WHITE PAPER

Low Latency Switching: The Race to Zero

TEST AUTOMATION

Overview

In today's highly competitive world, firms from every industry, both public and private, continually seek to leverage ever higher operational benchmarks in order to innovate and stay ahead of the pack. Today's businesses must operate at a lightning-fast pace when bringing new products to market, processing information, executing transactions, and servicing their customers.

This white paper focuses on technology that can be leveraged to reduce network latency, and lower the time to complete transactions. It highlights physical layer switching (PLS) as one such technology that can be employed to reduce latency.

Why Is Low Latency Important?

If there's one thing history has taught us over the years, it is that speed is king. One of the earliest examples of the quest for speed was illustrated when the founder of Reuters Press Agency cornered the financial information market by using a fleet of pigeons to quickly and accurately deliver news and stock prices to and from telegraph terminals – minutes mattered. Seconds became the focus by the 1980's as companies chose communications network service providers based on their SLA's (service level agreements). In the last decade, financial firms have learned that trade transactions were won or lost by milliseconds (ms) or microseconds (μs). Today, nanoseconds (ns) can make the difference between winning and losing business. This trend to reduce latency, and make it more consistent for all priority traffic, shows how firms must leverage the fastest communications technologies available to achieve success.

There are many examples of businesses that leverage high-speed communications technology as an integral part of their success. Product manufacturers leverage their test lab environments to bring products to market faster, complete regression testing, and support their customers by instantly recreating their dynamic topologies. High frequency quantitative trading firms, financial information exchanges, military battlefield information systems, CyberSecurity firms, media and news outlets, research networks, search engines, and Cloud Networking companies claim the network *is* their business. These businesses all have a common denominator: *speed.*

What Is Network Latency And How Is It Measured?

In the network context, latency is defined as delay - an expression of how much time it takes for a unit of data to get from one designated point to another. Latency is measured accurately from point A to point B with precision test equipment, such as a time synchronized traffic generator or analyzer, as shown in Figure 1. Alternatively, some companies such as service providers may measure and define latency in terms of total round-trip time (RTT).

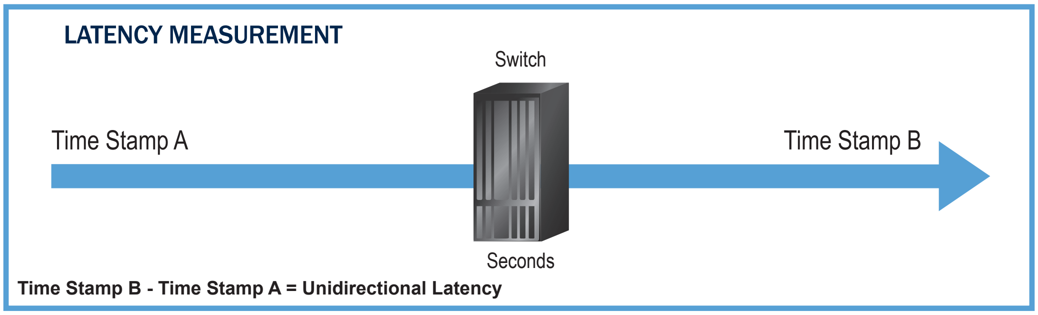

Figure 1. Latency measurement.

The main contributors to network latency include:

Transmission path delay: The medium itself introduces delay as data travels through it.

- High-performance network propagation speeds are approximately 5 nanoseconds (5 ns) per meter, or 5 microseconds (5 μs) per kilometer,so the shortest distance between nodes is critical.
- Although optical fiber and copper cables have similar propagation speeds, optical fibers can carry much higher data rates over much longer distances without the need for regeneration and further delay.

Processing delay: Intermediate devices such as switches, routers, firewalls, gateways, load balancers, etc. need additional time to inspect and/or switch traffic between networks. Figure 2 shows the latency by three common device categories.

- Physical Layer (Layer 1) switches with an all optical switching fabric offer pure fiber to fiber switching. These have the lowest latency in the industry at 5 ns.
- Layer 1 switches with an electrical switch fabric also add multicast and TAP (test access ports) functionality across fiber to fiber, fiber to copper, and copper to copper connections. These range between 50 and 100nsec.
- Layer 2/3 switches add delay on the order of 1 to 50 microseconds (µs) depending on the amount of packet processing, with an average of 30 µs.

