STeP-IN SUMMIT 2007

International Conference On Software Testing

IMPROVEMENT AREAS IN QOS FOR VOIP SOLUTIONS

By

Dheeraj Gupta & Manpreet Singh Gulati

Aricent, Gurgaon, Haryana

(dheeraj.gupta@aricent.com) & (manpreet.gulati@aricent.com)

Practices Experience Knowledge Automation

Produced By

STeP-IN Forum

Hosted By

QSIT The Knowledge Corporation

www.stepinforum.org

www.qsitglobal.com

IMPROVEMENT AREAS IN QOS FOR VOIP SOLUTIONS

Abstract

The Internet has shown tremendous growth and continuous evolution in order to accommodate multimedia services including both voice and video applications. In particular, Internet telephony, or voice over IP is one of the most promising services currently being deployed. The Voice over IP (VoIP) significantly reduces long distance telephone charges by transferring all the long-distance voice and data over Internet connection. Besides the potentially significant cost reduction, Internet telephony offers many new features and easier integration with widely accepted Web- based services. Since VoIP shares the Internet connection with other forms of traffic, it must compete with other applications for voice quality. There still exists a number of barriers to the widespread deployment of Internet telephony. One of the key challenges in implementing VoIP is to design and build an IP-based network that meets stringent QOS requirements. In order to deliver the high QOS that users demand, VoIP system designers must tackle various parameters specific to Packet network.

The aim of this paper is to highlight different QOS measurements used to improve the quality of voice against various network performance impairments which effect service quality. The QOS improvements are implemented on Media gateways which converges circuit switch (Wireline/Wireless) and IP networks. This paper emphasizes on increasing QOS in end to end call using VoIP network, so as to achieve equal or better than PSTN quality. The major functional components involved in end-to-end VoIP call includes PSTN nodes, Signaling Gateway , Media Gateway and IP cloud. This paper describes 6 different QOS mechanisms which improve the voice quality.

WHITE PAPER

Dheeraj Gupta (dheeraj.gupta@aricent.com)
Manpreet Singh Gulati (manpreet.gulati@aricent.com)
Aricent, Gurgaon, Haryana

