Study on Appropriate Voice Data Length of IP Packets for VoIP Network Adjustment

Hiroyuki OOUCHI, Tsuyoshi TAKENAGA[1], Hajime SUGAWARA, and Masao MASUGI
NTT Network Service Systems Laboratories
9-11, Midori-Cho 3-Chome, Musashino-Shi, Tokyo 180-8585, Japan
[1] Nippon Telegraph and Telephone West Corp. 2-27, Nakanoshima 6-Chome, Kita-Ku, Osaka-Shi, Osaka 540-8511, Japan

***Abstract*–This paper evaluates suitable voice-data length in IP packets for the adjustment of VoIP network systems. Based on measurements in a real environment, we examined the voice-quality level while varying the voice-data length of IP packets under various network conditions. We found that a VoIP system with long-voice data has high-transmission efficiency but there is a high deterioration in the voice-quality level in an inferior network. We also discovered that a VoIP system with short-voice data is tolerant to packet losses and preserves voice quality. Based on these results, we propose a VoIP system that sets the voice-data length of IP packets according to dynamically changing network QoS conditions to achieve both high-transmission efficiency and stable voice quality.**

I. INTRODUCTION

The range of computer network technology has expanded in size and diversity, and a wide variety of value-added applications for use on the Internet have appeared in recent years. Networks (e.g. the Internet) have become broadband, and multimedia communication environments where data, voice and images can be exchanged have been rapidly improving. IETF is studying communication services that connect existing telephone networks with IP networks, and ITU-T is deciding on a VoIP (Voice over Internet Protocol) protocol and studying the technologies for communication services where INs (Intelligent Networks) can collaborate with IP networks. There has been remarkable progress in the IP-based technology for computer-telephony integration, and this has resulted in lowering communication costs. In particular, packet voice applications such as Netmeeting, CuSeeMe and other conferencing multimedia applications have become widespread.

Due to the shared nature of current network structures, however, guaranteeing the quality of service (QoS) of Internet applications from end-to-end is sometimes difficult. Because voice transmission on the Internet is unreliable, the current best-effort technology cannot guarantee the QoS or reliability of VoIP services. Various studies on VoIP technology have tried to establish reliable services and evaluate QoS properties [1-4]. However, the QoS level of VoIP systems depends on many parameters including the end-to-end delay, jitter, packet loss in the network, type of codec used, length of voice data, and the size of the jitter-absorbing buffer [5], so that further investigations are needed to clarify the factors affecting QoS. For instance, it has been pointed out that the transmission efficiency of voice data carried by IP packets is not sufficient because of the high ratio of header length to voice-data length in VoIP packets. Hence, to provide efficient and reliable VoIP services, it is important to clarify the effect that changing the system's voice-data transmission rate has, among others. Also, there are some other issues such as less degraded voice quality due to VoIP packet losses, decreased number of UDP (User Datagram Protocol) connections for a voice gateway and decreased processing load for routing.

The transmission efficiency of VoIP can be enhanced by omitting redundancy in the IP/UDP/RTP (Real-time Transport Protocol) header information, compressing it by not resending header information that does not change after call connection setup, and multiplexing the voice data of two or more call channels in one IP packet with a sub-header identifying call channels. In ITU-T and IETF, IP-based multiplexing methods have been studied to enhance the transmission efficiency of VoIP [6-9].

We evaluated a VoIP system with the intention of designing optimal network services for various network conditions. Based on measurements in a real network environment, we examined the voice-quality level (PSQM: Perceptual Speech Quality Measure [10]) while changing the voice-data length of IP packets for various packet loss and jitter conditions. The above measure is widely used, and it provides an objective quality level for the voice over existing telephone voice bandwidths (300 Hz-3.4 kHz), and it is recognized as an appropriate method with relatively little error to determine the subjective quality level (MOS: Mean Opinion Score [11, 12]). The smaller the PSQM value, the better the voice quality. Using the results obtained from our experiment, we created a new VoIP mechanism that enhances transmission efficiency and provides stable voice quality. It sets the appropriate voice-data length for IP packets based on the dynamically changing network QoS conditions.

II. TRANSMISSION INEFFICIENCY OF VoIP

When packets transmit an analog voice signal, the VoIP-GW (VoIP-Gateway) digitizes the voice signal through a codec, such as G.711 (64 kbps) and G.729 (8 kbps), and it transmits voice data of a fixed length, where UDP and RTP are used to transmit voice data in real time. The structure of an IP packet over an Ethernet is shown in Fig. 1. Here the voice data length can be changed according to the VoIP transmission efficiency.

