BUILDING COMPETITIVE SYSTEMS: A PACKETCABLE PERSPECTIVE

Burcak Beser *
Juniper Networks

Abstract

PacketCable defines a network superstructure that overlays the two-way data-ready broadband cable DOCSIS 1.1 access network. PacketCable specifications define how PacketCable elements interact with each other and the protocols that are used between these elements. Since PacketCable certification/qualification only includes the protocol compliance of these elements, the certification/qualification does not suffice as the necessary means to provide a competitive service as provided by today's Public Switched Telephony Network.

This paper details some of the features that are necessary for competitive telephony service. Some of these features are not covered by PacketCable certification/qualification tests.

INTRODUCTION

The PacketCable project defines the protocols that are necessary for building competitive telephony.

Building a successful competitive telephony services over cable infrastructure requires the grade of service that is provided by Public Switched Telephony Network (PSTN) landline services to be met.

The grade of service provided by PSTN landline services has many dimensions that require different aspects of services to be engineered. Due to time and space limitations, this paper mainly focuses on the issue of perceived voice quality.

PUBLIC SWITCHED TELEPHONY NETWORK

Since the aim of the cable telephony is to match or exceed the service quality that is offered by the landline telephony systems, it is very important to understand the landline telephony of today.

Each phone call is carried as 64 kb/s bit stream, with bits flowing regardless of whether the sender talks or not. The speech signal is encoded at a sampling rate of 8 kHz, with eight bits per sample. The encoding is a simple lookup that maps sample amplitudes from 0 to 8159 to a 7-bit table entry, with a roughly logarithmic scale. There are two encoding schemes, A-law and u-law, where the former is found in European countries, while the latter is used in North America and Japan. The encoding is often also being referred to by its ITU Recommendation name, G.711 or called PCM coding [1]. The u-law encoding offers a signal-to-noise ratio of 39.3 dB for a full-range signal and a dynamic range of 48.4 dB. This is roughly equivalent to that of FM radio, except that the audio bandwidth is far lower.

The 8 kHz sampling period yields the basic clock period, 125 us, that is found throughout the digital telephone system, even when no voice is being transmitted. A number of these digital signals are then multiplexed into a single frame. For example, a T1 circuit consists of 24 voice channels, with one byte per channel. This packaging of channels into a single digital stream is called time-division multiplexing (TDM). A frame consists of these voice channels plus one or more synchronization bits.

Due to the TDM nature of the PSTN network the delay that is perceived by the users is mostly the propagation delay [2]. For North America the end-to-end delay worst case is calculated using maximum national distance of 6000 Km is found as 33 msec. In the same manner the long distance submarine fiber connection between San Francisco and Hong Kong can be found as 78 msec. It is important to note that even though the typical propagation delays are much less the impact of the PBX equipment, compression CODECs and multiplexers the given delays constitute good reference points.

GRADE OF SERVICE

The Grade of Service can be divided into two: the call connecting quality and perceived voice quality.

The call connecting quality depends on many factors including but not limited to call blocking, post-dial delay and accurate billing.

The perceived voice quality is generally more important than the call connecting quality, people tend to forget sporadic call connection problems, but when they have bad voice quality that they are paying for they tend to remember.

Perceived Call Quality

The voice quality in the PSTN networks was historically measured using 'mean opinion score'. The mean opinion score measures the subjective quality of a voice call. Historically the telephony providers invited people and used various call types (with delay, echo etc.) and recorded the results.

The MOS is a scale of 1-5 where the PSTN stands at 4.4 for local calls (perfect score). The score of national calls is generally above 4, which is considered as satisfactory. Anything below 4 may result with customer dissatisfaction with the service being received. Since the MOS is a subjective scale and requires subjective tests to be carried out which is not a good method of designing for a target. For design purposes ITU E-Model can be used [3].

The equation for the transmission rating factor *R* is:

$$R = Ro - Is - Id - Ie$$

Where,

- *Ro*, the basic signal-to-noise ratio based on send and receive loudness ratings and the circuit and room noise;
- *Is*, the sum of real-time or simultaneous speech transmission impairments, e.g., loudness levels, side tone and PCM quantizing distortion;
- *Id*, the sum of delayed impairments relative to the speech signal, e.g., talker echo, listener echo and absolute delay;
- *Ie*, the Equipment Impairment factor such as packet loss and CODEC loss, if CODEC being used is different than G.711.

