The Myth of Latency

CyrusOne

Datamation®
Executive Brief

Latency delays are a fact of life on the Internet, but within the continental U.S. they are so small as to have no effect on application performance in most cases. (Note: Times shown are round-trip.)

Any company seeking a colocation partner will put geography near the top of the list of selection criteria. Geographical location can affect everything from the cost of electricity to the probability of a natural disaster to the availability of a skilled workforce to handle maintenance. But there's one area where its importance is debatable, and that's the issue of latency.

Ultimately, the laws of physics dictate that latency in any network will be impacted by the physical distance between the sender of a packet and its receiver. The map above shows typical Internet latencies between some major U.S. cities as measured by AT&T.

The map reveals, for example, that a packet traveling the 1,372 miles from Austin to Los Angeles will take 38 milliseconds, while the same packet traveling the 939 miles between Austin and Chicago will take 32 milliseconds. But the real question is not what the precise latency time is, but rather, how much does it matter? In most cases, the answer is, *not at all.*
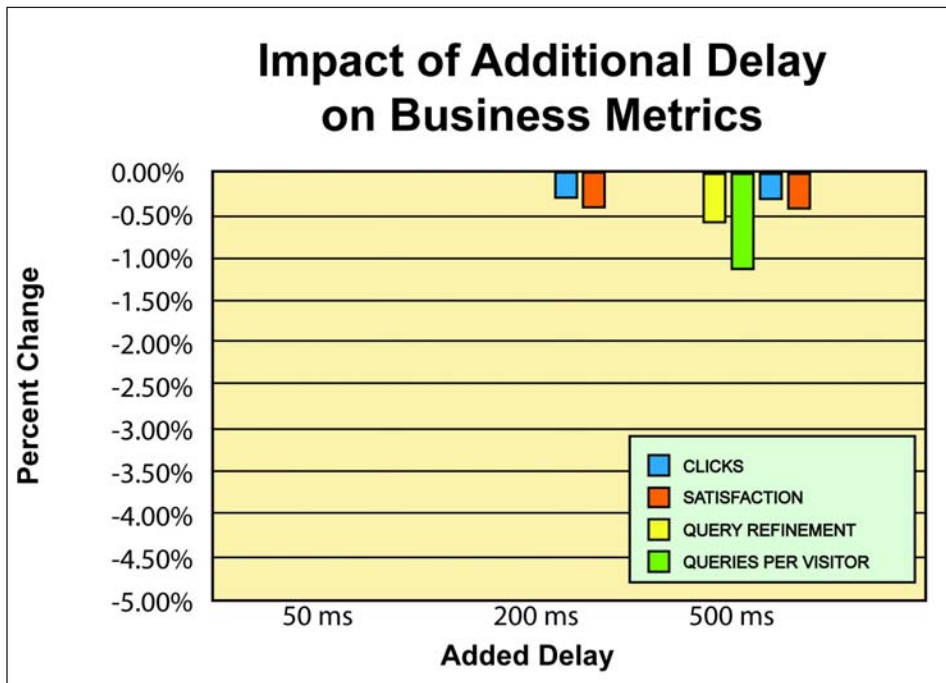
This paper will show that distance-based latency typically has a trivial effect on the user experience, and that in cases where latency might influence that experience, the cause lies not in physical distance, but in other technical factors.

How Fast Is Fast Enough?

According to one study,[1] if two individuals are having a conversation, a delay of 100 milliseconds isn't even noticeable. In fact, users are still engaged with delays of a full second. These are admittedly somewhat subjective measurements, but the same study also collected hard data that included the effect of delays on revenue, which is in no way subjective.

As the graph indicates below, no effects of delay are even measurable with 50 milliseconds of latency. At 200 milliseconds, satisfaction decreases – but only by one half of one percent. Revenue isn't affected until delays

[1] source: http://www.slideshare.net/bitcurrent/impact-of-web-latency-on-conversion-rates

Impact of Additional Delay on Business Metrics

Latency has no impact on business processes until 200 ms, while the average latency-based delay on U.S. networks is less than 100 ms

reach 500 milliseconds. From a pure business perspective, at least in the case of consumer-facing e- commerce sites, this indicates that latency delays typically encountered within the continental United States will have essentially no significant effect.

