, our work will show the reverse idea: the size of a voice packet needs to be maximized to the upper bound of the delay budget is not exceeded. Our measurement experiments will prove our idea.

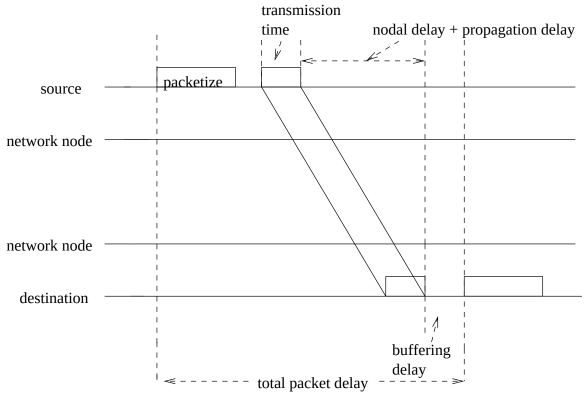

Fig. 2. Packet delay in network node

II. VoIP Performance Variations

This paper explores time-varying network delay characteristics for VoIP streams over a typical intercontinental path, the path from Korea to Eastern Canada. We analyze different sizes of the voice payload per RTP (Real Time Protocol) packet with respect to their effect on the VoIP performance. All the reports in this paper are based on experimental results measured using the end-to-end one-way measurements method.

A. Time-varying Effects

Knowledge of the time-varying characteristics of the VoIP performance is important because some solutions to compensate for the variance delays and jitter can be achieved only after the acquisition of the VoIP behavior data from the present-day Internet. As long as the Internet congestion varies continuously, network behavior fluctuates at all times. Moreover, VoIP traffic behavior is difficult to handle using the analytic method because the Internet is evolving as a composition of independently developed and deployed protocols, technologies and aggregated core applications. Therefore, a technology is required to develop advanced measurement methods in the new applications such as streaming media. This is why we use the end-to-end one-way measurements method.

B. Voice Payloads

A 64 kb/s, G.711 encoder produces a voice payload of 160 bytes every 20 *ms*. This is subsequently placed in an RTP packet, with 40 bytes of overhead, as shown in Figure 3.

As the compression technology continues to evolve, the size of a voice payload tends to become smaller. However, as requirements for secured traffic (IPsec) and/or Virtual Private Networks have been imposed, the additional headers have tended to increase the amount of

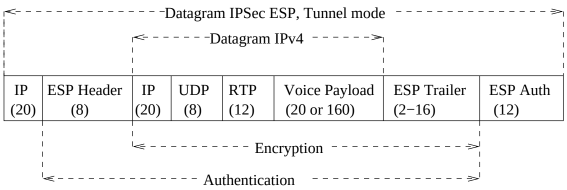

Fig. 3. IPSec ESP tunnel datagram and header size(in bytes)

overhead. Figure 3 shows the encapsulation of a voice packet by IPSec ESP in tunnel mode.

IPSec provides data confidentiality and integrity by using an Authentication Header and an Encapsulating Security Payload. The ratio of the overhead size over the IPSec packet size is between 35% for a 160 byte (64kb/s) G.711 voice packet and 81% for a 20 byte (8kb/s) G.729 voice packet. As this happens, voice packets will have negative influences on end-to-end latency and lateness loss at playout times. Obviously, with the conventional point of view, the overhead ratio becomes important with IPSec when voice packet sizes are limited and these added fields degrade the QoS performance of the VoIP. However, in this paper, we explain what happens, if a payload size of more than 160 bytes is used.

