TIERED DATA SERVICES TO ENHANCE CUSTOMER VALUE AND REVENUE: WHY ARE DOCSIS 1.1 AND ADVANCED QUEUING AND SCHEDULING REQUIRED?

Gerry White
Senior Director of Advanced Technology
Network Infrastructure Solutions
Motorola Broadband Communications Sector

Abstract

High-speed data services are making significant revenue contributions to Broadband Operators as service penetration rates are now averaging over 10% and exceeding 30% in some markets. Still, there are large potential customer segments that can be captured by offering data service tiers that provide a better match between customers needs and income levels and the features of the high-speed data service required. For example, business customers are willing to pay much higher rates for a service with greater bandwidth and service level guarantees. At the other end of the spectrum, customers who are paying $25-30 per month for dial-up services would jump at the chance to have an "always on" connection at speeds that are double standard dial-up rates.

This paper will discuss how measurable data service tiers can be delivered to customers through Quality of Service (QoS) control enabled by the use of the DOCSIS 1.1 specifications and advanced queuing, scheduling and congestion control techniques.

Additionally the paper will present a mechanism to support overbooking of network resources in a tiered environment, enaling cost effective deployment while ensuring that guaranteed service levels are met for each of the service tiers.

INTRODUCTION

Initial DOCSIS deployments have generally followed a "one size fits all" model. To date the success of this approach in terms of service penetration and revenue generation has been considerable such that cable modems have become the preferred means of Internet access for a significant section of the population.

In order for Multiple System Operators (MSOs) to attract a wider customer base and reap additional rewards from their investments in Hybrid Fiber Coax (HFC) and IP infrastructure it will be necessary to go beyond this initial audience and provide more tailored services to specific target populations. Thus high-end service offerings can be created for business customers offering guaranteed services at bandwidth (and price points) comparable to T1 and business DSL offerings. At the other end of the spectrum there are many dial up customers for whom current cable modem service is too expensive. Providing a rate limited entry-level service for these users can add to the customer base with an immediate revenue impact and also create a pool of broadband addicts, which can be targeted for later upgrade to premium services.

Ultimately all networks are a shared resource and rely on statistical multiplexing between users to provide cost effective service. In the case of the HFC network the shared resource extends to the customer premise, while for a DSL network the customer link may be dedicated copper pair. In both cases the infrastructure network is shared and statistical multiplexing is used to reduce costs. Thus to ensure fairness between users (even for a best effort single service tier network) QoS must be provided to some degree, although it may be relatively simplistic. When tiered service offerings are enabled QoS becomes more complex as it must not only provide fairness between users within a service tier but also differentiate between tiers.

The MSO community is in the fortunate position of having the tools to provide tiered services readily available. In fact in many cases they are already deployed, waiting to be enabled.

DOCSIS 1.1 provides the mechanisms to deliver tiered services over the HFC network and when this is combined with suitable backbone technologies the MSO can deliver end to end service tiers which compete effectively with any in the market.

TRAFFIC FLOWS

At the customer level the focus of interest is on individual applications, the user wants to make a phone call, connect to a corporate network or simply surf the web. In general each application will involve multiple traffic flows. For example a VoIP client will exchange control packets with a call management system and voice traffic with the far end called system. These traffic flows may take different paths through the network and each will require QoS to be provided. The concept of traffic flows and of providing QoS to each flow is central to both DOCSIS and backbone QoS mechanisms.

AGGREGATE VS PER FLOW QoS

Two distinct mechanisms are typically used to provide QoS on an end-to-end basis over large IP networks. These mechanisms have been the subject of research and standardization efforts resulting in the differentiated services (DiffServ) model for aggregated QoS and the integrated services (IntServ) model for per flow QoS [References 1 and 2].

At the edge of the network bandwidth is typically scarce and expensive to provide. Devices in this domain, such as a CMTS, deal with a bounded number of traffic flows (e.g. to a maximum of 8000 flows in each direction in DOCSIS). In this region QoS can be provided most effectively at a per flow level. For each flow a traffic specification and a flow specification are defined. The traffic specification defines a classifier to identify the packets that belong to a specific flow. Typically this is a masked set of fields based on the content of the packet header such as IP addresses, port numbers, DSCP markers, etc.. The flow specification defines the QoS parameters to be applied to the flow (bandwidth, latency…). This mechanism was defined by the IETF as the IntServ architecture and was adapted to provide the basis for DOCSIS 1.1 QoS [Reference 3]. DOCIS 1.1 provides for a signaling mechanism to set up new flows, for admission control functions and for isolation between flows. The CMTS has the primary responsibility to provide QoS in the DOCSIS realm by implementing admission control and providing isolation between flows based on the upstream and downstream scheduling mechanisms.

