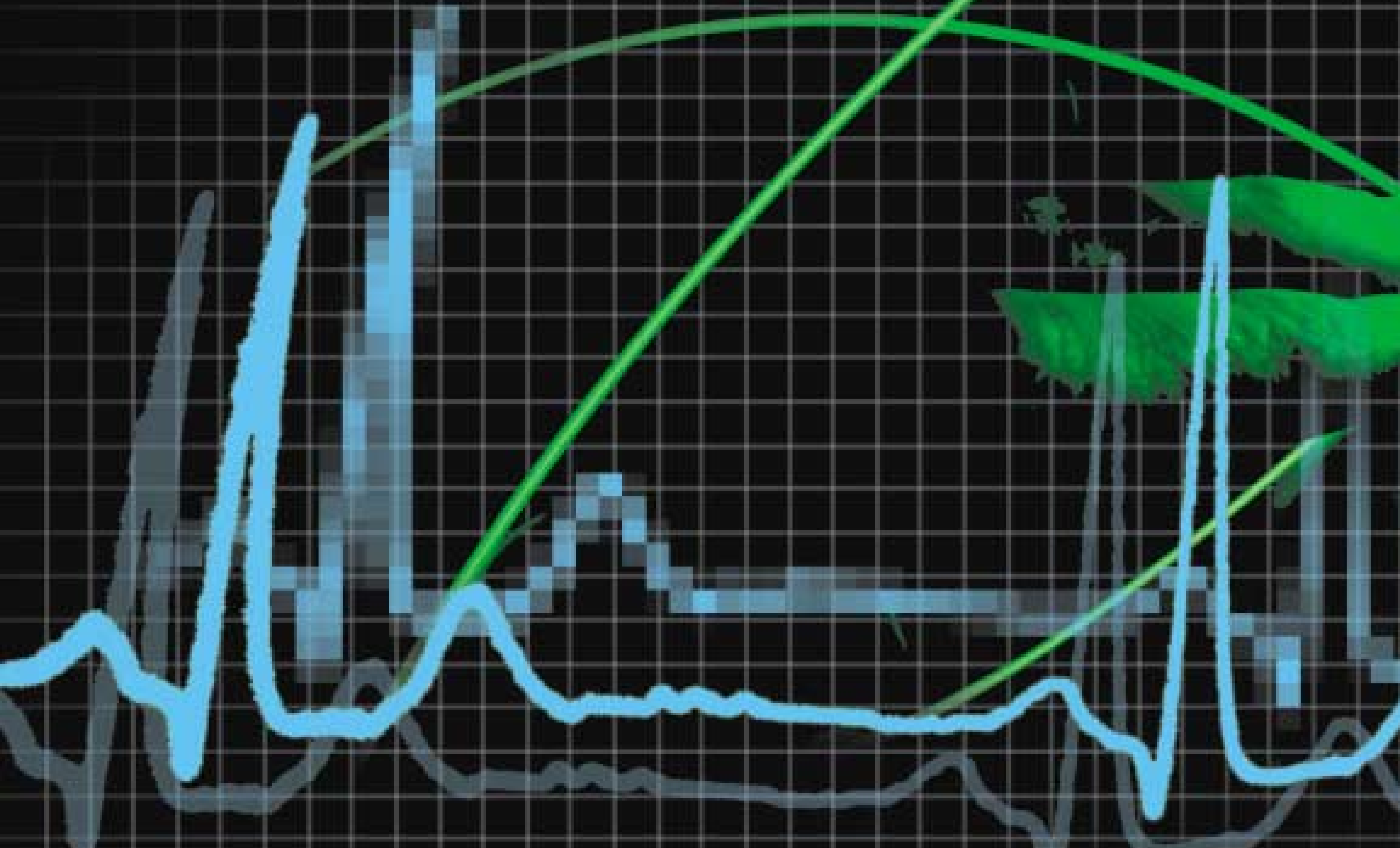
Multimedia

TRAFFIC ENGINEERING

Understanding the Bursty Data Model

By John T. Chapman, Cisco Systems

Networks that carry data, voice and video must be engineered if they are to operate properly. By modeling data traffic, engineers may predict how many modems their system can support.



