BANDWIDTH UTILIZATION ON CABLE-BASED HIGH SPEED DATA NETWORKS

Terry D. Shaw, Ph. D.[1]
CableLabs

Abstract

The CableLabs Bandwidth Modeling and Management project addresses the use, management, performance simulation, and network economics of cable high-speed data systems. On this project, we have analyzed consumer usage patterns based upon data collected on live cable-based high-speed data systems as well as network simulations.

Usage data on live cable networks indicate that traffic flows are fairly predictable over DOCSIS™ networks. Some of the primary patterns emerging include:

- ***Skewed distribution of bandwidth consumption**. As a general rule in many systems, 30% of the subscribers consume about 60% of the data.*
- ***Students drive seasonal characteristics**. System usage is appreciably higher during local school holidays and vacation periods.*
- ***Usage rapidly evolving.** In the data we collected, the average per capita use of data shows steady growth over time in both the upstream and the downstream, with upstream use growing more rapidly. This results in the average downstream/upstream symmetry ratio trending downwards, with a significant number of subscribers using more upstream than downstream data.*

These observations indicate the importance of deploying DOCSIS 1.1 in order to meet this increasing demand for upstream capacity. Preliminary simulation results indicate DOCSIS 1.1 will enable a significant increase in upstream system carrying capacity over the already substantial capacity of DOCSIS 1.0. These simulations indicate that DOCSIS 1.1 will allow almost 20% more upstream capacity than DOCSIS 1.0.

Simulations have also been used to study the characteristics of specific types of traffic found on cable networks. For example, simulations of peer-to-peer applications indicate that one user without rate limits can consume up to 25% of upstream capacity with a usage pattern resembling a very high-speed constant bit-rate application. These simulation results highlight the potential benefits for cable operators to manage their bandwidth using tools such as rate limits, service tiers, and byte caps on usage.

INTRODUCTION

The usage of DOCSIS high-speed data networks by a mass audience is expected to grow dramatically over the next several years. In order to gain a better understanding of the nature of growth to be expected, CableLabs has initiated the Bandwidth Modeling and Management project. This project has several aspects including the collection and analysis of bandwidth use on live cable high-speed data systems, the development of tools for the modeling and management of data traffic, and assessment of the strategic technical and economic implications of these measurements.

To date, this project has focused on the aggregate and individual characterization of the use of data by subscribers on DOCSIS and other cable-based high-speed data networks. We have studied patterns of behavior based on data collected from live cable systems and the performance of DOCSIS 1.0 and 1.1 systems in simulation. The Bandwidth Management project enjoys operational support from a number of cable operators in North America. The analyses in this paper are comprised of a composite view of these data collection efforts.

The primary focus of our data collection and analysis work during 2001 has been to characterize the data usage patterns over cable systems. As will be shown, usage data on live cable networks indicate that traffic flows are fairly predictable over DOCSIS and other cable-based high-speed data networks. However, the high-speed data market is rapidly evolving and continuing study is necessary in order to understand the network implications as new network applications and services are developed.

DATA COLLECTION AND ANALYSIS

Data has been collected from a variety of types of sources. The three primary sources used to date have been the web page MRTG[2] graphics used for bandwidth management by cable operators, data logs used to archive information by MRTG, and specially instrumented data collection systems.

A number of parameters can be measured and collected from high-speed data systems. However, the specific parameters collected vary from system to system depending on the available tools. Also, the means of presentation of the data are varied. Data collected on different systems must be normalized (converted to the same basis of measurement) so that results can be compared on an "apples-to-apples basis." Three parameters that are of particular use for the analysis of system performance and the analysis of user behavior are defined as:

- **Volume of data used.** The amount of data consumed (uploaded, downloaded, and uploaded + downloaded) during the measurement interval. This parameter can be collected for an aggregate (for example, all of the users on a given upstream or downstream frequency).
- **Data Rate.** Practically, the average data rate is merely an expression of volume of data used within a specified short time period, usually a second. It varies from volume of data rate primarily in the time units of the measurement interval. It is usually collected for an aggregate (for example, all of the users on a given upstream or downstream frequency).
- **Symmetry.** This parameter is the ratio of (Downstream)/(Upstream) where Downstream and Upstream refer to either volume of data used or data rate. It is used to characterize user behavior and has many system architectural implications.

