OPTIMIZING THE LAST MILE

Rice, D. and Schnitzer. J.
Stargus, Inc.

Abstract

Several sources of inefficiency common in current DOCSIS network deployments are identified and provided as motivation for cable operators to optimize their network and operations. Through an investigation of DOCSIS technology, two complementary methods for the optimization of last mile cable networks are proposed.

INTRODUCTION

Multiple System Operators (MSOs) can optimize their access networks for new and existing services by efficiently managing "last mile" network quality and capacity. Obtaining visibility into, proactively acting on, and quantifying service-affecting HFC issues can raise the quality of the broadband experience and ensure the success of advanced, IP-based services. These successes can be measured in customer satisfaction, revenue growth, efficient operational expenditures, and reduced capital expenditures.

Today many last mile networks are operating with impairments and inefficiencies that are tolerated by email, web surfing, and other less data intensive or less time critical applications. Because of the error correction built into DOCSIS™ (Data Over Cable Service Interface Specifications) and the resilience of Transmission Control Protocol (TCP), these errors are unobserved by and unaffecting to most Broadband customers. However, advanced, IP based services such as Voice over Internet Protocol (VoIP), video-conferencing and streaming audio/video are rendered inoperable by such errors and inefficiencies.

By using the capabilities built into DOCSIS, the inefficiencies can be minimized through configuration optimization, thereby allowing MSOs to defer capital expenditures, reduce operational expenses, and ready their infrastructure for advanced IP-based services.

SOURCES OF INEFFICIENCY IN THE LAST MILE

Congestion and latency have unnecessarily existed in some network segments since the first cable modem deployment. Conversely, due to a lack of visibility into real network issues and fast subscriber growth, many of today's networks have been over-engineered from a bandwidth capacity point of view, resulting in inefficient capital spending. Subscriber growth will accelerate with the addition of multiple Internet Service Providers (ISPs) and service tiers. Making optimal use of MSOs competitively superior resources of spectral bandwidth and a potentially high quality signal environment will be key to minimizing capital expenses for additional infrastructure and reducing operational expenses.

Error levels often get worse over time and are typically associated with transient noise and interference due to HFC plant problems. While email and web surfing mask these low error levels, these latent and worsening HFC plant problems are often undetected for months until customers become dissatisfied with the performance of their applications. A degraded subscriber experience results in operating inefficiencies such as subscriber churn, as well as heightened customer care and network maintenance costs.

MOTIVATIONS FOR OPTIMIZING THE LAST MILE

There are many motivations for optimization of the last mile network-

- **Harvesting bandwidth —** Models mapping packet size distributions to operational configurations and channel quality show that capacity can be increased 2 to 6 times over current commonly used DOCSIS 1.0 CMTS configurations. This additional capacity can be used to increase revenue by offering higher margin tiers of service and supporting more customers.

- **Deferring capital—**Capital expenditures for network elements such as CMTS equipment and transmission infrastructure are significant. These expenses can be postponed until absolutely needed through increased visibility and efficient use of current infrastructure and spectral resources.

- **Reducing operational costs—**Labor expenses associated with installing additional CMTS infrastructure, re-building RF combining and optical splitting networks, and transmission facility construction can also be avoided and delayed. Providing focused maintenance can result in reduced trouble call rate, reduced call handle time, and reduced truck rolls.

- **Increasing customer satisfaction—** Customer satisfaction is the result of managing capacity well and providing stable, reliable services.
  - Visibility of network issues and tracking of useful performance metrics can be used to continually manage and improve the quality of the product offering.
  - HFC networks change over time and faults will always occur. User experience can be optimized while fix agents expeditiously address problems thus repairing many faults before they become customer-critical.
  - As new services are deployed that require guaranteed bandwidth and knowledge of capacity resources, efficient management will be crucial
  - Errors due to degraded network quality will impact these advanced, IP based services and degrade them as much if not more than capacity or Quality of Service (QOS) issues. The error performance and the capacity of the network are inter-dependent and must be managed for a quality customer experience.

- **Enabling new services—**Technology is constantly changing. DOCSIS 1.1 and DOCSIS 2.0 contain many more methods for optimization than DOCSIS 1.0. DOCSIS 1.0 has many features that are not yet fully utilized by MSOs. These methods are exposed and should be used by MSOs for the purpose of optimizing the use of their spectral bandwidth and high signal quality network resources. To take advantage of the ability to offer QOS and symmetrical bandwidth, enabling new services, these methods for optimization need to be utilized efficiently. The underlying technology is becoming increasingly more complex and needs to be abstracted and automatically managed to avoid an increasingly expensive operations workforce.