In the core of the network bandwidth is relatively abundant (and cheaper) but the infrastructure is shared between tens of thousands of clients. Thus core switches and routers must support hundreds of thousands of flows. If these devices were to operate at an individual flow level the amount of data to be maintained would be massive and systems would simply not scale so that an aggregated mechanism is needed. To achieve the scaling required packets entering the DiffServ domain are classified into one of a limited number of behavior aggregates (64 max.). DiffServ defines a field in the IP header of the packet known as the DiffServ code point (DSCP). Systems at the edge of the DiffServ domain mark packets with the code point desired before transmission into the domain. All packets with the same DSCP to be transmitted on a given link are considered part of the behavior aggregate and are to be treated in the same manner. The DSCP defines the required behavior for each packet so that no per flow state must be maintained. Routers within the DiffServ domain use the DSCP to determine the QoS desired by the packet and apply the appropriate queuing and scheduling algorithms to achieve the required QoS.

Multi Protocol Label Switching (MPLS) [Reference 3] has recently emerged as another mechanism to provide aggregated QoS in metropolitan and core networks. It defines a mechanism to set up label switched paths (LSPs) between endpoints at the edges of the MPLS network. Packets entering the network are assigned to a particular path and a label added to the packet identifying the LSP to be used. This label is used by the MPLS routers in the core of the MPLS network to forward the packet to its destination endpoint. The core network does not need to examine the packet headers beyond the label and can therefore focus on switching traffic to its destination as quickly and efficiently as possible. Each LSP can be associated with a defined forwarding equivalency class (FEC) which defines specific QoS parameters so that the MPLS network can provide a mesh of QoS enabled paths.

In a typical MSO network environment the transition point between per flow and aggregated QoS domains occurs at the intersection of the HFC/DOCSIS and metro/IP networks. The HFC access network is based on the DOCSIS 1.1 protocol, which provides QoS to applications on a per flow basis. In the upstream direction an application flow is mapped to a service identifier (SID) and in the downstream direction to a service flow identifier (SFID). The cable modem (CM) and CMTS cooperate to assign the required QoS to each flow and to ensure that it is met. The MSO metro networks that connect the CMTS systems are typically based on gigabit Ethernet, Sonet or RPR physical infrastructure. Historically an IP routing infrastructure ran on top of this possibly providing QoS based on the DiffServ model. Newer networks replace the pure IP routing model with one based on MPLS infrastructure. Both types of network provide the MSO with the capability to provide aggregated QoS based on traffic engineering and provisioning. In either case the CMTS/ER provides the transition point between the QoS domains. The major issues, which must be resolved at this demarcation point, will be considered later in this paper.

QoS MECHANISMS

The ability to deliver QoS involves four key functions:

- Classification of packets to determine which flow a packet is part of and the appropriate service level for each traffic flow
- Policing of traffic to prevent flows from getting higher than agreed upon service levels
- Buffering to ensure that queues are created to contain packets during periods of congestion
- Scheduling to enforce packet transmission in accordance with QoS policy.

In order to provide QoS successfully the CMTS/ER must provide this functionality for both HFC and metropolitan network environments.

QoS IN THE HFC NETWORK

The DOCSIS 1.1 specifications provide QoS for the cable access network. They define enhancements to the Media Access Control (MAC) protocol of DOCSIS 1.0 to enable more sophisticated access methods over HFC access networks by adding the following:

- Packets are classified into service flows based on their content. Thus each application can be mapped to a unique service flow.
- Network access (upstream and downstream) is scheduled per service flow using one of a number of defined scheduling mechanisms including constant bit rate, real-time polling, non real-time polling and best effort.
- Service flows may be configured through management applications or created and deleted dynamically in response to the starting and stopping of applications.
- Fragmentation of large packets is required to allow low latency services to operate on lower-bandwidth upstream channels.

These features provide the basic tools for QoS management. They allow applications to request QoS changes dynamically and allow providers to isolate multiple data streams from each cable modem, set-top box or MTA. DOCSIS 1.1-based systems can therefore potentially deliver the ability to allow application-specific QoS treatment within the HFC access network for each traffic flow.