CABLE TELEPHONY TARGETS

Cable Telephony competing with landline services aims to have E-Model R-value of 80, which corresponds to MOS scale of 4. The cellular services offer a much lower R-value than 80 but they have an advantage factor that adds to the total and improves the R-value. For new services, like satellite phones a correction value A is intoduced, to take into account the advantage of using a new service and to reflect acceptance of lower quality by users for such services. It is assumed that the Advantage Factor will be reduced over time as the service improves and the customers get used to the benefits of the new service. It is not recommended to include a non-zero Advantage

	Framing	Look Ahead	Coding	Decoder	Total Delay
G.711/10	10 msec	5 msec	1 msec	1 msec	17 msec
G.711/20	20 msec	5 msec	1 msec	1 msec	27 msec
G.729/10	10 msec	5 msec	10 msec	10 msec	35 msec
G.729/20	20 msec	5 msec	10 msec	10 msec	45 msec

Table 1 Delay introduced by various CODECs.

Factor for IP telephony because it is a replacement for existing services, rather than a completely new service.

CABLE TELEPHONY PERCEIVED CALL QUALITY

The perceived call quality of the cable telephony will be discussed using the e-model. As with the e-model the contributing factors of absolute speech delay and packet drop will be discussed.

Absolute Speech Delay

The absolute speech delay known as mouth to ear or one-way delay is a very important factor in the perceived voice quality. Figure 1 shows the drop in the voice quality in e-model with respect to absolute speech delay. To find the voice quality drop one has to calculate the absolute speech delay look up from the graph to find the delay related impairment in E-model [2].

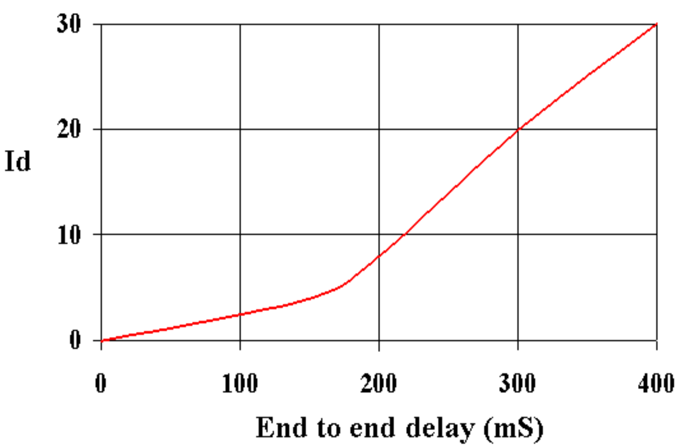

Figure 1 Impact of End-to-End delay using E-model

The absolute speech delay is contributed by many factors: Coding delay, Cable Access Delay, Network Side Delay and Jitter Buffer Delay.

Coding Delay

The MTA introduces a certain amount of delay due to framing, look ahead processing, and decoding. The delay introduced by the two PacketCable CODEC's are given in table 1.

Cable Access Delay

The delay introduced by the Cable Access network on the upstream direction depends on many assumptions. Some of the assumptions are listed below:

- The MTA's coding/de-coding clocks are slaved to CMTS DOCSIS master clock:

  If this assumption does not hold than the packets on the upstream-direction would experience a delay that is increasing/decreasing between 0 and 10 (the UGS interval) and then the same behavior will repeat as depicted in Figure 2.

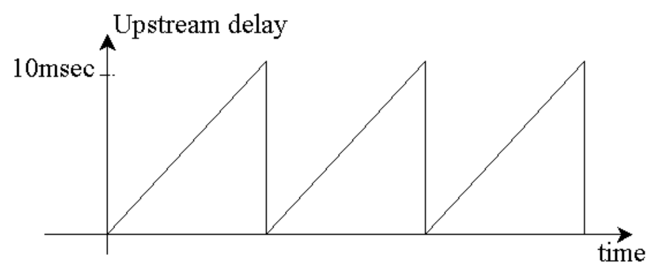

Figure 2 Upstream Delay when MTA's are not synchronized

- The MTA's framing interval will be aligned to UGS intervals:

  The first Dynamic Service Addition (DSA) message from the CMTS will include the time reference of the first UGS grant [4]. The MTA should align its framing interval such that the time between the end-of-framing and time to transmit upstream is minimized. If the time is not minimized or the framing is not aligned than there will be a constant value between 0 and 10 (UGS interval) upstream delay added.