A simple conservative cost model for an MSO considering only the savings in capital expenses as shown in row 1 Table 1 can defer significant capital infrastructure expenses over a 2-year period of $3.05 per CM per year. The operational expenses to perform the infrastructure upgrades or the unnecessary recombining and reconfiguration of networks before the current resources are actually fully utilized will easily outweigh the capital expense and increase these savings.

Some additional modeling based on quality optimization that improves customer experience reducing churn and call center costs along with more efficient use of truck rolls and fix agents is also shown in Table 1. The total savings per CM per year can be greater than $10.

Table 1: Operational Savings Through Capacity and Quality Optimization

Optimization Benefit	**Assumptions**	**Annual Value (subscriber/yr.)**
Deferred capital based on Capacity Optimization	• CMTS infrastructure retail cost: $32 to $36 per cable modem* • Large MSO purchase discount: 40% • Customer growth: 20% per year, 2 year average • Capacity optimization: 2X	$3.05
Churn reduction Improved customer satisfaction	• 5% of controllable churn • (.4% x $20/mo) value compounds • Number is 3 year average	$1.92
Truck roll reduction	• 50% call rate • 15% yield truck roll • 5% reduction x $50/truck roll	$2.25
Call center - Reduced calling rate	• 50% call rate • 5% reduction x $5/call	$1.50
Call center - Reduce call time duration	• 5% efficiency gain • (45% t/c rate x 5% x $5/call)	$1.35
Total (per data sub)		$10.07

*Based on recent vendor list price and approved re-seller quotes.

METHODS OF OPTIMIZATION IN THE LAST MILE

DOCSIS Cable Modems (CM), Multi-media Terminal Adaptors (MTA), Advanced Set Top Boxes (ASTB), and Cable Modem Termination Systems (CMTS) can all be utilized to detect and manage errors while providing bandwidth intelligence data, enabling the capacity optimization discussed in Method 1. The DOCSIS network can be configured to "four wheel drive" through most service affecting errors—errors that otherwise result in degraded or complete loss of service to subscribers--while maintaining optimal bandwidth capacity. This can be accomplished while notifying operators of degraded network quality before it becomes service impacting, as it occurs, and also providing isolation and identification of the faults.

The second complementary method to reduce inefficiency, discussed as method II below, addresses service quality optimization. In this case, we combat inefficiencies impacting broadband experience by considering all elements and conditions degrading service within the DOCSIS transport network. This method describes how user applications and services are mapped to all of these underlying network conditions to provide performance metrics that map directly to user experience.

Method I – Capacity Optimization

Communication Systems and the Capacity Optimization Trade Space

When designing or operating digital data communication systems, there are several goals that help drive system optimization:

Goal 1: Transmit as much data in the shortest amount of time possible through the system.
Goal 2: Transmit this high rate of data using as little of the physical resources (spectral bandwidth and power) as possible.
Goal 3: Transmit this data reliably at a much lower rate of errors than will impact the performance or reliability of any of the services.
Goal 4: Develop and operate this system with as little expense and complexity as possible

The challenge is that these four goals are not completely independent. To accomplish high transmission rate, it often requires more physical resources. Error performance can be traded off against physical resource utilization, but is best left fixed at very low error performance. High transmission rates at low error with less use of physical resources can be accomplished, but typically at much higher costs and complexity. [1]

Fortunately, the HFC network architecture has significant competitive advantage in its physical resources. The physical resources that an MSO has available to utilize are significant spectral bandwidth and power. The spectral bandwidth in the network can be up to 1 GHz and beyond in some cases. The power resources include the potentially high quality low noise power environment enabled by the use of optical transport and shielded coaxial transmission lines. These power resources also include the ability to transmit high-powered signals as long as they are within the dynamic range of the active components and the network meets FCC emission standards. Many of the other competitive transport technologies do not have anywhere near the advantage in these physical resources that MSOs have with HFC architectures.

A 3-dimensional model of the trade space that is used to optimize capacity and take advantage of the physical resources available for DOCSIS 1.0/1.1/2.0 is shown in Figure 1. The performance of the communication system is based on the trade off between error performance, channel quality, efficiency and capacity. The ability to send a high rate of data using as little spectral bandwidth as necessary is indicated by the Bandwidth Efficiency/Capacity axis. Sending this high rate of data and its relationship to signal and noise power is expressed on the channel quality axis. Sending this high rate of data reliably without errors can be expressed by the error performance axis. Choosing operating points within this space can often be determined by the cost and complexity of the communication system. [2]

In the middle of the last century, some Bell Laboratories information theory scientists determined the theoretical outer bound of what is physically possible in this trade space. [3] This outer bound of a digital communication system's capacity can be expressed simply using the Shannon-Hartley Capacity Theorem.

