

Practical Field Overview Voice Quality of RTP Packet Size Analyze on Codec G729 Annexb = no in Low Bandwidth Area of Bangladesh

N. A. Shafi, *Member, IACSIT*, Al Kawser, and O. Farrok

Abstract—In VoIP applications, packet loss causes a major impact on perceived speech quality. This impact is affected by some factors like packet loss locations, loss size and loss pattern. In this paper, we have investigated using perceptual-based objective measurement methods the impact on loss location on perceived speech quality and the relationships between convergence time and loss location for two different codecs (G.729 and G.723). Experimental result shows that loss location has a severe effect on perceived speech quality. The convergence time rely on the speech content e.g. voiced/unvoiced. In terms of unvoiced segments, the convergence time is stable as in voiced phases it varies. But it associate degree bound at the top of the segment. Our method allows a more accurate measurement of the exact effect of packet loss on perceived speech quality. As most of the internet subscribers of Asian countries use very low internet bandwidth. Hence, the goal of this analysis is to propose to change some parameters of this system so that the quality of voice may be kept in a tolerable limit using only 5kbps to 5.5kbps where this codec uses at least 6.4kbps. In a real life environment it is tested practically in research that it is possible to transmit voice satisfactorily using 5.3 kbps by changing some parameter described in this paper.

Index Terms—G.723, G.729, annex = no, RTP, UDP.

I. INTRODUCTION

Once a call has been set up between two or more VoIP devices, the caller starts speaking. At this point the voice signal has to be converted into a digital signal, formatted for TCP/IP transmission and sent along the network to the destination, where all of the preceding steps have to be reversed. PCM produces a 64 kbps stream of data with excellent voice quality. This process allowed long distance calls to be places on the T1 lines of the telephone company for transmission. One voice call takes up one channel, not a very efficient scheme. With VoIP, we want to cram as much voice data into as little digital signal as possible. And instead of diverting our digital voice signal directly onto a T1 line, we need to packetize it and send it over an IP network.

Encoding and compression techniques are published as standards by the International Telephone Union. Expect to see these when looking at specifications for VoIP equipment. Jitter buffers are memory areas used to store voice packets

arriving with variable delays so that it appears that each voice sample has arrived in the same amount of time. The steady output of the voice samples from the jitter buffer is called playout. The playout is steady and constant, and as long as the jitter buffer receives an ample supply of voice packets, the system appears to have a fixed delay. VoIP is inefficient for small voice packets while large voice packets lead to long delays. The VoIP packet will have overhead in the form of headers. The headers for IP, UDP and RTP add up to 40 bytes. If the data was as small as 40 bytes, the packet would only be 50% efficient. The largest size packet that can exist on an Ethernet system is 1500 bytes. Take away the 40 bytes for the header and you still have 1460 bytes available. That translates into 1460 samples of uncompressed voice or about one fifth of a second (182ms). If it is compressed with a ratio of 1 to 8, that represents about 1.5 seconds. If a packet with this much voice is lost or arrives out of turn, the conversation will be severely disrupted.

According to the Nyquist Sampling theorem, the sampling rate must be at least 2 times the highest frequency contained in the signal. The analog signal is sampled by the Nyquist sampling rate = $2 \times f_{\max}$ (where, f_{\max} is the maximum voice frequency = 4 KHz) [1]. Thus, the sampling rate is 8 KHz. The output of the sampling is converted into a series of amplitude pulses called Quantizing. After each sample is quantized, it is encoded to digital streams with an 8 bit rate known as Encoding. Thus, the required bandwidth for a telephone call that PCM produces is 8 KHz \times 8 bit = 64 KHz or 64 kbps stream of digital data with excellent voice quality, which is mostly used in telecom world. The digitalization process is shown in Fig. 1.

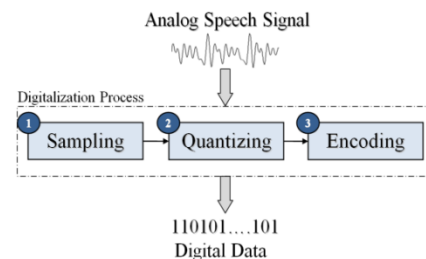


Fig. 1. Analog to digital conversion (ADC) process.

