Application Demands Outrun Internet Improvements

Net Forecasts – Peter J. Sevcik
BCR Volume 36, Number 1
January 2006

You keep hearing how performance on the Internet is getting better all the time. There was a lot of investment in bandwidth and connectivity during the boom years, and we're still reaping the benefit of these investments. Furthermore, many more users are accessing the Internet via broadband. So the foundations of the Internet economy appear sound. There are many parties interested in seeing that this health continues to improve.

One such group is the researchers who use the Internet as a foundation for very high-performance computing. The International Committee on Future Accelerators (these are particle accelerators for physics research) has a working group dedicated to measuring and monitoring the performance of the Internet. This group, led by R. Les Cottrell of the Stanford Linear Accelerator Center (SLAC), published a very interesting report: "January 2005 Report of the ICFA-SCIC Monitoring Working Group."

The report shows measurements of impressive performance improvements in throughput, latency and loss over the last decade (1995 to 2005). In fact, the report makes the bold statement that, "Internet performance is improving each year with packet losses typically improving by 40-50 percent per year, and Round Trip Times (RTTs) by 10-20 percent and, for some regions such as southeast Europe, even more."

Some of these improvements cannot continue indefinitely. For example, at some point RTT will approach the limits imposed by the speed of light over the fixed paths of the fiber that interconnects cities around the globe. I do not question the data presented by Cottrell et. al., showing great improvements in capacity. I think, however, that there is another side to the performance story--the demands posed by applications upon that capacity.

Task-Oriented Measurement

NetForecast has been measuring application demand by profiling how an application operates on a network in order to accomplish a task (see BCR, March 2005, pp. 8-10). A task is defined as the elapsed time required for an application system to respond to a human user input such that the user can effectively proceed with the process they are trying to accomplish.

I have been profiling applications and analyzing the effects of networks upon their performance for many years--well before the invention of the World Wide Web. But in 1995, Intel asked me to provide a comprehensive study of their business-to-business uses of the Web. One of the by-products of that study was a composite profile of the typical business Web page. The typical page required 20 turns and 50,000 bytes of payload. (A turn is each non-payload bearing application client/server software interaction needed to generate a user response or task. This is often called application "chattiness.")

Since then I have had the opportunity to use another source regarding general business-to-business Web pages--the Keynote Business 40 Index sites. I profiled them in September 1999 (see BCR, pp. 10-13), November 2002 (see BCR, pp. 8-9), and most recently November 2004 (see BCR, pp. 8-12). The composite profile of these business sites is shown in Figure 1 as a plot of Turns versus Payload for each of the measurement periods. The line within the chart is the trend line for these two parameters.

Turns and Payload per page a clearly climbing. Turns are increasing at 11 percent per year while payload is growing at 19 percent per year. These two factors combine to show a very significant increase in what a network must do to deliver constant performance.

Demands Outstripping Improvements

Much of the payload growth can be mitigated by the growth of broadband access. But most of the turn growth is not being compensated by better Internet performance. The bottom line is that bigger application demands are outstripping Internet improvements.

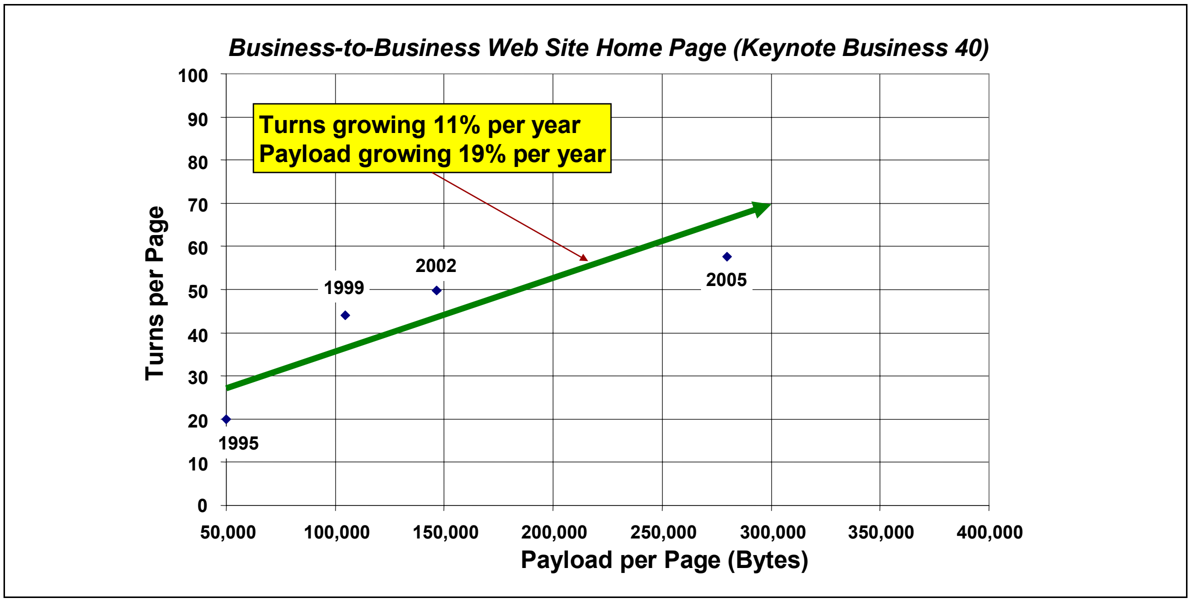

Figure 1 – Web Pages Keep Getting Bigger and More Complex

Today's typical website wants to accomplish 70 turns and move 300,000 Bytes. This 20-fold increase in demand over a decade is double the 10-fold improvement in overall Internet performance--including access lines.