Reprinted with permission from the April 2002 issue of *Communications Technology*

How many cable modems can a cable modem termination system (CMTS) support? This has to be one of the most often asked questions in our industry. Is it 250 cable modems per upstream? It might be, but only for a given set of circumstances. What if those circumstances, such as channel bandwidth or user bandwidth, change? How much bandwidth does each subscriber actually get? How much do they need?

Another approach to answering this question is to use one less cable modem than the total number it takes to break the system. While this may work in the beginning, it obviously doesn't scale for larger networks. The real answer is that it depends. If the total bandwidth available from the CMTS and the bandwidth required for the user could both be accurately described, then the answer would be:

$$\text{Number of CMs} = \frac{\text{Total BW}}{\text{BW per CM}}$$

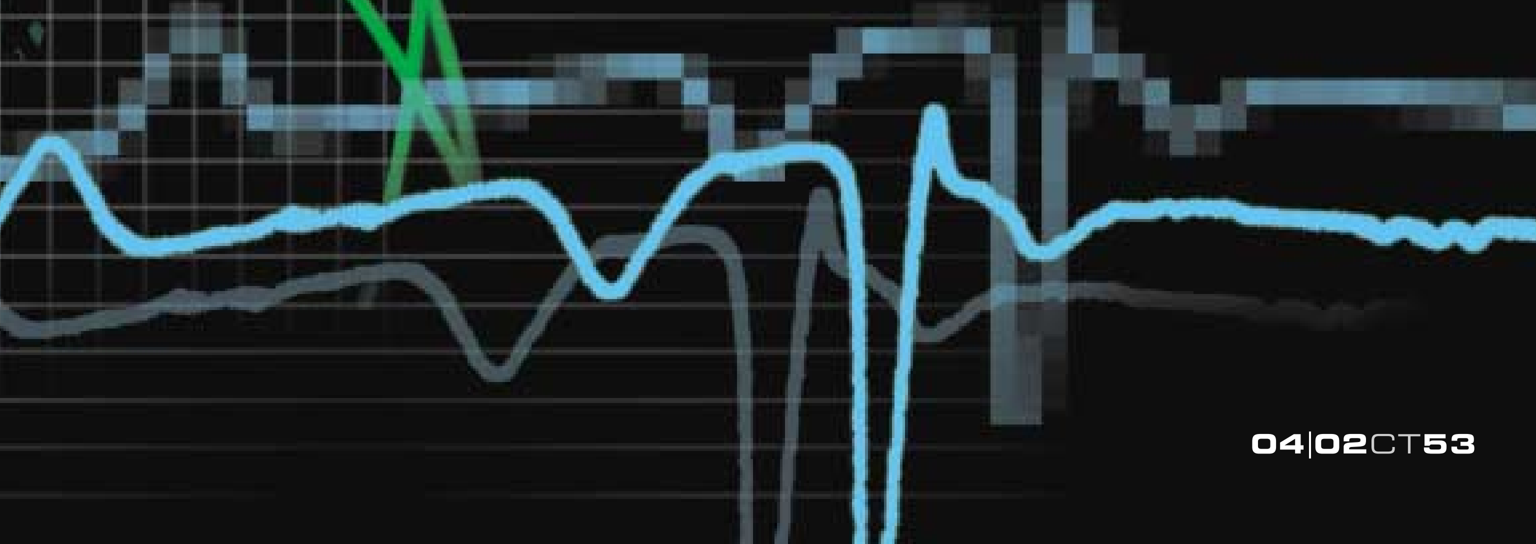
Although this equation is simple, the derivation of the bandwidth per cable modem is not. To perform this calculation, we need a model for the bandwidth usage of the user behind the cable modem. This bandwidth consumption depends on both the activity level of the user and the applications that are being run. These applications generally fit into three distinct categories:

- **Data:** This includes applications such as Web traffic, e-mail, file transfers and lower bit rate audio and video streaming.
- **Voice:** This includes all voice traffic that generally is constant bit rate traffic and is carried over Internet protocol (IP) or circuit-switching equipment.
- **Video:** This includes higher bit rate broadcast quality video, usually those encoded as Moving Picture Experts Group (MPEG)-2, which is of sufficient bandwidth that usually multiple downstreams-per-fiber node are required. The transport could be either native MPEG-TS or IP over Data Over Cable Service Interface Specification (DOCSIS).

