Coding algorithms in voice over internet protocol: A comparative study

Magdy Abd Elazim Ahmed
Computer Systems and Engineering Dept., Faculty of Engineering, Alexandria University, Alexandria, Egypt
magdy_azim@yahoo.com

Most emerging applications involve the delivery of images, voice, and video to end-user. One of these applications that has generated considerable recent interest is Voice Over Internet Protocol (VoIP). The objective of this paper is to address the issue of introducing VoIP in today's communication advancements. We demonstrate the problems facing this new technology which is expected to replace or at least work in parallel with the Public Switched Telephone Network (PSTN). Also a simple VoIP implementation is proposed in order to study the effects of compression, packet size, delays and jitter on voice transmission. An experimental analysis is presented of the delay components for a VoIP implementation.

تتضمن التطبيقات الحديثة تسليم الصور، الصوت والفيديو الى المستخدم، واحدي هذه التطبيقات التى ظهرت مؤخرا ووجدت اهتمام كبير هى نقل الصوت على بروتوكول الإنترنت. والغرض الآساسى من هذا البحث هو أولآ استعراض المشاكل التى تواجه هذه التقنية الجديدة لكى تحل محل شبكة الهاتف المحلية أو تعمل بالتوازى معها، ثم تصميم برامج لتطبيق نقل الصوت على بروتوكول الإنترنت عمليا و الذى يستخدم جميع خوارزميات التَشْفير المختلفة، كما أجريت العديد من التجارب المعملية لعمل مقارنة بين وظائف التشفير المختلفة فيما يتعلق بتأثيرها على التأخير فى نقل الصوت

Keywords: Voice over IP, Encoding algorithms, End-to-end delay, Jitter

1. Introduction

Voice over IP is an advancing technology that is used to transmit voice media over the internet or a local area network using Internet Protocol (IP) [1]. This technology provides more sophisticated and enhanced features such as cost-efficiency or cost-reduction compared to the traditional Public Switched Telephone Network (PSTN).

A packet switched network for VoIP costs as much as half as that of a traditional circuit-switched network (PSTN) in the field of voice transmission [2]. This is because of the efficient use of bandwidth requiring fewer long-distance trunks between switches.

There are many measurement studies that have been conducted by researchers to describe the end-to-end delay and loss packet of voice over the Internet [3-11]. But there was no specific research that has addressed the direct comparison between various coding functions with respect to its effect on the end-to-end delay. For example, [11] has shown the impact and feasibility of using VoIP in both Intranet and Internet environments. Three coding techniques were used: G.711, G.726 and u-law. Also a comparison was previously done between AMR and (see table 1) G.722 recommendations [12].

In this paper, we develop VoIP software that uses the major speech coding schemas in order to compare the delay in VoIP transmission. A dynamic jitter buffer discussed in [13] is considered in order to reduce jitter.

The rest of this paper is organized as follows: Section 2 discusses the benefits and factors that influence the adoption of VoIP technology. Section 3 summarizes the problems facing development and deployment of VoIP. Section 4 presents the VoIP system and its key components. Section 5 presents our experimental setup and implementation. Section 6 presents the experimental results, Finally, Section 7 gives conclusions and suggests some future work.

2. Benefits of VoIP

PSTN uses circuit switching networks which first were developed to handle voice traffic. Circuit switching is in a dominant position because it is well suited to the analog transmission of voice signals. But in today's digital world, circuit switching is rather inefficient since channel capacity is dedicated to the duration of a certain connection, even if no data are being transmitted. On the other hand, packet switching has a number of advantages over circuit switching. The efficiency of the connection line is greater. In addition to that, a packet switching network can perform data-rate conversion. Also calls are never blocked, but delayed in case of heavy traffic and priorities on packets transmitted can be used [14].

The cost of an IP network for VoIP could be as much as half that of a traditional PSTN for voice transmission. PSTN have to dedicate a full duplex 64 kbps channel for each call, whereas with VoIP networks the bandwidth is used only when something has to be transmitted. This enable more calls to be carried over a single link.