It has been estimated that as much as 60% of a voice conversation is silence. Deleting these empty bits decreases the amount of data needed for the voice transmission. Silence Suppression, also called Voice Activation Detection (VAD), is used in telephony to describe the process of not transmitting information over the network when one of the parties involved in a telephone call is not speaking [2]. In packetized voice, this gives us the opportunity to save

Manuscript received November 12, 2013; revised January 14, 2014.

N. A. Shafi is with the Network Operations Center of Inspire Systems Limited in Dhaka, Bangladesh (e-mail: nahid_apee@yahoo.com).

Al Kawser is with the Network Operations Center of Inspire Systems Limited in Dhaka, Bangladesh (e-mail: kawser.telecom@gmail.com).

O. Farrok is with the Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology (AUST) (e-mail: omarruet@gmail.com).

bandwidth because packets containing only silence do not need to be sent. In addition to Silence Suppression, Compression of voice traffic can save bandwidth which, in turn, reduces the time and cost of voice transmission. A CODEC (sometimes referred to as coder/decoder or compression/decompression) is basically a set of mathematical rules that define how an analog waveform will be digitized. The differences between the various CODECs are due in large part to the levels of compression and quality that they offer [3].

Several techniques have been invented for measuring the quality of the voice signal that has been processed by different compression algorithms (CODECs). One of the standard techniques for measuring quality of voice CODECs, which is also an ITU standard, is called Mean Opinion Score (MOS). MOS values, which are subjective and expressed by humans, range from 1 (worst) to 5 (perfect or equivalent to direct conversation). Table I displays some of the ITU standard CODECs and their corresponding bandwidth requirements and MOS values. MOS is an ITU standard method of measuring voice quality based on the judgment of several participants; therefore, it is a subjective method. Table I displays each of the MOS ratings along with its corresponding interpretation, and a description for its distortion level. It is noteworthy that an MOS of 4.0 is deemed to be Toll Quality [4].

TABLE I: MEAN OPTION SCORE

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible but slightly annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Very annoying and objectionable

G.729 is mostly used in Voice over IP (VoIP) applications for its low bandwidth requirement. Standard G.729 operates at 8 Kbit/s, but there are extensions, which provide also 6.4 Kbit/s and 11.8 Kbit/s rates for marginally worse and better speech quality respectively. Without silence suppression and other bandwidth efficiencies, the requirement can be reduced to approximately 5 Kbps average bandwidth during a conversation. Bandwidth at any location is finite, and the number of simultaneous conversations any particular connection can carry is directly related to the type of CODEC implemented. Once the voice data has been digitized, compressed and the silence suppressed, it has to be divided into sections for placing into IP packets. All VoIP packets are made up of two components: voice samples and IP/UDP/RTP headers. Although the voice samples are compressed by the

Digital Signal Processor (DSP) and can vary in size based on the CODEC used, these headers are a constant 40 bytes in length. When compared to the 20 bytes of voice samples in a default G.729 call, these headers make up a considerable amount of overhead. With cRTP, these headers can be compressed to two or four bytes. This compression offers significant VoIP bandwidth savings shown in Fig. 2.

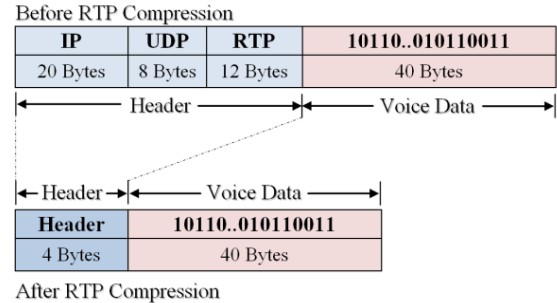


Fig. 2. RTP compression.

Typically, 10ms to 30ms (average 20ms) of voice is placed inside one packet. 20ms of uncompressed voice takes 160 bytes. Compressed at 4 to 1, 20ms would take 40 bytes. If a packet with this much voice is lost or arrives out of turn, the conversation will be severely disrupted [5].

