QUALITY OF SERVICE - IT IS WHAT IT IS, AND IT AIN'T WHAT IT AIN'T

Doug Jones
YAS Broadband Ventures, LLC.

Abstract

What is Quality of Service (QoS) really? It impacts the network and the business. It promises the ability to offer new revenue-generating services. But QoS is not a panacea. To offer QoS services, more than just the DOCSIS network will need to be upgraded to QoS technology.

The specific QoS technology is not so important as knowing why and where QoS should be deployed in the first place. This talk defines what QoS is, and then how an operator can use it in the network.

INTRODUCTION

As with any new business, it's important to both understand the technical issues and have a business strategy for deployment. It's generally not good to get excited by a new technology without fully understanding the issues.

From an engineering perspective, QoS is a set of technologies that can be used to control the delivery of services over various types of connections. Services based on QoS could be categorized as follows:

- Guaranteed throughput
- Guaranteed latency
- Guaranteed packet jitter

QoS is what it is, which is a way to offer new revenue generating services. QoS is not a panacea to all things. For example, deploying a QoS technology is not a substitute for proper network engineering. Deploying QoS will require new expenditures for technology and operations, and the promise is new services revenue will recover these costs.

These are valuable services and subscribers may be willing to pay for them. Deploying QoS technology holds the promise of additional revenues. But only in the presence of a realistic business plan does it make sense to begin upgrading the network for QoS. However, the promise is a QoS-enabled network and services are the basis for broadband interactive services going forward.

Standard Internet QoS protocols and technologies are developed within the Internet Engineering Task Force (IETF). However, many of these protocols are new and multivendor interoperability is not always guaranteed. There are also many proprietary QoS technologies offered by suppliers as a means of holding market share

QoS technology is specific to the underlying type of connection. Cable networks use many types of connections, from DVB-ASI hops, to 100Base-T hops, to a trunk connection to a telephone network. Each type of hop can have its own QoS technology. One size does not fit all. In order to offer end-to-end QoS services, the network may need to deploy more than one QoS technology in order to have QoS on all the different hop types. Having a single network for all types of traffic would make deploying QoS simpler as fewer technologies may be needed.

QoS technology is included in most CableLabs™ projects including DOCSIS™, PacketCable™, and CableHome™. It's well known that DOCIS™ 1.1 provides QoS, and it is one intent of this paper to describe how this fits into the big picture of a QoS-enabled network.

DOCSIS QoS is only available over one hop of a network, the cable connection between a CMTS and a CM. To offer an end-to-end QoS-based service, DOCSIS 1.1 will not be enough. Operators will need to consider adding QoS technology to other elements of their network to provide true end-to-end QoS.

WHAT IS QoS

For the purpose of this paper, QoS is the ability to guarantee network resources on a particular hop through a network for a sample of data. Network resources generally center on bandwidth, latency, and jitter guarantees.

A QoS technology is generally specific for the particular technology used for the hop on the network. In this case, a hop is a point-to-point connection. Examples of these include:

- a 100Base-T connection between a CMTS and a router on the network
- a downstream connection between a QAM modulator and a set top box
- The DOCSIS connection between a CM and a CMTS

In some cases, the same data may be subject to multiple QoS technologies because it is carried over multiple hop technologies in the network. Figure 1 shows an example voice call that uses three QoS technologies. The voice connection between the CM and CMTS uses DOCSIS QoS. That same voice call between a CMTS and a gateway still needs QoS, but this QoS will be based on the technology of that particular hop, probably a 100Base-T. From the gateway to the PSTN, that phone call is generally placed on a constant bit rate 64 kbps connection, either on a GR-303 or a T1. To offer one phone call, each of these three hops needs underlying QoS that guarantees throughput, low latency, and low jitter. Without QoS on all three hops, the voice quality could not be guaranteed. In this example, only having DOCSIS QoS may not be enough to ensure the call would have acceptable voice quality.

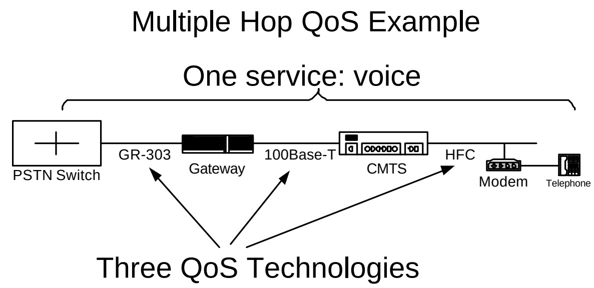

Figure 1

Up to this point, QoS has been described as a particular technology to guarantee a network resource on a network hop. Not only is this “low layer” QoS technology needed, but a higher layer technology is also needed to signal QoS on an end-to-end basis. An example is the Resource reSerVation Protocol (RSVP).

