



How to Make Ethernet Ready for Converged Enterprise Data Center Networks

An Advanced Systems Group White Paper

Over the past several decades, Ethernet has emerged as the dominant networking protocol. Although organizations add other networking protocols for special purposes, Ethernet has become pervasive.

Unfortunately, Ethernet doesn't supplant the need for other networking fabrics, most notably Fibre Channel (FC). The main reason for this is it lacks support for quality of service (QoS), which is probably its most significant limitation. QoS refers to the ability to create and manage networks to deliver different levels of service for different types of traffic. Currently, Ethernet QoS doesn't do enough to distinguish between types or classes of traffic. While it's true that Ethernet can achieve some level of QoS, native Ethernet gives all classes and/or types of traffic equal access to bandwidth.

QoS also refers to lossless data transmission. Ethernet networks lose data when oversaturated with traffic because its buffers overflow and it has to drop data packets. Ethernet tries to compensate by resending dropped packets, but that often only exacerbates the problem. Although these resends happen quickly (<25 milliseconds), they contribute to the lack of consistent response times.

Overcoming these limitations is the key to using Ethernet as the foundation for efficient converged data center networks and for delivering robust QoS.

Data Center Bridging

Ethernet utilizes upper level layer protocols (TCP) to manage end-to-end data delivery and integrity. FC, by comparison, provides a buffer-to-buffer credit that ensures packets aren't dropped due to network congestion, making it lossless. As Ethernet is an 802-based network, we can only make Ethernet lossless by adopting higher-level protocols such as TCP/IP, which have adapted to the intent of IEEE 802-based networks by incorporating end-to-end congestion avoidance and flow control algorithms.

Data center bridging (DCB) is an architectural extension to Ethernet designed to improve and expand its role in the data center. With DCB, organizations can logically manage networks end-to-end with QoS through:

- Quantized Congestion Notification (QCN)— provides end-to-end congestion management mechanisms to enable throttling of traffic in the event of traffic congestion.
- Priority-based Flow Control (PFC)— enables control over individual data flows on shared lossless links
- Enhanced Transmission Selection (ETS)— Control/Negotiation protocol to negotiate QoS parameters and queuing and provides discovery exchange protocol to ensure consistent configurations across the network

Ethernet's pervasiveness delivers a key benefit: Product vendors and network administrators alike widely support it, Ethernet skills are readily available, it's quite flexible, and compatibility is rarely an issue.

Currently, Ethernet QoS doesn't do enough to distinguish between types or classes of traffic. While it's true that Ethernet can achieve some level of QoS, native Ethernet gives all classes and/or types of traffic equal access to bandwidth.



With these enhancements, we now only need one type of network adapter eliminating the redundant infrastructure cost by using one cable infrastructure to support both storage and IP traffic.

It should be noted that Ethernet does support storage, but mainly through iSCSI, the server-to-storage protocol designed to transport SCSI block storage commands over Ethernet using TCP/IP. iSCSI adoption is growing and has found considerable acceptance in small and midsize enterprises. (At Advanced Systems Group, we've documented an increase in 10Gbps iSCSI deployments over the past 18 months.)

Data Center Bridging enhancements to Ethernet networks also enable Fibre Channel over Ethernet (FCoE), a new protocol specification developed in the INCITS T11 committee. With FCoE, we can transmit FC commands natively over DCB networks. FCoE specifies the mechanism for encapsulating FC frames onto DCB networks, allowing FC and Ethernet to coexist with a converged infrastructure.

By deploying a converged network, data center managers gain broad benefits such as:

- Lower costs from a reduced need for dedicated single-use components such as FC-only HBAs and switches
- A common management platform, personnel, knowledge and training across the data center
- A platform with a future that looks to provide a performance upgrade path to 100Gbps

Since DCB now can provide an adequate transport for storage protocols, it still needs to overcome its shortcomings with the Spanning Tree Protocol (STP). In traditional Layer 2 (L2) Ethernet networks, organizations create highly available networks by designating paths through the network as active or standby using STP. While this provides an alternate path, only one path can be used at a time, which means that network bandwidth isn't well utilized. Since one of the goals of server virtualization is increased utilization of the physical server, we can also expect increased utilization of network bandwidth.