The MAC (Media Access Control) header, IP header, UDP header, RTP header and FCS (Frame Check Sequence) are necessary for transmitting voice data over the Ethernet, and preamble and IPG (Inter Packet Gap) should be considered as occupying bandwidth on the transmission line. For instance, the total occupied bandwidth is 98 bytes including IPG, preamble, MAC header, IP header, UDP header, RTP header and FCS when transmitting 20-byte voice data. The 78 bytes thus correspond to the overhead of IP transmission, so the ratio of voice data to the total is less than 25%.

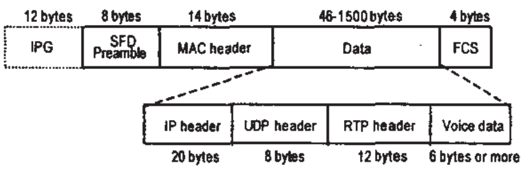


Fig. 1. VoIP packet structure for Ethernet.

The voice-data length of an IP packet usually depends on the method used for the VoIP-GW. Eighty-byte voice data is often used for G.711, whereas 20-byte voice data is used for G.729 in conventional VoIP communication. However, it is also permissible to change the voice-data length per packet by changing the VoIP-GW set up.

Table I shows the relationship between the packet transmission cycle and voice-data length, and Fig. 2 shows the relationship between the packet transmission cycle and the bandwidth occupied by the VoIP frames of an Ethernet. The longer the transmission cycle becomes, the longer the voice-data length. Moreover, the longer the voice data in an IP packet becomes, the more the transmission efficiency increases because the VoIP packet has overheads for the MAC header (in the case of the Ethernet), IP header, UDP header, and RTP header (Fig. 3). However, the longer one packet becomes, the more packet errors are likely to occur, so it is important to evaluate how the network traffic conditions affect the packet behavior and QoS in VoIP systems.

TABLE I

Relationship between packet transmission cycle and voice data length

Transmission cycle	Voice-data length (bytes)	
	G.711	G.729
20 ms	160	20
40 ms	320	40
60 ms	480	60
80 ms	640	80
100 ms	800	100

A high traffic load can cause packet loss and jitter, and a poor transmission line can cause bit errors. Larger jitter than the jitter absorbing buffer size may cause packet loss, and bit errors may also result in packet loss if the data link layer (such as HDLC (High-level Data Link Control procedure)) has a function for dropping irregular frames. Given 40 bytes of voice data in an IP packet, e.g., when the bit error rate is 0.01%, the reproduction rate of voice data in the worst case may be 96.8%. Moreover, given 800 bytes of voice data in an IP packet, the reproduction rate of voice data in the worst case may be 36%. This explains why the VoIP quality decreases drastically with bit error rate. The above relationship is expressed by the following formula,

$$Err\_f = L * 8 * e \quad (1)$$

where Err_f (%) is the errored frame rate, L (bytes) is the voice data length per IP packet, and e (%) is the bit error rate. Note that this does not take into consideration the possibility of bit errors in the packet header.

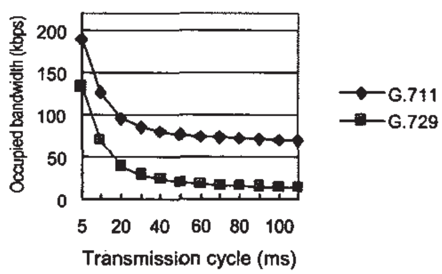


Fig. 2. Bandwidth occupied by VoIP frames.

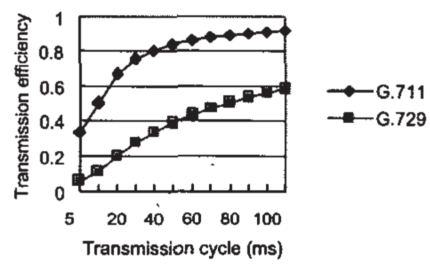


Fig. 3. Transmission efficiency

III. EXPERIMENT

We evaluated the effect of changing the packet loss rate and the voice data length of IP packets on the setup described below. The background traffic generator fixed the frame length and the amount of background traffic.

A. Test-bed network

Fig. 4 shows the measurement setup we used to evaluate the voice quality of a VoIP system.

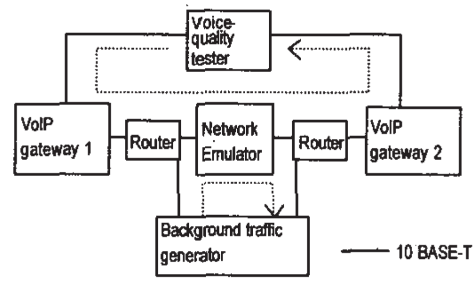


Fig. 4. Test-bed network.

