Impact of Web-Based Applications on Enterprise Traffic

By John Bartlett and Peter Sevcik
October 2002

NetForecast Report 5063

When the economy and business are slow, people think that there is no need to change infrastructure. It is for these times that the saying, "If its not broken don't change it" applies. Nothing could be farther from the truth when it comes to networks. There are three inescapable reasons why networks are always in need of an upgrade. First there is a natural growth to traffic that keeps increasing even in bad times. Second, there are always new applications being introduced, whose traffic is predictable. Finally, there is the need to be ahead of demand in order to cope with the unpredictable.

Enterprises must start planning for higher bandwidth today. The data in this report shows that just as 10 Mbps LAN technology needed to be upgraded to 100 Mbps, the cycle of the next upgrade is upon us. The next demanding application – Web – is predictable; it is in the development plans of many enterprises.

This report looks at the historic growth of bandwidth demand in the enterprise data center. It works from the bottom up to determine how much bandwidth is required to support the enterprise. The resulting model shows a significant growth in bandwidth requirements if new Web-based applications are introduced, or if current applications are converted to a Web-based approach.

We therefore, recommend that enterprises seriously consider gigabit scaled data center infrastructure to deliver Web-enabled enterprise applications.

Applications Are Demanding More Bandwidth

Our appetite for information increases every year. According to Hal Varian, Dean of the School of Information Management at UC Berkeley, "The recent economic slowdown is a tiny blip relative to the overall information growth trend. Information continues to double every year …" [1].

The NetForecast models confirm that demand for Internet bandwidth is doubling every year, and continues to do so even in the face of the current business slow-down. The demand for bandwidth within the enterprise is rising at a commensurate pace. Companies have learned that their information is a key strategic tool in managing their business, and information systems are being continually updated to better provide the right data to the right user at the right time.

A review of application history reveals that each application class demands significantly more bandwidth than its predecessors. In the late '70s and early '80s, Telnet was the primary network application. Later in the 1980s, as computers became more reliably connected, email and ftp transfers came into play, and consumed more bandwidth than simple terminal traffic. In the early '90s the client-server model became the dominant method of connecting workstations to computing and data resources, and contributed to the bandwidth demand.

In the late '90s came the introduction of the World Wide Web, and the implementation of the Web browser. This graphically oriented view of data quickly became popular for the Internet. Today we find much of our on-line interaction taking place through the browser. The browser has become the industry standard client to which most applications are being designed.

NetForecast maintains a library of application profiles that characterize most popular applications. These profiles are derived from traffic captured on the network while an experienced user drives the application. We find that these profiles do an excellent job of

portraying the network signature of each application, and can be used to predict performance along with traffic load in varying network environments.

Figure 1 shows the bandwidth consumption for the four primary application types referred to above in the history of network usage, with the year in which those applications were first dominant. The blue bars represent average demand during an active session, while the red bars represent peak demand. Note that this graph uses a log scale on the vertical axis. An increase from one tick mark to the next represents an order of magnitude growth in network traffic demand.