Each of these categories requires a different traffic model. The final model is the sum of these three models. This article focuses on the data model. The development and use of these models for IP networks, including hybrid fiber/coax (HFC) DOCSIS networks, is a practice I refer to as multimedia traffic engineering, or MMTE for short.

Bandwidth modeling

Modeling data traffic is not easy. While voice and video traffic tend to be somewhat predictable and linear, data traffic is completely the opposite. Just when you think you have begun to understand the applications and data flows on your



network, a new data application arrives that changes everything. Such was the case with file sharing programs such as Napster, and the newer Morpheus and KaZaA.

Modeling all the different data applications can get quite complex. Some fields of research are exploring fractal mathematics to predict usage patterns. To an extent, all models are inherently flawed because they are models and are not reality. However, if the limitations are understood, models become quite useful for understanding and sizing networks.

Modeling what already exists is interesting, but modeling what does not yet exist is what is most important. The real value of modeling is the promise of predicting the future. A good model allows current measured results to be combined with theory and growth projections to predict future requirements.

The need for a basic model

What really is needed is a model that is both simple and useful. The measurement of a simple model would be that the equations must:

- Fit on the back of an envelope
- Be easy to put into a spreadsheet
- Be easy to use by all

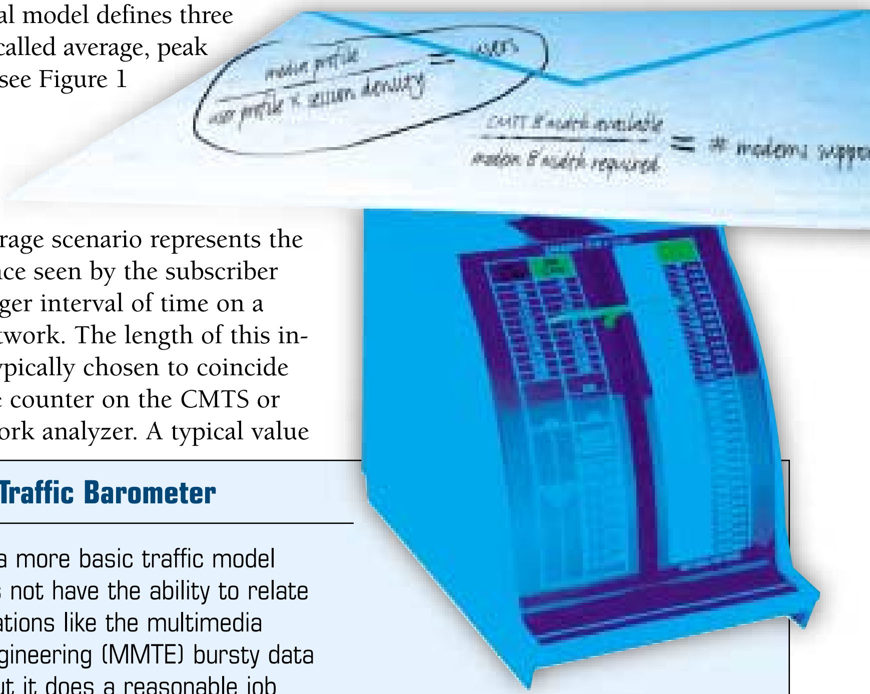
Approximations are fine as long as they are understood and complications are avoided. The measure of usefulness would be that the model must relate to measured parameters and be usable for bandwidth calculations. A model would not be of much use if it could not be verified by field data, and if it could not be used with traffic engineering to generate actual results.

The bursty data model

The bursty data model is an attempt to satisfy the previously mentioned criteria. This model is derived in large part from a behavioral description of what the subscriber and operator see when looking at the network.