Contents

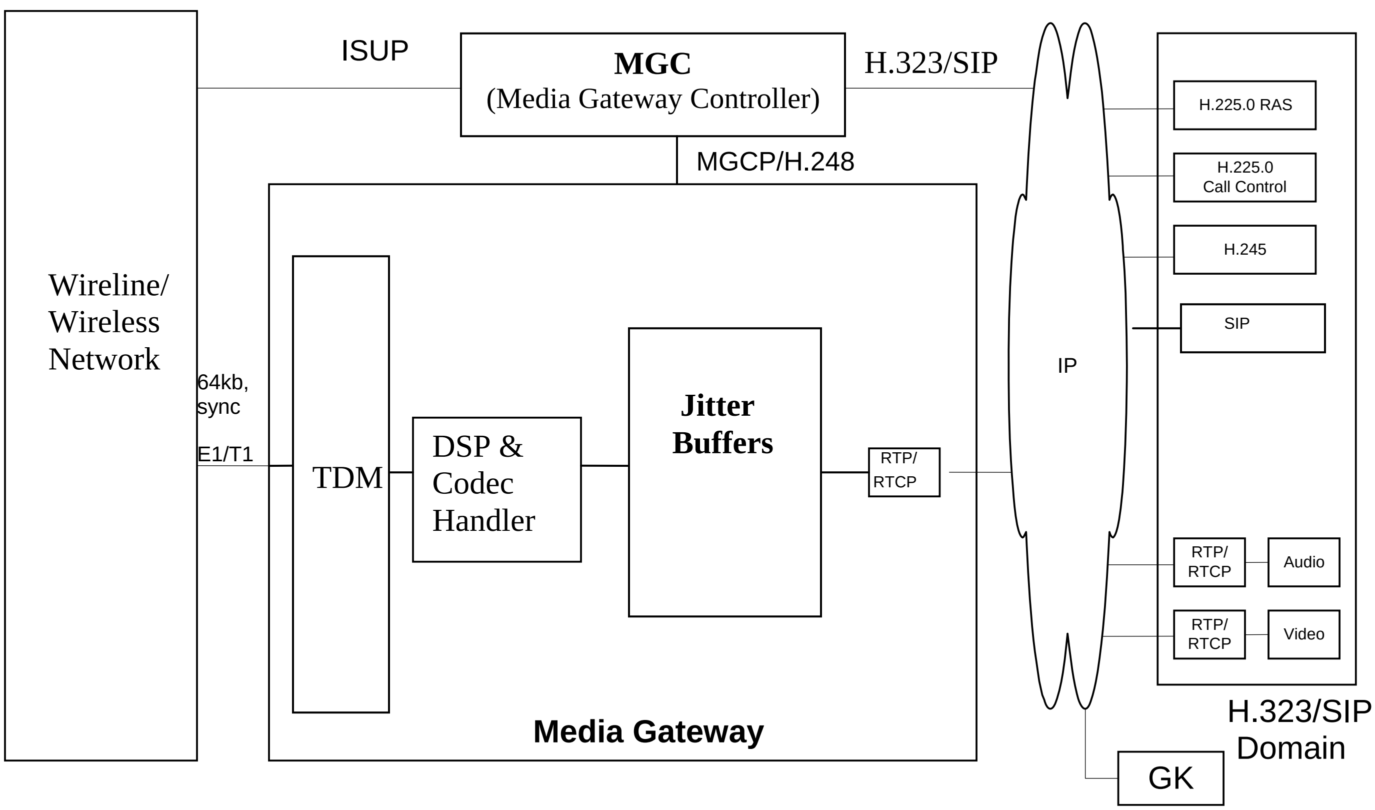
1. INTRODUCTION

Service providers are migrating to VoIP in response to today's highly competitive business environment, in order to quickly introduce new services and efficiently utilize their existing bandwidth. Enterprises are using VoIP to provide better services to their customer and employees, along with better utilization of their existing IP network infrastructure. In order to accomplish these goals, both organizations must choose products that effectively address issues like jitter that compromise voice quality.

This white paper discusses about various QOS improvements to be applied on Media gateway/IP clients which handles voice calls. The main emphasis of this paper is on tuning the parameters in reference to Media Gateway and IP cloud so as to increase QOS in VoIP call to a much better extent The RTP/RTCP protocols carry parameters which are useful to derive the QOS mechanisms. Using some of the RTP parameters, it is possible to estimate the inter-arrival jitter of the packet and inter-media synchronization. The following improvements shall be discussed in this paper with implementation specifics.

- Static Jitter Buffer
- Adaptive Jitter Buffer
- Adaptive Jitter Buffer with error concealment
- Lip synchronization (Used in Multimedia calls)

The following diagram shows sequence in which packet processing takes place with respect to jitter buffer handling in generic VoIP Call.



2. WHY QOS IMPROVEMENT IS NEEDED IN VOIP

2.1. Network Delays

The transit time for VoIP packets to travel from the sender to the receiver (i.e. from one IP phone to another) is defined as Network Delay. Voice/Video quality is affected by the latency or delay of IP networks. It may take about several seconds to hear what the other person talking. Users of a poorly implemented VoIP solution will hear a similar delay. A typical IP packet will transit through multiple hops or routing devices, with each device imposing a variable amount of delay. There are both fixed delay sources such as propagation, coding/decoding, packetization, etc., and variable delay sources such as queuing, buffering, etc.

The Internet Engineering Task Force (IETF) has attempted to address these issues by utilizing Real Time Protocol (RTP). Network managers can use various bandwidth reservation schemes in conjunction with RTP to minimize the overall end-to-end delay. However, these schemes are based on emerging technology that is not available across all networks.

2.2. Out of Sequence Conditions

VoIP packets that arrive out of their original transmission sequence, often resulting in "late" arrival and discard by the VoIP receiver. In an IP network, individual packets may take different routes to a specific destination. Since they can arrive at different times, the packets may be out of sequence. At any point in time, one path can be either faster or slower than the other paths. Voice requires that these packets be placed into the correct sequence.

2.3. Loss Packets

Packets Loss is the number of VoIP packets that are discarded by the network due to congestion or packet corruption. An individual packet will often be lost within the IP network and may never make its way to its final destination. This methodology is acceptable in data communication as the transport/application level protocol takes care of resending the dropped packet. Lost voice packets, however, will translate into breaks in the speech and could result in overall loss of speech quality. In multimedia call, when video frames are lost it results in blurred image or severe degradation of video quality.

2.4. Duplicate Packets

An individual packet traveling in IP network may take different routes and in some cases it is sent out along two paths towards its destination. This result in destination node receiving two copies of the same RTP packets marked with the same timestamp.