To increase network utilization, Multiple Spanning Tree Protocol (MSTP) and similar protocols allow for separate spanning trees per VLAN. While this improves bandwidth utilization, it still leaves the STP limit of one active path between switches. And, because traffic paths are manually configured with MSTP, complexity increases.

Another challenge with STP is network behavior when links fail. Links do fail, and when that occurs the spanning tree needs to be redefined. This can take anywhere from five seconds with Rapid Spanning Tree (RSTP) to several minutes with STP—and this situation can vary unpredictably even with small topology changes. Finally, when a redefined spanning tree is reconverging, broadcast storms can occur causing more congestion. These STP limitations are why L2 networks typically are kept small in the data center.

These limitations become more important when applications run as virtual machines (VMs) instead of on physical servers. VM mobility now occurs within a cluster of physical servers, and with the limitations of STP, the sphere of VM migration is constrained to L2 subnets. The solution for flexible VM mobility is a more scalable and available L2 network with higher network bandwidth utilization.

What we need is a L2 network that:

- Is highly available
- Guarantees high-bandwidth utilization over equal-cost paths
- Doesn't stall traffic when links are added or removed due to failure or network reconfiguration
- Makes latency deterministic and is lossless



QoS also refers to loss-less data transmission, particularly regarding storage networking. Networks lose data when over-saturated with traffic, which forces buffers to overflow and data packets to drop. Ethernet tries to compensate by resending the dropped data packets, but that often only exacerbates the problem.



DCB provides these features through the use of TRILL (Transparent Connection of Lots of Links). The basic premise of TRILL is to replace STP as a mechanism to find loop free trees within layer 2 broadcast domains. It is an IETF standard that allows every switch to know the MAC address and World Wide Name (WWN) of all devices located on all the other switches. The shared view of the network lets you add a new physical link between two switches with no software configuration. The switches see the new link and add it to the network.

Ethernet DCB and TRILL allows data centers to broaden the sphere of application mobility and optimize server resources for applications just as it improves networking for storage. Unfortunately different vendors are introducing different adaptations of TRILL. Brocade implements Virtual Chassis Switch (VCS) whereas Cisco uses its own FabricPath implementation. Juniper is approaching the problem with a TRILL-less implementation known as QFabric.

With DCB Ethernet data centers can now:

- Logically eliminate the management of multiple switching layers
- Apply policies and manage traffic across many physical switches as if they were one switch
- Scale network bandwidth without manual reconfiguration of switch ports and network policies
- Provide a single, customized view of network status available to server, network, and storage administrators

With the new DCB fabric, IT can simplify network architecture, more rapidly scale their networks, easily adopt higher speed Ethernet standards (40Gbps is already defined with 100Gbps soon to follow) and significantly reduce management overhead.

About the Author

Mark Teter is the Chief Technology Officer at Advanced Systems Group. He is an internationally recognized authority on information technology who regularly advises IT organizations, vendors, and government agencies on a broad range of information management issues. Each year, Mark conducts dozens of seminars and training programs for corporate and government institutions. He sits on several financial industry advisory boards and has recently published *Paradigm Shift: Seven Keys of Highly Successful Linux and Open Source Adoptions*.

About Advanced Systems Group

Since 1981, Advanced Systems Group (ASG) has been providing comprehensive consulting services, successful storage and data management solutions, assessments, and implementation services to help customers meet today's IT & business challenges. In particular, ASG focuses on customer needs, customizing unique solutions and successfully addressing companies' particular IT challenges. As a consistent member of the VAR Business Top 500, ASG pursues active involvement in the industry, maintaining the highest level of engineering certifications with partners and the vendor community.



Call us at 800.894.3619
www.virtual.com

Denver . Baton Rouge . Boise . Colorado Springs . Dallas . Houston . Los Angeles . New Orleans
Oklahoma City . Orange County . Phoenix . Portland . Salt Lake City . San Diego . Seattle