Obviously, not all companies interested in colocation are engaged in e-commerce, and different types of applications are more or less affected by latency. But in general, a number of evolutionary developments in application design and implementation have combined to eliminate latency concerns.

The rise of Web-enabled applications. After the turn of the century, as the Internet gained traction on its way to becoming the world's primary

communication network, companies began to tout "Web enabled" as a feature for applications. Today, it's unusual to find an application that is not designed to take advantage of the Web. Applications that are browser-based, as opposed to depending on an application-specific client, are unlikely to have a problem with latency.

The rise of Web-savvy design. Application design principles that maximize efficiency in respect to the Internet are now becoming well established. Most round trips for a website consist of HTTP requests and responses. If the number of these requests is reduced, and if they are executed in parallel wherever possible, the actual response time for the user can be improved to the point where latency is no longer an issue. In

addition, through the elimination of unnecessary decoration and focus on relevant data, the signal-to-noise ratio of many sites has been significantly improved, which translates into an ability to deliver the same functionality with lower bandwidth requirements.

Virtual desktops. The use of virtual desktop implementations such as Citrix can have a dramatic impact on bandwidth usage, because the traffic between users and the data center consists only of commands.
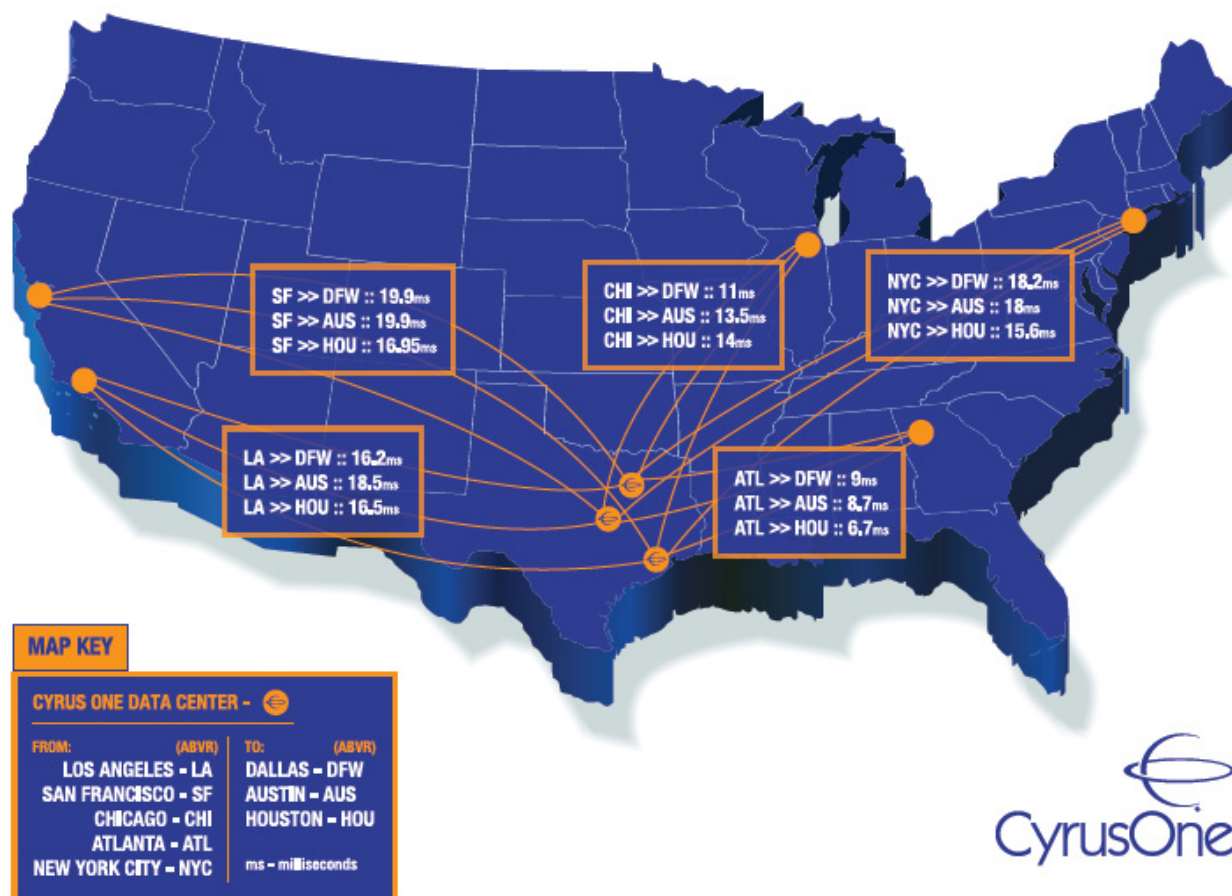
Network Developments
For companies that want to take advantage of the cost savings and workload reduction that colocation offers, there are three important factors related to the network itself that should be considered.