$$C = W * Log_2\left(1+\frac{S}{N}\right)$$

Where,

$C = Capacity\,(bps)$

$S = SignalPower$

$N = NoisePower$

$W = Bandwidth$

$$W = \frac{1}{H^2(0)}\int_0^\infty H(F)\,df$$

$H(F) = Frequency\ \mathrm{Re}\,sponseofChannel$

This theorem for the maximum capacity bound stipulates that there exists a Forward Error Correction (FEC) and modulation technique to allow information to be transmitted up to the capacity C, at an arbitrarily small error rate approaching 0. It also stipulates FEC or modulation technique does not exist that will allow information to be sent at a higher rate than C without errors.

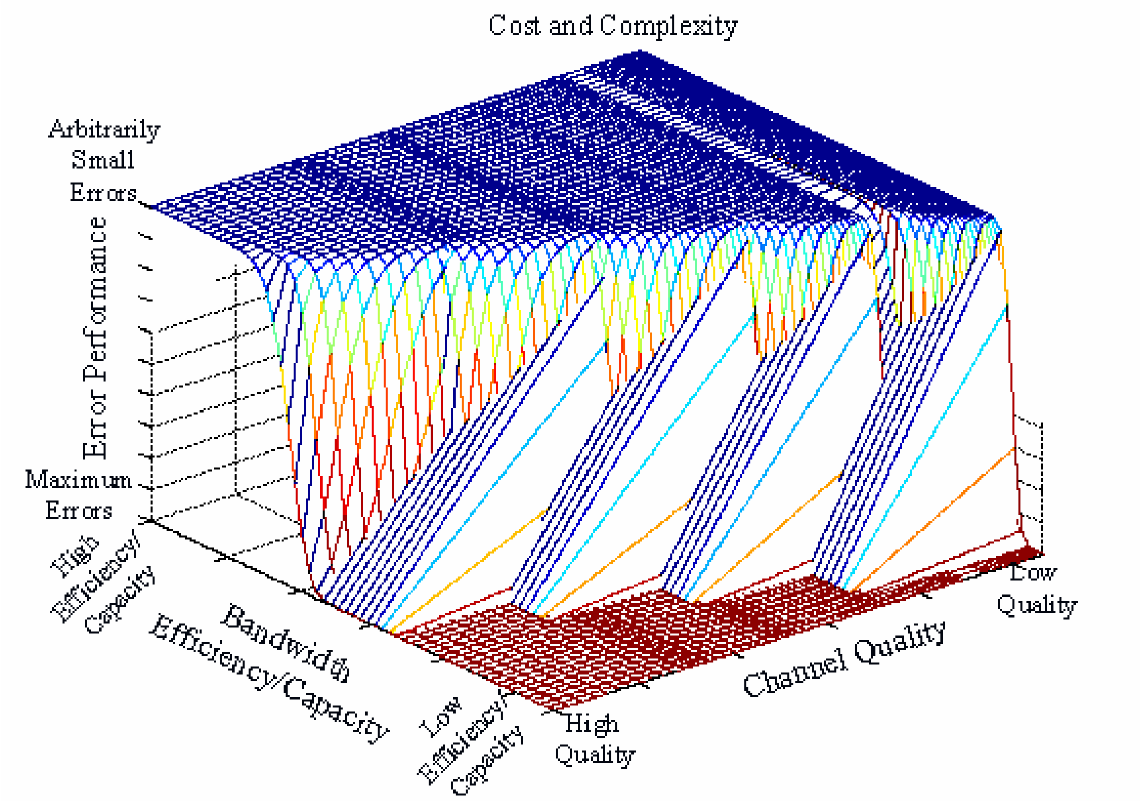

Figure 1: Capacity Optimization Trade Space

Mapping to DOCSIS

The DOCSIS 1.0/1.1/2.0 specifications developed at Cable Television Laboratories™ comprise a progression of techniques that allow operators to optimize capacity within the trade space of Figure 1 based on the services offered and the quality of the network. [4] If we refer back to the goals of designing a communication system, they are consistent with the goals of the DOCSIS program. The maximum data rates available have increased through time. The optimization against the use of physical resources has seen increasing efficiency. The complexity of operating the technology has increased significantly, while conversely, the cost of network equipment such as cable modems (CMs) and CMTSs have decreased due to production volume and competition. Due to this increasingly complex technology, capacity optimization methods will become important to deploy new IP-based services successfully.