A voice quality tester measured the PSQM+ between two VoIP-gateways, and a network emulator varied the packet loss rate within a range from 0 to 5% and average jitter time within a range from 0 to 50 ms. Background traffic was generated in the same direction as the voice data packet flow.

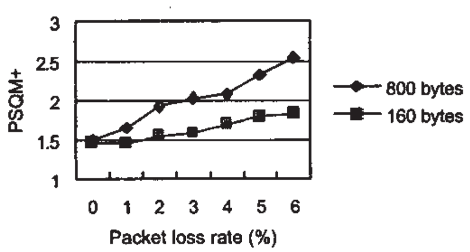
B. Voice samples

The speech database [13] contained eight voice sample files (four samples from two men speaking in Japanese, four samples from two women speaking in Japanese), which was based on ITU-T recommendation P.80. The PSQM+ value in the results is the average of these eight samples. We also calculated the standard deviation from these.

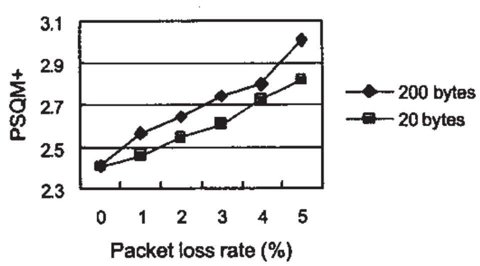
C. Results

Fig. 5 plots the measured PSQM+ for G.711 and G.729 codecs versus packet loss rate, where we chose a maximum and minimum that the VoIP-GW could assign for each codec as the voice-data length of the VoIP packets. We concluded the following from these results:

- Short-voice data and long-voice data have the same voice quality when packet loss rate equals zero.
- The higher the packet loss rate, the lower the voice quality (= higher PSQM+), and the longer the voice-data packet length, the lower the tolerance to voice-data packet loss.
- Considering that the PSQM+ value for actual communications should be less than about 2.5, the G.711 codec can tolerate packet loss rates of up to 5%, and the G.729 codec can tolerate losses up to 2%.



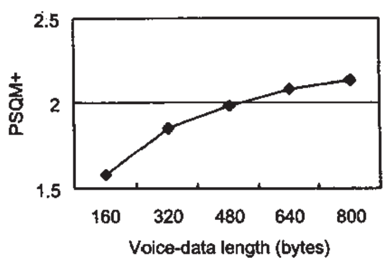
(a) 160 and 800 bytes (G.711)



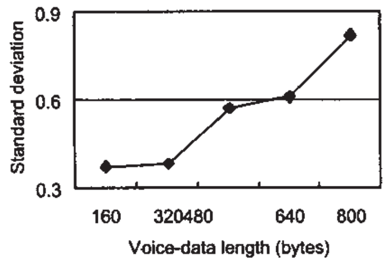
(b) 20 and 200 bytes (G.729)

Fig. 5. Examples of measured PSQM+ for two-voice-data lengths.

Fig. 6 shows the measured PSQM+ and standard deviation for five voice-data lengths (160, 320, 480, 640, 800 bytes) given a constant packet loss rate of 3%. Here, the G.711 codec was selected and the eight voice-sample files were transmitted. The 160-byte voice-data packets had the highest tolerance to packet loss and the 800-byte voice-data packets had the highest fluctuation in PSQM+, indicating that the longest packet was affected most.



(a) Measured PSQM+



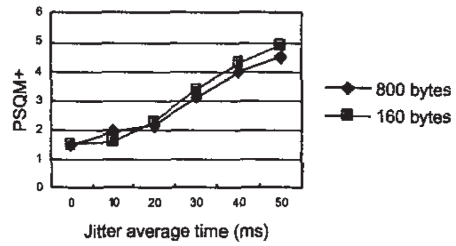
(b) Standard deviation

Fig. 6. Examples of measured results for several voice-data lengths (G.711, packet loss rate: 3%).

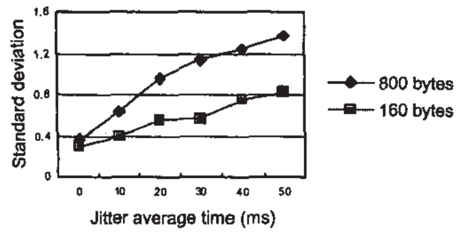
Fig. 7 shows the measured PSQM+ and standard deviation influenced by jitter. Here, we also selected the G.711 codec. We concluded the following from these results:

- There was no difference between long-voice data and short-voice data, though the more the jitter the lower the PSQM+ value, when jitter occurred in the network.
- Longer voice-data packets had higher fluctuations in PSQM+, showing that the longer packet was affected more.
- Considering that the PSQM+ value for actual communications should be less than about 2.5, the G.711 codec can tolerate jitter up to about 20 ms.

