MOTOROLA
Broadband Communications Sector

SCTE 2004 Advanced VoIP Development Symposium

Engineering CMTS and HFC for VoIP with Capital and Operating Expenses in Mind

Jay Strater
Senior Systems Engineer
Motorola Corporation
101 Tournament Dr
Horsham PA, 19044
215 323 1362
jstrater@motorola.com

Steve Nikola
Senior Systems Engineer
Motorola Corporation
111 Locke Drive
Marlborough, MA, 01752
508 786 7571
Steve.Nikola@motorola.com

1. Introduction

Cable operators are increasingly looking to offer voice service (VoIP) in addition to video and data services as a strategy for wining subscribers from ILECs and keeping subscribers from defecting to satellite plus DSL offerings. However cable operators who make the decision to employ VoIP are often faced with a daunting learning curve. This is because VoIP introduces fundamentally new requirements.

VoIP introduces new traffic loads and traffic that is synchronous in nature, versus asynchronous like typical residential high speed data. VoIP also requires support for low delay and low packet loss. This is because VoIP is a streaming service where retransmission is not feasible. In addition VoIP requires high service availability. The level of availability depends on whether the VoIP service is intended to be a primary line service, available after a power outage or a secondary line service where service availability is not as critical.

Accompanying the VoIP requirements are new design choices too. For just the CMTS and HFC portions of a cable operator's network, VoIP design choices span:

- Multimedia Terminal Adaptor (MTA) configuration parameters
- HFC node and DOCSIS parameters
- CMTS and MTA configuration parameters
- CMTS and MTA features

As with any engineering decision, it is important to consider all costs and benefits before making a choice. Design choices should provide an appropriate balance between capital costs and operational costs. Considering just the CMTS and HFC portions of a cable network, this means balancing design goals for low capital cost such as:

- Maximum utilization of CMTS and HFC resources
- Limited CMTS and MTA features

with low operational cost such as:

- High QoS and availability to avoid issues from poor service
- Large traffic margin to avoid frequent CMTS / node reconfiguration

This paper addresses the engineering trade-offs for VoIP design choices in the CMTS and HFC portions of the cable network. The trade-offs span design choices for MTA configuration parameters, HFC node and DOCSIS parameters, CMTS configuration parameters, and CMTS and MTA features. Brief consideration is also given to CMTS and HFC related standards and future directions. The paper concludes with a summary and references.

2. MTA VoIP Configuration Parameters

There are many MTA configuration parameters but those with the largest impact on VoIP service include VoIP codec, packetization period, voice activity detection (VAD) state, and jitter buffer size. The VoIP codec defines the method by which voice is encoded and is often referenced in terms of its output bit rate. The VoIP packetization period is the period over which encoded voice bits are collected for encapsulation in packets. The VAD state refers to whether VoIP packets are sent all the time or only during talk periods, with VoIP packets not being transmitted during quite periods. The VoIP jitter buffer size defines the maximum delay and nominal play-out delay of a jitter buffer.

For those not aware, a VoIP jitter buffer is used at the receiving ends of a VoIP connection. This includes at MTA and Media Gateway (MG, the VoIP trunk interface to PSTN), which are need for "off-net" calls. A jitter buffer stores received, time-jittered VoIP packets that arrive within its time window. It then plays stored packets out, in sequence and at a constant rate for subsequent decoding. A jitter buffer is typically filled half way before playing out packets to allow early or late packet-arrival jitter compensation. We call this delay the play-out delay. If a packet is received outside a jitter buffer's time window it is dropped. As a consequence of a packet being dropped, a gap will occur in a jitter buffer's packet play-out sequence.[1]

The choice of the MTA configuration parameters impacts capital cost from a CMTS utilization perspective and operational cost from a VoIP QoS perspective. The following table highlights the trade-offs for the VOIP parameters.