Optimally, the network must operate at an error level small enough not to degrade the different applications and services. The communication system trade space concept can be simplified by bisecting this space in Figure 1 near the top of the Error Performance axis into a two-dimensional plane as shown in Figure 2. Based on the techniques for optimization in this space that are dictated by the DOCSIS specifications, a system can operate in the areas show in Figure 2 for the different versions of DOCSIS. Using the "knobs" provided by DOCSIS the capacity and efficiency can be increased for a given channel quality as shown by the vertical arrow. If the channel quality is sufficient it can be further increased to the maximum provided by DOCSIS. If this plane were a map of actual capacity, instead of efficiency (bits per second per Hz), the increases would appear much more pronounced since DOCSIS allows for use of increasingly wide frequency channels in the upstream to increase capacity. As DOCSIS evolves and the cost of more complex technology continues to drop, it will continue to extend closer to the theoretical boundaries described by Shannon and shown in Figure 2. This evolution will allow MSOs to make optimal use of the infrastructure they have in place as technology improves.

The ideal operating points within this space are those along the edge of the space yielding the highest capacity and efficiency relative to the quality of the channel, while maintaining errors at no-impact to performance levels.

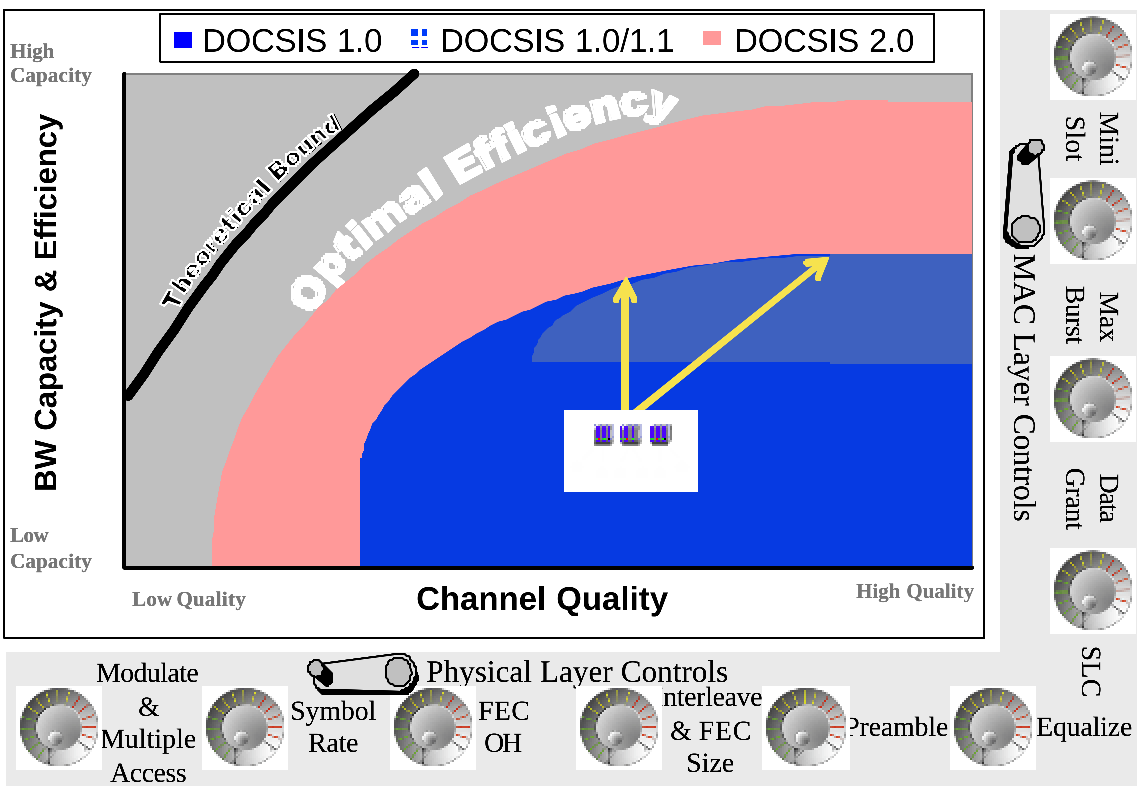

Figure 2: Efficiency and Capacity vs. Channel Quality

There are many different "knobs" and "levers" that are available in DOCSIS systems that can be tuned to allow this capacity optimization and movement in the plane show in Figure 2. They can be divided into three categories: Physical layer Capacity, MAC layer efficiency, and traffic scheduling. All three of these categories are related to each other and optimizing all aspects can result in the significant gains in capacity discussed above. While short-term traffic scheduling and power adjustments are the domain and responsibility of the network infrastructure, longer-term quality control through optimal selection of network operating parameters based on performance over longer time periods is the domain and responsibility of MSOs' management systems.