If the MTA has implemented both the clock and framing synchronization than the delay on the upstream direction consists of:

Framing to transmit delay	<1 msec
Cable propagation delay	<0.8 msec
Cable receiving delay	<0.2
CMTS internal delay	<3 msec

A total of 5 msec is assumed for the upstream direction.

For the downstream direction cable delay consists of

CMTS internal delay	<3 msec
Interleaving/transmit delay	<1 msec
Cable Propagation delay	<0.8 msec
Reception to buffer delay	<0.2 msec

Making a downstream direction cable delay of 5 msec.

Network Side Delay

The network side delay depends on many factors such as the distance, the number of routers between end-points, the traffic and the connection technology. The end-to-end delay number for a national network is around 60-90 msec with jitter as large as 50 msec or more. When packet prioritization is being used the average delay remains about the same but the jitter reduces.

Jitter Buffer

The other delay factor is the jitter buffers on the de-coding section. The predicament with the jitter buffer is it is one of the items that is left for vendor differentiation.

The jitter buffer implementations can either be adaptive or static. On static jitter buffer implementations the buffer generally holds one or two packets, which spans one or two framing interval delays.

The goal of adaptive jitter buffer management is to remove the jitter while minimizing the amount of delay or incremental latency that is added to what's already been provided by the network. The adaptive buffer management schemes use interpacket arrival time variations, doing statistical analysis on it, then adapting the mean holding time of the packets or the jitter buffer length.

The problem with the static buffer management algorithms is that they tend to be conservative and assume the worst network cases. The MTA buffer management should be designed for end-to-end IP transport jitter values no matter where the call is connected. That means, if the worst-case jitter is 15 msec end-to-end then all the calls experience twice the jitter value delay of 30 msec.

The problem with the adaptive buffer management is twofold: First most of the adaptive buffer management statistical analyses schemes are not designed against the bursty nature of the network delay that is experienced today. Second, the buffer management is generally carried out during silence periods, which most probably does not coincide with the events that require jitter buffer changes.

The incorrect jitter buffer assignment has a two different impacts: When the jitter buffer is set to a value that is too low than the packets that arrive later then the buffered time will be dropped, and if the buffer is set to a too high value then the delay will be too high.

Depending on the network configuration and load the jitter on the VoIP packets may vary. In some cases some packets are so much delayed that setting the jitter buffer will result in an overall drop in voice quality.

Packet Loss

Almost all IP networks exhibit Packet Loss. Figure 2 shows the e-model impact of packet loss on G.711 CODEC, which is the only mandatory CODEC in PacketCable specifications [5]. As can be seen in figure below, the impact of packet loss is tremendous on the voice quality, if 1% of the voice packets are lost than the speech impairment due to packet loss is 25. Combining this with the starting point of the G.711 the quality of no-delay VoIP system with 1% packet loss results with a e-model rating of 70 which is equivalent to cellular phone quality.

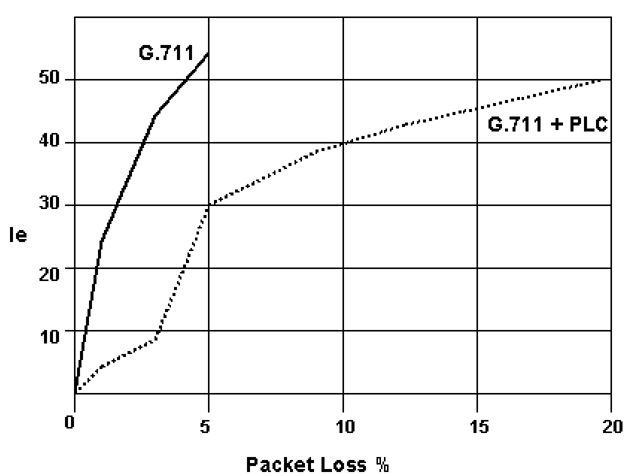

Figure 3 Impact of Packet Loss on E-model

The Packet Loss in IP networks can be attributed to several sources: queue overflow, synchronization, jitter buffer overflow/underflow, damaged packets. The impact of packet loss on the e-model impairments depends directly on the CODEC being used and whether Packet Loss Concealment (PLC) is implemented.

