QoS Adaptation in Voice over IP

MIROSLAV VOZNAK
Department of Electronics and Telecommunications Technology
Faculty of Electrical Engineering and Information
Technical University of Ostrava
17. listopadu 15, Ostrava-Poruba, 708 33
CZECH REPUBLIC
miroslav.voznak@vsb.cz

Abstract: This paper explains many of the issues facing Voice over IP (VoIP). Many of these issues, such as compression/decompression of the speech frame and propagation delay, are inherent to VoIP. With careful planning and solid network design these effects on VoIP networks can be minimized. This paper details these various issues and explains how they can affect packet networks.

1 Introduction

VoIP delay or latency is characterized as the amount of time it takes for speech to exit the speaker and reach the listener. The ITU-T recommendation specifies that for good voice quality, no more than 150 ms of one-way, end-to-end delay should occur. In an unmanaged, congested network, queing delay can add up to two seconds of delay. This lenghty period of delay is unacceptable in almost any voice network. Queuing delay is only one component of end-to-end delay. A packet-based networks generate various delay for various reasons.

The main parts of the total delay are following:

- propagation delay
- handling delay
- queuing delay
- jitter

On the other side is no less important the type of using voice coding/decoding. Codecs are developed and tuned based on subjective measurements of voice quality. Standard objective quality measurements, such as total harmonic distortion and signal-to-noise ratios, do not correlate well to a human's perception of voice quality, which in the end is usually goal of most voice compression techniques. A common subjective benchmark for quantifying the performance of the speech codec is the *Mean Opinion Score (MOS)*. MOS tests are given to a group of listeners. Although MOS scoring is a subjective method of determining voice quality, it is not the only method for doing so. The ITU-T put forth recommendation P.861, which covers ways objectively determining voice quality using *Perceptual Speech Quality Measurements (PSQM)*.

3 Propagation Delay

Four types of delay are inherent in today's telephony networks: *propagation delay, handling delay, queuing delay and jitter.*

Propagation delay is caused by the speed of light in fiber or cooper-based networks. A fiber network stretching around the world induces a one-way delay about 70 ms. Although this delay is almost imperceptible, propagation delays in conjunction with further delays can cause noticeable speech degradation.

3 Handling Delay

Handling delays can impact traditional phone networks, but these delays are a larger issue in packetized environments. This delay is caused by devices that forward the frame through the network and it is related with the voice processing as actual packetization, compression and switching. The packet headers that include protocols IP, RTP and UDP add 40 bytes to each frame. RTP header compression is used to reduce the large percentage of bandwith consumed by a voice call, cRTP enables to compress the 40-bytes IP/RTP/UDP header to 2 to 4 bytes. For an explanation of packet sizes and bandwith consumed, see Table 1.

$$BW = \frac{Ps + Os}{Ps} \cdot Cr \cdot k \qquad (1)$$

BW .... Bandwith [Kbps]
Cr .... Codec rate [Kbps]
Ps .... Payload size [bytes]
Os .... Overhead size [bytes]

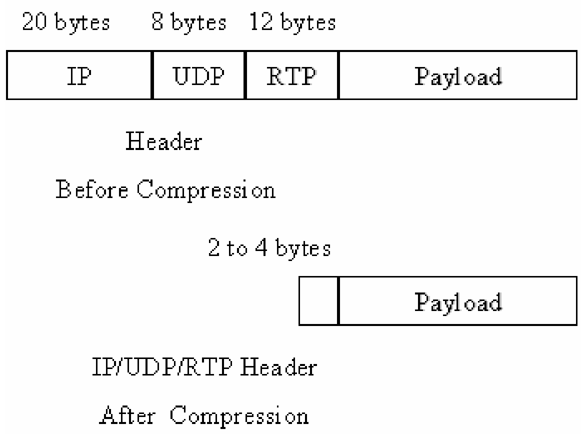

Figure 1. *IP/UDP/RTP Header.*

Compression Technique, Recommendation ITU-T	**Payload size [Bytes]**	**Bandwidth at full rate [Kbps]**	**Bandwidth with cRTP [Kbps]**
G.711 (64 Kbps)	240	74,7	64,8
G.711 (64 Kbps)	120	85,3	65,6
G.728 (16 Kbps)	80	24	16,6
G.728 (16 Kbps)	40	32	17,2
G.729 (8 Kbps)	40	16	8,6
G.729 (8 Kbps)	20	24	9,2
G.723.1 (6.3 Kbps)	40	12,6	6,8
G.723.1 (6.3 Kbps)	20	18,9	7,2
G.723.1 (5.3 Kbps)	40	10,6	5,7
G.723.1 (5.3 Kbps)	20	15,9	6,1