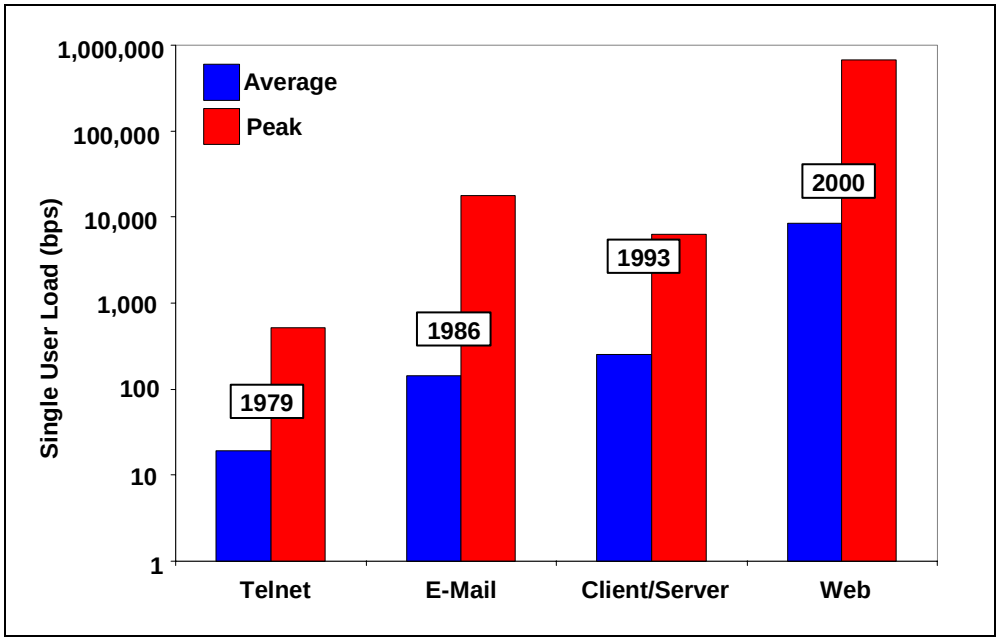

Figure 1 - Applications Demanding More Bandwidth per Active User

The increase in average traffic from the primary new application is doubling every 29 months, or just less than 2½ years. It is clear that the functionality of Telnet, email, client-server applications and now web browsers, produce much more information at each step of the user interaction. Furthermore it is not hard to imagine that we would like to have the Web interface bring us more information both in terms of the speed with which it is presented, and its complexity, such as movie clips or interactive video images. Thus we can expect this growth in application demand to continue at this pace.

Basic Average vs. Peak Demand of Each Application

Figure 1 shows that both average and peak demand are growing at similar rates. Average demand in this chart indicates the average during an active session of the user. Peak demand represents the demand incurred by a user as he/she hits the enter button or clicks on the next Web link. This is the instantaneous demand (averaged over 1 second) of each application. Often network utilizations are averaged over fairly long periods of time, which tend to hide the peak demand. To maintain the best employee productivity, servers and networks should be able to handle this peak demand. Productivity drops quickly when the expected response time of an application is not met [2].

The busy period of the day is when the highest number of users are on the network, trying to get work done. On the Internet we can slow these folks down a bit and support the average load, hoping that they will come back later if the service is too slow. In the enterprise environment, however, the goal is to maximize the productivity of employees

so their work can be accomplished. Enterprises should not expect users to abandon the service with the hope that they will return later.

Summing up peak demands for the number of active users would give an inaccurate view of the bandwidth requirement for the data center, since everyone does not click on a Web link at the same instant. To determine the relationship between peak and average bandwidth consumption as user counts increase, we ran these applications through a NetForecast traffic model. The results are seen in Figure 2.