Packets transmitted into the network from a host system are treated as follows. The interaction between the CM and the CMTS upstream scheduler is shown in Figure 1:

1. The CM filters each packet and classifies it into a service flow identified by a unique SID.
2. The CMTS schedules upstream transmission for the SID based on its QoS parameter set and the traffic history of the SID. This is the most complex and the critical step for providing upstream QoS.
3. The CM transmits the packet to the CMTS
4. The CMTS receives the packet and reclassifies it based on CMTS configuration. In general the CMTS is the first trusted device in the network and should not rely entirely on user or CM packet classifications.
5. The CMTS maps the packet into the QoS scheme for the MAN (mark traffic for differentiated services forwarding, map traffic to MPLS LSP tunnels, map traffic to physical interface)
6. The CMTS queues the packet to the egress link based on the required QoS
7. The CMTS implements its network interface scheduling algorithms and transmits the packet to the link when it reaches the head of queue.

Packets received by the CMTS from the network are treated as follows and shown in Figure 2

The CMTS receives the packet and classifies it into a downstream service flow identified by a unique SFID.

1. The CMTS enforces policing on maximum rate if required.
2. The CMTS queues the packet to the egress link based on the required QoS
3. The CMTS implements its downstream scheduling algorithms and transmits the packet to the link when it reaches the head of queue.
4. The CM forwards the packet to the host.

QoS IN THE MAN/WAN

There are two primary methods for providing QoS control in the regional network, packet based and connection-based.

In the packet based case, individual flows are policed and marked at the edge of the network using a DiffServ DSCP marker in the IP header, so that aggregated flows are delivered to the network core with each flow tagged for the appropriate QoS treatment. The DiffServ standard defines the code points to use and the per-hop forwarding behavior to be applied to each marked packet.

Connection-based QoS can be implemented using MPLS. In an MPLS network a number of paths are established between the end points of the network. Each path can be traffic engineered to provide a defined level of QoS. All packets on the path share the same forwarding equivalency class (FEC) consisting of the MPLS end point and the QoS parameter set. As with the DiffServ case the packets from the individual flows are policed and marked at the edge of the network. In this case the marker is an MPLS label that is prepended to the packet and identifies the path through which the packet will traverse the MPLS network. Each path is referred to as a label switched path or LSP. A single LSP, with a defined FEC can support multiple flows and thus provide the necessary aggregation mechanism.

QoS AT THE BOUNDARY

In order for applications to see real benefits, QoS must be provided on an end-to-end basis. Thus the QoS-enabled traffic flows from the HFC access network must be mapped to the QoS mechanism(s) used in the regional or backbone networks. The CMTS/ER at the boundary must be able to perform this mapping and implement the QoS mechanisms for the two domains. To deliver QoS at a per service flow level this must take place for multiple flows at wire speed. The mechanisms employed within the CMTS/ER must maintain the QoS during this transition. The problem is complex due to QoS requirements constantly changing as service flows are created and deleted dynamically.

The metro to HFC boundary is also a capacity transition point. In the downstream direction packets received from gigabit speed optical links must be transmitted onto megabit capacity DOCSIS networks. Thus the CMTS/ER must implement congestion management based on the conformance of the subscribers and applications to their Service Level Agreements (SLA's).

The queuing and scheduling mechanisms used within the CMTS/ER to implement the transition between the per-flow HFC and aggregated metro domains will determine how successfully QoS can be delivered. The key concepts required to provide this transition successfully are:

- Queuing and Scheduling
- Congestion Control

QUEUEING and SCHEDULING

Three queuing and scheduling mechanisms will be considered; FIFO, class based and per flow.

FIFO Queuing

First-In-First-Out (FIFO) queuing is both a queuing and a scheduling mechanism. In FIFO queuing, all packets are stored in a single queue and are transmitted in the order that they are received. FIFO queuing is easy to implement and it requires little configuration. Unfortunately it does not provide any support for the differing QoS levels required by diverse applications. In a FIFO scheme packets for a low-latency service can be queued behind those from high bandwidth services and must wait for these to be transmitted.

Class-Based Queuing

Class-based queuing (CBQ) attempts to avoid this problem by sorting the traffic into different classes by examining the packet and trying to determine the type of traffic to which it belongs. Once the packet has been classified, it is placed in a FIFO queue that contains only other packets of the same type. Each per class FIFO queue can be serviced according to configured policy to provide the behavior required. Thus a FIFO that is serviced frequently so that it is usually empty of packets could provide a low latency, low loss service such as VoIP. Similarly a FIFO that is kept full can provide a service such as bulk file transfer, which requires high bandwidth but can tolerate moderate packet loss.