The impact of packet loss can be prevented if the packet loss concealment algorithms are implemented on the receiver side. As depicted in figure 3 the quality drops to 5 when PLC is implemented, which is one fifth of the non-PLC G.711 system. Unfortunately the PacketCable specifications do not mandate the implementation of PLC when using G.711 CODEC[1].

Packet Loss due to Queue Overflow

The queue overflow results when a certain interface receives more packets than it can send for a certain time duration. The solution for queue overflow is two fold: prioritization of packets, control of the queuing available to each priority.

The prioritization of packets is now a well-established standard in the IP world and is called Differentiated Services (DiffServ) [6,7]. The PacketCable standards already support the DiffServ packet marking. The DiffServ in general is the scheme that when a high priority packet is received that packet is being sent before the lower priority packets. Even when DiffServ marking is being used, it is still possible that at some funneling points, the router would receive a larger amount of high priority packets then it can handle and has to queue the high priority packets. In this case the packets either have to be queued for a long time or should be dropped. The issue with funneling is that it would cause all calls to be

[1] The term used for G.711 is ‘RECOMMENDED’ which is much weaker in terms of testing/certification viewpoint than use of the term ‘MUST’.

impacted. If a certain link can handle at most 1000 calls and 1001st call is being connected all 1001 calls will be impacted not only the one call that is added.

The issue of packet dropping in DiffServ environment can be prevented by over-designing the network with no funneling points.

Packet Loss due to Jitter Buffer Overflow/Underflow

The jitter overflow/underflow is caused by the jitter experienced by the packets. As stated before every MTA has a jitter buffer, which can be static or dynamic. Since at any instant the jitter buffer is perceived as being static, only the static jitter buffer case will be considered.

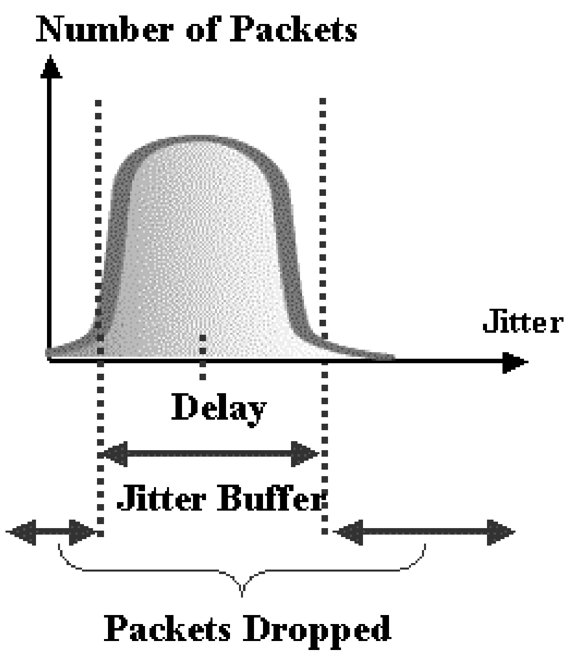

Figure 4 Packets Dropped and Jitter Buffer

Lets assume that the MTA has a jitter buffer of 20 milliseconds, which is actually a +/-10 milliseconds jitter buffer. What this means is that if a typical packet is received by time t0 then it will start to be played at time t0+20 milliseconds. Now assuming that the second packet is being received with a jitter of +10 milliseconds than it will be played without any jitter buffer induced delay. If a packet is early by 10 milliseconds (-10 millisecond jitter) than the packet will be delayed by 10 milliseconds. In short the jitter buffer regulates the delay variation on the packets and makes the impression that there is no jitter experienced by the packets only a constant delay of 20 milliseconds added.

The real issue comes to play when a packet experiences a jitter that is more than that of the jitter buffer. If a packet were earlier or later than the jitter buffer allocation, the MTA would drop the packet since it cannot handle the packet with in the jitter buffer. As shown in figure the jitter in a IP network can be depicted as a bell shape and the jitter buffer as a band within/engulfing this bell shape as shown in figure 4.