II. DESCRIPTION OF THE SYSTEM

A wide range of factors can affect the perceived quality of VoIP services, such as coding scheme, packet loss, noises, network delay and its variation, echoes, and handovers. High bandwidth region voice quality, we get better than low bandwidth region [6]. This is our target how to improve the voice quality in low bandwidth region [7]. Recent studies reveal that packet loss constitutes the principal source of perceived quality degradation of VoIP calls. The negative effect of missing packets is more disturbing especially when packets are removed in bursts, i.e., multiple media units are consecutively dropped from the original media stream. This means that the probability of missing a given packet is much higher when the previous ones have been dropped [8].

Packet loss is a major source of speech impairment in voice over IP (VoIP) applications. Such a loss could be caused by discarding packets in the IP networks due to congestion or by dropping packets at the gateway/terminal due to late arrival. The impact of packet loss on perceived speech quality depends on several factors, including loss pattern, codec type, and packet loss size. It may also depend on the location of loss within the speech.

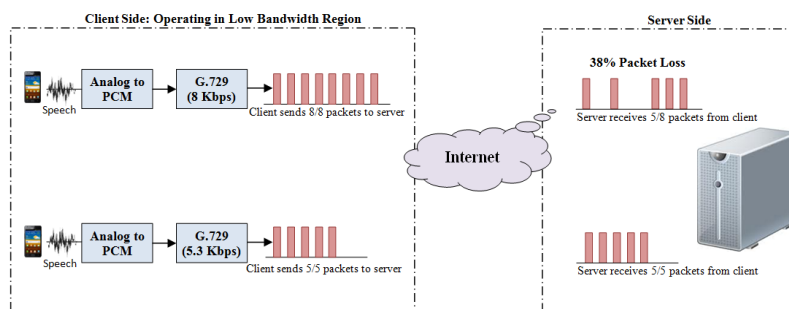


Fig. 3. RTP packet stream analysis using G.729 both in client and server side.

For example, it has been shown that, using G.729 implementation for VoIP; let's say that each of the red bars in Fig 3 represents a packet at client side operating in a low bandwidth region for two different G.729 bit rates 8 Kbps and 5.3 Kbps. The resulting packets from voice speech coming from mobile devices or any other VoIP devices are transmitted to the server side over the internet. For G.729 (8 Kbps), packets 1, 3, 6, 7 and 8 reach the server side successfully, but packets 2,3 and 4 is lost somewhere in transmission. For G.729 (5.33 Kbps), packets 1, 3, 6, 7 and 8 reach the server side successfully, but packets 2, 3 and 4 is lost somewhere in transmission.

This concealment strategy replays the lost packet received by CODEC G.729 (8 Kbps), so the listener does not hear gaps of silence. Because the lost speech is only 20 ms, the listener most likely does not hear anything. Basically, rather than the packet loss pattern itself, theoretical and representative

models measure that using CODEC G.729 with approximate bit rate 5 Kbps can reduce the loss up to 38%.

III. DISCUSSION AND RESULTS

We have investigation about RTP packet loss issue and found that our dialer sent 8 packets per second to byte saver but our problematic region dialer sent 5 packets per second to byte saver. i.e. We got loss around 3 packets (120×3=360 bytes) for this cause byte saver to switch got 18 packets loss that effect on voice quality. It is caused randomly not continues which is explained in Fig. 4a and Fig. 4b. There are a wide range of vital network-related operations where the accurate assessment of time-varying perceived quality is desirable and helpful [9]-[10]. This voice quality is mostly depending without network packet loss.

No.	Time	Source	Destination	Protocol	Length	Info
2557	21.582128000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2564	21.700129000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2582	21.850152000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2589	21.942167000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2600	21.062148000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2610	22.212158000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2617	22.332218000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2625	22.421615000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2630	22.540016000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2636	22.661159000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2642	22.808591000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2648	22.926205000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2661	23.640202000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2678	23.167173000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2684	23.288187000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2688	23.407212000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2703	23.527184000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2717	23.647188000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2721	23.768186000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2733	23.890298000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508
2757	24.006211000	192.168.110.10	172.16.0.72	UDP	165	Source port: 53352 Destination port: 18508

Fig. 4a. Call-trace report from originating side for codec G729 annexb = no.