But more importantly, application demands increase unabated while the ability of the Internet to improve is starting to slow down. We have just started to see the effects of the balance tipping towards degraded performance as seen by the user.

A good way to understand the effects of these two forces over time is to plot them using the Apdex methodology (see www.apdex.org and BCR, March 2005, pp. 8-10). Apdex lets us view the performance of different applications (the Web page circa 1995 versus 2005) over different delivery systems (the Internet circa 1995 versus 2005).

To get started, we must select a uniform application task target time. For this analysis, I have chosen 9 seconds, which means that the typical user who looks at about 3 things in the page and is in a low repetitiveness process (a.k.a. browsing) will be "satisfied" with response times under 9 seconds. It also implies that the same user will be "frustrated" with response times above 36 seconds.

I modeled the business-to-business Web page profiles for the four years in Figure 1 over the improving Internet as seen in three regions of the world. The model places the server in San Francisco, connected to the Internet with very high speed links (1 Gbps). It assumes that the user is connected with a consistent broadband connection of 1.5 Mbps. Back in 1995 this was a high-end T1/E1 line. Today it is the typical effective bandwidth of a DSL or cable connection.

I kept the server processing time constant at 1 second. That is, although the applications and databases behind them keep getting more complex, Moore's Law is keeping their performance constant. The effect of these assumptions is to constrain the changes to shifts of Internet performance.

So the user wants to see the page load in 9 seconds. The computers use up 1 second in processing, leaving 8 seconds for the Internet. In our model the Internet keeps improving; the question is, does it improve fast enough to keep up with the increasing turn and payload demands?

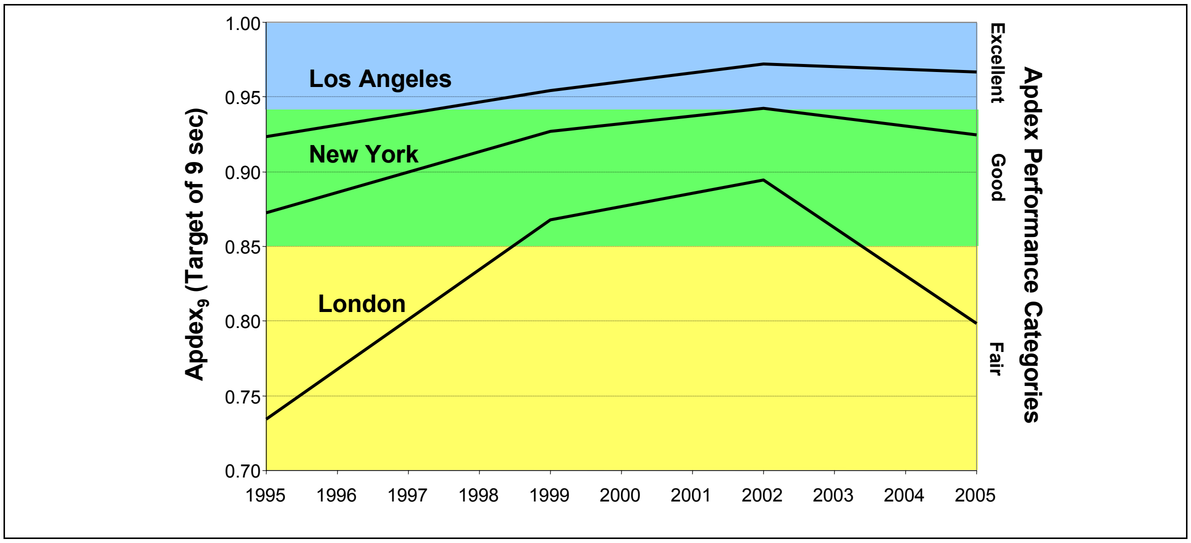

Figure 1 – Web Pages Keep Getting Bigger and More Complex

Figure 2 shows the results in Apdex terms for the users in three key cities accessing the then-typical business Web page--Los Angeles, New York and London. As expected, performance was not great in 1995. But it was good enough to let the Web catch on as a business tool.

It is also clear that the farther you are from the San Francisco server, the slower things get, so the Apdex score declines. The Apdex curves for all the cities improve from 1995 to 2002. The Internet was getting better, and was doing so faster than the applications were getting more demanding.

But by 2005, the tide has turned. Degraded performance is not yet visible in the United States but performance is no longer improving despite improvements in the Internet. However, the London users who were dependent upon much more Internet distance got out of Fair performance into Good only to see themselves slipping back into Fair performance. The more Internet you need, the less it can keep up with your needs.

Here's another interesting point: If you improve bandwidth on the user's access line to 10, 50 and 100 Mbps in 1999, 2002 and 2005 respectively, the results are exactly the same. Contrary to popular opinion, more broadband does not help transactional applications.

Conclusion

In this analysis, London is just the bellwether for the global user population. The 2005 performance drop is much more dramatic from the 2002 peak in most oversees locations. If you use the Web to sell fresh hot pizzas to customers in your neighborhood, you can ignore this column. But if you are trying to leverage the Internet on a global scale with the latest application techniques, your lever just broke!

There is no fundamental fix for the divergence of these two trends. Applications will keep demanding more. The Internet will not be able to use increased bandwidth alone to improve performance enough to keep up. This is not a disaster, but rather an opportunity for measurement vendors and application delivery system vendors.

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

NetForecast helps change delivery systems to improve the performance of networked applications. This includes advising enterprises on how to evaluate, improve and manage the performance of business applications, as well as advising vendors about customer requirements, technology issues, and adoption trends.

NetForecast

Smart Strategies From Hard Data

www.netforecast.com