VoIP can offer other telephone services such as caller ID and call forwarding that can be added to VoIP networks at minimal costs. Also, VoIP can allow Internet access and voice transmission over a single phone line eliminating the need for two telephone lines. In addition to that, IP network can be easily integrated with the existing PSTN infrastructure and networks. Further benefits of VoIP are discussed in [11, 15].

3. Problems of VoIP

Implementing and adopting voice over IP solutions into networks has not yet been widely accepted. This is because VoIP is still lacking some deployment and communication characteristics.

Quality of Service (QoS) issues, reliability, time delay of received voice packets, along with requirements that both end users have similar equipment/software are major problems of VoIP that need to be addressed and resolved.

Voice transmission applications must meet strict requirements on packet delay because it is an important factor that affects the quality of calls. New traffic may keep entering the network even beyond the network capacity limit, consequently making both the existing and the new flows suffer packet loss and/or significant delay. As [16] stated: “To prevent these occurrences and provide QoS guarantees, a Call Admission Control (CAC) mechanism has to be introduced in IP networks.” However, [16] admitted that none of current VoIP systems can really provide QoS guarantees to VoIP because none of them are able to well supply and support CAC mechanisms. Now VoIP systems use resource reservation protocols to do explicit reservation on all routers along the path of the traffic in the network.

4. System Components of VoIP

The basic idea of VoIP involves the transmission of voice as data packets using IP. Fig. 1 shows a VoIP system structure. The voice is an analog signal, and must be converted into a digital signal before it is compressed and broken down into a series of packets. These packets are transmitted over IP networks to another destination where the signal will be reassembled and decoded.

The above shown VoIP system could be divided into three parts: voice collection or playing, voice coding or decoding and socket communications. We are going to discuss each part in the following sections.

4.1. Voice collection or playing

This process is performed by the sound card connected to a headset with a microphone. A Graphical User Interface (GUI) software program is used here to make an interface between the user and the hardware (sound card).

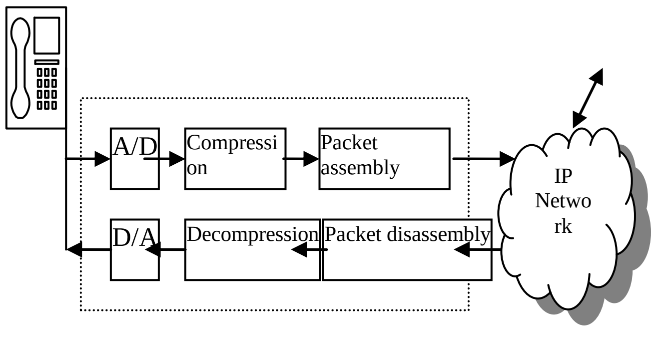

fig. 1. VoIP architecture.

4.2. Voice coding or decoding

Analog signals must be converted into binary or digital data before transmission. So voice coders are used to encode speech samples into digital data. Compression is required to reduce the number of bits transmitted in case of link errors, jitter and burst transmission. Coders and decoders are used at sending and receiving ends respectively. Pulse Code Modulation (PCM) at 64 kbps is not the only technology available nowadays. Coding in which conservation occupies 32 and 16 bps has been developed, and so has vocoders that only require 4.8 kbps or less. We can reach 800 bps and still have a clear speech. Knowing that Adaptive Differential Pulse Code Modulation (ADPCM) enables the transmission of voice with a minimum decrease in quality at 32 kbps. G.726 is the ITU-T's recommendation for ADPCM [17]. In our implementation and measurement tests, we experimented with different CODEC types to investigate the impact of the current Internet on voice quality perceived by end users. The different speech coding algorithms were downloaded from Intel® IPP website [18]. Table 1 shows coding algorithms being used [13].

The choice of codec is to some extent dictated by the bandwidth on offer which determines the maximum bit rate for the codec and the maximum speech quality that the system will achieve under ideal conditions. The lower the bit rate, the lower the quality of speech received by the listener.

4.3. Socket communication

We have two types of socket communications, connectionless and connection-oriented service communication. A connection-oriented service provides the establishment, maintenance and termination of a logical connection between users. This service generally implies that the connection is reliable. Transmission Control Protocol (TCP) is the Transmission Control Protocol/Internet Protocol (TCP/IP) suite transport protocol. UDP is another transport protocol that provides a connectionless service for application programs. In this paper, we choose RTP/UDP as our real-time transport protocol because there is no need to acknowledge sent packet, the sender will receive a feedback from the receiver himself by another voice signal.