III. Measurements

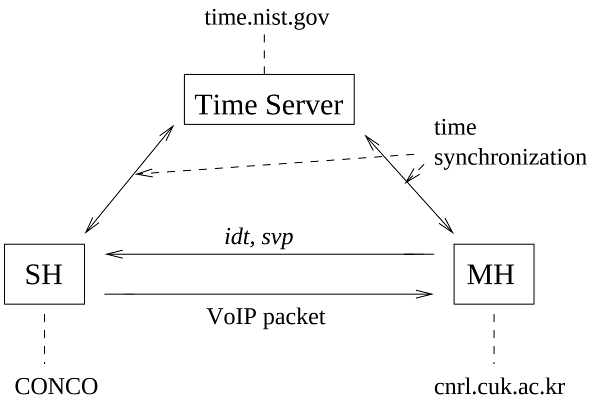

Fig. 4. Time synchronization and delay measurement

So far, there exist two categories of internet measurements: passive and active measurements [11]. The end-to-end one-way measurements method we proposed falls into a new category. The end-to-end one-way measurements method basically needs to maintain logical time synchronization between the measuring host (MH) and the sender host (SH). They are synchronized to the third time server, which corresponds to a class 1 time server. The synchronization algorithm in the end-to-end one-way measurements method is the Network Time Protocol (NTP) described in [12]. Every *N* seconds, hosts equipped with the NTP daemon send messages to the time server asking for the current time, and they adjust their local time based on the time information the time server provides. Under the condition that $N = 64$ *sec*, it took about an hour for both clocks to enter stable states in logical synchronization, which for our purposes was precise enough to begin measurements. At the measurement phase, the real-time streaming protocol (RTSP) compliant audio software we developed begins to run at each side. At first, the MH sends commands containing two parameters: *idt* and *svp* to the SH, where *idt* and *svp* are defined as packet inter-departure time in *ms* and the size of a voice payload in bytes, respectively. Then, the SH generates corresponding VoIP packets, which contain timestamps and sequence numbers. The timestamp values correspond to their departure time from the SH. When a packet arrives at the MH, its network delay is computed by a way of subtracting its timestamp value from the arrival time. By examining its sequence number, the MH monitors whether there is a lost packet or not. The hosts chosen were The Catholic University of Korea (cnrl.cuk.ac.kr) serving as MH, Concordia University (fir.cs.concordia.ca) as SH, and the National Institute of Standards and Technology (time.nist.gov) as the third time server.

IV. Results

Starting at 2:00 PM, Mar 22, 2006, we successfully obtained $5000 * 5$ measurement data for *svp* = 20, 320, 640, 960 and 1280. We repeated the $5000 * 5$ measurement 30 times for each payload size. As a result, we obtained $5000 * 30$ measurement data for each size of payload. We calculated the average variance in delay for 5000 measurements. Figure 5 shows the number of measurement data for a certain level of average variance in all measured delay values. Then we divided the measurements data into five groups. Measurement data whose average variance values are between 0 and 1999, between 2000 and 3999, between 4000 and 5999, between 6000 and 7999, between 8000 and 10000 are considered as group-1, group-2, group-3, group-4 and group-5, respectively. Here, group-1 corresponds to the measurements data that shows the lowest jitter condition and group-5 corresponds to the worst jitter condition.

Figure 6 shows the lateness loss ratio for group-1 measured data as a function of *tloss*, for five different payload sizes. *tloss* is defined as the average additional buffering (jitter removal) delay (in *ms*) in the receiving host. In our lateness loss calculation, the packets that have longer delay values than [the value of average delay in a certain group + *tloss*], are considered to be lost.

Figure 7, Figure 8, Figure 9 and Figure 10 show lateness loss performances for group-2, group-3, group-4 and group-5 data, respectively.