RSVP is used to signal QoS from one end of a connection to the other end of the connection. RSVP is not a “hop” QoS technology, but a method to signal what QoS is needed to allow QoS to be configured on each hop of the connection. Figure 2 gives an example of this.

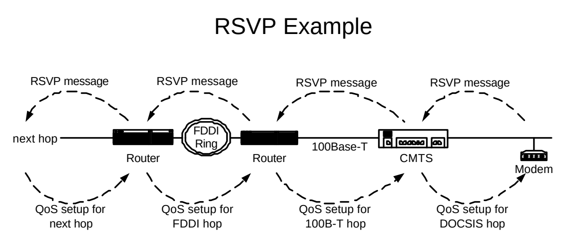

Figure 2

An RSVP message contains a high level description of what QoS is needed for that session. On each hop across the network, that high level description of QoS is translated into the lower layer QoS parameters needed to actually guarantee throughput, latency, and jitter on that particular hop. Since each hop has its own specific QoS, a higher layer messaging is needed to indicate what is needed, and then the individual hops translate that to the actual QoS parameters needed.

QoS On Specific Hops

As introduced, there is generally a specific type of QoS for each hop through the network. This section will look specifically at several hops in a cable network as illustrative examples. In some cases QoS is easier that might be thought.

Digital Video over QAM

Consider a 6 MHz wide 64 QAM channel that has a nominal capacity of 27 Mbps. Digital video streams can be added to this hop one at a time, until the 27 Mbps is used up. QAM hops are generally operated at or below total capacity. If the hop were filled to capacity, some of the digital video would have to be dropped. This, of course, would have a devastating affect on the service.

This illustrates an interesting point about QoS. If the hop is not planned to be overloaded, QoS may not be needed. If the hop is planned to be overloaded, then some bits are going to get through and some others are not. This is when a QoS technology is needed. A 27 Mbps hop cannot carry 28 Mbps, or even 27.01 Mbps. QoS technologies are needed on hops where the hop can be overloaded to guarantee that some bits get preferential treatment. When the hop is intended to be overloaded, a QoS technology can guarantee service for some of the bits but not all of them.

While always having enough (or too much) bandwidth and capacity available on the network would solve many issues, it would probably be economically prohibitive to over-engineer a network like this. Faced with that reality, the engineers developed QoS technologies to ensure that some bits can be guaranteed better service even when the hop is overloaded.

Data over DOCSIS

The DOCSIS network is defined on the hop between a CM and CMTS. DOCSIS 1.0 offers Class of Service (CoS), but not QoS. DOCSIS 1.1 does offer QoS.

DOCSIS 1.0 CoS is essentially a best-effort service that offers bandwidth limits. When the hop is full, the bandwidth limits won't come into play and the service is just best-effort. The network will do the best it can to be fair to the users, but there are no guarantees. When the hop is full, the network has no choice except to drop packets. For DOCSIS, it's the CMTS that decides which packets will be transmitted, and which will be dropped. There are many algorithms designed to drop packets randomly, such as Random Early Detection (RED) and Weighted Fair Queuing (WFQ). These could be considered QoS protocols, but they are really a means of choosing packets to drop as opposed deciding which packets will get through.

The DOCSIS hop is generally expected to be overloaded at peak usage hours and packets will get dropped. However, data services such as email and web browsing are tolerant to packet loss, unlike digital video or voice. In the case of email and web, if a packet gets dropped there are "higher layer protocols," such as TCP, that cause the dropped packet to be retransmitted. So while there may be a delay of a few tens or hundreds of milliseconds, the packet will eventually get through. Internet protocols are designed to be very forgiving and can recover gracefully from packet loss.

DOCSIS 1.0 offers what's known as Class of Service (CoS), which is different than QoS. DOCSIS 1.0 CoS can be used to implement bandwidth limits, both upstream and downstream, on modems. These are bandwidth limits, not guarantees, hence the difference between CoS and QoS. With CoS, there is no guarantee.