2.5. Echo

With the migration of telecommunication services from Time Division Multiplexing (TDM) to Voice over IP (VoIP), traditional voice quality impairments have combined with IP-based impairments to seriously degrade the overall quality of voice calls. Echo is one such problem that is being increasingly discovered in VoIP networks. Audio distortion in which a speaker or listener hears audio repeated with noticeable delay; echo can be either line-induced (i.e. through analog/digital conversion) or acoustically induced through faulty speaker phones, headsets or handsets.

2.6. Jitter

Jitter is defines as the inter-packet arrival variation that results from variation in transit delay. Excessive jitter results in "early" or "late" packet delivery and results in discard of packets at the receiving jitter buffer.

2.7. Noise Level

Noise Level is defined as the excessive background level that can be measured during periods of silence.

2.8. Signal Level

The speech energy level contained in VoIP packets which, if excessively high or low, can negatively impact the listener's experience.

3. ADDRESSING THE JITTER PROBLEM

The major concern of both service providers and enterprises when migrating to VoIP, is the need to maintain the same voice quality as that offered by current circuit-switched network. When an RTP voice packet reaches a media gateway, the preparation and conversion required for transmitting over the PSTN can be broken down into three major steps; storage, sorting and decoding/playing.

3.1. Storage

Mitigating the effect of jitter in packet communication is one of the major challenges being faced by VoIP vendors. Removing jitter requires collecting packets and storing them long enough to allow the slowest packets to arrive in order to be played in the correct sequence. The storage area used by those devices is known as the "Jitter Buffer". The network device increases the delay as it waits for the slowest packet to arrive. In order to achieve voice quality, the vendor must balance the need to minimize delay with the need to remove jitter. The bigger the buffer, the more delay, but if the buffer is too small, then voice quality can be compromised. Therefore, the ideal solution would adapt to the characteristics of the network, storing only the required amount of buffered voice traffic.

Most vendors use one of the following two methods to manage the size of the jitter buffer: Packet time variations in the jitter buffer are measured over a period of time, and the buffer size is incrementally adapted to match the calculated jitter. The number of packets that arrive too late to be processed are counted and compared to the number of packets that were successfully processed. This ratio is then used to adjust the jitter buffer to target a predetermined allowable late packet ratio.

3.2. Sorting

As the data is stored, it must also be sorted into the original sequence to accurately reproduce the original audio. RTP packets can arrive in any sequence, at any time or not at all. The Jitter Buffer algorithm sorts the voice/video frames according to a sequence number supplied in the RTP packet. The jitter buffer algorithm leaves open slots for those packets that have not yet arrived. The sampling size used by the voice/video coder determines the size of the slots. The Jitter Buffer algorithm dynamically determines the average holding time of a packet and thus the jitter buffer size. RTP protocol carries the payload type and sequence number of the packet, using this payload type, it is possible to determine the sampling rate of the codec and determine the holding time of a packet and expected jitter size.

3.3. Decoding and Playing

Another source of delay is the decoding of voice packets followed by the playout of audio to the receiving party. In order to decode the packets, the device must support codec that was used to encode the packets. The playout is the actual playing of the media sample. Closely related to the Jitter Buffer is "frame erasure", which is used to compensate when a packet has not arrived in time for playout. In the best case, frame erasure will not be necessary since it can hurt voice quality. The goal of a good jitter buffer is to use as little frame erasure as possible.

Most vendors use one of the following schemes for frame erasure:

- The voice packet preceding the missing packet is replayed during the interval of missing packet. This method is effective when the packet loss is infrequent.
- Both the N^{th} packet and the $(N+1)^{th}$ packet are transmitted every time, which results in every packet being sent twice. This method greatly increases the chance that the data will be received but uses a double the required amount of bandwidth.

4. JITTER BUFFER DESIGN

A jitter buffer is required to make sure that packets are available when needed for play-out. It removes the jitter in the arrival times of the packets at the cost of an increase in the overall delay. The objective of a jitter buffer algorithm is to keep the buffering delay as short as possible while minimizing the number of packets that arrive too late to be used. A large jitter buffer causes an increase in the delay and decreases the packet loss. A small jitter buffer decreases the delay but increases the resulting packet loss. The following two approaches are followed to design the jitter buffers.

4.1. Static Jitter Buffer

The traditional approach is to store the incoming packets in a packet buffer before sending them to the decoder. Because packets can arrive out of order, the jitter buffer is not a strict first-in-first-out buffer, but also reorders packets if necessary. The most straightforward approach is to have a buffer with a fixed number of packets. This results in a constant system delays and requires no computations and provides minimum complexity. The drawback with this approach is that the length of the buffer has to be made sufficiently large that even the worst case can be accommodated. This approach is called static jitter buffer.