No.	Time	Source	Destination	Protocol	Length	Info
7	0.59881	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
8	0.959266	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
9	1.159926	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
10	1.319536	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
11	1.518879	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
12	1.679670	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
13	1.879221	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
14	2.039646	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
15	2.239562	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
16	2.439342	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
17	2.599290	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
18	2.759954	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
19	2.959475	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
20	3.358857	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
21	3.400521	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
22	3.559080	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
23	3.719164	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352
24	3.919193	172.16.0.54	192.168.110.10	UDP	167	Source port: 42308 Destination port: 50352

Fig. 4b. Call-trace report from terminating side for codec G729 annexb = no.

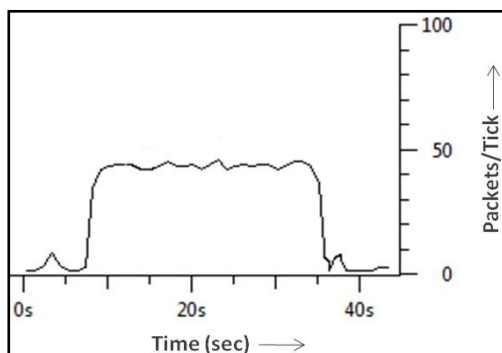


Fig. 5a. Voice quality bandwidth graph for codec G729.

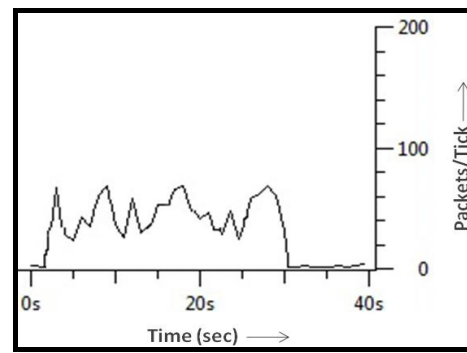


Fig. 5b. Voice quality bandwidth graph for codec G723.

Fig. 5a and Fig. 5b show bandwidth characteristics graph that is proved voice quality performance response activity. Those figures are between it Codec G729 annexb=no smooth bandwidth graph but Codec G723 has jig gage bandwidth

graph. Codec G729 annexb=no is good voice excellence at low bandwidth area.

Fig. 6 represents various bit rate analyzed at codec G729 annexb=no. We checked that Bit rate 5.3 Kbps is better than

others Bit rates (8Kbps, 6Kbps and 4.8Kbps) at low bandwidth area. We are generally used Bit rate 8Kbps at practical field. We hope that if we used bit rate ~5Kbps at worldwide then we will save bandwidth per call $[(12.27-8.17)/12.27] \times 100\% = 33\%$ or approximately 33%. Our results confirm the analytical results that quality is better with a larger packet interval [11].

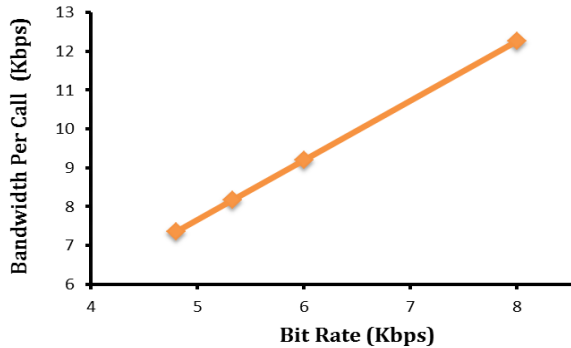


Fig. 6. Bandwidth utilization of G729 annexb=no for various bit rates.