The jitter buffer may not be an issue if it can be set to a value that would always be able to accommodate the worst-case jitter in the network.

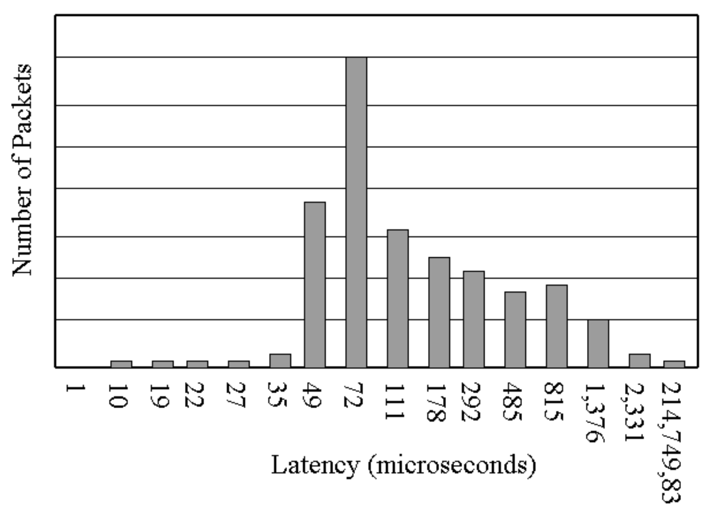

Figure 5 Example Core Router Jitter

Unfortunately this is not an easy task. Figure 5 shows a core router performance on a two-interface situation that no queuing is necessary. As can be noticed the spread of the delay is very wide [8]. If the delay from this one router is to be accommodated then a jitter buffer of 214 milliseconds will be required. The better approach is to set an acceptable packet-drop level and set the jitter buffer to this value around 2 milliseconds to get a packet drop around 1%.

Using similar analyses it can be shown that the best settings for IP network jitter buffer are around 1% packet drop. Any settings above the 1% packet drop will cause a larger delay introduced by the jitter buffer thereby dropping the voice quality faster than the voice quality improvements coming from the decreased packet drop rate. Any value above the 1% packet drop would result with a decreased voice quality due to the fact that the voice quality drop induced by the increased packet drop will not be compensated by the reduced delay.

Packet Loss due to Errored Packets

The packet drop due to errored packets is generally due to RF impairments in the cable plant. This is due to the fact that the modern IP transmission equipment provides reliable transmission with errored packets less than one in 10000.

In the downstream direction the packet-drop rate is in the order of 10^{-5} due to the fact that on the downstream the transmitter is more powerful and the bandwidth used for downstream transmission has better SNR (Signal to Noise) characteristics than the upstream direction [4].

On the upstream direction a provider has many possibilities that would impact the packet loss. Almost all of the countermeasures against the packet loss would have an impact on the perceived bandwidth on the upstream side. For example using a 1% packet loss a typical CMTS would be able to use a 2Mbps upstream channel whereas for a 10^{-5} packet drop rate a 612 Kbps will be achieved.

BUILDING AN EXAMPLE SYSTEM

Lets assume that the objective is to design a Voice over Cable system that would be competing against the PSTN landlines in North America.

Form the viewpoint of TDM based PSTN equipment the end-to-end performance of the PSTN network is impressive to say the least:

- Less than 1 in 1000 samples are dropped
- The end-to-end delay in North America is less than 35 msec.

When the cellular phones are taken as a base point these characteristics change as:

- As much as 3% packet loss.
- More than 200 msec of delay

The perceived quality of the cellular phone calls suffers from these characteristics.

CODEC Decision

Since the PSTN network will carry the voice as G.711, it is assumed that the G.711 will be used on the VoIP portion(s) of the networks. If this assumption is not valid than the initial coding loss and transcoding loss should be taken into account.

Absolute Speech Delay

Assuming that the aim is to be as close to PSTN landline services as possible the worst case has to be considered. The worst-case scenario for the Cable Telephony is that the call starts on Cable hops into PSTN and ends at Cable. As depicted in figure 1 the call starts in a user calling via a PacketCable certified MTA in Sunnyvale, CA to another user that has PacketCable certified MTA at the Providence, RI. The call goes to PSTN on the Gateway in San Francisco, CA, and then exits from PSTN on the Boston, MA. The delay on the PSTN segment between San Francisco and Boston is 30 msec.