The bursty data model uses scenarios. Each scenario has an interval of time known as the measurement interval. During that interval, the number of users and their bandwidth usage is determined. By multiplying these two numbers together, you get the bandwidth for that scenario. The scenario then may be used for calculating how a larger network will operate and how many cable modems it will support. That's about it.

You may define any number of scenarios, each with its own time interval. The general model defines three scenarios called average, peak and max (see Figure 1 page 55).



The average scenario represents the performance seen by the subscriber over a longer interval of time on a loaded network. The length of this interval is typically chosen to coincide with some counter on the CMTS or on a network analyzer. A typical value might be five minutes.

For example, let's say that for the average scenario, the downstream rate per user is 80 kbps. That means the user has received three million bytes over a five-minute interval. Those bytes may have come in multiple random bursts from multiple applications, but over a five-minute interval, the user received three million bytes, which equates to 80 kbps. This is ob-

> The Traffic Barometer

There is a more basic traffic model that does not have the ability to relate to applications like the multimedia traffic engineering (MMTE) bursty data model, but it does a reasonable job of indicating growth. This model is referred to as the MMTE traffic barometer.

In general, it is very easy for a CMTS to measure the number of cable modems per upstream and downstream and the average bandwidth in terms of Mbps and PPS per u/s, d/s, and WAN port.

With these numbers, usually available through simple network monitoring protocol (SNMP), you may calculate the following statistics for the upstream and the downstream:

- Average Mbps per cable modem: for example, 20 kbps
- Average PPS per cable modem: for example, 5 PPS
- Average packet size: for example, 400 bytes

These metrics may be tracked as the network grows, and used to measure trends and predict new growth. Other trends, such as the number of cable modems per upstream over time, are also useful trends to follow. MRTG, the Multi Router Traffic Grapher, from http://people.ee.ethz.ch/~oetiker/webtools/mrtg/, is a popular graphing program available on the Internet, and is often used for plotting these trends.

Note that the traffic barometer is useful for measuring trends such as growth, but because of its low sampling, is not granular enough to notice traffic peaks reliably. Although a good rule-of-thumb, the traffic barometer must be combined with theory and the MMTE bursty data model to correctly predict new traffic patterns.

viously more of a measured statistic than a real-time performance number.

The peak scenario is the performance seen by the subscriber over a shorter time on a loaded network. A typical value may be one second. It is typically chosen to relate to some measurable user experience.

For example, let's look at a downstream channel that has a payload bandwidth of 26.25 Mbps. How many users would share those 26.25 million bits, and how many bits would each one get? One second of 26.25 Mbps would be 2,161 packets if each packet were 1,518 bytes.

How would those packets get allocated to your users?

The applications demanding bandwidth during the peak scenario such as file transfer protocol (FTP) or video streaming typically use large packet sizes. By using a different packet size for the peak calculation, the packets per second (PPS) requirement of the network may be partially relaxed.

The max scenario is the rate seen by the user when the network is not loaded. This is the value that the CMTS uses to rate-shape the traffic to a cable modem and is the only scenario that is enforced by the CMTS. The max scenario is more for completeness and to check performance when rate-shaping. The average and peak scenarios are the two main scenarios that are used. >

FIGURE 1 | Three Scenarios for the Bursty Data Model

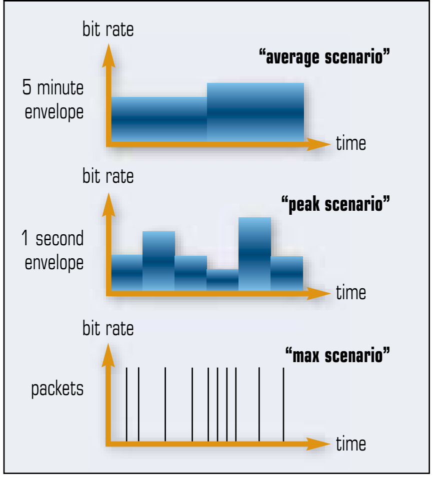


FIGURE 2 | The Bursty Data Model Calculations

Multimedia Traffic Engineering -- Bursty Data Model — v1.3

Up Stream	User Profile: Bit Rate kbps	User Profile: Pkt Size bytes	User Profile: Pkt Rate pps	Upstream: Pkt Rate pps	Upstream: Sessions	Session Density (SD): Relative %		Session Density (SD): Direct %	Users per U/S	Final PPS per U/S
Average	24	64	47	2424	52	25%	of users	25%	207	2245
Peak	100	1518	8	150	18	20%	of avg	5%	365	79
Max	384	1518	32	150	5	20%	of peak	1.0%	475	61