General Observations

The access to live network traffic data provided to CableLabs by its members provides insight to a number of facets of network performance and operational methodologies. These include:

- **Daily (diurnal) usage patterns**
- **System capacity loading**

- **Seasonality of use**
- **Overall trends in volume and symmetry of network use**

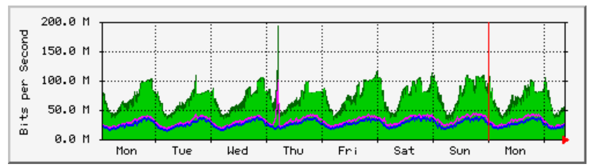
All of the findings in this report should be regarded as preliminary because of the limited amount of data analyzed, and the way individuals are using cable-based high-speed data systems is constantly evolving. Continuing study is necessary in order to understand the evolution of system use. The mathematical characteristics of several of these parameters are discussed in the appendix.

Diurnal Usage Patterns

Our traffic measurements indicate that regular or predictable diurnal, or daily, variations of the volume of traffic flow over a network are present. These variations are basically sinusoidal in nature with a period of 24 hours, with possible variations of behavior between weekdays, weekends, and holidays.

While the specific pattern diurnal variation observed will vary from system to system, the pattern shown in Figure 1 is fairly typical. This graph, taken from a MRTG site used by a system operator for on-going system monitoring, represents the aggregated traffic of cable modem subscribers during the period 11/26/01 through 12/4/01 in a major metropolitan market. The green area represents the downstream data use and the magenta line represents the upstream data use. Each point graphed represents the average aggregate data rate for a one-hour interval. In this market, traffic begins to build in mid-afternoon on school days and typically reaches its peak around midnight local time. It then rapidly falls to a minimum at about 5 a.m. local time. In this case, the daily usage peak "busy hour" is from 8 p.m. to midnight on Friday evening. Weekend use is similar to weekday use with more traffic seen earlier in the day (the spike occurring on Thursday morning is caused by a rollover in the counter used to collect this data).

Figure 1. Daily system usage pattern for a metropolitan system.



Indications of Seasonality

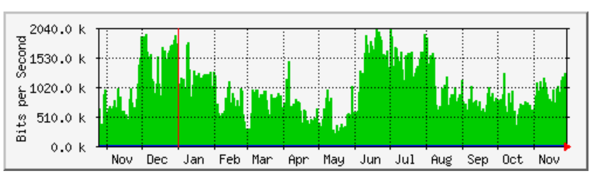
Seasonal variations are the macroscopic variations in the way a system is used throughout the course of a year. During the early stages of system deployment, the seasonal variations are usually obscured by the overall increase in traffic due to the addition of new subscribers on existing network segments. This behavior can also be masked as users are re-allocated to different network systems as the subscriber traffic increases. Hence, seasonality can only be observed in systems where the same user population has been fairly stable or slowly increasing over a longer (more than six months) period of time. In systems where this has been the case, however, the seasonal effects can be clearly visible. However, Internet usage is continuously evolving, and these variations must be tracked over a long period of time so that the effects can be clearly understood.

The seasonal variations are clearly visible in Figure 2. In general, high traffic times appear to correspond well with school vacations and major holidays[3]. This traffic data shows a marked increase about Thanksgiving (end of November), slows during mid-December, and upsurges again near the end of December (Christmas). After the December school break, traffic decreases with a brief upsurge in early April (Spring Break) and early May (Finals Week). When school lets out for the summer (early June), there is a tremendous sustained increase in the traffic until early August.