Table 1 – MTA configuration parameter trade-off

Configuration parameters	CMTS utilization	VoIP QoS	Typical configuration today
Codec	Lower rate will reduce VoIP payloads allowing higher channel utilization but may impact MTA and MG processing capacity	But lower rate may degrade voice quality and degrade voice-band data performance	G.711, but G.728 and G.729E are optional recommendations iLBC and BV16 are mandated in PC1.1
Packetization period	Higher period will reduce VoIP overhead allowing higher channel utilization	But higher value will increase VoIP path delay and can increase PER	10ms or 20ms (20ms and 30ms only options for iLBC)
VAD state	VAD will reduce VoIP load allowing higher channel utilization	But VAD clipping will degrade voice quality	Avoided in cable applications
Jitter buffer size	No impact	Choosing a large jitter buffer reduces packet dropping from jitter but increases VoIP path delay	Variable, but not uncommon to see 15 ms play out and 30ms maximum

As seen in Table 1, a lower codec rate produces a lower rate voice stream thereby allowing higher channel utilization. However a lower rate codec also results in lower voice quality, and may degrade FAX and modem performance and prevent the passing of inband DTMF tones.[2] A lower rate codec also imposes a higher processing burden on the MTA as well as the MG. This may result in MTA call-mixing limitations and lower MG capacity.[3] Most cable VoIP trials and early deployments today use G.711 encoding with its 64kbps pulse code modulated output. But PacketCable, the cable VoIP standards forum, has several low-rate codec recommendations and requirements that are categorized as providing "toll-grade" voice performance with much lower output rates.[4]

The following table provides an example of the reduced-rate benefits of low rate encoding. It shows raw output rates and DOCSIS service flow rates (including overhead) for G.711, G.728, and G.729E encoding.[5] The example assumes 10 ms packetization periods and use of DOCSIS PHS, BPI+, and FEC.[6]

---

[1] A packet play-out gap will also result if a packet is received in error and dropped because it can not be corrected. This will occur prior to the jitter buffer. Whatever the cause for a packet gap, a packet loss concealment algorithm is typically used to help mitigate gaps in played out packets.

[2] T.38 FAX relay, V.151/2 modem relay, and RFC2833 DTMF signaling are currently being considered by PacketCable, in part, as means for allowing VoIP service to operate reliably with low rate encoding.

[3] CMTS utilization benefits from low rate encoding also may not be fully realized if low-rate encoding causes CMTS packet per second (PPS) limits to be reached. This scenario is most probable with small packetization periods.

[4] PacketCable 1.0 and 1.1 codec specifications ([1] and [2]) include options for G.728 and G.729E encoding, and PacketCable 1.1 codec specification includes mandatory requirements for iLBC and BV16 encoding. G.728 requires 16kbps and G.729E requires 11.8 kbps. iLBC requires 15.2kb/s with 20ms packetization and 13.3 kb/s with 30ms packetization. BV16 requires 16 kbps. A primary benefit of iLBC and BV16 is lower processing requirements than G.728 or G.729E.

[5] Service flow overhead includes overhead from RTP, UDP, IP, Ethernet, and DOCSIS protocols.

Table 2 – Example of codec impact on call rate

Codec	Codec output rate	QPSK upstream call flow rate	Downstream call flow rate
G.711	64 kbps	115.2 kbps	109.6 kbps
G.728	16 kbps	57.6 kbps	61.6 kbps
G.729E	12 kbps	57.6 kbps	57.6 kbps

Referring back to Table 1, we next see that a higher packetization period produces less packet overhead thereby allowing higher channel utilization. However a higher packetization period also causes higher VoIP path delay and can increase packet error rate PER.[7] Packetization delay can be codec dependent and nominally contributes around 1.5 times the packetization period to VoIP path delay.[8] The path delay contribution is the result of the packetization delay itself plus DOCSIS unsolicited grant service (UGS) grant uncertainly. UGS grant uncertainty is the time a newly generated voice packet must wait at the MTA's CM before receiving a UGS grant transmission opportunity. Nominally this delay is half the packetization period.[9] Because of these concerns, most cable VoIP deployments today use 10ms or 20ms packetization periods.

The following table provides an example of the trade-off between packetization-period delay and rate reduction. It shows nominal path delay and DOCSIS service flow rates for 10, 20, and 30ms packetization periods. The example assumes G.711 encoding and use of DOCSIS PHS, BPI+, and FEC.[10]

Table 3 – Example of packetization period trade-off

Packetization period	Nominal path delay contribution	QPSK upstream call flow rate	Downstream call flow rate
10 ms	15 ms	115.2 kbps	109.6 kbps
20 ms	30 ms	89.6 kbps	86.8 kbps
30 ms	45 ms	85.3 kbps	79.2 kbps

Referring back to Table 1, we next see that enabling VAD will reduce VoIP load since voice will only be sent when talking occurs. This allows more calls and/or data to exist concurrently, resulting in higher channel utilization. However enabling VAD will also degrade voice quality as the result of a voice clipping at the beginning of each talk period. All cable VoIP trials and early deployments that we know of today avoid VAD for this reason.