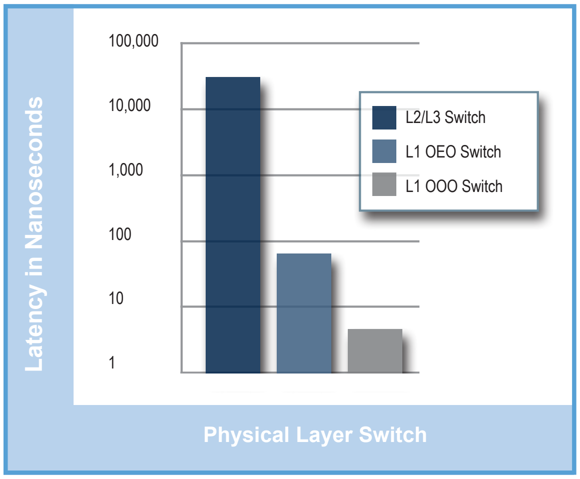

Figure 2. Average latency comparison. *Source, Tolly Report 2010*

Architecting data applications to run at wire speed is critical for many of today's business needs. While bandwidth (the bit capacity through a given media) does not represent the speed at which data travels from one point to another, it does affect latency when the utilization of that bandwidth increases. Many of today's L2/L3 devices have the benefit of sharing network bandwidth and the disadvantage of being oversubscribed. Oversubscription represents having less data capacity than the sum of the I/O (input/output) interfaces. Network engineers should be cognizant of the capacity of networking devices in their design and try to avoid oversubscription of such devices.

How Can Physical Layer Switching Help?

There are two options traditionally accepted in the industry for network infrastructure interconnection or switching:

Manual Patch Panels - are cheap and easy to start small, but offer no automation or intelligence.

Layer 2/3 Switches - are accepted for general networking, however they can add significant latency and cost.

When you need more than a patch panel but less than a Layer 2/3 switch, there is another option often overlooked – the Physical Layer Switch. Unlike a L2/L3 switch, which is burdened with packet processing decisions, a Physical Layer Switch (PLS) is an OSI Layer 1 switch in which all interfaces can transmit and receive simultaneously at the maximum line rate without dropping traffic or processing packets.

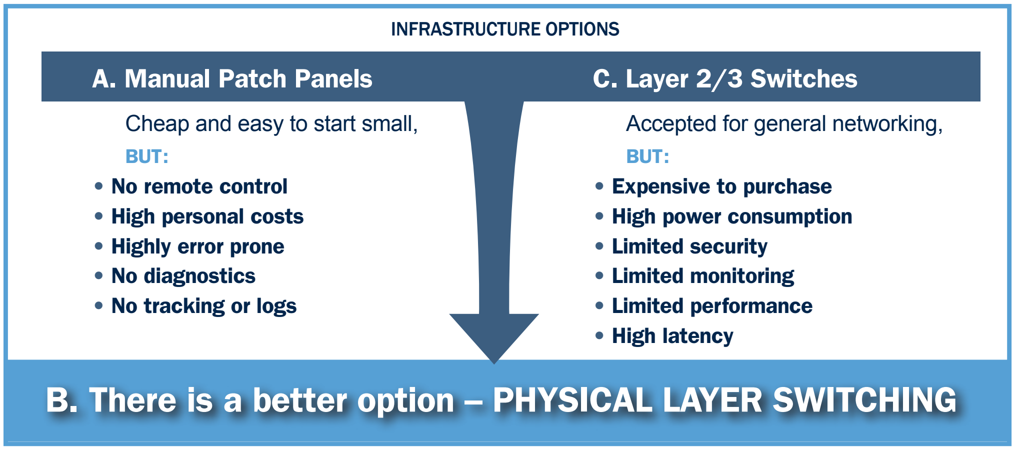

Figure 3. Traditional infrastructure options.

LATENCY BENEFITS FROM LAYER 1 SWITCHING

- **Up to 1,000x lower latency than layer 2 switches or layer 3 routers**
- **Always 100% line rate performance and no oversubscription**
- **High consistency of latency across the switch or network paths**
- **High consistency of latency across speeds (ie GE and 10GE)**
- **High consistency of latency across protocols (ie GE, Fibre Channel, SONET)**
- **Easy TAP connection across networks and measurement tools**

Figure 4. Primary latency benefits from layer 1 switching

PLS allows moves, adds, and changes to be executed precisely and in an automated fashion, with very high consistency across the network. In addition, physical layer switching is also protocol independent, so the latency is consistent across data rates and protocols. The consistent low latency is generally kept in a very small range of +/- 1 ns for all-optical switches and +/- 5 ns for electrical physical layer switches.