After schools reconvened for the Fall term, traffic reverted to a level similar to that seen in the Spring. There is one minor upsurge on Columbus Day 2001 (early October) that corresponds to a similar upsurge on Columbus Day 2000 (not shown on this graph).

Figure 2 is a MRTG graph of the upstream traffic on a well utilized upstream for the time late-October 2000 through November 2001. This upstream serves an area in which cable modem service has been available for several years, and has had fairly stable (but slowly growing) user population over the course of the period May through November 2001. Each data point represents the 24-hour average traffic on the upstream for this aggregate of users.

Figure 2. Seasonal high traffic appears to correspond to school vacations for late-October 2000 through November 2001.



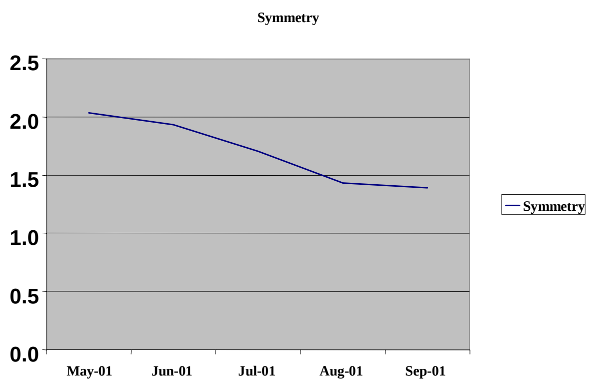
Overall Trends in Volume and Symmetry

The overall trends of the use of data (volume of data consumed per user, symmetry of data flow on the network) will vary depending on a number of factors, including:

- The demographics of a particular region,
- User sophistication,
- "Neighborhood effects" such as the popularity of an interactive game played among residents of a neighborhood,
- Geographic and weather factors, and
- System management policies (such as tiers of service based on data consumed).

These effects will vary from system to system and from node[4] to node within a system. Overall, across a large aggregation of users, the volume of data consumed per user appears to be increasing, and the downstream/upstream ratio of data consumed appears to be decreasing. Figure 3 shows the average symmetry, the ratio of downstream to upstream use, calculated for the months of May through September 2001 for a very large aggregate of subscribers. During this period of time, the symmetry ratio decreased from just over 2 (2 times as much downstream data as upstream) to 1.4. This value was calculated based on the data flow on the network. Other measurements taken on an individual basis on a node indicates that individuals vary widely in their symmetry of use. The decrease in the overall average symmetry of traffic has been observed on a number of different systems. Early indications are that, on an individual node basis, the symmetry is relatively high (about 3) when a system is first deployed and decreases as the users on the system begin to more fully understand the capabilities of the cable high-speed data communications system.

Figure 3. Overall symmetry of network use based on volume of traffic flow for a large aggregate of users.

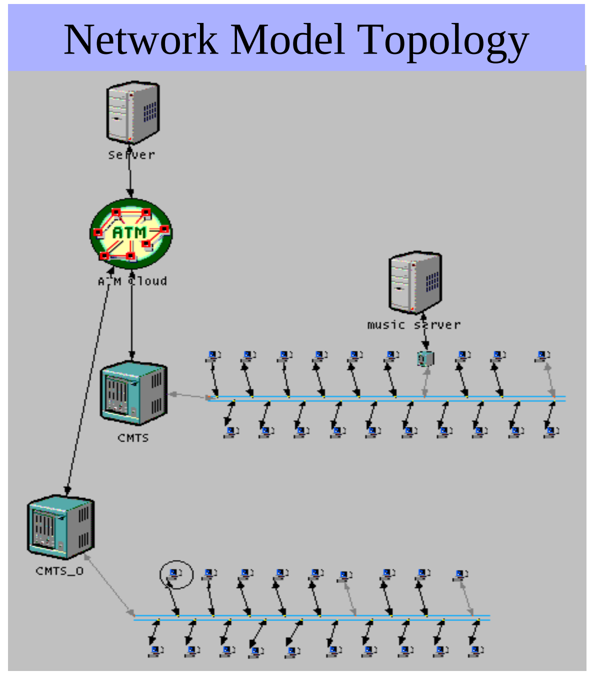


SIMULATION of DOCSIS NETWORKS

An important reason for performing data collection and analysis is to have a basis of empirical data to use for the construction of predictive tools. Analysis of measurement data, such as the probability distributions discussed in the Appendix, are essential for making simulations realistic and accurate. Predictive tools can simulate network performance given the behavior of specific network elements based on these data. These tools can also be used to study the behavior of new applications and services on the network. One such predictive tool is the Modeler software developed by OPNET.