Additionally, when we compared the PSQM+ values of jitter-absorbing buffers A (large) and B (small), we found that the jitter tolerance of the former was higher than that of the latter.



(a) Measured PSQM+



(b) Standard deviation

Fig. 7. Examples of measured influence by jitter (G.711).

From these results, we found there was a relationship between voice quality and voice-data length in VoIP systems as follows.

TABLE II

Characteristics of voice-data length

	Long-voice data	Short-voice data
Occupied bandwidh	Narrow	Wide
Influence of packet loss	Tend to degrade More fluctuation in voice quality	Tolerant to packet loss Less fluctuation in voice quality
Influence of jitter	More fluctuation in voice quality	Less fluctuation in voice quality
Advantage	High transmission efficiency	Stability of voice quality

IV. Variable Voice Data Length VoIP System

In the previous section, we obtained the characteristics of a VoIP system operating under inferior network conditions. They can be used to create a VoIP system that offers optimal voice quality and transmission efficiency by varying the voice data length in IP packets based on network QoS conditions. Fig. 8 shows the architecture of the variable voice-data-length VoIP system we proposed. The VoIP-GW of the system works as follows:

- The VoIP system monitors network conditions (response time, packet loss rate, jitter time) by periodically pinging the destination VoIP-GW after a call connection is set up. Here, we regard a late response time, high packet loss rate and a large jitter time as an inferior network.
- The VoIP-GW assigns long voice data to IP packets as long as little delay is experienced in telephone communications, because longer voice data increase end-to-end delay. If the response time exceeds a certain threshold, the VoIP-GW reduces the voice-data length. If the response time is less than the threshold, the VoIP-GW increases the voice-data length.
- If the IP network is stable (packet loss rate of nearly 0%), the VoIP-GW assigns long voice data. If the packet loss rate exceeds a certain threshold, the VoIP-GW reduces the voice-data length to preserve communication quality. The VoIP-GW then increases voice-data length when network conditions return to a stable state.
- If the jitter time is less than a certain threshold, the VoIP-GW assigns long voice data. If the jitter time exceeds a certain threshold, the VoIP-GW reduces the voice-data length to preserve communication quality. The VoIP-GW then increases voice-data length when jitter time returns to a low level.

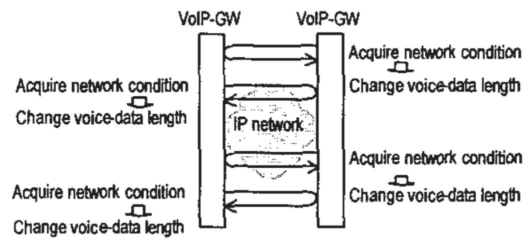


Fig. 8. Variable voice data length VoIP system.

Next, we provide the threshold for network condition values (response time, packet loss rate and jitter time) to change the voice-data length. When network condition values exceed the threshold, the VoIP-GW varies the voice-data length and jitter absorbing buffer size.

A. Response time

ITU-T defines the guidelines for one-way transmission time in G.114 [14] as follows.

- 0 to 150 ms: Acceptable for most user applications.
- 150 to 400 ms: Acceptable provided that Administrations are aware of the transmission time impact on the transmission quality of user applications.
- above 400 ms: Unacceptable for general network planning purposes; however, it is recognized that in some exceptional cases this limit will be exceeded.

The delay time for the source and destination VoIP-GW used in our evaluation was about 90 ms (voice data length: 160 bytes) – 160 ms (voice data length: 800bytes) under G.711. Considering the delay time in the VoIP-GW, the threshold for the delay time in an IP network should be less than 200 ms.

B. Packet loss rate

From the results in Section III.C, the threshold for packet loss rate should be about 5% under G.711.

C. Jitter time

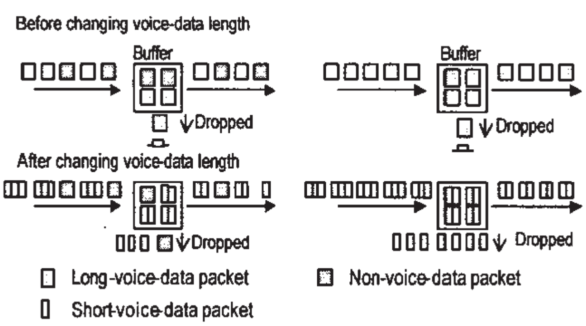
From the results in Section III.C, the threshold for jitter time should be about 20 ms under G.711.