In Voice/Video IP applications, static jitter buffer helps in setting the jitter value to maximum allowable delay and queuing the incoming packets before playout at other end of the device. The static jitter settings are done when the system is up and then reused for all the calls until new values are set. Media gateways maintain jitter buffers separately for audio and video streams and these buffers are used when sending media from IP to TDM.

The jitter buffers are implemented as queues of packets, and the maximum size of the queue (jitter buffer) can be provisioned using configuration interface. The audio buffer size depends on the call type and codec type. And again the audio buffer size (in terms of milliseconds) also depends on the number of audio frames in the incoming RTP packets. The media gateways audio jitter buffers are divided into 'N' number of slots with each slot of size 'S'. Each slot will contain exactly one RTP packet. If the size of the RTP payload exceeds the slot size, the RTP packet is dropped. When static jitter buffers are used in the system to store the media packets, the codec type and the expected number of frames per RTP packet need to be known.

The following table shows different delays with respect to different voice codecs.

Codec	frame size (bytes)	frame duration (ms)	frames/ packet	packets/ sec	payload size (bytes)	packet size (bytes)	latency (ms)
G723.1	20	30	1	33.33	20	60	30.00
G723.1	20	30	2	16.67	40	80	60.00
GSM	33	20	1	50.00	33	73	20.00
GSM	33	20	2	25.00	66	106	40.00
G.711	240	30	1	33.33	240	280	30.00

G.711	240	30	2	16.67	480	520	60.00

4.2. Adaptive Jitter Buffer

In order to keep the delay as short as possible, it is important that the jitter buffer algorithm adapt rapidly to changing network conditions. Therefore, jitter buffers with dynamic size allocation, so-called adaptive jitter buffers, are implemented most commonly. The adaptation is achieved by increasing/decreasing the buffer depth to match the jitter in RTP packets arrival. Using RTP timestamp parameter in RTP protocol, it is possible to estimate the relative difference of packet arrival at destination. The adaptive jitter buffer controls the media packet flow according to network conditions by adjusting jitter depth.

The calculation of Adaptive Jitter buffer as follows:

- An estimate of the statistical variance of the RTP data packet interarrival time measured in timestamp units and expressed as an unsigned integer. The interarrival jitter J is defined to be the mean deviation (smoothed absolute value) of the difference D in packet spacing at the receiver compared to the sender for a pair of packets. As shown in the equation below, this is equivalent to the difference in the "relative transit time" for the two packets; the relative transit time is the difference between a packet's RTP timestamp and the receiver's clock at the time of arrival, measured in the same units.

  Si = RTP timestamp from packet-N at Source
  Sj = RTP timestamp from packet-(N+1) at Source
  Ri = RTP timestamp from packet-N at destination
  Rj = RTP timestamp from packet-(N+1) at destination

  For two successive packets, the delay is expressed as:

  $$D(i,j)=(Rj-Ri)-(Sj-Si)=(Rj-Sj)-(Ri-Si)$$

The inter-arrival jitter is calculated continuously as each data packet-*N* is received from source SSRC_N. The jitter is calculated using difference D for N^{th} packet and $(N-1)^{th}$ packet in order of arrival (not necessarily in sequence), according to the formula

$$\text{Relative Delay} = J=J+(|D(N,N-1)|-J)/16$$

Relative delays are calculated for each incoming RTP packet from network and store it as an array of values for about 10 to 20 RTP packets. Take Average of relative delay of 10 to 20 packets and then use this value as an Adaptive jitter value. The above mechanism is to be repeated for next 10 to 20 packets and continue to use this value to adjust the jitter.

4.3. Adaptive Jitter Buffer with Error Concealment

A new approach which combines an advanced adaptive jitter-buffer control with error concealment has been developed to control the severe packet losses. Experimental results indicate that multimedia data streams over internet are affected by high packet loss rates. Hence, control of errors is essential to keep the communication more robust and reliable. Combining adaptive jitter control and packet loss concealment into one unit makes this unique algorithm capable of adapting the buffer size on a millisecond basis. The approach allows it to quickly adapt to changing network conditions, and to ensure high speech quality with minimal buffer latency. This can be achieved because the algorithm is working together with the decoder and not in the packet buffer. In addition to minimizing jitter buffer delay, the packet loss concealment part of the algorithm is based on a novel approach, and is capable of producing higher quality than any of the standard methods. Experiments show that with this type of approach one-way delay savings of 20 – 100 ms are achievable in a typical VoIP environment. The approach works as follows.