The Physical Layer and MAC Layer knobs which should be adjusted by operators to maximize the performance and provide a solid basis for the novel scheduling algorithms being developed by CMTS infrastructure vendors are listed below. Different knobs are available for the downstream and upstream direction and setting of these parameters needs to be considered as a collective. For example, increasing the symbol rate without optimally setting the mini-slot size will result in much less capacity gain than would be expected. Additionally, setting mini-slot size incorrectly can make large PDUs unable to be transmitted.

- Physical Layer capacity
    - Spectrum Bandwidth/Symbol Rate, Modulation order, Multiple access type, FEC Type, FEC Overhead, FEC Codeword size, Interleaver mode and depth, preamble length and unique word, Equalization
- MAC Layer Efficiency
    - Mini-slot Size, Maximum Burst Size, Long Data Grant, Short Data Grant, Shortened Last Codeword, Extended header size, Upstream Channel Change
- Scheduler tools
    - Unsolicited Grant Service, Dynamic Service methods, Polling Services, Fragmentation, Concatenation, Payload Header Suppression, and contention vs. unicast periods.

There are also many inputs listed below that must be considered and utilized when optimizing capacity. DOCSIS provides a rich set of standard management information through the Simple Network Management Protocol (SNMP) and a management information base (MIB). Other methods of network information collection also exist and are provided for in the DOCSIS specifications.

A rich set of vendor proprietary information also exists for harvesting. This information on network quality and performance along with current configuration of the knobs discussed above needs to be normalized across the network elements and analyzed to determine the optimal configuration for the network elements. With this analysis, MSOs can maximize capacity while maintaining the arbitrarily small error performance.

This analysis should consist of real time data collection from the network, complex analytical computer models that can be parameterized for real time use, empirical evidence from laboratory experimentation, and field experience all based on vendor implementations. All of these techniques should be used to determine the optimal configuration for maximum capacity.

- Physical Layer Inputs
    - Channel Quality –
        - Noise Type and statistics
    - Channel frequency magnitude and phase response
    - Power Levels
    - Available Spectrum
    - Error Performance
        - Packets, channel symbols, and codewords
    - CMTS Configuration

- MAC layer efficiency Inputs
  - MAC Frame Structure
  - MAC Header
    - EHDR including BPI, Req, zero pad, PHS, etc
  - Packet Size Distribution
    - VOIP, WEB, FTP, PHS
  - Channel Utilization
  - Total customers and active customers
  - CMTS Configuration

As shown in **Figure 3**, the inputs listed above and statistical analysis of the data, along with the creation of performance metrics, can be utilized to optimize Capacity. These analyses and calculations are based on the physical layer inputs described above, using the appropriate knobs to optimize for channel quality along with efficiently packing the data into the MAC layer based on the type of traffic and the mapping of DOCSIS MAC frames to the optimal physical layer, security, and MAC layer configuration. Using a capacity optimization system such as this will ensure excellent customer experience and will also ensure that the MSO is efficiently utilizing their physical resources and efficiently investing capital.