CableLabs developed a model of the behavior of the Gnutella peer-to-peer file transport protocol on a DOCSIS 1.0 access network using the OPNET Modeler software program in order to determine the types of traffic characteristics that can be expected from these applications.

Figure 9. Network topology used to examine the effects of peer-to-peer file sharing on a DOCSIS access network.

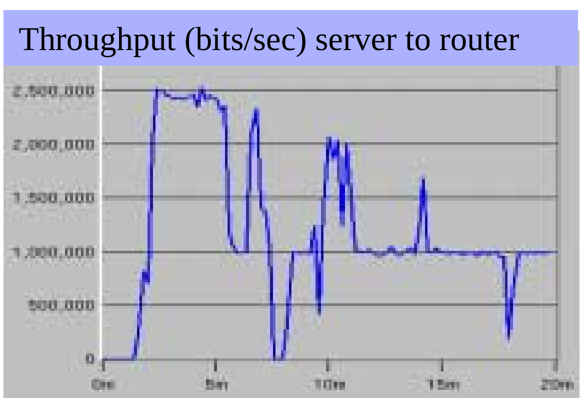


In this simulation, two DOCSIS access networks were configured. Figure 9 provides a diagram of the network topology studied in this simulation. One of these access networks (connected to ***CMTS*** in Figure 9) hosted one cable modem that supported a music server. In the simulation, this modem-server (***music server*** in Figure 9) combination would replicate the performance of a single subscriber that was acting as a source of files (such as MP3 audio files) to other users of the peer-to-peer application. The MP3 file transfer was simulated as a 5 Mbyte HTML file download from this server to users of the peer-to-peer application. The users of the peer-to-peer application were simulated by 12 of the cable modems connected to ***CMTS_0*** in Figure 9, each of which would download five of the simulated MP3 files per hour. These modems served as data sinks (or destinations) for the files transferred from ***music server***. The primary role that these modems played in the simulation was to send requests for file transfers to ***music server***. The module ***server*** acted as the source of typical background traffic for this simulation.[5]

In this simulation, the DOCSIS access networks were configured without any rate limits in order to study how much traffic could be generated by the single subscriber hosting the ***music server***. The DOCSIS upstream data rate simulated was 5.12 Mbps. The simulation was configured to simulate the activity generated in 20 minutes of network activity.

The effect on the upstream traffic generated by the ***music server*** in this limited simulation is shown in Figure 10.

Figure 10. Upstream traffic load generated by a single user serving MP3 files on a DOCSIS 1.0 access network. In this simulation, the CMTS is configured as a router.