U/S Bit Rate	2.56	Mbps
U/S Admission	80%	for Data

Max Users per Upstream: 207
Final CM per Upstream: 192

Down Stream	User Profile: Bit Rate kbps	User Profile: Pkt Size bytes	User Profile: Pkt Rate pps	Downstream: Pkt Rate pps	Downstream: Sessions	Session Density (SD): Relative %		Session Density (SD): Direct %	Users per D/S	Final PPS per D/S
Average	80	400	25	7185	287	25%	of users	25%	1150	7185
Peak	256	1518	21	1931	92	20%	of avg	5%	1832	1212
Max	2000	1518	165	1931	12	20%	of peak	1.0%	1173	1893

D/S Bit Rate	26.25	Mbps
D/S Admission	90%	for Data

Max Users per Downstream: 1150
Final CM per Downstream: 1150

Direction	Domain Ratio	Users Allowed	Users per Domain: max	Users per Domain: final	Users per Dir	Max HHP per Direction
Upstream	6	207	1241	1150	192	1916
Downstream	1	1150	1150		1150	11496

Users per HHP	1
MP of data	10%

This system is downstream limited.

Scenario B: At 2000 HHP per u/s, the max market penetration for data is 10%

Scenario C: At 4 domains per CMTS, the CMTS nominal packet rate is 82,630 pps.

Probe: on

Legend: inputs — Yellow cells with underlines indicate input cells.
outputs — White cells are output cells.

The average, peak and max scenarios are repeated separately for the upstream and the downstream, then all six cases vote to see which is the worst case scenario. The worst case scenario then sets the operating limit for the CMTS.

Wall Street analogy

To give a feel to the differences between the three scenarios, there is an interesting analogy that may be made with Wall Street. The analogy would be: The average scenario is the equivalent of quarterly sales. The peak scenario is the equivalent of weekly sales. The max scenario is the equivalent of daily sales.

One may predict quarterly sales reasonably well. Weekly sales have large variations, whereas daily sales may be anything. Yet, a good factory must be able to respond well to daily and weekly fluctuations to be efficient.

And so it is with this model. The average scenario is somewhat predictable and measurable. The peak scenario is more difficult to predict, and the max scenario could be anything. Still, the network must have the headroom to be able to respond to a variety of packet arrival rates and peak rates, just as manufacturing must be able to respond to daily and weekly variations.

Getting results

The three scenarios, their bit rate per user, and the corresponding packet size produce a user data profile as shown on the left side of Figure 2 (page 55). This section explains the math behind this table and uses the upstream average scenario as an example. The abbreviation used in the formulas are:

us = upstream
ds = downstream
cx = c ⇒ channels (us, ds),
= x ⇒ scenario (avg, pk, max)

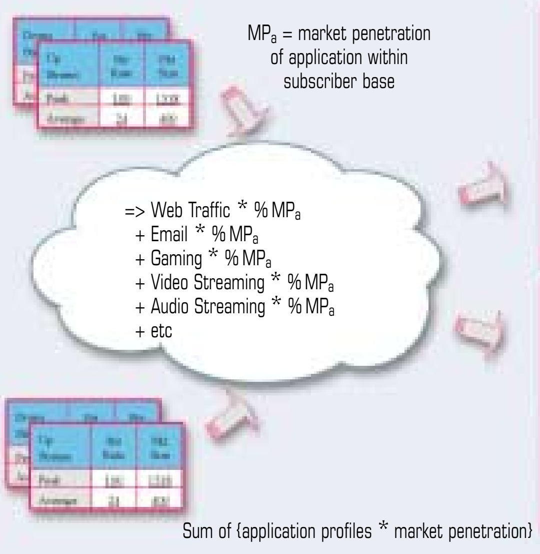
For each scenario, the user packet rate is found by dividing the bit rate by the packet size. The example uses a bit rate of 24 kbps and a packet size of 64 bytes.