---

[6] PHS assumptions are for 41 byte savings in upstream and 12 byte savings in downstream. Upstream FEC assumptions are for long data grants with RS K=220, T=8, short data grants with RS K=78, T=5, and use of shortened last code words. In the example G.711 packets are assigned to long data grants and G.728 and G.729E packets are assigned to short data grants.
[7] PER may increase significantly with packetization period if error conditions tend to be random in nature, such as from thermal noise or short impulse noise.
[8] G.729E has an addition 5ms look ahead-delay and iLBC has an additional 5ms and 10ms look-ahead delay for 20ms and 30ms frames respectively.
[9] UGS grant uncertainty can be reduced by grant synchronization in the MTA. But grant synchronization effectiveness drops as more calls are added and channel changes occur.
[10] Lower rate encoding would realize a larger rate reduction for increased packetization periods. For example G.728 realizes nearly a 50% flow rate reduction between 10ms and 30ms packetization compared to around 25% for G.711.

Finally from Table 1, we see that MTA jitter buffer size has no impact on CMTS utilization and should be chosen based on VoIP QoS considerations. Jitter buffer size should provide an appropriate balance between jitter-induced packet drops and its contribution to VoIP path delay. In particular, its selection should be coordinated with VoIP PER considerations (addressed in Section 3) for a total packet loss rate target, and with the choice of VoIP packetization-period to meet a VoIP path delay target.

To determine an appropriate VoIP path delay target, first consider an appropriate maximum round-trip VoIP delay. 300ms is often used as a target for maximum round trip delay between an MTA and PSTN phone over a long distance connection. From this delay, 150ms is often targeted for end-to-end delay, with around 100ms assigned to PSTN long-distance propagation delay. The result in this case is a remainder of 50ms for local VoIP path delay.

Jitter buffer size, packetization period, and packetization-related grant uncertainty combine with other delays to form a VoIP path delay. The following figure provides an example of one-way VoIP path delay (upstream and downstream for an "on-net" call). The delay is broken down by component and sub-component contribution and assumes 10ms packetization and 15ms jitter buffer play-out delay.

Table 4 - Example one-way VoIP path delay budget

Network	Component	Function	Nominal Delay (ms)
Upstream HFC Access	MTA	Voice Packetization	10.00
		DSP Operations	5.00
		Packet Encryption	0.50
		Grant Uncertainty	5.00
	HFC	Upstream Transmission	0.50
	CMTS	CMTS Forwarding	1.00
Local IP Network	Routers		5.00
	Trunking GW	IP Processing	2.00
		RTP Decryption	0.50
		DSP Processing	3.00
		Jitter Buffer	15.00
Total			47.5

As seen in the example jitter buffer size and packetization period generally provide the largest delay contributions. Moreover, their selection will determine whether a VoIP path delay target can be met.

3. HFC Node and DOCSIS Parameters

HFC node parameters that impact VoIP service include average node size and the maximum number of nodes per CMTS receiver group. Average node size defines the average number of homes served by a fiber node. The maximum number of nodes per CMTS receiver group defines the maximum number of nodes that can be connected to one or more CMTS receivers.

The choice of these parameters impacts capital cost from a node and CMTS utilization perspective and operational cost from a VoIP QoS perspective. The following table highlights the trade-offs for the node parameters.

Table 5 - Node parameter trade-off

Node parameter	Node & CMTS utilization	Upstream VoIP QoS	Typical configuration today
Average node size	Larger node sizes require fewer fiber terminating nodes and fibers	Smaller nodes generally result in less ingress noise in an upstream channel	500 to 1000 homes, but with movement to <500 homes
Number of nodes per CMTS receiver (Rx) group	Higher node allowance per Rx group permits more nodes to be supported per Rx when service take rate is low	But larger allowance will also funnel more ingress noise in an upstream channel	Unclear, 1 to 2 seen

As seen in Table 5, larger node sizes reduce node equipment and fiber requirements but generally result in higher upstream ingress noise. In addition, higher node allowance per CMTS receiver permits greater design flexibility when take rates are low. However more nodes per receiver also generally results in more ingress noise being funnel into a receiver. Typical node sizes range from 500 to 1000 today. It is less clear what the typical node allowances are per CMTS receiver group.