5. Experimental setup and implementation

The experimental configuration is illustrated in fig. 1. We used two identical workstations running our VoIP implementation on Windows XP operating system. We traced and monitored the network using ethereal software [18]. A GUI windows application shown in fig. 2 represents our system. This program is just used to encode/decode the voice signal received from the microphone and then forward/accept UDP voice packets from the network interface card. The client requests a call, and then he begins speaking after a response is sent by the listener. The voice gets recorded and encoded using the CODEC specified by the client in the dropdown list and then gets transmitted to the same decoder technique on the listener side where the voice will get decoded, put in the play-out buffer of the sound card, played back and listened by the receiving host.

In parallel with the above running application, we run ethereal in order to catch the User Datagram Protocol (UDP) traffic being sent and received by each machine. Our measurements are based on computing the end-to-end delay of the voice packets between the two end points for each coding algorithm with respect to the number of packets being transmitted.

The delay for VoIP system was calculated by the sum of delays caused by recording, compression, transmission, decompression and playback. The recording and playback delays were considered constant for all calls in case of a uni-threaded system. The compression and decompression delay depends on the CODEC being used. Finally, the transmission delay depends on the performance of the underlying network. This latter delay was measured by letting the receiver of the voice packet echo it back to the sending host. Using filters in ethereal; the timestamp of the sent and received packets on the sender machine is recorded. So the transmission delay is half the difference of the two timestamps. The jitter is

Table 1
Coding techniques

Voice coder	Coding technique
G. 711	Pulse Code Modulation (PCM)
G.726	Adaptive Differential PCM (ADPCM)
G.728	Low-Delay Code Excited Linear Prediction (LD-CELP)
G.729	Algebraic Code Excited Linear Prediction (ACELP)
G.723.1	Multi-Pulse Max Likelihood Quantization (MP-MLQ)
GSM FR	Regular Pulse Excited Long Term Predictor (RPE-LTP)
GSM EFR	Algebraic Code Excited Linear Prediction (ACELP)

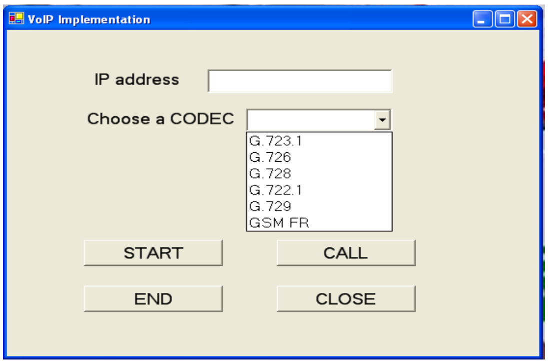

Fig. 2. GUI implementation.

the variation in the inter-packet arrival rate. In other words, it is the average difference between the arrival times of two consecutive received packets.

Our program interface, shown in fig. 2, works as follows: First the user must input the destination's IP address. Then he clicks the CALL button to setup the socket connection. When the connection is established, the user clicks on START to begin speaking. The END and CLOSE buttons are used to end the call and close the application, respectively.

6. Experimental results and analysis

Using various speech coding functions that have different bit rates produces varying levels of intelligibility and fidelity in voice transmitted over the Internet Protocol. Table 2 shows some information about the speech coding techniques used in our implementation. The compression percentage is the average percentage of coding or compressing a certain input wav file to produce a compressed output file.

The speed in MHz is the average frequency of the corresponding coding algorithm measured on a Pentium 4 1.6GHz processor. The

Table 2
Coders specifications

Coding algorithms	Compression percentage	Speed in MHz	Bit Rate (Kbps)
G.723.1	5%	24.83	6.3
G.726	50%	10.00	16
G.728	20%	49.34	16
G.722.1	200%	4.22	32
G.729	100%	24.38	8
GSM FR	50%	5.63	13

bandwidth in Kbps for each algorithm is shown in table 2. These numbers are the results of our measurements that are done on each coding algorithm separately.