$$\text{cxuser pps} = \frac{\text{cxuser kbps} * 1{,}000}{\text{cx pkt bytes} * 8}$$

$$\text{us avg user pps} = \frac{24 * 1{,}000}{64 * 8} = 47 \text{ pps}$$

We mentioned earlier that the bandwidth on the wire might be shared between the bursty data model and other models that describe signaling, VoIP and video traffic. The admission variable in the next equation allows the traffic engineer to specify the maximum amount of that bandwidth to be used by this model.

FIGURE 3 COMBINING MULTIPLE APPLICATIONS INTO ONE DATA PROFILE



Down Stream	Data Profile Bit Rate kbps	Pkt Size bytes	Pkt Rate pps
Max	2,000	1518	165
Peak	256	1518	21
Average	80	400	25

Up Stream	Data Profile Bit Rate kbps	Pkt Size bytes	Pkt Rate pps
Max	384	1518	32
Peak	100	1518	8
Average	24	64	47

BOTTOMLINE

> **Modeling Data Traffic**
>
> Networks that carry data, voice and video must be engineered if they are to operate properly. Multimedia traffic engineering (MMTE) is a series of techniques and models for engineering data, voice and video over an Internet protocol (IP) network. The bursty data model is a particular technique for engineering a network for data. With it, you may calculate how many modems a cable modem termination system (CMTS) can support.
>
> The real value of modeling is the promise of predicting the future. A good model allows current measured results to be combined with theory and growth projections to predict future requirements.

The DOCSIS protocol has both per-channel and per-packet overhead. The per-channel overhead is accounted for by choosing the appropriate value for the payload bandwidth. The per-packet overhead is accounted for by the following functions:

$$F(\text{ds pkt bytes}) = \text{pkt bytes} + 11$$
$$F(\text{us pkt bytes}) = (\text{pkt bytes} + 22) * 1.1 + 11$$

The total packet rate of the channel is calculated by dividing the bits in the channel by the bits in the packet.

$$\text{cx pps} = \frac{\text{c mbps} * \text{adm level} * 10^6}{F(\text{cx pkt bytes}) * 8}$$

$$\text{us avg pps} = \frac{2.56 * 80\% * 10^6}{((64 + 22) * 1.1 + 11) * 8} = 2{,}424$$

Dividing the PPS per channel by the PPS per user will provide the number of data sessions the channel will support.

$$\text{cx sessions} = \frac{\text{cx pps}}{\text{cxuser pps}}$$

$$\text{us avg sessions} = \frac{2{,}424}{47} = 52$$

The session density indicates how many users will be active *during* the measurement interval. This is an important input to the model and is highly subjective. The input is best thought of as relative to the other scenarios, but the calculations use the absolute value.

$$\text{cx users} = \frac{\text{cx sessions}}{\text{cx session density}}$$

$$\text{us avg users} = \frac{52}{25\%} = 207$$

Each of the three scenarios now vote to see who is the worst case.

$$\text{users} = \text{MIN (avg, peak, max)}$$
$$\text{us users} = \text{MIN } (207, 365, 475) = 207$$

Then the upstream and downstream channels vote for the worst case.

$$\text{us subs} = \frac{\text{MIN}\left(\frac{\text{ds user max}}{\text{ration us ds}}, \text{us users max}\right)}{\text{users per hhp}}$$

$$\text{ds subs} = \text{us subs} * \text{ratio us ds}$$

$$\text{us subs} = \frac{\text{MIN}\left(\frac{1{,}150}{6}, 207\right)}{1} = 192$$

The surprise in this example is that this system is limited by its downstream performance, not its upstream performance. You then divide this number by the market penetration of data services to specify the maximum households passed per upstream or downstream.

$$\text{c hhp} = \frac{\text{c subs}}{\text{mp data}}$$

$$\text{us hhp} = \frac{192}{10\%} = 1{,}920 \text{ hhp per us}$$

The result of the user profile, the session density and the market penetration is the maximum number of households passed per upstream and households passed per downstream.