Testing and monitoring of the network traffic becomes seamless with physical layer switching since all ports can be multicast or tapped. Utilization of all equipment, within the test lab, data center, and cloud environments increases as well, which in turn saves money and increases productivity.

When network architects or test lab managers look for obvious places to begin seeking lower latency and higher throughput advantages, the access layer becomes an area of focus in identifying these measured latency improvement gains. For every added network layer beyond the access layer, it is important to consider just how much latency or data transmission lag-time is added, in contrast to a flatter network design using PLS technology.

Using a Physical Layer Switch to Reduce Latency by up to 1,000x

As discussed before, L2/L3 switches and routers process packets, and hence contribute to increased network latency. In addition, small scale switches require multiple network hops which further amplifies latency. On the other hand, PLS technology does not involve packet processing and can scale to thousands of high-speed ports. As companies migrate to higher speed host interfaces, replacing some or all of the access-layer L2/L3 switching with a Physical Layer Switch could decrease latency by as much as one thousand times.

Conclusion

Businesses that care about speed need to be designing high-performance networks with Physical Layer Switching. Layer 1 switching offers an innovative solution for reducing latency and cost, while increasing performance. This unique technology offers more than a patch panel and less than a layer 2 switch to provide the optimal performance for latency sensitive applications.

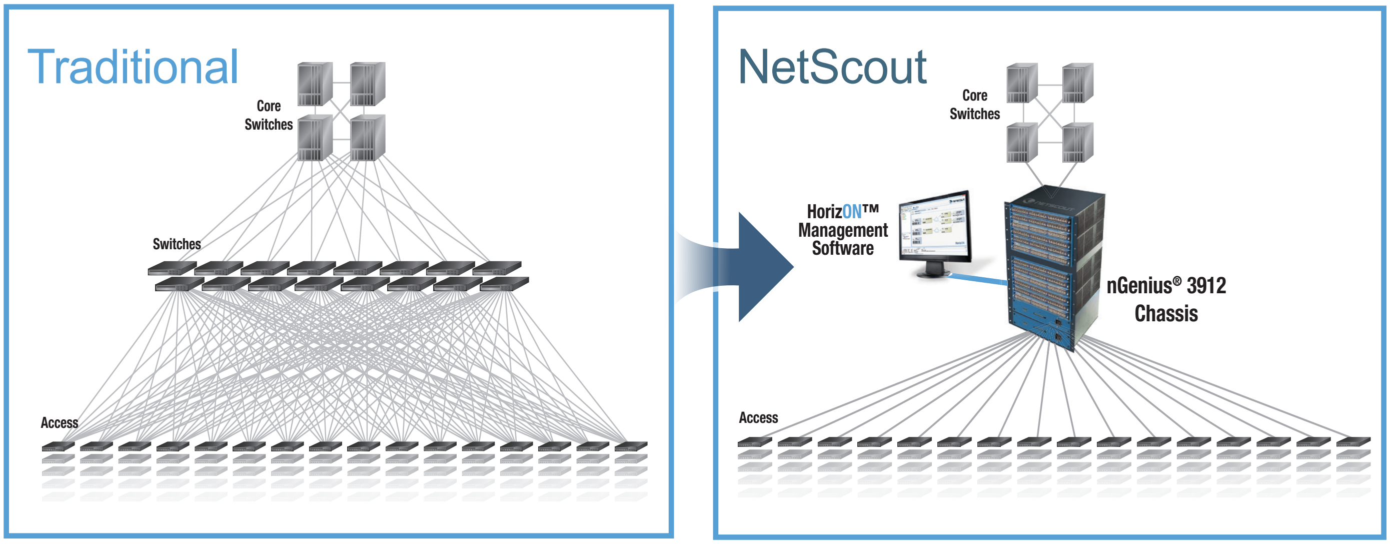

Figure 5. Using a physical layer switch to reduce latency.

Americas East
310 Littleton Road
Westford, MA 01886-4105
Phone: 978-614-4000
Toll Free: 800-357-7666

Americas West
178 E. Tasman Drive
San Jose, CA 95134
Phone: 408-571-5000

Asia Pacific
17F/B
No. 167 Tun Hwa N. Road
Taipei 105, Taiwan
Phone: +886 2 2717 1999

Europe
One Canada Square
29th floor, Canary Wharf
London E14 5DY, United Kingdom
Phone: +44 207 712 1672

NetScout offers sales, support, and services in over 32 countries.

For more information, please visit www.netscout.com or contact NetScout at + 1 877-468-5784 or + 1 609-518-4100

EWP_000-13