In theory, this allows the FIFO of each class to provide the desired type of service, but in practice there are a number of problems with this approach. It requires a constant, heavy configuration burden because the operator has to configure the allocation of service to the different classes e.g. 1/10 for e-mail, 1/10 for voice, 1/3 for Web traffic, etc. In a dynamic network environment, the class-based queuing method of allocating service is impractical because the allocation is independent of the number of users of a given class. CBQ does not provide application isolation, as although each queue contains traffic from a similar application type (e.g.VoIP), flows from multiple users are mixed within the queue.

Per-Flow Queuing

Per-Flow Queuing (PFQ) solves the problem of providing isolation between application flows by assigning each packet stream to its own queue. Those queues for flows with QoS reservations are served at their guaranteed rate while flows without reservations are served in a round robin or fair-share manner. Thus the queue for each flow is served at the rate defined in the service level agreement.

To assign flows without reservations to a queue a method known as stochastic queuing is used. In stochastic queuing the parts of the packet header that are the same for all packets of a flow, such as the source and destination IP addresses and source and destination port numbers are fed to a hash function that is used to map the packet to a queue. This ensures that all packets for the same flow are mapped to the same queue (to avoid miss ordering). It also eliminates the need to configure bandwidth shares per-class, and consequently avoids the miss-allocation caused by varying usage patterns. If the system can support more queues than there are flows then most flows either have their own queue or share it with a small number of other flows. To support stochastic per-flow queuing the CMTS/ER must support thousands of flows (with DOCSIS 1.1 a CMTS can support 8000 flows in each direction per DOCSIS domain).

CONGESTION CONTROL

Given that the data rate in the MAN will be significantly greater than the capacity of an HFC link it is also important to consider congestion control mechanisms, especially as applied to the downstream traffic. Three mechanisms to handle congestion will be discussed; tail drop, Random Early Detection (RED) and Longest Queue Pushout (LQP).

Tail Drop and RED

(Refer to Figure 3)

With all packets sharing the single queue for simple FIFO or all packets of the same type sharing the same queue for CBQ there are limited options for congestion control. The simplest scheme is simply to drop packets from the tail of the queue when no buffers are available for them. This takes no account of the content of the packet and hence it's potential value or of whether the flow to which it belongs is in compliance with its service level agreement. Mechanisms such as RED attempt to solve this problem by randomly dropping some of the arriving packets when the queue starts to fill. The intent is that the end systems using a windowed flow control, such as TCP will notice the packet loss and slow down. Thus it is dependent on the end systems of the flow to slow down transmissions to solve the congestion problem. While this might work in a well-controlled environment such as an enterprise network it is unlikely that the end systems will be as cooperative in a public network. FIFO queuing does not provide a mechanism to ensure isolation for well-behaved flows from miss-behaved applications generating heavy loads.

As the number of flows sharing the FIFO queue increases this problem becomes worse. The limitations of class-based queuing and RED often result in the inappropriate discarding of packets. Since traffic flows are lined up in shared queues, it is impossible to isolate and discard those flows that are exceeding service level agreement (SLA) guarantees before discarding traffic flows that are staying within their SLAs. Packets are therefore discarded randomly from shared queues as they become full.

Longest Queue Pushout

(Refer to Figure 4)

In a per-flow queuing environment longest queue push out (LQP) is the mechanism that best meets the congestion control requirements. LQP allocates buffers to the individual flows as required until 100% of the buffer pool is used. When no buffers remain and a new packet is received, LQP discards traffic from the flows, which are the longest queues. The scheduling system is transmitting from these per-flow queues at a rate that matches the QoS assigned for each flow. Thus by definition the longest queue is that which is exceeding its allocation by the greatest amount so that traffic is automatically discarded from those applications which are non-compliant with their SLAs. This occurs without the need to configure congestion control.

Further details on PFQ performance can be found in [Reference 5].

OVERBOOKING TIERED SERVICES

In order to create a commercially viable network the operator must rely on statistical multiplexing. Not all users are active at the same time so that the network can be over subscribed to reduce the cost per user. In a single tier best effort network overbooking is relatively simplistic. Subscribers are added to the network until a local heuristic, such as the number of cable modems per upstream or a defined traffic load, is reached. In a tiered network overbooking becomes a more complex process and should be applied at each service tier as well as on each network interface. The amount of over booking is dependent on the types of services to be offered (e.g. guaranteed vs. best effort) and on the traffic patterns of the users within each service tier.