Table 1. *Codec type and payload size effects on bandwith.*

Compression Technique, Recommendation ITU-T	Payload size [Bytes]	Bandwidth with VAD [Kbps]	Bandwidth with cRTP & VAD [Kbps]
G.711 (64 Kbps)	240	52,3	45,4
G.711 (64 Kbps)	120	59,7	45,9
G.728 (16 Kbps)	80	16,8	11,6
G.728 (16 Kbps)	40	22,4	12
G.729 (8 Kbps)	40	11,2	6
G.729 (8 Kbps)	20	16,8	6,3
G.723.1 (6.3 Kbps)	48	8,8	4,8
G.723.1 (6.3 Kbps)	24	13,2	5
G.723.1 (5.3 Kbps)	40	7,4	4
G.723.1 (5.3 Kbps)	20	11,1	4,3

Table 2. *Codec type and VAD effects on bandwith.*

Next bandwith saving method takes advantage of the breaks and pauses in a speech patterns, known as *Voice Activity Detection (VAD)*. When the VAD detects a drop-off of speech amplitude, it waits a fixed value of time and stops putting speech frames in packets. This time is known as *hangover* and is typically 200 ms. VAD experiences certain inherent problems in determining when speech end and begins, and in distinguishing speech from background noise, the speech beginning of a sentence is cut or clipped. VAD can reduce about 30 % requirements on bandwith, see the Table 2.

Depending upon which codec is used and how many voice samples is wanted per packet, the amount of bandwith per call can increase drastically. Digital Signal Processor (DSP) usually generates a speech sample every 10 ms when using G.729, an initial look-ahead of 5 ms, it is giving an initial delay 15 ms for the first speech frame with only one sample G.729, two samples within one packet 25 ms and four samples 45 ms. Two of these speech samples consume 20 bytes per frame, which works out to 8 Kbps, including IP/UDP/RTP header works out to 24 Kbps, see Table 3.

Number of 10-ms samples per frame	Bandwidth consumed [Kbps]	Sample latency [ms]
1	40	15
2	24	25
4	16	45

Table 3. *Codec G.729/8 Kbps and sample size effects on bandwith.*

4 Queuing Delay

Queuing delay is caused by congestion on a outbound interface, it is occurs when more packets are sent out than the interface can handle at a given interval. Various tools are available to achieve the minimal latency and the *Quality of Service (QoS)* for a packet flows. These tools content different queuing techniques, such as:

- WFQ – Weighted Fair Queuing
- CB-WFQ – Class-Based Weighted Fair Queuing
- PQ - Priority Queuing
- CQ - Custom Queuing
- WRED - Weighted Random Early Detection
- RSVP - Resource Reservation Protocol

5 Jitter

Simply stated, *jitter* is the variation of packet interarrivel time. Jitter is one issue that exists in packet-based networks. While in a packet voice environment, the sender is expected to reliably transmit voice packets at a regular interval. These voice packets can be delayed throughout the packet network and not arrive at that same regular interval at the receiving station. The difference between when the packet is expected and when it is actually received is *jitter*. It is cause, why is necessary *jitter bufer*. The jitter buffer is considered a dynamic queue. This queue can grow or shrink exponentially depending on the interarrival time of the RTP packets.

3 Conclusion

Many skeptics do not believe IP can give the proper QoS for such a real-time application, but with the proper network design and the right tools, it is possible. Each network is different and requires not only attention to detail but also a knowledgeable administrator who knows how to tune the network to provide optimal QoS. QoS can help solve some of these problems, namely packet loss, jitter and queuing delay. Some of problems QoS cannot solve are propagation delay, codec delay and sampling delay.

References

[1] Schulzrinne H., Casner S., Frederick R., Jacobson V. , *RTP: A transport protocol for real-time aplication,* RFC 1889, 1996

[2] Peters J., Davidson J. , *Voice over IP Fundamentals,* Cisco Press, Indianopolis USA, 2000, ISBN 1-57870-168-6