DOCSIS parameters that impact VoIP service include upstream channel bandwidth, upstream and downstream modulation, upstream and downstream packet header suppression (PHS), and upstream forward error correction (FEC). DOCSIS allows for a variety of upstream bandwidth and upstream and downstream modulation choices. DOCSIS 2.0 offers higher upstream bandwidth and modulation options and, therefore, higher upstream capacity than DOCSIS 1.x. Please refer to [3] for a listing of DOCSIS 1.1 choices and [4] for a listing of DOCSIS 2.0 choices. PHS defines the amount of packet overhead savings in the upstream and downstream channels accompanying VoIP packets using UGS. FEC defines Reed Solomon coding parameters used in upstream channels.[11]

The choice of these parameters impacts capital cost from a CMTS utilization perspective and operational cost from a VoIP QoS perspective. The following table highlights the trade-offs for the DOCSIS channel parameters.

[11] Upstream FEC is defined for both short and long data grant intervals on an upstream channel. VoIP is assigned to the short data grant interval if the VoIP packet size with overhead is small enough to fit within the maximum burst interval of the short data grant. Otherwise it is assigned to the long data interval.

Table 6 - DOCSIS channel parameter trade-off

DOCSIS channel parameter	CMTS utilization	VoIP QoS	Typical configuration today
Upstream bandwidth	Choosing 3.2 MHz provides double the capacity of 1.6 MHz channel but requires support for a contiguous chunk of 3.2 MHz BW within upstream band	It also requires a 3 dB higher signal to noise ratio (SNR)	1.6 or 3.2 MHz
Upstream modulation	Choosing 16QAM provides double the channel capacity of QPSK	But 16QAM also has ~7 dB higher SNR requirement than QPSK at 10^{-6} BER, plus somewhat greater amplitude and phase noise sensitivity	QPSK or 16QAM (only choices)
Downstream modulation	Choosing 256QAM provides over 40% higher channel capacity than 64QAM	But 256QAM also has ~6 dB higher SNR requirement than 64QAM at 10^{-6} BER, plus significantly greater amplitude and phase noise sensitivity	64QAM or 256QAM (only choices)
PHS	Choosing PHS reduces VoIP overhead allowing higher channel utilization	No impact	~40 bytes upstream, not clear for downstream
Upstream FEC	Limiting the amount of FEC will help limit VoIP overhead and allow higher channel utilization	Increasing the amount of FEC improves noise and interference robustness, possibly permitting support for higher BW and modulation choices	Dependent on upstream conditions

As seen in Table 6, larger bandwidth and modulation choices provide higher channel capacity at the expense of increased noise and interference sensitivity. Doubling upstream bandwidth may have less impact on noise and interference than changing from QPSK to 16QAM. However it requires that adequate bandwidth be available. By contrast use of PHS generally has only positive benefit on VoIP applications.[12] Its benefit is more significant at lower voice coding rates. Finally increased upstream FEC increases VoIP overhead but it also improves error mitigation robustness. Typical DOCSIS bandwidth and modulation choices are shown in the table.

We recommend coordinating the choice of node and DOCSIS parameters so as to reach a target maximum PER for VoIP service while minimizing aggregate node and CMTS costs (as driven by utilization considerations). The maximum VoIP PER target should be chosen from a total VoIP packet loss rate target that also accounts for jitter-induced packet drops.[13] The total VoIP packet loss rate target should be chosen according to voice service needs.

The following table provides a “rule-of-thumb” for packet loss rates as a function of service conditions. The service conditions in the table range from sub-toll quality to toll-quality with voice-band data support (FAX and modem service).

[12] Utilization benefits of PHS may not be fully realized if the added processing requirements for PHS cause MTA or CMTS processing limits to be reached.
[13] Some margin could also be added for packet loss from network congestion. However this loss should be negligible if voice packets are given appropriate packet priority.