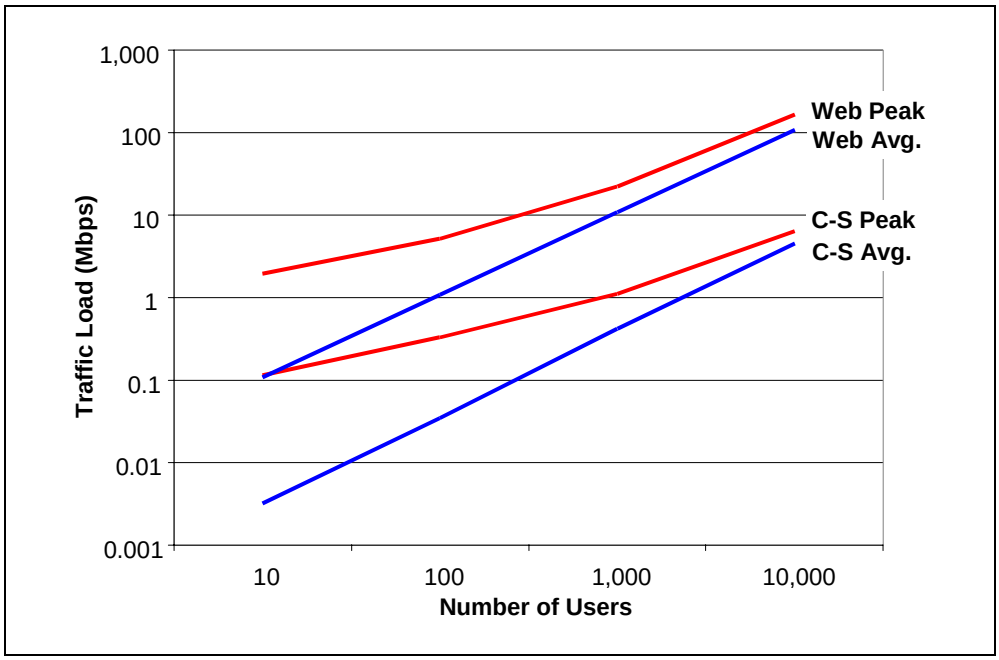

Figure 2 - Average and Peak Bandwidth Utilization

The lower two curves in Figure 2 are the demand curves for client-server traffic, while the upper two curves represent the demand from Web traffic. Note again that the scale is logarithmic.

Table 1 - Web Peak to Average Ratio

# of Users	*Peak to Avg.*
10	18.0
100	4.8
1,000	2.0
10,000	1.5

The curves indicate that the peak traffic demand approaches the average traffic demand above 10,000 active users. Even at the 10,000-user level, the peak demand is about 50 percent higher than the average demand. Table 1 shows the peak-to-average ratio for each of the four active user populations.

The higher curve for each application in Figure 2 should be the design point for best application response when sizing the enterprise network or the data center equipment.

Growth Within an Application

The next phenomenon we notice is that there is a need for more bandwidth over time within each application class. The easiest example to understand is email. In 1985 people were sending text messages back and forth, which required relatively few bytes on the network. Today users are attaching large files and using fancy headers and trailers in most communications. A model of this email bandwidth growth shows that email is doubling its bandwidth demand about every 3 years.

Client-server applications are likewise growing in their bandwidth usage. Early client-server applications, modeled after IBM 3270-type interactions, were modest consumers of bandwidth. More recent applications fetch data as the mouse rolls over parts of the screen, preload drop-down boxes, and create a more interesting screen with graphics, generating more demand on the network in the process. Client-server applications appear to be doubling their bandwidth demand every 100 months, or 8 years.

Web applications are showing the most dynamic bandwidth growth. The average Web page size has grown from about 50,000 bytes of payload in 1995 to 115,000 bytes of payload in 2001. This growth curve represents a doubling of Web bandwidth demand every 30 months, or 2½ years. This value only represents Web page loading; it does not count streaming media such as audio and video clips, which contribute significant additional payload to the Web applications.

Total Bandwidth Requirements

So how much bandwidth is really needed? And how will it change over time? To make this forecast, we built a model that estimates the usage of three major application types over time. Telnet drops out because it has such a small impact.

The history and forecast shown in Figure 3 is based on a two-step process. It begins with a table showing the growth of bandwidth by each application over the years. Then an estimate of the application mix is calculated for each two-year period depicted.