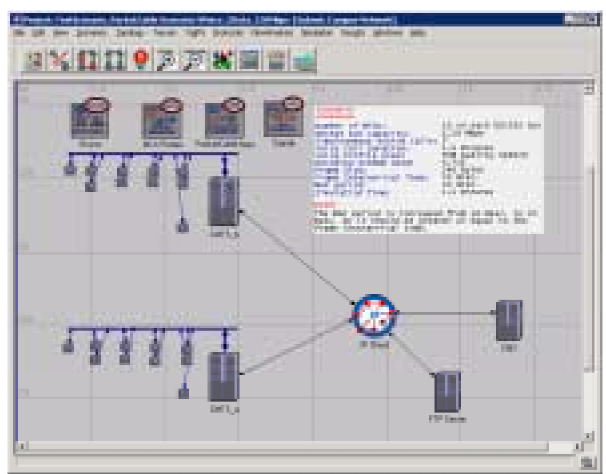


As can be seen in this graph, this single user consumed nearly 50% of the upstream bandwidth available for a substantial proportion of the simulated period, and nearly 25% for the rest of the period. Therefore, a single user of peer-to-peer applications will have a persistently high data rate resembling a very high-speed constant bit rate application. The results of this simulation correspond closely with network behaviors that have been observed in the real-world environment. This simulation highlights the importance of bandwidth management tools such as rate limits, volume limits, and traffic prioritization and differentiated services (from DOCSIS 1.1 and PacketCable™) for assuring a well-functioning network and good customer experience for all subscribers.

PACKETCABLE SIMULATION

CableLabs developed PacketCable protocol extensions to the commercially released DOCSIS 1.1 OPNET model library. A number of scenarios were developed in conjunction with the PacketCable project to test the capabilities of the model and explore the efficiency of the DOCSIS 1.1/PacketCable protocols. The basic scenario topology used for testing is shown in Figure 11. In the scenarios tested, the upstream channel data rate was set to 1.28 Mbps.[6]

Figure 11. PacketCable™ Scenario Topology



In these scenarios, two CMTS supporting 10 multimedia terminal adapters[7] (MTA) each were configured that were able to communicate through an IP cloud. A combination of voice (using the G.711 codec) and data traffic was modeled in order to study specific load conditions. In the simulation, calls originated on an MTA on a bus supported by one CMTS were terminated at an MTA supported by the other CMTS, and all of the voice calls were initiated and terminated simultaneously in order to produce "worst case" results. Due to the low upstream data rate (1.28 Mbps) and the high data rate required by the G.711 codec (nominal line rate is 64 kbps, but encoding inefficiencies create an actual line rate closer to 128 kbps), the upstream channel was saturated with relatively few simultaneous calls. In fact, in simulation, the upstream channel could only support 9 simultaneous calls; a tenth call specified in the simulation was not completed due to lack of available bandwidth at a quality level needed to support the call. Figure 12 provides a graph of the upstream channel utilizations expressed as a percentage of the total possible load throughput rate for the scenarios with 8, 9, and 10 specified voice calls. (Note that the number of calls was limited due to the low data rate of the underlying DOCSIS system. Much higher data rates, with resulting call carrying capacity would be used in actual system deployments).

Figure 12. Upstream channel utilization for PacketCable™ voice usage simulation.