When RTP packets are received from IP network, determine the latency using adaptive jitter buffer scheme and then analyze whether there are any packet losses. To control the packet loss, replace the loss packet by inserting appropriate noise or silence frames or form a redundant packet based on two successive packets.

5. ADDITIONAL QOS IMPROVEMENTS

The jitter buffer design schemes explained so far will not be helpful to achieve better quality for multimedia calls. As there will be synchronization issues when dealing with video calls. The audio frame may be played before the video frame reached to destination. In order to achieve synchronization between audio and video streams in a multimedia call, the following design schemes helps in achieving good synchronization.

5.1. Lip Synchronization

Synchronization between audio and video signals is very much needed in multimedia applications because they have very strict timing constraints. Multimedia applications with audio and video need a synchronization scheme if they are transmitted and processed independently. The synchronization scheme is responsible for ensuring that the audio and video streams are synchronized after processing. The video and audio packets are sent as separate streams from single video call. Each stream is transmitted with its own payload type and timestamp is generated based on codec clock rate. This RTP timestamp information is necessary to correlate the audio packets to the video frames. Hence, integration is achieved without the use of video information, such as lip movements.

Implementation strategy:

In order to implement, lip synchronization it is required to know the real clock time when RTP packet generated at source. RTCP protocol facilitates proving real clock time as NTP timestamp. This parameter is sent in RTCP Sender reports. The proposal is to use the RTCP sender reports from the source to estimate the relative timing difference between the Audio and Video RTP streams. This difference will be used to adjust the Playback target to achieve synchronization between the two streams. The solution is as follows:

- Start with constant audio and video playback targets
- The Video playback target will remain the same and we will adjust only the audio playback target for any lip sync adjustments. This is because video stream is contiguous and temporally dependent. And so we can't remove or insert data into video stream.

- From the Video RTCP packets, note servers NTP wall clock time. Wait for the next audio RTCP packet. From this packet, take the RTP Timestamp.
- Calculate the audio rtp timestamp interms of video rtp timestamp using video ntp timestamp.
- Determine the relative time difference between audio and video RTP packets. The difference value is nothing but delay between audio and video RTP packets. Use this value to adjust the audio jitter.

6. CONCLUSION

A designer of VoIP system will face many challenges, some similar to what have been experienced in traditional telecommunications design, and some very specific to VoIP. The most prominent of these challenges have been discussed in this paper and solutions to overcome them were presented. Some of the most important aspects specific to VoIP are related to the characteristics of the transport media – IP networks. The design has to be able to cope with packet loss and transmission time jitter with a minimum of latency and high voice quality. Different jitter buffer mechanisms should be implemented on the system and use it based on different call scenarios. Static jitter buffer is helpful when there is a situation to tackle the bursty media traffic and adaptive jitter buffer is helpful for person to person communication. Designer of the VoIP QOS would need to choose appropriate jitter scheme to handle different call scenarios.

7. ACRONYMS

DSP	Digital Signal Processor
IP	Internet Protocol
ISDN	Integrated switched Digital Network
PSTN	Public switch telephone network
QOS	Quality of Service
RTP	Real time protocol
RTCP	Real time control protocol
VOIP	Voice over internet protocol

8. REFERENCES

RFC 1889 RTP: A Transport Protocol for Real-Time Applications
[IETF standard for Audio-Video Transport Working Group]

RFC 1890 RTP Profile for Audio and Video Conferences with Minimal Control
[IETF standard for Audio-Video Transport Working Group]

9. ABOUT THE AUTHORS

Dheeraj Gupta has more than six years of experience in Software Development, Testing, Automation and is currently playing the role of Technical Leader in **Aricent**. He is working upon VOIP (Media Gateways, Softswicth) for more than 3 years and has rich experience in various VoIP Protocols. Prior to this, He has worked upon various 2.5 GPRS Protocols.

Dheeraj holds a M.S degree from Bits Plain, India and has done his graduation in computer science from Punjab technical University. He holds overseas experience of more than 3 years.

Manpreet Singh Gulati has more than three years of experience in Software Testing, Automation and is currently playing the role of Senior Software Engineer in **Aricent**. He is working in VoIP Domain (Media Gateways, Media Gateway Controller, and Signaling Gateway – Soft switch Architecture) for more than 3 years and has rich experience in various VoIP Protocols.

Manpreet holds B.Tech Degree in Electronics and Communication, from Guru Gobind Singh Indraprastha Unviversity, Delhi, India. He also holds overseas experience.