Service Classes

Responsibility for providing QoS to a service flow resides with the CMTS, which must be configured with the parameters necessary to control this operation. In order to simplify this configuration and help operations staff retain their sanity the concept of service classes has been introduced. DOCSIS 1.1 has defined a set of QoS parameters, including maximum sustained and minimum reserved traffic rates, and a way for associating specific QoS parameter values to service flows. It has further incorporated the concept of a service class name so that service flows, when being created, may be assigned their QoS parameters by referencing a service class name.

Thus an operator may define a number of service classes. Individual service flows will be assigned to a service class and all flows belonging to that class provided with a defined quality of service. Service classes can be supported for both downstream and upstream directions.

In order to facilitate overbooking the concept of the service class can be extended by introducing two additional service class parameters, Maximum Assigned Bandwidth (MAB) and Configured Active Percent (CAP). With these additional parameters the operator can explicitly control overbooking for each service tier and for the network interface in total. With this scheme all service flows must be assigned to a service class. If a service class name is not present in a Registration Request message generated by a CM, then the CM service flows are assigned to a default service class.

Maximum Assigned Bandwidth

Maximum Assigned Bandwidth (MAB) specifies the amount of bandwidth a service class is permitted to consume on an interface. It is expressed as a percent of the total interface bandwidth capacity. The MAB of a service class is applied during admission control to determine whether to admit a new service flow and again by the scheduling algorithms to provide a class-based weighting to the scheduler. Any unused portion of a class' bandwidth may be used 'on demand' by other classes which have a traffic load in excess of their own MAB.

Configured Active Percent

Since not all service flows are active simultaneously the service level classes feature permits customers to overbook service classes. Overbooking means admitting service flows to a service class such that the sum of their guaranteed minimum reserved rates are in excess of the configured MAB for the service class. To control the amount of overbooking, a configurable overbooking factor the Configured Active Percent (CAP) is provided. The CAP is an estimate of how many service flows, expressed as a percentage, are likely to be active simultaneously. For example, if the CAP for a service class is set to 20 percent then it is estimated that only 20 percent of the service flows belonging to that class will be active simultaneously. Therefore, 5x (1 / 0.2) overbooking would be allowed. A CAP of 100 percent means that no overbooking will be allowed. A CAP of zero percent means that unlimited overbooking is allowed.

ADMISSION CONTROL

Admission control is a process wherein the bandwidth requirements of a service flow are checked to verify that admission of the service flow to a service class does not exceed the class' MAB after accounting for the allowed level of overbooking. Service flows are created during modem registration or through dynamic service messaging. A CM registering with primary service flows should be permitted to register regardless of whether the admission of its service flows would exceed its service class' MAB. In this case however the service flow would be admitted in a 'Restricted' state meaning that the service flow will not be provided any guaranteed minimum reserved rate. Service flows created via dynamic service messaging will be rejected if admission of the service flow would cause its service class to exceed its MAB.

Examples of MAB and CAP operation can be seen in Figure 5 and Figure 6. In both examples the bandwidth on the interface is shared between two service tiers a best effort class and an enhanced service class such as a business service tier. In the first example three flows are active in the enhanced class with no flows active in the best effort class.

Initially the total bandwidth is below the MAB for the class so that all flows receive their guaranteed bandwidth. At this point the MAB is reached so that requests for additional flows to be set up would be rejected by admission control (or could receive best effort bandwidth as defined by operator policy). As flow 3 bursts to a higher data rate, above the guarantee, it will share available bandwidth from the best effort service class if this is available.

The second example shows the same 3 flows active with all MAB consumed so that no further flows could be admitted. Flow 3 then terminates at which time bandwidth consumption falls below the MAB for the class and additional flows could be added.

SUMMARY

Providing tiered service offerings has the potential to extend the target audience and revenue stream for high-speed data services.

Extending existing best effort DOCSIS to support service tiers with QoS guarantees requires

- The use of the DOCSIS 1.1 protocol extensions
- Sophisticated scheduling and congestion control mechanisms in the CMTS.
- A means of overbooking, which is service tier aware.

All of these features are available in CM and CMTS systems, including much of the currently installed equipment base.