Improved network performance. A number of technologies have been implemented over the past decade that have made the Internet "faster," in the sense that latency has decreased.

Cost of Bandwidth. While performance in general has sped up, the cost of bandwidth has decreased. This is a positive factor in favor of colocation when it comes to the total cost/benefit analysis.

Carrier options. The backbone routers used by all the major carriers are faster than the routers located at the edge. If a packet travels all the way from origin to destination on one carrier, it will move at maximum speed.

CYRUS ONE ESTIMATES FOR:
NETWORK LATENCY BETWEEN CITIES

SF >> DFW :: 19.9ms
SF >> AUS :: 19.9ms
SF >> HOU :: 16.95ms

CHI >> DFW :: 11ms
CHI >> AUS :: 13.5ms
CHI >> HOU :: 14ms

NYC >> DFW :: 18.2ms
NYC >> AUS :: 18ms
NYC >> HOU :: 15.6ms

LA >> DFW :: 16.2ms
LA >> AUS :: 18.5ms
LA >> HOU :: 16.5ms

ATL >> DFW :: 9ms
ATL >> AUS :: 8.7ms
ATL >> HOU :: 6.7ms

MAP KEY

CYRUS ONE DATA CENTER -

FROM: (ABVR)	TO: (ABVR)
LOS ANGELES - LA	DALLAS - DFW
SAN FRANCISCO - SF	AUSTIN - AUS
CHICAGO - CHI	HOUSTON - HOU
ATLANTA - ATL	
NEW YORK CITY - NYC	ms - milliseconds

CyrusOne®

Cyrus One already offers many low-latency connections between major U.S. cities, and more are soon to be available. (Note: All times shown are one-way.)

If it must transfer from carrier A to carrier B somewhere between origin and destination, it will have to pass through two edge routers, and this will increase its latency. The practical impact of this reality is companies may be able to improve latency numbers by moving to one single carrier.

Summary: Latency Is Not an Issue

For companies seeking a colocation partner, the bottom line is that the latency associated with geographical location is in most cases a trivial factor.

- Hard data is available that indicates that the latency numbers between U.S. cities fall below the annoyance threshold for end users. Noticeable effects don't appear until latency is > 200 ms, and typical latency numbers are < 100 ms.

- The rise of Web-enabled applications and Web-savvy application design have dramatically reduced latency issues.

- Companies have options which can provide higher bandwidth and better performance at a lower cost.

Colocation's Promising Future

As the map indicates above, many low latency opportunities already exist, and more are on the way. CyrusOne is specifically focusing on minimizing latency, with the strategic goal of offering services with latencies of less than 20 ms for companies located in every major metropolitan area in the United States. This in turn will make colocation is more attractive than ever as a cost-cutting strategy for IT. ◼