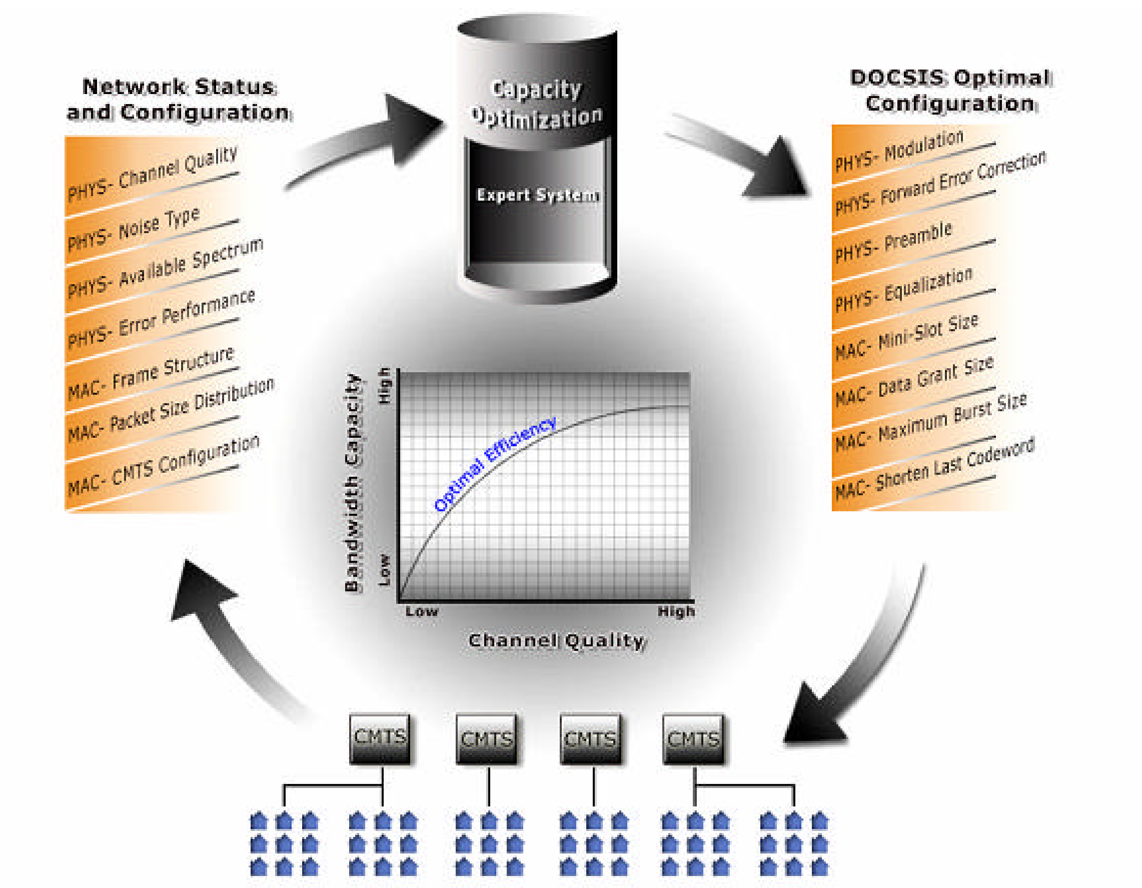

Figure 3: Capacity Optimization System.

Method II – Quality Optimization

Complementary to capacity optimization in the last mile, the second method to reduce inefficiency addresses service quality optimization. In this case, we combat inefficiencies impacting broadband experience by considering all elements and conditions degrading end users' perception of the service within the DOCSIS transport network.

Current approaches to monitoring quality using E/NMS (Element/Network Management System) technologies are limited in the following ways-

- Network element centric—Visibility into the condition of infrastructure elements provides a network-layer perspective of quality. Missing is a correlation between overall network status and its impact on the service experience of the subscribers.

- Undetected problems–Issues caused by complex or composite impairments affect the user experience but remain undetected by the monitoring system.

- Fault and performance centric--Problems detected by the monitoring system may have no discernable impact on subscriber experience.

To address problems unique to optimizing service quality management in DOCSIS networks a novel approach is required. The methodology described uses passive measurement to gather specific data related to network and infrastructure condition then correlates the value of each point of measurement to service quality. In turn, each correlation of network measurement to service quality is combined into a single composite figure that indicates the quality of the subscriber's experience. In this way, the root or composite cause of service degradation in the DOCSIS transport network can be quickly identified, topologically isolated, then cured thus optimizing service quality.

Quality optimization is ultimately a function of many points of measurement made in the DOCSIS network. The two general mechanisms for gathering of network measurements are active and passive-

- Passive monitoring relies on management instrumentation furnished by SNMP (Simple Network Management Protocol) agents embedded in network elements (CM and CMTS). These agents implement the suite of Management Information Bases (MIBs) required by the DOCSIS Operations Support System Interface (OSSI) standards [4].

- Active monitoring, the competing approach, is based on the introduction of synthetic traffic in an attempt to emulate service behavior. In this way the results of test traffic provides a sample of application and network quality.

Because the addition of traffic to the network through active measurement potentially exacerbates degraded conditions, passive measurement has been selected as the preferred mechanism for data collection. In addition, due to the "bursty" nature of traffic and impairments in DOCSIS networks, stochastic events may go undetected by scheduled active measurement.

The quality of the overall broadband experience is heavily weighted by the condition of the last mile infrastructure supporting it. Where DOCSIS provides this last mile transport of IP based services, two general areas reflect overall network quality-

- Connectivity—The physical condition of the DOCSIS connection between each CM and the CMTS. Contributors affecting connectivity include the health and configuration of the HFC plant as well as the state of the hardware resources terminating the connection.

- Capacity—The impact of traffic over the connection on latency and resources sensitive to network scheduling and loading. Contributors include downstream and upstream interface utilization, Media Access Control (MAC) domain loading, NSI utilization, queue depths, processing resources, and framing efficiency.

There exists no single point of measurement within a DOCSIS network that can be used as a proxy for overall network condition or service quality. Instead, each of the measurement points becomes a contributor to an estimation of connectivity or capacity. For each contributor, the opportunity exists to correlate the value of the raw network measurement to the effect it has on higher protocol layers. If a relationship between the measurement and overall service quality can be described, this correlation can be applied to the monitoring of an operational network transporting the service.