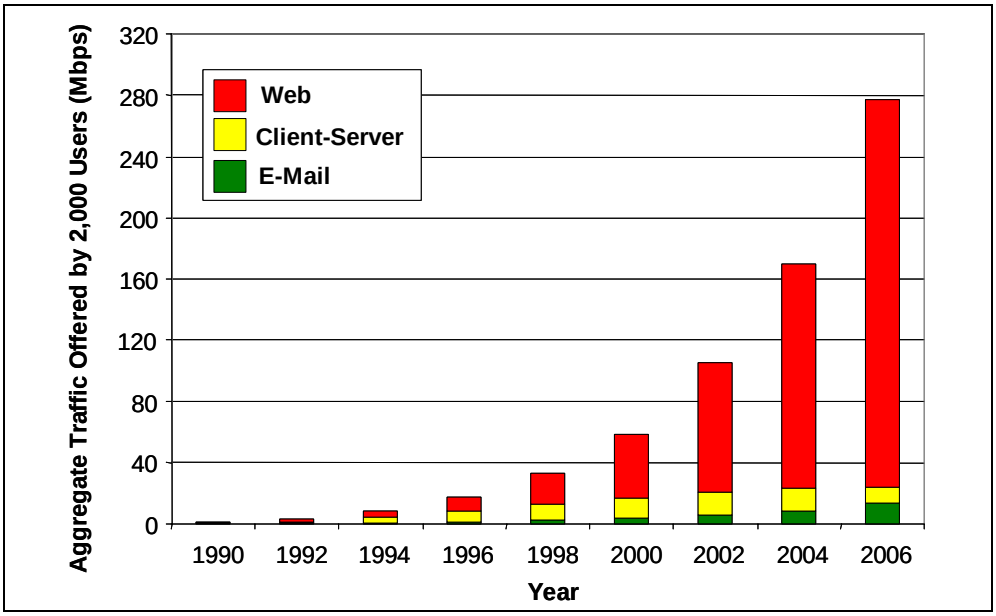

Figure 3 – Average Traffic Generated by 2,000 Active Users

The percentage of users using client-server applications peaked in the late 1990s, and is now dropping as these applications are replaced by Web-based applications. Email continues to be used by nearly everyone in the organization. Web traffic continues to grow in usage because more and more applications are being implemented through an HTML interface, making the Web browser the default interface client.

Equipment Specification and Traffic Engineering

Web growth is clearly much larger than the other types of traffic on enterprise networks, and it is growing at a faster rate. Note that Figure 3 shows the traffic growth for a fixed

number of active users. If an enterprise is experiencing user growth as well, traffic growth is commensurately higher.

So how much bandwidth is really needed in a data center? Data burstiness causes peak loading, which shows up in network equipment as increased queuing. NetForecast modeling work has shown that for Web traffic, queues start to build up once the network reaches 30 percent utilization. Increasing queue depth means that the packets are not being serviced as quickly, and the application performance is starting to degrade.

Figure 4 is calculated from Figure 3, to show the required bandwidth for 2,000 users over time. Figure 4 assumes the enterprise wants to maintain optimal performance for the applications, so average utilization is not allowed to exceed 30 percent of the line or equipment capacity.