Table 7 - Packet loss rate service impact

Feature	Total packet loss rate		
	>= 1 percent[14]	>= 0.1 percent	<= 0.01 percent
Voice quality	Sub toll quality	Toll quality	Toll quality
Call completion rate	98%	99.5%	99.95%
FAX	Dropped connections and errors on page	Dropped connections and errors on page	Supported
Dialup modem	Dropped connections	Dropped connections	Supported

4. CMTS Configuration Parameters

CMTS configuration parameters that impact VoIP service include the maximum bandwidth to allocate to VoIP service and the ratio of node-to-CMTS receiver assignment. The maximum bandwidth allocation for VoIP defines the maximum upstream and downstream bandwidths allowed for VoIP service at peak VoIP loading. This allocation is essential when data is sharing the same DOCSIS channel resources because it prevents VoIP calls from monopolizing all available bandwidth.[15] The ratio of node-to-CMTS receiver assignment defines the number of nodes assigned to a CMTS receiver group and the size of the receiver group in terms of number of receivers. For example the ratio may be 1 node per receiver group of 2 receivers.

The choice of CMTS parameters impacts capital cost from a node and CMTS utilization perspective and operational cost from a CMTS and node reconfiguration and VoIP QoS perspective. The following table highlights the trade-offs for the CMTS configuration parameters.

Table 8 – CMTS configuration trade-off

Configuration parameters	Node and CMTS utilization	Reconfiguration and VoIP QoS	Typical configuration today
Maximum VoIP bandwidth allocation	Optimal choice is based on current VoIP and data traffic conditions can maximize CMTS utilization	But padded traffic margin is needed to avoid frequent reconfiguration	Dependent on data traffic load, but typically between 40% to 60%
Ratio of node-to-CMTS Rx assignment	Allowing a ratio with CMTS Rx group size > 1 can prevent the need for node splitting when service take rate is high and even improve VoIP resource utilization but it also requires >1 upstream channel to fit in available US band	Allowing CMTS Rx group size >1 has indirect benefit of allowing greater design flexibility[16]	Unclear, but Rx group size >1 is likely to increase as VoIP take rate grows

As seen in Table 8, the use of current versus padded traffic margin when choosing a maximum VoIP bandwidth allocation provides a trade off between CMTS utilization and frequent reconfiguration. In addition, a node-to-CMTS receiver assignment that allows more than 1 receiver per receiver group can avoid the need for node splitting

[14] As a rule of thumb we choose 1% as the maximum packet loss rate to still allow communication in a VoIP call. This value may actually be higher. The PacketCable 1.0 Codec specification [1] indicates: "Anecdotal codec research suggests initial 3% packet loss rate results, on average, in a reduction in Mean Opinion Score (MOS) scores of 0.5 point, on a scale of 5. Above 3%, codec performance falls off rapidly, and resulting voice quality is unacceptable."

[15] Naturally data is allowed access to VoIP bandwidth allocations when VoIP service load is less than its peak load.

[16] This can include allowing "m" nodes to be assigned to "n" Rx in a CMTS Rx group.

at high service take rates. In this case dynamic load balancing is required within a receiver group to effectively "pool" Receiver resources and avoid an uneven distribution of VoIP calls and data sessions among receivers.[17] Aside from preventing utilization issues, "pooling" of receiver resources has the advantage of improving VoIP utilization.[18] Assigning "n" receivers per receiver group, however, comes at the expense of adding upstream bandwidth from "n" upstream channels.

The optimal choice of maximum VoIP bandwidth allocation and the ratio of node-to-CMTS receiver assignment requires an iterative computation to determine the maximum take rate for a given node size (or node size maximized for a given take rate). The computation requires voice and data traffic load assumptions along with traffic margin (typically added to take rate but could involve setting aside some receivers). In addition the computation requires the choices for MTA configuration parameters and node and DOCSIS parameters, as well as CMTS hardware design and performance constraints. The following figure illustrates this process.[19]