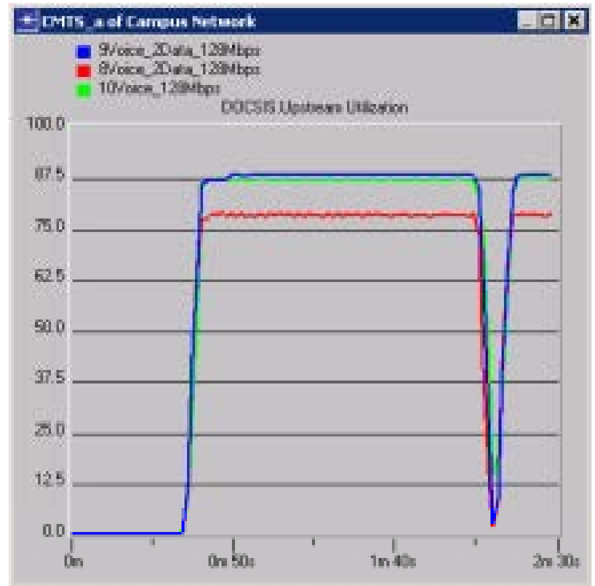


Figure 12 shows that in this simulation, the DOCSIS upstream has about 78% utilization (or about 1 Mbps in this case) with 8 voice calls and 87.5% utilization with 9 and 10 specified calls (Call 10 was not completed). The simulation was also tasked with simultaneously collecting information on the performance of the data applications (modeled as uploads of file transfer protocol, or FTP, traffic from two cable modems to the CMTS).[8] Figure 13 provides a graph of the amount of data uploaded to the CMTS for the same time interval. Note that in the case with 8 voice conversations and two simultaneous data uploads, the performance is fairly constant. With 9 voice conversations, the data rate peaks sharply in the brief interval between voice conversations. Another view of the performance of the data transfer is provided by Figure 14, the response time for data traffic uploaded to the CMTS. This graph shows that in the case of 8 speakers, the file transfer takes place with a fairly constant delay of about 1 second.[9] In the case of 9 speakers, the data is buffered at the cable modem until enough bandwidth is available to transfer it to the CMTS.

Figure 13. Rate of data uploaded to CMTS while voice conversations are taking place..

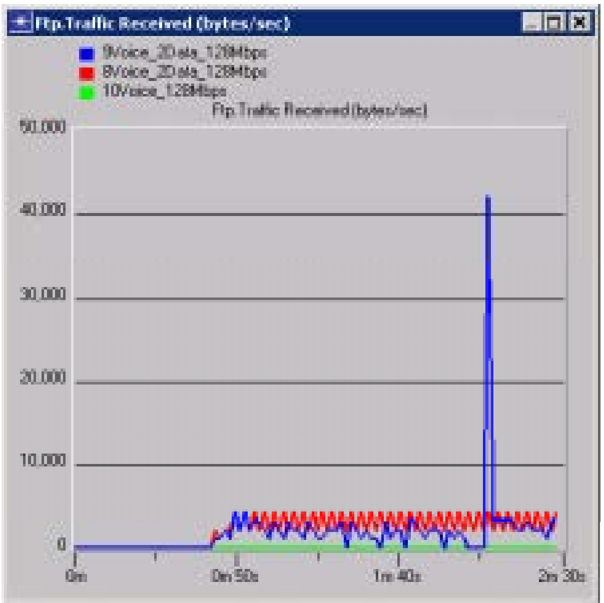
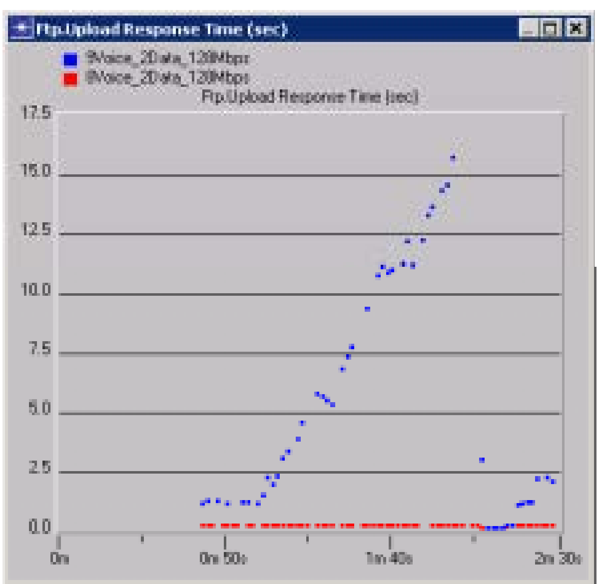


Figure 14. Data upload response time for various loading conditions.



Taken all together, these preliminary results indicate that the DOCSIS 1.1 system with PacketCable protocol can operate at a utilization level of over 78%. This can be compared to earlier simulation results (circa 1999) that indicated that the DOCSIS 1.0 system performance would degrade rapidly at a utilization level of 65%. Thus, these preliminary simulation results indicate that DOCSIS 1.1 will have a 15% increase in carrying capacity over DOCSIS 1.0. Furthermore, Figures 13 and 14 indicate that DOCSIS 1.1 allows graceful degradation of the upstream under peak loading conditions in that bandwidth for support of lower priority applications is cannibalized to support higher priority applications.[10]

CONCLUSION

The use of cable-based high-speed networks has evolved substantially over the course of the last year, and will continue to evolve as new applications and services are deployed on the networks. This paper marks the first, preliminary steps in understanding this evolution of use. In order to manage this evolution, the cable industry needs to continue the collection of data, not only on the access network, but also at higher levels of aggregation so that we can determine the nature of local content management that will be required. The cable industry also needs tools ranging from specialized data collection and analysis techniques to innovative ways to manage service deployments in a tiered service environment. To this end, CableLabs will continue analysis of bandwidth use on live cable high-speed data systems, the development of tools for the modeling and management of data traffic, and assessment of the strategic technical and economic implications of these measurements in the coming year.