V. Behavior Of The Proposed VoIP System

We studied how the relay node (e.g. router) of an IP network was affected by the VoIP-GW behavior above. Under stable network conditions, because the long voice data of a VoIP packet is assigned, transmission efficiency is high, and excess bandwidth can be used for non-voice data communications. The relay node incurs lower processing loads for long voice-data packets than for short voice-data packets.

When the relay node switches IP packets, it accumulates IP packets in its buffer, then routes them to a line. If there are too many processed packets and the buffer is full, packets overflowing from the buffer may be dropped. As the processing load for the relay node becomes larger and the congestion level of the IP network increases, packet loss occurs. Then, the VoIP-GW assigns shorter voice data and VoIP packets increase. However, the shorter data means that the lost voice data will be distributed among many voice data packets. Thus, there is less lost voice data in the short packets than in the long packets even if the packet loss rate becomes higher as a result of using the shorter packets.

In current network environments, the amount of non-voice data transmitted over IP networks exceeds that of voice data. Therefore, shortening the voice data of the VoIP packets has little effect on congestion, because the number of non-voice data packets basically determines the level of congestion. Thus in most cases, a variable voice-data length VoIP system is useful for ensuring voice quality (Fig. 9(a)). If, however, the IP network carries only voice data, shortening the voice data length per packet will cause more packet loss than in the mixed packet case, so our VoIP system may be limited for use in networks carrying mainly voice data (Fig. 9(b)).



(a) Mixed packet case (useful) (b) Only voice packet case (not useful)

Fig. 9. Comparison of dropped-packet behavior.

VI. Conclusion

In this paper, we introduced some problems with VoIP systems and some methods of efficiently transmitting VoIP packets by multiplexing transmission and other means.

We evaluated the voice quality of a VoIP system while changing the voice-data length under various network conditions in a real network environment. We found that a VoIP system with short voice data in IP packets is superior to one with long voice data in terms of tolerance to packet losses and the stability of voice quality. To solve the problems of inefficiency in a VoIP system, we proposed a new system that changes the voice-data length in IP packets based on dynamically changing network QoS conditions. This system achieves both high transmission efficiency and stable voice quality. We also described the behavior and the influence on IP networks of varying voice-data length under changing network conditions, and demonstrated the effectiveness of the proposed system.

References

[1] M. Masuda, and K. Ori, "Network Performance Metrics in Estimating The Speech Quality of VoIP ", Proc. of APSITT2001, pp. 333-337, Nov. 2001

[2] L. Zheng, L. Zhang, D. Xu, "Characteristics of Network Delay and Delay Jitter and its Effect on Voice over IP (VoIP)", Proc. of ICC 2001, pp. 122-126, 2001

[3] L. F. Sun, G. Wade, B. M. Lines, E. C. Ifeachor, "Impact of Packet Loss Location on Perceived Speech Quality", Proc. of IP Telephony Workshop 2001, Apr. 2001. [http://www.fokus.gmd.de/events/iptel2001/]

[4] ITU-T Recommendation, G.114: "One Way Transmission Time", 1996.

[5] M. H. Sherif, A. Crossman, "Overview of Speech Packetization", IEEE Symposium on Computers and Communications, pp. 296-304, 1995

[6] S. Casner and V. Jacobson, "RFC2508: Compressing IP/UDP/RTP headers for low-speed serial links," IETF, Internet Draft, Feb. 1999.

[7] K. El-Khatib, G. Luo, G. Bochmann and F. Pinjiang, "Multiplexing Scheme for RTP Flows between Access Routers," IETF, Internet Draft, draft-ietf-avt-multiplexing-rtp-01.txt, Oct. 1999.

[8] ITU-T Study Group 15, Delayed contribution D.510 (WP2/15), "A Result of Performance Evaluation of IP-based Circuit Multiplexing Schemes," Geneva, Jun. 1999.

[9] ITU-T Study Group 15, Temporary Document TD.24 (WP2/15), "Baseline document of Rec. G.ipcme (IP-based Circuit Multiplexing Equipment)," Geneva, Feb. 2001

[10] ITU-T Recommendation P.861, "Objective Quality Measurement of Telephone-band (300-3400 hz) speech codecs", Aug. 1996.

[11] ITU-T Recommendation, P.800: "Methods for Subjective Determination of Transmission Quality", 1996.

[12] ITU-T Recommendation, P.830: "Subjective performance assessment of telephone-band and wide-band digital codecs", 1996.

[13] NTT-AT CD-ROM, "Multi-lingual Speech Database for Telephonometry 1994", 1994.

[14] ITU-T Recommendation, G.114: "One way transmission time", 1996