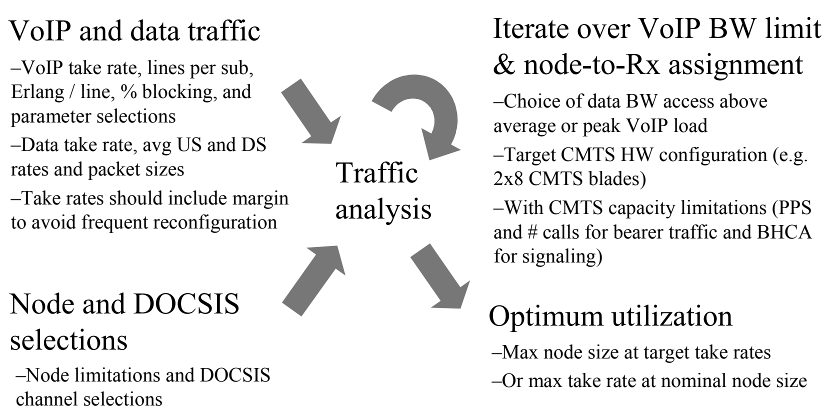

Figure 1 - CMTS configuration optimization

If or when VoIP and data traffic loading grows beyond the capacity of a current CMTS configuration, CMTS reconfiguration will be required. In this case optimal selection of maximum VoIP bandwidth allocation and the ratio of node-to-CMTS receiver assignment should be recomputed using new traffic load assumptions and parameter choices.

5. CMTS and MTA Features

CMTS features that impact VoIP include redundancy for rapid failover and hitless software download, synchronization to BITS clock, ingress noise cancellation, and advanced spectrum management. Redundancy for rapid failover and hitless software download refers to the ability of the CMTS to avoid cutting off existing VoIP calls (or data sessions), and to have minimal impact on new VoIP call establishment when a CMTS failure or download occurs. We distinguish between the timing for rapid failover and hitless download because failover typically requires a detection period that makes failover noticeable. Synchronization to BITS clock refers to having the CMTS connected to a BITS clock for more accurate timing than by using its internal clock. Ingress noise cancellation refers to the CMTS employing upstream signal-processing to cancel ingress noise. Advanced spectrum

[17] Dynamic load balancing allows a CMTS to shift VoIP calls and data assignments amongst receivers (upstream channels) in a Rx group at any time during a call or session. It employs DOCSIS dynamic channel change (DCC) messaging for hit-less channel changing.
[18] Pooling of Rx resources has the advantage of improving VoIP utilization because it decreases the ratio between peak and average VoIP call requirements within a resource pool. This decrease is dictated by the size of the resource pool, the blocking probability assigned to VoIP service, the algorithm used to compute maximum call capacity for a given trunk group size within the pool.
[19] Details of the computation would require another paper given their extensive nature so are not provided here.

management refers to the CMTS providing the ability to make accurate signal-to-noise and impairment measurements over any portion of the upstream band, and using an idle CM to avoid impacting service.

MTA features that impact VoIP are the use of FAX and modem relay and primary line service. FAX and modem relay refer to implementation of ITU T.38 and V.151/2 standards respectively for sending FAX and modem signals as encoded content over a VoIP connection, and reconstructing them at the terminating endpoint. Primary line service refers to the ability for the MTA to continue providing VoIP service during a power outage.

The choice of CMTS and MTA parameters impacts capital cost from a component perspective and operational cost from a QoS and service availability perspective.[20] The following table highlights the trade-offs for the CMTS and MTA features.

Table 9 - CMTS and MTA feature trade-off

Feature	Components	VoIP QoS and availability	Typical configuration today
CMTS redundancy for rapid failover and hitless software downloads	Additional CMTS blades are required to provide hardware redundancy	Redundancy with rapid failover is essential for “toll grade” VoIP service	Rapid failover typically employed, hitless failover typically not employed yet
CMTS synchronization to BITS clock	Cost of BITS source is required (typically a GPS BITS) but cost can be amortized over multiple CMTS that share a source	BITS sync is needed for “toll grade” FAX and modem service if FAX and modem relay are not employed	See use of BITS sync in some deployments and not in others
CMTS ingress noise cancellation	Feature is typically available in DOCSIS 2.0 CMTS but employing it may permit higher modulation selection / CMTS utilization	Mitigating ingress will benefit both QoS and availability	Not employed yet
CMTS advanced spectrum management	An additional Rx per CMTS blade is required	Constant monitoring of HFC plant without impacting service will benefit system availability	Not employed yet
MTA FAX and modem relay	Will avoid need for CMTS BITS clock sync but may impact MTA and MG processing capacity	Will avoid FAX and modem performance issues with low rate encoding	Typically not employed yet in cable applications
MTA primary line service	Requires UPS for in-home MTA or plant powering for side-of home MTA	Allows service availability during power outages	See in some deployments and not in others

As seen in table 9, CMTS redundancy for rapid failover and software downloads introduces the expense of additional CMTS components. Rapid failover is essential, however, if “toll grade” VoIP service is to be offered. Hitless software upgrade may not be as critical because it can be conducted during early morning hours when service impact is minimized. In addition, CMTS BITS synchronization introduces the expense of a BITS source but this feature is needed if “toll grade” FAX and modem service is to be offered and a receiving MTA does not use an adaptive jitter buffering algorithm, or FAX and modem relay are not employed. An appropriate BITS source will provide a CMTS and its MTA, which derive their clock sync from the CMTS’s downstream sync messages, with an accurate enough clock to permit long FAX and modem connections.