With bandwidth limits, the user is throttled to that amount of bandwidth even if there is additional capacity on the network that could be used. Conversely, users are not guaranteed that amount of bandwidth. Bandwidth limits are particularly useful on the upstream because this lightens the load on the hops that connect from the CMTS to the Internet. For downstream rate limits, there is an argument that if a packet comes all the way to the CMTS, it should be delivered regardless of any rate cap that may be in place. If that packet is dropped at the CMTS, then higher layer protocols will probably cause it to be retransmitted, which means additional load on the backbone. This is arguable unnecessary traffic and an issue where engineering and marketing can respectfully disagree.

DOCSIS 1.1 QoS it is possible to provide QoS guarantees for bandwidth, latency, and jitter. However, if the DOCSIS 1.1 hop is oversold, it cannot guarantee the delivery of all QoS intended to be sent over it. If the DOCSIS 1.1 upstream channel is configured to be 5 Mbps, then only 5 Mbps can be carried. The services guaranteed for delivery over that hop should not total more than 5 Mbps; it's simply not possible. Rather, QoS for guaranteed services can only be offered to a level below the total capacity of the hop. There should be some "wiggle room" on the hop, probably best-effort services that can be degraded in order to meet peak QoS loads.

DOCSIS 1.1 only offers QoS between the CM and CMTS. Just because there is a guarantee to deliver a packet on the DOCSIS hop does not mean there is a guarantee to carry that packet over the rest of the network with QoS.

Voice over DOCSIS

Voice service using DOCSIS can be offered using either of two methods. One is through a gateway to the PSTN, and the other is through a softswitch.

In the first case, there is a DOCSIS connection between the CM and CMTS, and then an IP connection to a voice gateway, and finally a constant bit rate connection to the PSTN.

In this case, there is DOCSIS QoS to the CMTS, and then generally an underutilized 100Base-T hop between the CMTS and voice gateway. If this hop is underutilized (that is, never over subscribed), then there may be no need for using a QoS technology on this hop. In cases like this, just a lot of available bandwidth is a viable way to provide QoS. If there is more available bandwidth (duplex 100 Mbps) than phone call traffic, then all the traffic fits nicely on the 100Base-T hop, and no QoS is needed. Once at the gateway, the voice traffic is converted onto a normal telco hop, like a GR-303 hop, that provides constant bit rate QoS for the voice call. So one phone call, while going from CM to PSTN switch, will traverse at least 3 different types of hops, will get 3 different types of QoS, but all those hops will offer the QoS necessary to maintain good voice quality.

In the second case, there is DOCSIS QoS between the CM and CMTS, but from the CMTS into the soft-switch network, the connection is all IP. The IP connection between the CMTS and the softswitch will probably require a QoS technology, and here there are many choices. If it's a 100Base-T connection, or Gigabit Ethernet (GigE), the QoS technology could be either Differentiated Services (DiffServ), which provides for priority forwarding through router hops, or Multiprotocol Label Switching (MPLS), which provides for a switching through router hops. If the connection from the CMTS to the softswitch is ATM-based (but still IP), the ATM link layer will provide QoS. There are many technologies available, and operators should make choices with strong input from the business case.

QoS Over All Hops

To provide an end-to-end session with QoS, all hops along that path need to offer QoS. It's a simple statement, but it has big implications.

As mentioned before, DOCSIS QoS is only between the CM and CMTS. This is just one hop in the network where there will be multiple hops to get end-to-end. If the particular data stream needs QoS from say a computer on the east coast to a computer on the west coast, then more than just DOCSIS QoS is needed. The connection between the computer and the CM can use DOCSIS QoS, but the connections between the CMTSs will entail several router hops and these will need QoS. If these router hops occur on different networks, for instance the cable operator network and a 3rd party network, then QoS will need to be coordinated through a peering agreement, which is a business issue.

If the connection requires a hop over the public Internet, then it may not be possible to offer end-to-end QoS. The public Internet provides only best effort-service, and that's probably all it will ever offer.

Operators will have to study their networks closely to decide where to deploy QoS technology. In some cases, maybe all the hops in their network will need QoS, but maybe only certain paths through the network will need QoS. These decisions will be key as operators continue to build their networks for both IP and MPEG services. Having a single network for voice, data, and video services will have benefits of using fewer technologies.

QoS Peering Agreements

Once the operator owned network is enabled for QoS, the next decision will be to peer (interconnect) with additional 3rd party networks that support QoS. Such arrangements would make it possible to offer more services to a larger number of subscribers.

Peering requires more of both technology and business. First, business agreements will be needed to monitor the connection, to ensure that subscribers get the QoS for which they are paying, etc. Also the network being peered with may use a different QoS technology, therefore, gateways and translators will be needed to ensure that QoS on one network gets carried with the proper QoS on the other network. Eventually, there should be coast-to-coast and worldwide networks that offer QoS. These QoS networks can run in parallel with the public Internet and will provide an alternative for new services.