Since the design is made for worst case, it will be assumed that the 10 msec packetization interval with G.711 coding will be used. Looking from the table 1 the coding and decoding delays of G.711 CODEC with 10 msec framing is 16 msec for coding and 1 msec for the de-coding, a total of 17 msec CODEC induced delay is found.

Since the coding is carried out on the originating MTA and de-coding on the ingress to PSTN, and coding on the egress from PSTN and de-coding on the terminating MTA, there two occurrences of coding-decoding in the Voice sections making 34 msec of CODEC delay.

The delay on the network side between the CMTS and the PSTN egress point is assumed to be 10 msec; when the cable access delay of 5 msec is added the total IP network side delay becomes 15 msec.

The Jitter Buffer delay is assumed to be 10 msec. The jitter buffer value of 10 msec should be sufficient for access to a local PSTN egress point but for the end-to-end VoIP call through the backbone this value may be too low, causing too much packet drop. Since the PacketCable does not have any means of setting the Jitter Buffer Size per call, the provider has to make a compromise between PSTN quality and end-to-end voice quality.

The total absolute speech delay consists of many pieces:

Origination	
Coding/Decoding	17 msec
Cable/Network Delay	15 msec
Jitter Buffer Delay	10 msec
PSTN end-to-end Delay	30 msec
Termination	
Coding/Decoding	17 msec
Cable/Network Delay	15 msec
Jitter Buffer Delay	10 msec
	+
Absolute Speech Delay	114 msec

Using figure 1 to resolve the E-Model quality drop the drop can be found as 4.

Packet Loss

The amount of packet loss contributed can be partitioned as:

Router queuing	0.1%
Jitter buffer	1%
Cable Access (RF)	0.01%

Making 1.11% packet loss in origin and 1% packet loss in destination. Looking at the impact of 2.22% packet loss on figure 3, the e-model quality drop can be found as 7.66.

E-Model Result

When the G.711 coding (just sampling the voice with 8000 times a second) is being used, the base for voice quality would start from e-model score of 94.2. Calculating the drop due to delay (4) and packet drop (7.66) would result with an end-to-end quality of the 82.6 which is barely above the desired limit of 80.

CONCLUDING REMARKS

Even though at first glance it looks like that the desired of score 80 can be achieved, some points are worth mentioning:

- The calculation is for the worst case and for most of the geographical locations the score would be higher.
- The packet loss of 1% would only be seen on cases where severe jitter is observed.

- The overall experience with respect to PSTN landline services would be lower due to the fact that the introduced absolute delay is higher[1].

- When connected to high delay endpoints such as cellular phones, PBX's or international calls with long delays, the call quality may drop to an unacceptable level.

- The use of CODEC's that provide better bandwidth utilization will cause the voice quality to drop further.

- The use of a bigger Jitter Buffer to accommodate the connections to other end-points would cause voice quality to drop further.

- Any additional packet loss would cause the voice quality to drop further.

REFERENCES

[1] CCITT Recommendation G.711 (11/88), *Pulse code Modulation (PCM) of voice frequencies.*

[2] ITU-T Recommendation G.114 (05/00), *General Recommendations on the transmission quality for an entire international telephone connection.*

[3] ITU-T Recommendation G.107 (05/00), *The E-Model, A Computational Model for use in Transmission Planning.*

[4] CableLabs SP-RFIv1.1-I09-020830 (08/02), *Data-Over-Cable Service Interface Specifications Radio Frequency Interface Specification*

[5] ITU-T Recommendation G.113 (02/96), *Transmission impairments.*

[6] IETF Request For Comments RFC2475 (12/98), *An Architecture for Differentiated Services*

[7] IETF Request For Comments RFC2474 (12/98), *Definition of the Differentiated Services Field in the IPv4 and IPv6 Headers*

[8] Light Reading (7/01); The Internet Core Routing Test: Complete Results; *http://www.lightreading.com/document.asp?site=testing&doc_id=6411*

[1] Due to added delay most of the long distance calls will sound like high quality international calls.