[20] By availability here we mean VoIP service availability following a CMTS or MTA hardware or software failure. VoIP service availability can also be effected by CMTS resources being saturated from peak traffic loading. This condition is minimized by selecting a small VoIP blocking probability that factors into CMTS configuration selections. VoIP blocking probability is an input to CMTS configuration selection as shown in Figure 1 in Section 4. VoIP service availability can also be effected by VoIP signaling packet loss caused by errors or dropped packets. Parameter selection to constrain PER and jitter-induced packet drop losses was addressed in Sections 2 and 3.

When considering what BITS clock accuracy to use for CMTS, we recommend considering the time required to over or underflow a jitter buffer and accounting for jitter buffer size and the clock accuracy on the far end of VoIP connection.[21] Assuming an over or underflow results in loss of a FAX/modem connection, then the time between such events will represent the maximum supported FAX/modem connection time.

The following table provides an example of the estimated timing slip interval for different Stratum clock sources on the local end of a connection. It includes different jitter buffer play-out delays and assumes that the remote clock source is only a Stratum 3 clock.[22]

Table 10 – Example of timing slip intervals with a Stratum 3 Clock

Stratum Level of Remote Timing Source	Expected Slip Interval		
	10 ms Jitter Playout	15 ms Jitter Playout	20 ms Jitter Playout
1	36 minutes	54 minutes	72 minutes
2	36 minutes	54 minutes	72 minutes
3E	30 minutes	45 minutes	60 minutes
3	18 minutes	27 minutes	36 minutes

In this example if it is deemed critical to avoid having a timing slip of less than an hour, to allow most FAX calls to operate without interruption, then a Stratum Level 3E or higher clock with 20ms jitter play-out will be required at the local endpoint. However if it is deemed critical to avoid having a timing slip of less than several hours, to allow most modem calls to operate without interruption, then a more accurate remote-endpoint clock should be considered.

Referring back to Table 9, we next see that CMTS ingress noise cancellation and advanced spectrum management, will provide additional levels of QoS and availability for mitigating ingress and allowing plant conditions to be constantly monitored. The ingress noise cancellation feature comes at the expense of requiring a DOCSIS 2.0 CMTS that supports this feature. It is only found in certain 2.0 CMTS chipsets. Advanced spectrum management comes at the expense of requiring an advanced CMTS that includes an extra receiver and appropriate capability for constant, uninterrupted service monitoring.

Referring back to Table 9, we next see that MTA FAX and modem relay will avoid the need for CMTS BITS clock sync and avoid FAX and modem performance issues with low rate encoding. However FAX and modem relay may impact MTA and MG processing capacity, much like that of low rate encoding. Consequently, their cost on MTA and MG capacity needs to be weighted against the savings in avoiding CMTS BITS sync (when adaptive jitter buffering is not employed) and avoiding low rate encoding issues. T.38 FAX relay and V.151/2 modem relay are currently being considered by PacketCable as noted previously.

Finally from Table 9, we see that MTA primary line service permits service availability during power outages but this comes at the expense of an indoor UPS or plant powering. We see most MTA interest today with indoor MTA. When supporting primary line UPS, the latest indoor MTA include embedded UPS. Moreover they utilize Lithium-Ion batteries to simplify customer-battery replacement.

---

[21] Any clock source will introduce a clock sync error that will periodically result in a loss of multiple VoIP packets from jitter buffer overflow or underflow if an adaptive jitter buffer algorithm that provides gradual Pulse Code Modulation (PCM) sample addition or removal is not employed. Without adaptation, a jitter buffer will under or overflow, requiring the jitter buffer to re-center is nominal play-out delay. Such an event may be often enough to cause a FAX or modem connection to terminate.

[22] Jitter buffer size can be selected independently for FAX and modem calls versus voice calls. In this case a call may start with a jitter buffer setting for voice and adjust to a FAX and modem jitter buffer setting upon FAX or modem detection.