IV. CONCLUSION

We have investigated the impact of loss positions on perceived speech quality and the relationships between the convergence time and loss locations. Packet size has in general no obvious influence on perceived speech quality for a given packet loss rate, but the deviation in speech quality depends on packet size and codec. The impact of loss position on perceived speech or the concealment performance of two modern codecs [G.729 (annexb=no) and G.723] have also been compared and analyzed. The research department of VOIP update is G.729 codec annexb=no if the audio packet size is approximately 5kbps then a good voice quality is achieved in a low bandwidth area that saves average bandwidth of approximately 33%. Further research will focus on a more extensive analysis of the impact of packet loss on speech content.

APPENDIX

VOS switch, VPS switch, various dialer, SQL database, firewall system, STM device and CISCO router & switch, OS-Windows server and Linux server etc.

REFERENCES

[1] B. A. Forouzan, *Data Communications and Networking*, 4th ed., Tata McGraw-Hill, 2009, p. 69.
 [2] N. Wittenberg, *Understanding Voice over IP Technology*, 1st ed., Delmar Cengage Learning, 2009, pp. 34-36.

[3] J. V. Meggelen, L. Madsen, and J. Smith, *Asterisk™: The Future of Telephony*, 2nd ed., O'Reilly Media, Inc, 2007, pp. 185-206.
 [4] J. V. Meggelen, L. Madsen, and J. Smith, *CCNP ONT Official Exam Certification Guide*, 2nd ed., Cisco Systems, Inc, 2008, pp. 24-25.
 [5] CISCO. (February 2006). Voice over IP – Per Call Bandwidth Consumption. [Online]. Available: http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_not_e09186a0080094ae2.shtml
 [6] *Avaya IP Voice Quality Network Requirements*, Issue 3.0, April 2004.
 [7] ITU-T Recommendation, "G.729 Annex B: A silence compression scheme for use with G.729 optimized for V.70," *Digital Simultaneous Voice and Data Applications*, vol. 35, no. 9, pp. 64-73, Sep. 1997.
 [8] S. Jelassiz and G. Rubina, "A study of artificial speech quality assessors of VoIP calls subject to limited bursty packet losses," *Jelassi and Rubino EURASIP Journal on Image and Video Processing*, 2011.
 [9] A. Rix, J. Beerends, D. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality: Technology and applications," *IEEE Trans Audio Speech Language Process*, vol. 14, no. 6, pp. 1890-1901, 2006.
 [10] S. Jelassi, G. Rubino, H. Melvin, H. Youssef, and G. Pujolle, "Quality of experience of VoIP service: A survey of assessment approaches and open issues," *IEEE Communications Surveys and Tutorials*, vol. 14, issue 2, pp. 491-513, 2012.
 [11] W. Y. Jiang and H. Schulzrinne, "Perceived quality of packet audio under bursty losses," *IEEE INFOCOM 2002*, Columbia University.



N. A. Shafi was born in Rajshahi, Bangladesh on July 23, 1985. He received the B.Sc. and M.Sc. degrees from the Department of Applied Physics and Electronic Engineering, Rajshahi University, Rajshahi, Bangladesh, in 2006 and 2007 respectively. He has been working as a system engineer in the Network Operations Center of Inspire Systems Limited in Dhaka, Bangladesh since June 2011. He is a member of the Science and Engineering Institute (SCIEI) & International Association of Computer Science and Information Technology (IACSIT). His current research interests include VOIP, microcontroller and spintronics.



Al Kawser was born in Dinajpur, Bangladesh on 08 March, 1989. He received bachelor degree from Hajee Mohammad Danesh Science & Technology University, Dinajpur in the Department of Telecommunication & Electronic Engineering (TEE) and master degree from Dhaka University in the Department of Information Technology. He has been working as a NOC support engineer in the Network Operations Center of Inspire Systems Limited in Dhaka, Bangladesh since March 2013. His interests are in telecommunications and VoIP communications.



O. Farrok was born in Bangladesh on 16 July, 1982. He received bachelor and master degrees from Rajshahi University of Engineering & Technology in the Department of Electrical and Electronic Engineering (EEE), Rajshahi-6204, Bangladesh. He has been joined since 2007 and serving as an assistant professor in the Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology (AUST). His current research interests include power electronics, digital electronics and communication engineering. He is a member of The Institution of Engineers, Bangladesh.