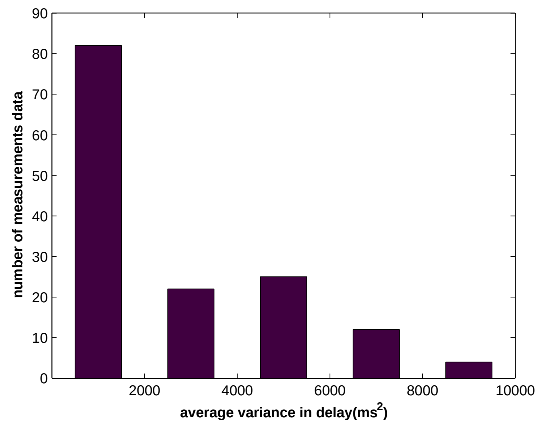

Fig. 5. Number of measurements data

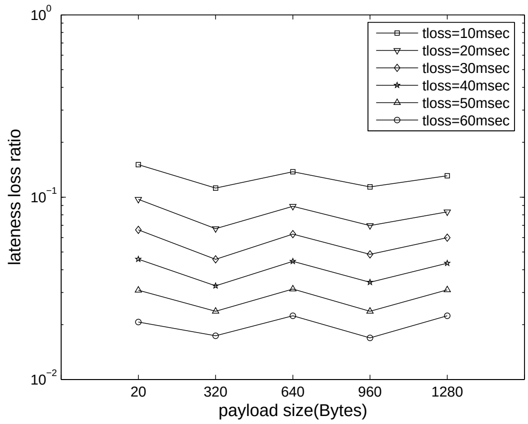

Fig. 6. Lateness loss for group-1 measured data

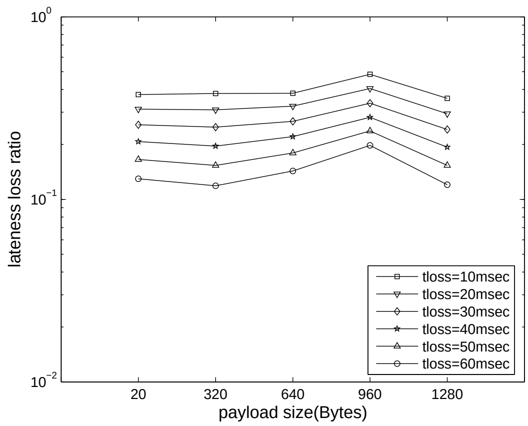

Fig. 7. Lateness loss for group-2 measured data

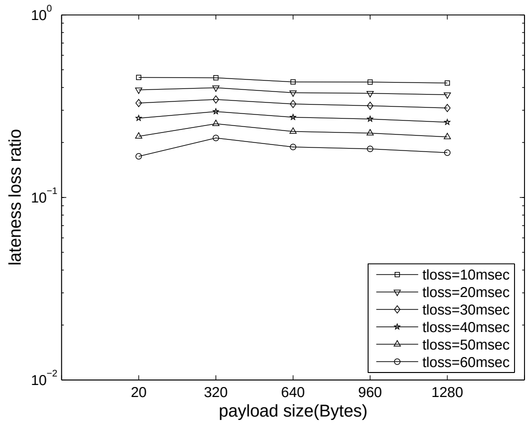

Fig. 8. Lateness loss for group-3 measured data

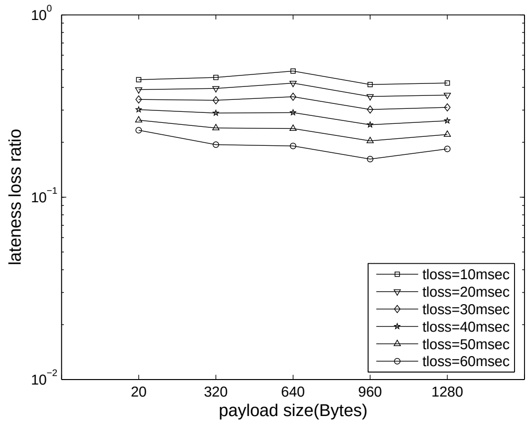

Fig. 9. Lateness loss for group-4 measured data

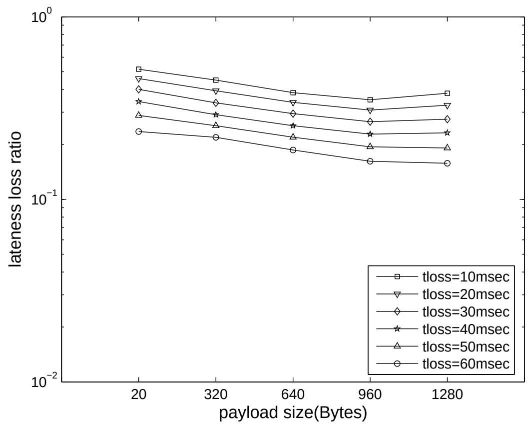

Fig. 10. Lateness loss for group-5 measured data

V. Discussion

As can be seen in figure 6, the lateness loss for the low-delay-variation case is strongly influenced by the amount of jitter buffering allocated. However, the effect of payload size is relatively minor, indicating that the benefits of larger packets (fewer packets in the network, less congestion) are not negated when larger payload sizes are used.