APPENDIX: Observed Parameters and Their Distributions

Appendix A describes the parameters used for the detailed analysis of data on the Bandwidth Management Project. This appendix also reports preliminary research findings on the probability distributions of these parameters based on live network traffic data.

A.1 Volume of data use

A.1.1 Definition

The amount of data consumed (uploaded, downloaded, and uploaded + downloaded) during the measurement interval. Specifically, this parameter corresponds to the number of bytes of information delivered to, or originating from, a single cable modem as represented by the MAC address[11]. This parameter can be collected for an aggregate (for example, all of the users on a given upstream or downstream frequency) and an individual (per MAC address) basis.

A.1.2 Use

This parameter is useful for examining clusters of users and pattern of use. It can also be used as a measurement mechanism for volume-based tiering (where a usage tier is defined in terms of, for example, number of Gigabytes per month).

A.1.3 Units of Measurement

This value is recorded in terms of bytes, and typically measured in Mbytes/(unit period) for cable modem networks. The unit period may be seconds, hours, days, or months.

A.1.4 Distribution

If **V** = Volume of data used by an individual user, the probability

$$P[\mathbf{V} < v]$$

has been observed to have an exponential distribution for upstream, downstream, and total volume of data used.

A.2 Data Rate

A.2.1 Definition

Theoretically, the data rate is the derivative of the volume of use with respect to time. Practically, the average data rate is merely an expression of volume of data used within a specified short time period, usually a second. It is usually collected for an aggregate (for example, all of the users on a given upstream or downstream frequency). The use by an individual is difficult to instrument at this time and is usually derived from a longer term volume measurement for the user.

A.2.2 Use

This parameter is typically collected by network management tools such as MRTG, and is the fundamental parameter for system capacity planning due to the small time interval of measurement and use in the definition of communication systems. The time trajectory of this parameter for successive measurement intervals for an aggregate of users, particularly to those users assigned to the same upstream and downstream frequency of operation, is a fundamental tool for use in capacity analysis. The distribution characteristics of this parameter are useful in predicting the performance and QoS characteristic of a network segment.

A.2.3 Units of Measurement

Typical measurement units for this parameter are bits/second (bps) and bytes/second (Bps) (8 bits = 1 byte).

A.2.4 Distribution

For aggregates of users over typical measurement intervals (5 minutes, 15 minutes), the time trajectory of this parameter has the characteristics of a normally distributed, or Gaussian, white noise. That is, individual observations of this parameter are normally distributed with mean equal to the average data rate. The standard deviation of this parameter can be used in conjunction with the mean as the parametric basis for system capacity estimation. The diurnal variation $\mathbf{X_t}$ in the data rate load on an upstream or downstream has the formal functional form (at time t) as a stochastic process:

$$\mathbf{X_t} = \mathbf{R(t)} + \mathbf{n_t}$$

Where

R(t) is a sinusoidal function expressing the mean data rate, and

$\mathbf{n_t}$ is a zero mean Gaussian white noise with standard deviation much less than **R** since the data rate will always be positive.

The expected accumulated volume of use is related to data rate by the expression:

$$\mathbf{V(T)} = \int_{[0,T]} \mathbf{R(t)}\, dt$$

Since the expected mean of $\mathbf{n_t}$ is 0.