To illustrate, consider the quality of IP telephony service as a function of DOCSIS 1.1 Physical (PHY) layer response to HFC signal impairments. In terms of the protocol stack, upstream telephony data is packetized by first encapsulating it in a Real-Time Protocol (RTP) based UDP datagram. In turn, the datagram is wrapped in an IP packet and encapsulated in a DOCSIS frame. The DOCSIS frame is then embedded in a Reed-Solomon (RS) codeword and transmitted over the HFC network.

Packet loss is the most damaging form of impairment for IP telephony [5]. In the DOCSIS segment this loss is a function of either degraded capacity or connectivity. An operationally common cause of degraded connectivity is DOCSIS frame loss resulting from HFC impairments. This is expressed through an SNMP agent in either the CM (downstream) or CMTS (upstream) and reported as uncorrectable Reed-Solomon (RS) FEC. The theoretical relationship assuming ergodicity between Codeword Error Ratio (CER) and DOCSIS Frame Loss Ratio (FLR) is illustrated in Figure 4. The example VOIP Packet would reside in these bounds based on things such as EHDR length and PHS that impact packet size.

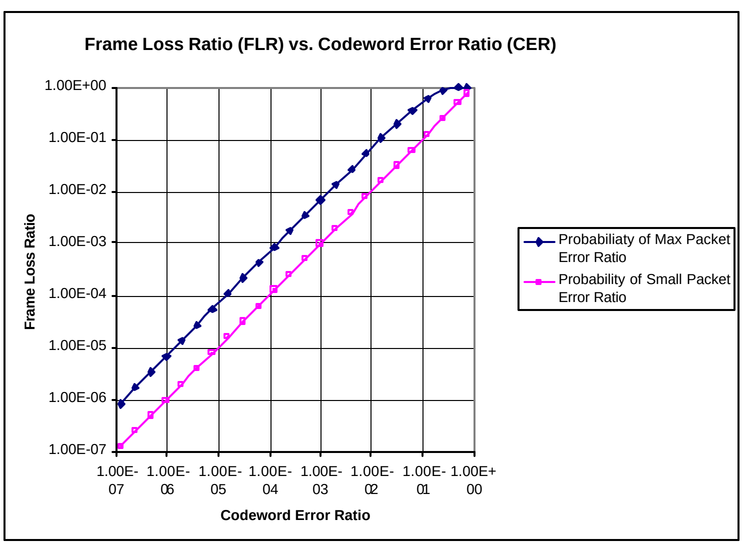

Figure 4: Frame Loss Ratio vs. Codeword Error Ratio

Holmes, Aarhus, and Maus [6] provide experimental results comparing frame loss to MOS (Mean Opinion Score) in an effort to describe a relationship between underlying network transport and subjective assessment of IP telephony service quality. Their findings suggest that with as little as 3% packet loss, audio quality degrades to an estimate of "fair".

Assuming that through capacity optimization, IP telephony packets in the upstream will occupy short-data DOCSIS frames, a one-to-one relationship between layer 3 IP packets and layer 2 DOCSIS frames exists. Through understanding this relationship, the correlation between an SNMP-based network measurement of physical layer impairment (CER) and the subjective quality of an IP telephony service on the application layer can be made.

In order to quantify the result of this correlation, we introduce the Degraded Modem (DM) metric. A DM event occurs whenever the value of a contributor is correlated to an instance of degraded service quality provided by a single CM. In the example provided, a degraded IP telephony event (of "fair") was defined at a contributed CER value of 3.0x10E-2.

We capture this relationship between the contributor and Degraded Modem event using the correlation function illustrated in Figure 5. In this example, the correlation function coupling CER and DM is based on both theoretical relationships between CER and DOCSIS frame loss, as well as experimental data correlating IP packet loss with a subjective score of IP telephony audio quality.

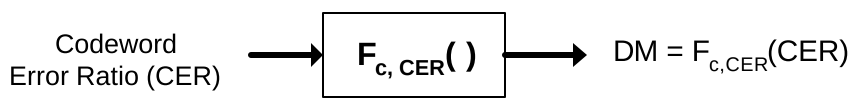

Figure 5: Correlation Function for CER Contributor

Likewise, for each contributor collected in the network a function can be described that correlates the state or value of the contributor to the degraded modem figure. In this way, all contributors and their associated DM figure can be combined using a logical OR operation to provide a composite estimate of service quality provided by the DOCSIS transport.