6. Other Considerations

Clearly standards create competition which lowers capital cost. They also create common interfaces which help to lower operating cost. At a minimum, CMTS and MTA should support DOCSIS 1.1 [5] and PacketCable 1.0 [6] standards for VoIP over cable plants. Consideration should be given for support of DOCSIS 2.0 standards [7]. As noted previously DOCSIS 2.0 allows for higher upstream capacity and in some CMTS chipsets, provides improved noise handling.

In addition, consideration should be given to choosing a CMTS that currently supports the PacketCable 1.1 Electronic Surveillance Protocol (ESP) specification [8]. PacketCable ESP provides a specification for achieving Communications Assistance to Law Enforcement Act (CALEA) in a PacketCable application.[23]

PacketCable Multimedia standards [9] should also be considered when choosing a CMTS since they permit access to DOCSIS QoS for multi-media services, such as video telephony. A typical application will have a video telephony endpoint interfacing with its application manager to establish a connection. At this point the application manager will contact a policy server to authorize DOCSIS QoS for the endpoint. The policy server will then contact the CMTS to establish an appropriate set of DOCSIS service flows with needed QoS.[24] Naturally additional multi-media applications will introduce new CMTS / DOCSIS channel engineering considerations for optimizing resource utilization.

Finally we anticipate CMTS having more modular designs in future. We expect them to evolve with decoupled downstream transmitter and upstream receiver elements. As such they should allow for more resource-optimized downstream to upstream configurations. When these solutions become available, the same approach to optimizing CMTS configuration that was outlined in Section 4 can also be applied. The only difference is that downstream transmitters and upstream receivers will be variables in the optimization.

7. Summary

This paper addressed engineering trade-offs for VoIP design choices for CMTS and HFC portions of the cable network. The trade-offs addressed: MTA configuration parameters, HFC node and DOCSIS parameters, CMTS configuration parameters, and CMTS and MTA features.

The MTA configuration parameters that were addressed included: VoIP codec, packetization period, VAD state, and jitter buffer size. The choice of these parameters provided a trade-off between CMTS utilization and VoIP QoS.

The HFC node parameters that were addressed included: average node size and number of nodes per CMTS receiver group. The choice of these parameters provided a trade-off between node and CMTS utilization and upstream VoIP QoS.

The DOCSIS parameters that were addressed included: upstream bandwidth, upstream modulation, downstream modulation, PHS, and upstream FEC. The choice of these parameters provided a trade-off between CMTS utilization and VoIP QoS.

The CMTS configurations that were addressed included: maximum VoIP bandwidth allocation and the ratio of node-to-CMTS receiver assignment. The choice of these parameters provided a trade-off between node and CMTS utilization and reconfiguration and VoIP QoS.

The CMTS and MTA features that were addressed included: CMTS redundancy for rapid failover and hitless download, CMTS synchronization to BITS clock, CMTS ingress noise cancellation, CMTS advanced spectrum management, MTA FAX and modem relay, and MTA primary line service. The choice of these parameters provided a trade-off between added component costs and VoIP QoS and availability.

The paper also provides a brief consideration of CMTS and HFC related standards and future directions.

---

[23] ESP specifies call and data content capture and the protocol for conveying this information from a cable operator delivery function (DF) to a law enforcement collection function (CF).
[24] For video telephony this will likely be a DOCSIS real time poling (RTP) flow in the upstream channel.

8. References

1. PacketCable™ Audio/Video Codecs Specification, PKT-SP-CODEC-I05-040113
2. PacketCable™ Audio/Video Codecs Specification, PKT-SP-CODEC 1.1-I01-040621
3. Data-Over-Cable Service Interface Specifications, DOCSIS 1.1 Radio Frequency Interface Specification, SP-RFIv1.1-I10-030730
4. Data-Over-Cable Service Interface Specifications, DOCSIS 2.0 Radio Frequency Interface Specification, SP-RFIv2.0-I05-040407
5. http://www.cablemodem.com/specifications/specifications11.html
6. http://www.packetcable.com/specifications/specifications10.html
7. http://www.cablemodem.com/specifications/specifications20.html
8. PacketCable™ Electronic Surveillance Specification, PKT-SP-ESP-I03-040113
9. http://www.packetcable.com/specifications/multimedia.html