A.2.5 Derivation

The MRTG tool is commonly used to collect and report data flow rates. In order to collect the data, MRTG will poll periodically a router to collect the value of a byte counter to get a value V(n) (the value of the counter at polling interval n). In order to determine the data flow in bytes per second during the period, MRTG computes

F(n) = [V(n)-V(n-1)]/(number of seconds in reporting period)

In order to get the flow value in bits per second for the reporting period, it is enough to compute 8*F(n). The normality of this distribution is to be expected from the Central Limit Theorem which states, in effect , that if one takes random samples of size ***n*** from a population of mean ***m*** and standard deviation ***s***, then, as ***n*** gets large, then the distribution of the average of these samples ***X*** will approach the normal distribution with:

- Mean = ***m***
- Standard Deviation = $\boldsymbol{s/\sqrt{n}}$

A.3 Symmetry of Data Use

A.3.1 Definition

Downstream Data used per measurement interval/Upstream data used per measurement interval.

A.3.2 Use

This parameter has many implications for system architecture and a fundamental indicator of system usage by subscribers.

A.3.3 Units of Measurement

This parameter takes a dimensionless positive value. Values in the range of 0.5 to 4 are typical. It can be expressed as a ratio R or in Decibels (DB, 10 $\log_{10}$(R)). The data used can be measured in volumetric or flow-rate terms, and can be used to measure populations and individuals. The key item is to maintain the same interval of measurement for comparison.

A.3.4 Distribution

If **R** = symmetry of data use by an individual then

$$P[\mathbf{R} < r]$$

Has been observed to have a log normal distribution when expressed as a ratio. When expressed in DB, the value $\mathbf{R_{DB}} = 10 \log_{10}(\mathbf{R})$ follows a normal distribution.

REFERENCE NOTES

[1] The views expressed in this paper are solely those of the author. The author appreciates detailed manuscript comments from Simon Krauss, Dorothy Raymond, and David Reed and manuscript preparation assistance from Kathy Mitts. The author, however, bears full responsibility for any errors of fact or interpretation in the paper.

[2] MRTG (Multi Router Traffic Grapher) consists of a Perl script which uses SNMP to read the traffic counters of routers and a C program which logs the traffic data and creates graphs representing the traffic on the monitored network connection. These graphs are embedded into web pages that can be viewed from any modern Web-browser. MRTG is available as a download from: http://mrtg.hdl.com/

[3] A working hypothesis for this is that, in many cases, the primary residential users of the system will be of school age. A holiday from school provides an opportunity to use the system. This effect is also seen in the diurnal variations. It has also been observed that systems serving communities with 12-month schools do not exhibit seasonal fluctuations to the same extent as communities with 9-month schools.

[4] In this paper, the term **node** refers to the group of subscribers that receive their cable modem service from the same blade in the CMTS.

[5] Typical background traffic consisted of a mixture of simulated Email, web page requests and downloads, and file transfers.

[6] The model is capable of using upstream channel data rates of 1.28, 2.56, 5.12, and 10.24 Mbps. 1.28 Mbps was used in order to reduce complexity during scenario development and testing and to explore the inefficiencies inherent in lower data rate systems. In general, DOCSIS systems work better when run at higher data rates due to the benefits of statistical multiplexing.

[7] The MTA is specified in the PacketCable specifications. In the simulation, it is modeled as a cable modem that can support both data and PacketCable protocol sessions.

[8] The file transfers used in this simulation were made at a simulated data rate of 1.28 Mbps (the label stating 128 Mbps in the figure legend is a misprint. The legend will not support the use of a decimal point).

[9] This delay is an artifact of the structure of the simulation.

[10] The issue of when an upstream DOCSIS 1.0 upstream channel is nearing its capacity limit is a difficult one to address. Most of the evidence in this area is anecdotal. The capacity limit depends on a number of factors including: The sophistication of the scheduling algorithm in the CMTS; the upstream channel line rate; the types of applications run by network users; and the number of MAC addresses (cable modems) that a CMTS blade can efficiently support

[11] The MAC address (Media Access Control Address) is the unique hardware number assigned to network connection devices such as cable modems.