Figure 6 illustrates a simple combiner structure as an aggregation of outputs from an array of contributor correlation functions. In this case, the estimation of service quality (DM) is based on input from a total of six contributors. The combiner structure is extensible in order to accommodate contributors and correlation functions associated with the evolution of DOCSIS technology and the introduction of new IP based applications and services.

The combiner structure can be imposed on network topology in order to isolate the areas of the network introducing degradation. DM can be calculated and applied for any physical, logical, or organizational node within a MSO's DOCSIS network. The combiner provides a holistic approach to analysis of DOCSIS network quality by providing a method to estimate the health of the subscriber experience while abstracting the complexity of the underlying DOCSIS infrastructure.

If applied within the operational environment, the combiner provides a promising approach to quickly isolating and identifying conditions within the last mile infrastructure that is driven by the subscriber's experience.

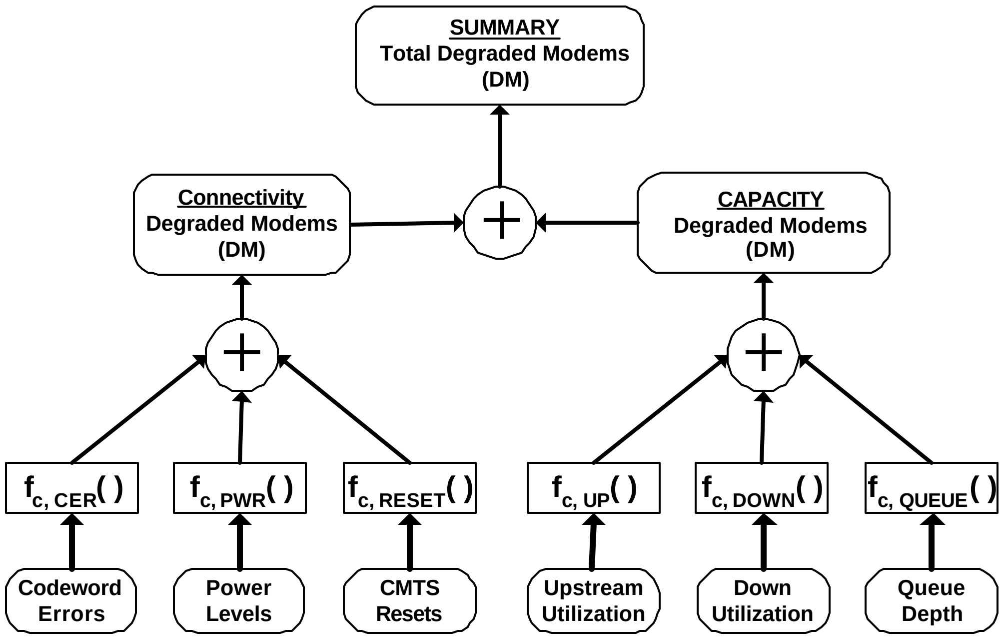

Figure 6: Simple Combiner Structure

SUMMARY AND CONCLUSIONS

Utilizing DOCSIS 1.0 and existing infrastructure, significant gains in capacity enabling new revenue and significant reduction in capex and opex can be obtained with a system that automates capacity and quality optimization. As DOCSIS technology evolves with versions 1.1 and 2.0 along with new DOCSIS devices, the customer base will experience exponential growth in the number of intelligent network elements. A scalable and reliable system that provides visibility, performance metrics and automated capacity optimization will become critical to deploying new IP-based services over the HFC network. Technology will continue to evolve, allowing new services and higher capacity and efficiencies. Costs of infrastructure will drop with volume, but complexity of operations will continue to increase and methods such as those discussed here will be essential to continued IP services product evolution.

REFERENCES

[1] B. Sklar, Digital Communications Fundamentals and Applications, Prentice Hall, 2001

[2] Performance Measures for Cable Data Transmission, NCTA technical papers, Rich Prodan, Dan Rice, Majid Chelehmal, 1999

[3] Shannon, C.E., “A Mathematical Theory of Communication,” BSTJ Vol. 27, 1948, pp. 379-423,623-657

[4] http://www.cablemodem.com/specifications.html

[5] “Modeling the Effects of Burst Packet loss and Recency on Subjective Voice Quality”, A.D Clark, Ph.D, 2001 IP-Telephony Workshop

[6] “Network Tolerance of Highly Degraded Conditions for an H.323-based VoIP Service”, Peter Holmes, Lars Aarhus, Eirik Maus, Norwegian Computing Center, P.O. Box 114, Blindern, Oslo, Norway

The authors can be contacted at Stargus, Inc. Andover, MA 01810, or {dan,jason}@Stargus.com