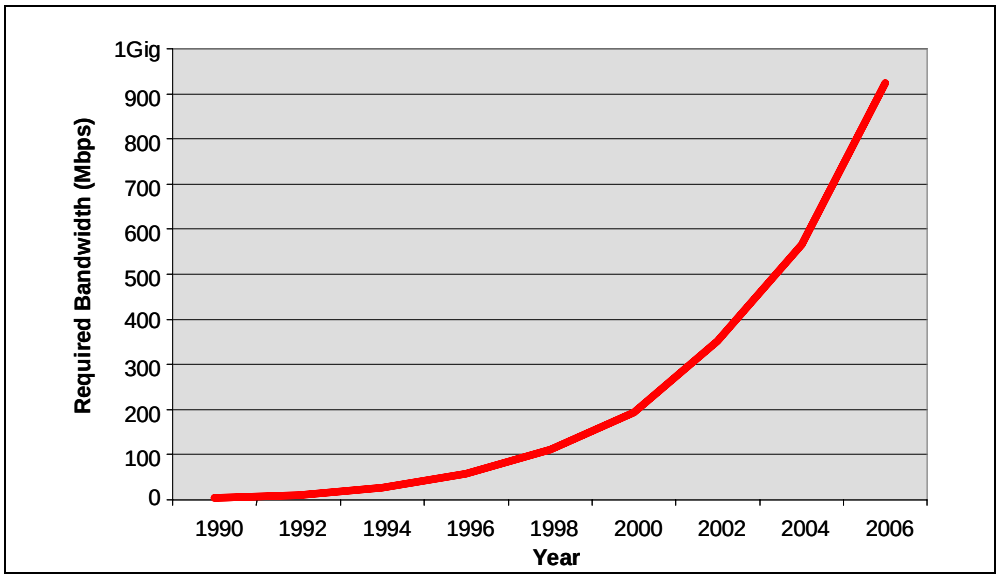

Figure 4 - Required Bandwidth for 2,000 Active Users

Many of today's enterprise data centers are built around 100 Mbps Ethernet technologies, using equipment (routers, load balancers, switches and servers) that will support those data rates. Figure 4 shows that today 2,000 active users require 360 Mbps of system bandwidth in order to provide optimal performance within proper safety margins. Systems that support 2,000 active users with this application mix are operating below the margin of safety. The number of Web based applications, and their bandwidth requirements, will continue to drive this growth higher in the future.

Correlating Active Users to The User Population

It is important to size networks and equipment to support the needs of the active users as described above. The number of active users varies as a percentage of the total user population as shown in the following two scenarios.

The typical work environments where most people are engaged in activities that are not specifically computer oriented, people spend much of their time *not* typing on a keyboard (e.g., hospitals, construction, transportation, teaching). Many of the employees of a campus in such a profession are busy performing their primary role, which often involves dealing with people in person. Additionally, some are on vacation, sick leave, traveling, etc. Even when in the office and not attending to "clients or patients," they are often in meetings. The net result is that even though this typical office environment is very

computer literate, the staff is not using a computer all the time. A good rule of thumb is that about 10 percent of the staff are active users in the busy period of the network day.

On the other side of the spectrum are sites that are completely engaged in computer activity. A good example is a telephone call center. Even in this environment, not all the users are typing all the time. The activity is governed by shift schedules and timed breaks. Some of the employees are also off-site or absent. Furthermore, many of the employees are often in training sessions. We have seen that even in this highly computer driven environment, about 70 percent of the user population is actually active in the network busy period.

This implies that the curve in Figure 4 must be scaled to the appropriate office scenario in order to convert active users to the total number of employees. Using the examples above the 2,000 active users translates to a call center of about 2,800 employees and an office campus of about 20,000 employees.

Summary

The availability of networking and the advancement of software applications have led to an increased demand on the network. Each generation of applications have increased their bandwidth utilization over time. The most recent move to Web-based applications has created a significant increase in the bandwidth demand per user. Combining these factors leads to the conclusion that bandwidth demand per user is increasing at an exponential rate, doubling every 2 years. The introduction of multimedia traffic exacerbates this problem. Streaming audio and video, and voice over IP will push this curve up significantly, as multimedia becomes a mainstream business tool.

Data center architects aiming to maintain user productivity need to accurately predict the bandwidth demand of their applications and user population in order to properly specify networking equipment capable of supporting the demand. Currently installed data center equipment is either already slowing down application performance, or will do so in the near future. In large organizations, the current crop of 100 Mbps equipment will need to be upgraded to gigabit-capable equipment, especially if new Web-based applications are introduced, or older applications are converted to a Web-based approach. Business use of multimedia accelerates the need for infrastructure upgrade. ¤

References

1. "Fixed Content" Marks Next Explosion of Information Growth, MassEcomm.org, http://www.massecomm.org/news/news.asp?NiID=364

2. Understanding How Users View Application Performance" by Peter Sevcik, July 2002 http://www.netforecast.com/PeterArticles.htm

John Bartlett is Vice President of NetForecast, and has 25 years of experience designing at the chip, board, system, and network levels, with a focus on performance. John led the team that built the first VLAN implementation, one of the first ATM switches, and he is a leading authority on the behavior of real-time traffic on the Internet. He can be reached at john@netforecast.com.

Peter Sevcik is President of NetForecast and a leading authority on network technology. He specializes in performance analysis and holds the patent on application response time prediction. Peter has contributed to the design of more than 100 networks including the Internet. He can be reached at peter@netforecast.com.