Fig. 3 represents the relation between each call duration and the number of packets or frames that this call would generate. The amount of speech placed in a packet is important for the network efficiency. VoIP is inefficient for small voice packets while large voice packets would lead to long delays. From fig. 3, a packet could contain a 20 ms of speech which provides a trade off between network efficiency and increased delay.

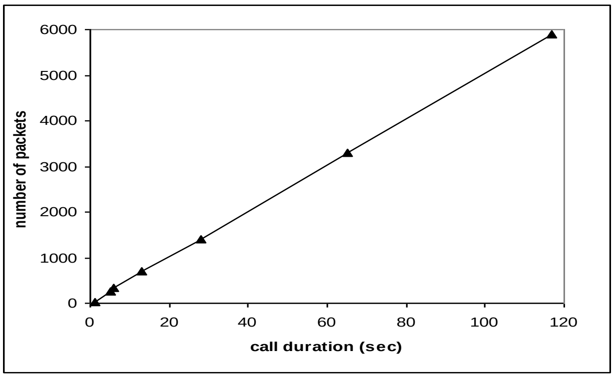

Fig. 3. Call packets.

Fig. 4 demonstrates the behavior of different speech coding algorithms with respect to the number of packets generated from varying calls duration. This graph reflects what have been measured and recorded in table 2 which is the average speed of each coding technique. As another means of comparison, we calculated the packet size in bytes being transmitted using each of the coding algorithms. Fig. 5 shows the obtained results. This figure proves what has been previously demonstrated in fig. 4. As the packet size increases the delay increases causing low quality of the received voice. According to this figure, we notice that G.726, G.722 or GSM FR are appropriate coding algorithms for voice transmission.

One may argue as to why G.728, for example, should be used at all if G.726 (ADPCM) is preferable in these two respects. It is worthwhile pointing out that although the voice quality of both are acceptable, in terms of output speech quality, PAMS scores from previous experiments [13] found that a slightly better voice quality is obtained with G.711 than ADPCM.

We noticed from the packets obtained at the receiver host that the average jitter value is around 8 milliseconds. This value is approximately the same for any CODEC algorithm being used. According to [13], a jitter buffer is needed to smooth over the distribution of packet delay that is a characteristic of IP.

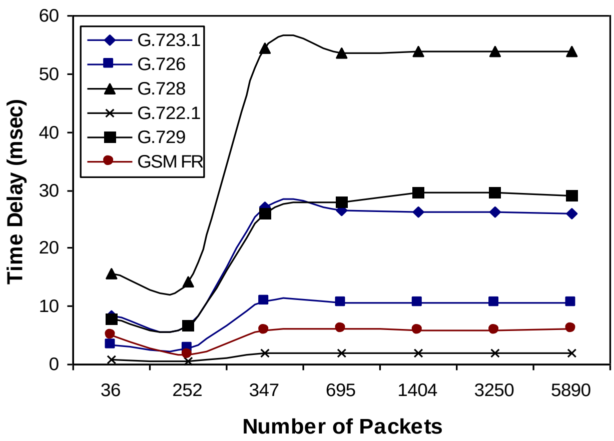

Fig. 4. Codec delay versus number of packets transmitted.

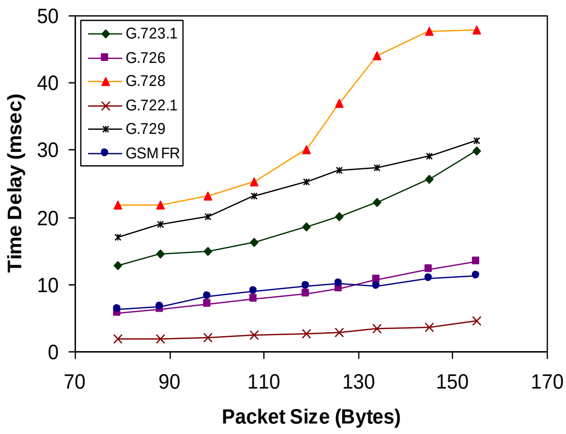

Fig. 5. Codec delay versus packet size.

