ARISTA
at-a-glance

Solace Systems®

Data Center Class

Ethernet: the Best Choice for Low Latency

INSIDE

SUBJECT
can Ethernet compete with Infiniband in the low-latency trading and high performance computing markets?

WHY
administrators and IT professionals face a choice when deciding whether to invest in Infiniband or Ethernet for their low-latency networks. This paper addresses many of the characteristics of Infiniband that have made their way into Ethernet

WHO CARES
High Frequency Traders, operators of High Performance Computing Systems, and anyone who wants to build a stable and scalable low-latency network

Why not Infiniband? This question is often asked when a discussion about low-latency network architectures is being had. Infiniband is fast, there is no arguing that, and no getting around it. Infiniband has a very low latency.

However, it is also difficult to bet against Ethernet. Ethernet has received several orders of magnitude more investment than Infiniband and has thousands of engineers at many of the largest and best companies in the networking industry developing Ethernet and developing on Ethernet. Ethernet has also been a very 'fluid' standard, one that has evolved from Half-Duplex, 10Mb, Carrier Sense Multiple Access Collision Detection busses, to switched, full-duplex, 10Gb structured topologies. From networks where spanning-tree blocked every possible path but one, to networks where multi-chassis link aggregation makes all available paths in a structured topology useable for data transmission.

Ethernet also enjoys tremendous supplier and silicon diversity with many companies actively investing in switching, systems, and NIC technologies. Infiniband has very few primary silicon suppliers, and those are also now hedging with Ethernet offers.

This is all good for someone purchasing a low-latency cluster in the future, but what about someone who needs to build a low-latency system today - whether for a high Frequency Electronic Trading system, or for High Performance Computing applications, what should they do?

Let's start by decomposing what makes Infiniband fast. In understanding the core reasons for low-latency forwarding we can then analyze recent, shipping, stable Ethernet enhancements.

Host Latency

Infiniband Host Channel Adapters (HCAs) take data from the wire and directly put it into the operating systems user-space without context switching. IB does this with very low latency.

Today, with Ethernet Network Interface Cards with Kernel Bypass and TCP acceleration in hardware Ethernet can deliver the same capability of bypassing the O/S kernel and delivering a low latency host interface.

Couple this with developments such as Remote Direct Memory Access over Convergence Enhanced Ethernet (RDMAoCEE) and Ethernet will have equivalent capabilities.

Serialization Delay
The second technology that makes Infiniband quick is serialization delay - the time it takes to clock bits onto the wire. It used to be that 1GbE was prominent in the HPC space, but now with 10GbE being very affordable it reduces serialization delay by a factor of ten.

For instance a 100 byte market data message transmitted on 10GbE, with all necessary headers takes about 1.2 microseconds to serialize at 10Gb/s and with one switch 'hop' in the transit path would take 1.8-2.0 microseconds to transmit. Infiniband would still be a bit faster here, approximately 500-600 nanoseconds, however Ethernet has clearly closed this gap considerably from the 1Gb days. Recently some custom commercial silicon vendors have announced they were able to deliver end-to-end latency over 10GbE of 1.2usec.

Cut Through Switching
When Infiniband was introduced it was all one speed and had cut through switching where Ethernet supported multiple speed variants and was store and forward based. In a store-and-forward architecture the entire packet has to be received in the Ethernet switch, examined, and then forwarded out the egress port. Cut-through switches receive just the header, make the forwarding decision, and then pipeline the rest of the packet directly to the egress port for much lower latency.

The Arista 7100 Series is based on a cut through switching architecture that enables port-to-port switching latencies of less than 1usec in an 10Gb Ethernet switch.

These three components that made Infiniband fast now exist within Ethernet.

Why is Ethernet a better choice?
With Ethernet a customer does not need a separate network investment and separate technology they need to learn, staff, and support. Every engineering student and MIS professional learns about Ethernet, everyone in IT has experience with Ethernet. People understand Ethernet.

Getting off of an Infiniband island out to other trading, storage, or computing systems or even out over a Wide-Area Network is complex and involves multiple gateways that slow down each transaction and create choke-points in the infrastructure. The same operation is trivial with Ethernet, and is supported by every networking vendor.

The large majority of all traffic in the world is IP traffic - it is the basis for today's Internet economy, and almost every application is designed to run on IP. By contrast the Infiniband MPI acceleration is not effective for IP and thus Infiniband has poor performance for IP traffic. Any application based on TCP/IP, UDP, or IP Multicast should use 10GbE.

Ethernet easily supports IP, Multicast, and TCP - all necessary in many HFT and HPC environments, and required for globally addressing any host, server, or storage node.

Summary
With the right network interfaces deployed in the server and the right switching choices Ethernet is more than capable of supporting any Infiniband workload.