QoS Over DOCSIS

In order to provide QoS, the technology on the hop has to provide an underlying mechanism that allows for controlling throughput, latency, and jitter. Not all hop technologies or configurations support QoS.

This final section of the paper is intended to provide technical detail on how DOCSIS provides QoS. Again, this QoS is only on the hop between the CM and CMTS, but it's interesting to know how it works. DOCSIS QoS is also compared to DSL QoS.

Upstream DOCSIS QoS

The cable data return path has the unique property of having many attached CMs that may all want to transmit at the same time. Clearly this needs to be controlled, and in fact the CMTS is the "traffic cop" on the hop that tells each CM when it can transmit, and for how long.

On the return path, the time is divided into periods called minislots. The minislot size is based on a number of parameters, but can be any of 1, 2, 4, 8, 16, 32, 64, 128, or 256 bytes of data. The size of the minislot is fixed on the return path, but will be one of the above.

Several of the minislot sizes just do not make practical sense and are rarely used, e.g., the smallest and largest sizes. For example, if the minislot size were 256 bytes and the data to be transmitted were only a 64 byte TCP ACK, then a lot of bandwidth would be wasted.

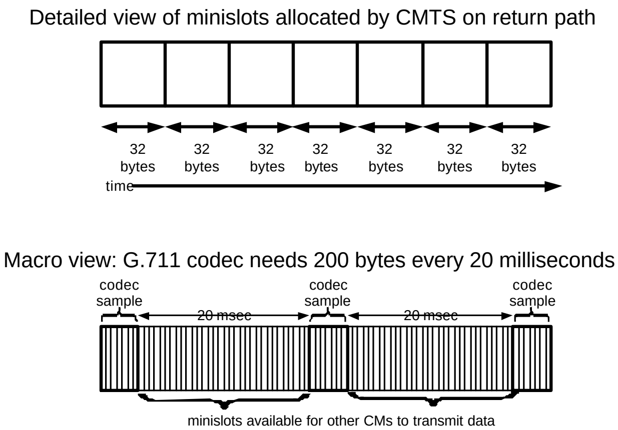

Figure 3

The CMTS is generally configured to use a minislot size in the "sweet spot" of from 8 to 64 bytes. Note that the chief competitor to cable data, DSL, is based on ATM which uses a 53-byte cell. DOCSIS provides comparable QoS as compared to ATM; in fact, DOCSIS can provide even finer granularity by using the 8, 16, or 32 byte minislots.

Figure 3 gives an example of how minislots can be allocated on a return path. In the top part of the figure, the conceptual view of 32 byte minislots is shown. In the lower part of the figure, the example shows how groups of minislots can be allocated for a single phone call. The CMTS assigns groups of minislots to the CM for upstream data transmission. The remaining minislots could be assigned to other CMs for upstream data transmission.

Downstream DOCSIS QoS

The cable data forward path is different from the return in that only one device transmits, the CMTS. Therefore, on the forward path the CMTS internally queues and transmits packets as necessary to meet all the QoS guarantees. There are no minislots on the forward path; these are available only on the return path. But because there is only one transmitter on the forward path, there is no need for minislots to allocate bandwidth. All the bandwidth is allocated to the CMTS to transmit packets.

CONCLUSION

QoS provides opportunity, but only if the network is upgraded to support it. Upgrading the network probably means more than just upgrading to DOCSIS 1.1. In addition, new operations and business systems will probably be needed too. But once the network is QoS enabled, there is the opportunity to claim additional revenue.

QoS is not a substitute for good network design. Once services offer real guarantees for throughput, latency, and jitter, how the network is configured and monitored becomes even more important to be sure subscribers are receiving the service being paid for.

REFERENCES

1. Quality of Service: Delivering QoS on the Internet and in Corporate Networks, Paul Ferguson and Geoff Huston, Wiley Computer Books, 1998.

2. Data Over Cable Service Interface Specification Radio Frequency Interface specification version 1.1 (DOCSIS 1.1), SP-RFIv1.1-I07-010829, CableLabs, August 2000.

3. GR-303, Integrated Digital Loop Carrier System Generic Requirements, Telcordia, December 2000.

Author Contact:
Doug Jones, Chief Architect
YAS Broadband Ventures
voice: (303) 661-3823
doug@yas.com