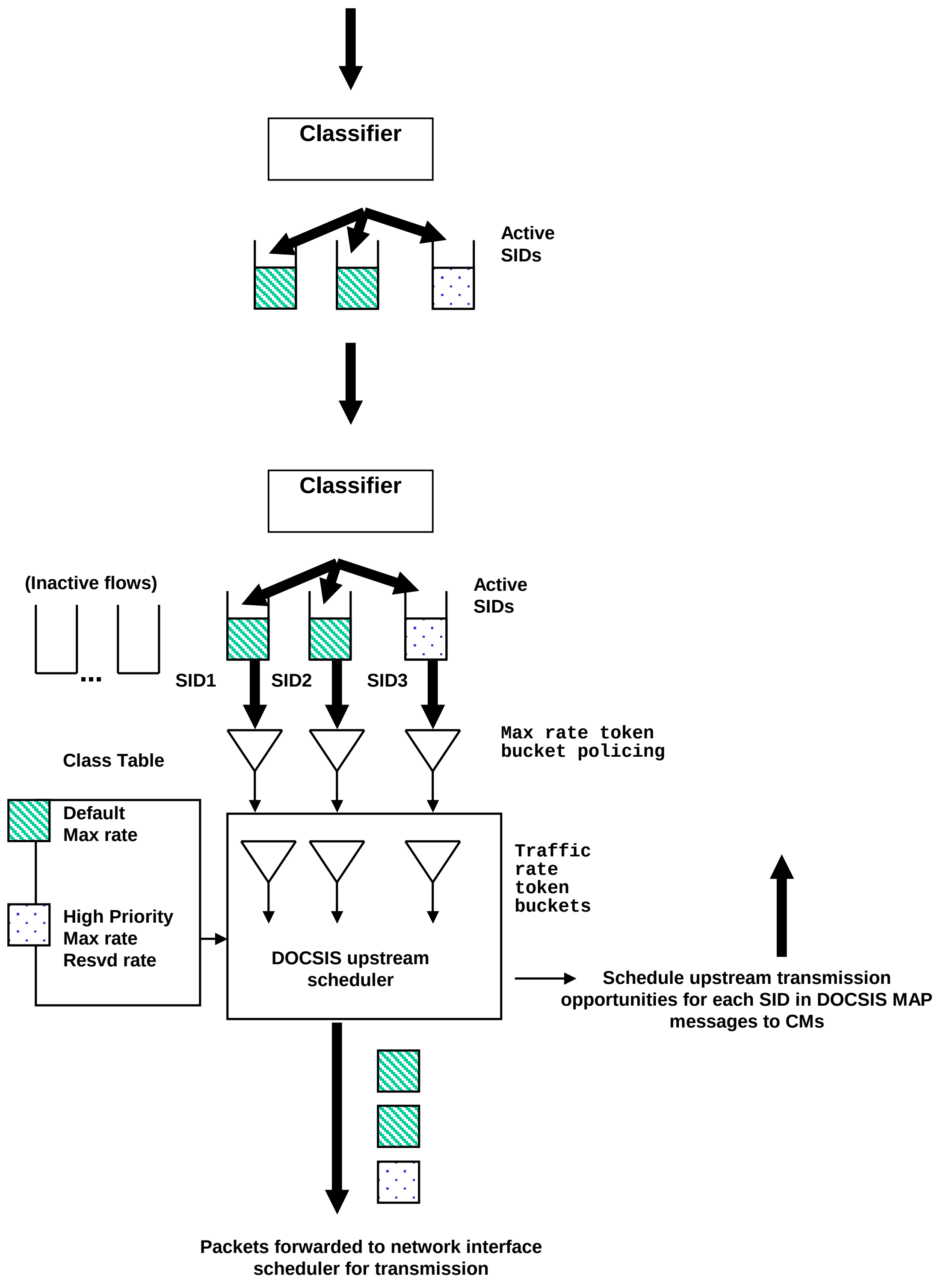

Figure 1 Interaction between CM and CMTS upstream scheduler

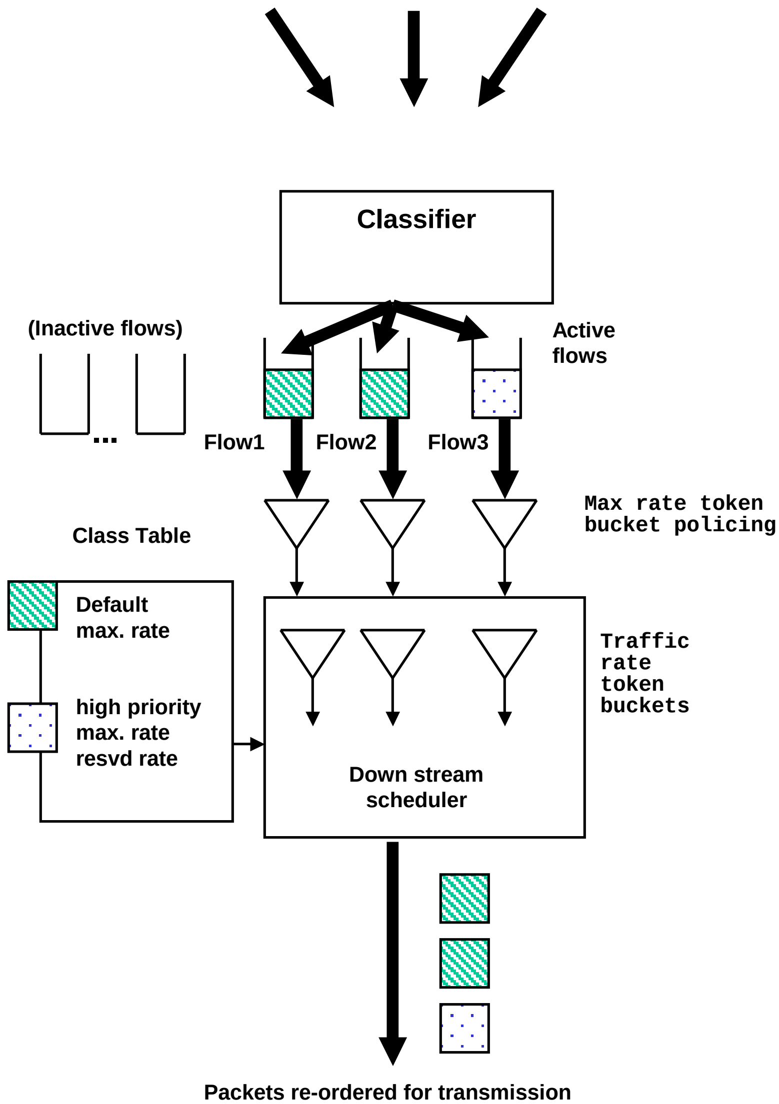

Figure 2 Downstream Packet Scheduling

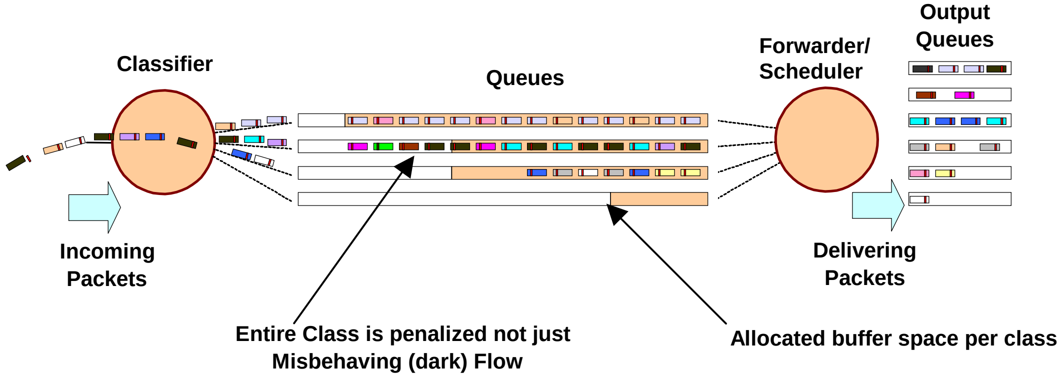

Figure 3 Class Based Queuing with RED

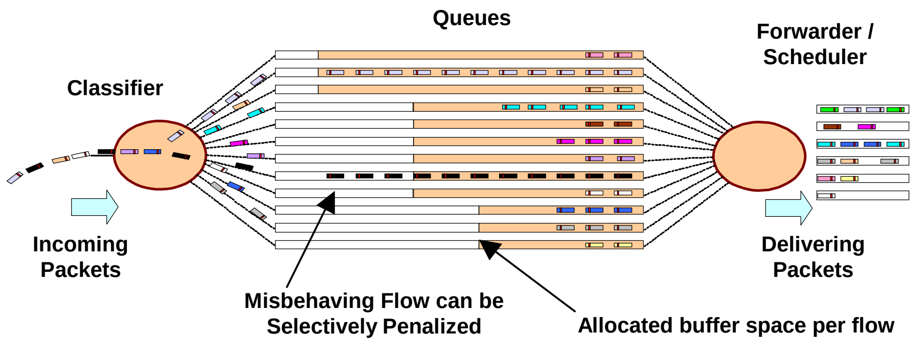

Figure 4 Per Flow Queuing and Scheduling with Longest Queue Push out

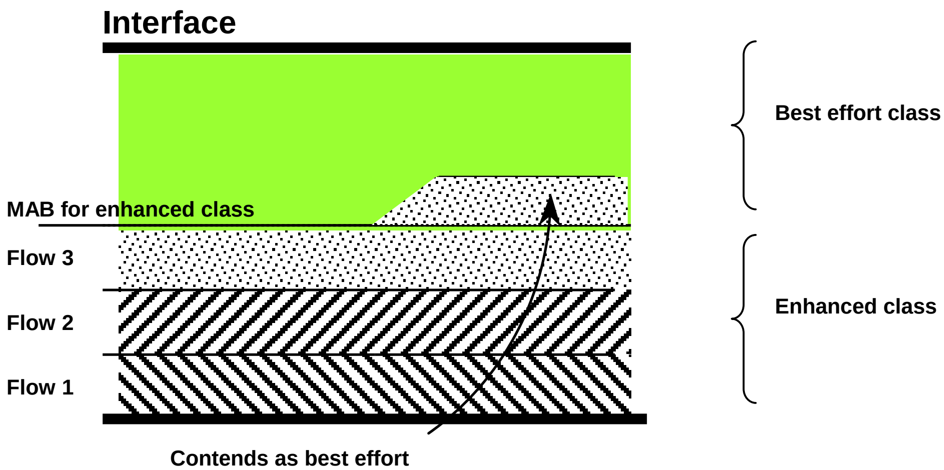

Figure 5 MAB Operation -flow 3 exceeds reserved rate

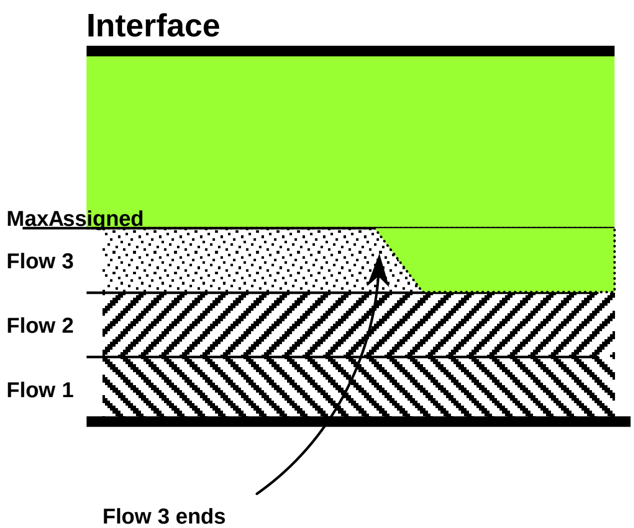

Figure 6 MAB Operation -reserved rate flow 3 ends

REFERENCES

1. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss. An Architecture for Differentiated Services (RFC 2475).
2. R. Braden, D. Clark, S. Shenker Integrated Services in the Internet Architecture: An Overview (RFC1633).
3. DOCSIS 1.1 Radio Frequency Interface Specification.
4. E. Rosen, A. Viswanathan, R. Callon. Multiprotocol Label Switching Architecture (RFC 3031)
5. J.C.R. Bennett and H. Zhang Hierarchical Packet Fair Queuing Algorithms

###

About the Author

Gerry White is Senior Director of Advanced Technology within the Network Infrastructure Solutions Group of the Motorola Broadband Communications Sector, and he was previously CTO of RiverDelta Networks, Arris Interactive, and LanCity. He has presented papers and spoken at dozens of cable industry events worldwide and has published many trade press articles on cable industry issues. White is the co-author of several patents and articles on data communications technology, and he holds a BSc. (Honors) from University College in London.

Author Contact Information

Gerry White
Senior Director of Advanced Technology
Network Infrastructure Solutions Group
Motorola Broadband Communications Sector
3 Highwood Drive
Tewksbury, MA 01876
(978) 858-2300
gwhite@motorola.com