Using measured results to generate the user profile

The user profile that contains the bit rate and packet size for each scenario may either be specified or measured. You may attain the following information by monitoring IP packets on the network during peak and average measurement intervals for both upstream and downstream:

- The mixture of applications by examining the transmission control protocol/user datagram protocol (TCP/UDP) port number;
- The number of users by looking for unique IP destination addresses (DA) in the downstream and unique IP source addresses (SA) in the upstream; and
- The packet size per application.

You may take these measurements over the two measurement intervals of one second and five minutes. From these measurements, you may calculate the following for each scenario:

- Bandwidth per user
- Nominal packet size per user
- Nominal PPS per user

These measurements provide the basis for the bursty data model.

Accommodating multiple applications

The bursty data model user profile has allowed us to describe a user and calculate how many users may fit onto a CMTS. We have also discussed how to generate this profile both intuitively and from field measurements. To meet the requirement of being useful, the bursty data model must relate somehow to the applications, such as e-mail and Web traffic, that the subscriber will use. The model must be applicable to both existing applications and new applications that may arise.

You've found us! Boonton Electronics 4400A and 4500A Peak Power Meters and Analyzers assure your radar stays on target.

Boonton Electronics' 4400A and 4500A Peak Power Meters deliver the widest peak power bandwidth, fastest rise time, and most advanced triggering capabilities on the market today. They also provide automatic waveform capture, automatic Pulse parameter measurements, and Statistical measurements including PDF, CDF, and 1-CDF. The 4400A and 4500A bring superior measurement capabilities to all radar and avionics applications.

When you're looking for someone who can meet the most demanding testing needs, set your sights on Boonton Electronics.

Boonton Electronics
A Wireless Telecom Group Company
PO Box 465
Parsippany, NJ 07054-0465

Phone: (973) 386-9696
Fax: (973) 386-9191
Email: sales@boonton.com
Web: www.boonton.com

BOONTON

Here is the concept. If the user data profile can reasonably describe a user, then it ought to be able to describe a single application as well. Each application would use the same model but with different parameters. You may then easily combine the models with a weighted average to generate the user data profile.

For example, if e-mail is used 10 percent of the time and Web traffic is used 90 percent of the time, then the e-mail profile is multiplied by 10 percent, the Web traffic profile is multiplied by 90 percent and the results are added. Because users run more than one application, the total weighting may be more than 100 percent. Some number between 100 percent and 200 percent is a good choice. This concept is illustrated in Figure 3 (page 56). You may further extend this approach to accommodate tiers of service by building separate user profiles for bronze, silver and gold, and then combining them with a weighted average that is equal to the market penetration of each service.

Conclusion

Networks that carry data, voice and video must be engineered if they are to operate properly. Multimedia traffic engineering is a series of techniques and models for engineering data, voice and video over an IP network. The bursty data model is a particular technique for engineering a network for data.

If you know the user data profile and session densities, then you may determine the number of users a network will support. You also may use the model in reverse. If you know the number of users and the user data profile, you may calculate the session densities. Likewise, if you know the number of users and session density, you may calculate the user data profile.

The MMTE bursty data model provides a simple and useful method for establishing a profile for a data user. This methodology allows traffic engineering to be based on service tiers, applications and real usage. Real-time and historic monitoring may be used to see trends and predict future requirements. The user data profile is intuitive, and may be calculated and measured. This profile may then be divided into the CMTS bandwidth to provide the number of cable modems per CMTS. CT

John T. Chapman is a distinguished engineer with Cisco Systems. He may be reached at jchapman@cisco.com. A spreadsheet with these formulas can be found at www.johntchapman.com/mmte

Did this article help you? Send your comments to snayalkar@pbimedia.com