Examining Figures 7 through 10, the lateness loss percentage increases with increased delay variation (as would be expected), but the effect of increasing the payload size remains small. The change in VoIP performance is not significant enough to affect the Quality of Service.

That is, as *svp* varies from 20 to 1280, the VoIP performance does not depend on the payload size. Figure 6, Figure 7, Figure 8, Figure 9 and Figure 10 show exactly the same pattern relating to payload size-varying VoIP performance. (However, it is worth noting that the sample size in Figure 10 is too small to permit the results to be considered as anything other than a directional indication.)

VI. Conclusion

The conventional approaches to increase VoIP QoS are based on the idea that the delay for the CODEC encoding should be minimized to reduce whole end-to-end delay. In this paper, we proposed the reverse idea that the payload size need not be limited to the level of 160 bytes, which corresponds to the encoding interval of 20 *ms*. According to our measurements experiment, up to 640 bytes of the payload size shows similar VoIP QoS as 160 bytes of it can provide. The CODEC encoding interval of 80 *ms* is tolerable for real voice streams under current Internet conditions. An implementation could easily measure the overall round-trip time, by putting some form of measurement packet in the data stream. It could then increase the size of the transmitted packet (by placing multiple 20-*ms* units in a single outgoing packet), up to the limit of the agreed-to delay budget. If the actual propagation time is long, fewer speech units would be put into a packet. We have also argued that a VoIP packet containing more payload at a time, will result in a smaller number of packet generations, thus reducing network congestion. As an extension of this paper, we did a Listening test to find how the lateness loss ratio affects speech quality. The speech material for the test was generated by superimposing the lateness loss patterns on the encoded speech material. We are preparing another paper which shows loss percent vs MOS (Mean Opinion Score).

Acknowledgements

This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2006-(C1090-0602-0011)), and by the Natural Sciences and Engineering Research Council of Canada, under its Discovery Grants program.

References

[1] Cherry, S.: Seven Myths about Voice over IP. Spectrum. IEEE. **42** (2005) 52–57

[2] ITU-T Recommendation H.323.: Packet-Based Multimedia Communications Systems.

[3] Moon, S.B., Kurose, J. and Towsley, D.: Packet Audio Playout Delay Adjustment: Performance Bounds and Algorithms. Multimedia Systems. **6** (1998) 17–28

[4] Aman Kansal and Abhay Karandikar: Adaptive delay estimation for low jitter audio over Internet. Global Telecomunications Conference. GLOBECOM'01. IEEE. **4** (2001) 2591–2595

[5] Miroslaw Varbutt and Liam Murphy: Improving voice over IP subjective call quality. IEEE Communications Letters. **8** (2004) 308–310

[6] Y.C.Jung and J.William Atwood: Dynamic adaptive playout algorithm using interarrival jitter and dual use of α. IEE Proceeding on Communications. **153** (2006) 279–287

[7] Y.C.Jung and J.William Atwood: Switching between fixed and adaptive playout: A per-call playout algorithm. IEEE Internet Computing. **9** (2005) 22-27

[8] Pulse Code Modulation (PCM) of Voice Frequencies. ITU-T Recommendation G.711. (1988)

[9] Coding of Speech at 8kbit/s Using Conjugate-Structure Algebraic-code-excited Linear-prediction. ITU-T Recommendation G.729. (1996)

[10] Haas, Z., Gitlin. R.D.: On the packet size in integrated network. INFOCOM '91. Proceedings. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking in the 90s. IEEE. **2** (1991) 732–740

[11] K.C.Claffy: Measuring the Internet. Internet Computing. **4** (2000) 73–75

[12] Network Time Protocol (version 3) Specification. Implementation and Analysis. Network Working group. RFC 1305. (1992)