7. Conclusions and future work

So, as a conclusion, the aim is to compress voice data in order to use shared bandwidth more efficiently; this can be done with little loss in the quality of voice. Further, the decompression process can be sophisticated enough to smooth out some of the problems associated with latency, jitter, and loss of voice data over a packet data network.

We believe that VoIP will be the premier in the future because of the following reasons:

a. It is inexpensive.

b. Many quality issues can be overcome by implementing better and faster network infrastructure and communication.

c. VoIP offers tremendous advantages since voice and data can be integrated over the same IP-based network. Therefore, additional phone services could be added.

It is important to note that Microsoft announced VoIP support in its recent release of Windows XP. “Imagine being able to exercise complete call control through your PC, share files with colleagues while discussing the contents in real-time, and enjoy a truly collaborative mode over the same network at very low cost”.

Finally, Voice over IP technology, as tested [19], performs the advertised functions of transporting voice, with reasonable quality, over an IP network. Note that the tests did not generate a large volume of traffic, so QoS mechanisms were not tested.

For the Future, our interests will be focused on implementing VoIP and testing it in the presence of network traffic and addressing the issue of mobility for VoIP users.

In addition to that, a study could be conducted studying the effect of introducing packet shaping software or hardware into an IP network.

References

[1] J. Walter Goralski and C. Matthew Kolon, IP Telephony, McGraw-Hill, 1st Edition (2000).

[2] U. Varshney, A. Snow, M. McGivern, and C. Howard, "Voice Over IP", Communications of the ACM, January (2002).

[3] O. Hagsand, K. Hanson, and I. Marsh, "Measuring Internet Telephony Quality: where are we Today?", Proc. IEEE GLOBECOM, Vol. 3, pp. 838-1842 (1999).

[4] A. Lakaniemi, J. Rosti, and V. Raisanon, "Subjective VoIP Speech Quality Evaluation Based on Network Measurements", Proc. IEEE Int. Conf. on Communications, Vol. 3, pp. 748-752 (2001).

[5] P. Denisowski, "How Does it Sound?", IEEE Spectr., pp. 60-64 (2001).

[6] ITU-T Recommendation "Objective quality measurement of telephone-band speech CODECs", p. 861 (1998).

[7] B. Duysburgh, S. Vanhastel, B. De Vreese, C. Petrisor, and P. Demeester, "On the Influence of Best-Effort Network Conditions on the Perceived Speech Quality of VoIP Connections", Proc. 10th Int. Conf. on Computer Communications and Networks USA, pp. 334-339 (2001).

[8] W. Kampichler and K. Goeskha, "Plain end-to-end Measurement for Local Area Network Voice Transmission Feasibility", Proc. 9th IEEE Int. Symp. on Modeling, Analysis and Simulation of Computer and Telecommunication Systems", USA pp. 235-240 (2001).

[9] W. Kampichler and K. Goeskha, "Measuring Voice Readiness of Local Area Networks", Proc. IEEE GLOBECOM, Vol. 4, pp. 2501-2505 (2001).

[10] W. Li and R. Sankar, "Real-time Voice and Data Integration in Token Ring LAN", Proc. IEEE Southeastcon'93 (1993).

[11] S. Zeadally, F. Siddiqui, P. Kubher, "Voice over IP in Intranet and Internet Environments", Proc. IEE Communications, Vol. 151 (3), June (2004).

[12] VoiceAge, "Wideband Speech Coding Standards and Applications", September (2001).

[13] R.J.B. Reynolds and A.W. Rix, "Quality VoIP-an engineering challenge", (2001).

[14] W. Stalling, Data and computer Communication, Prentice Hall 6th Edition.

[15] Mockingbird Networks, http://www.mockingbirdnetworks.com/pdfs/ wp_intro_IPtel.pdf

[16] S. Wang, Z. Mai, W. Magnussen, D. Xuan, W. Zhao, "Implementation of QoS-Provisioning System for Voice over IP", 8th IEEE Symposium (RTAS'02), pp. 266

[17] http://www.ericsson.com , June 2005.

[18] http://developer.intel.com/software/products/perflib/, June 2005.

[19] http://www.more.net/technical/research/voip/overview/#steps , June 2005.

Received July 25, 2005
Accepted July 30, 2005