A TRILLing subject

A new network protocol, called Transparent Interconnection of Lots of Links (TRILL), is expected to be completed in the second half of 2010, writes **Mike McNamara, (right), FCIA Marketing Chair, NetApp and Ahmad Zamer, (far right), FCIA member, Brocade.**

Overview of Transparent Interconnection of Lots of Links (TRILL)

Many IT organizations operate purpose-built multiple networks to connect to servers. Such networks are dedicated to IP networking, shared storage, and for Inter-Process Communication (IPC) for high-performance computing environments (figure 1). Most often overlapping networks contribute to IT organizations added cost in numerous ways; such as additional capital equipment, cable complexity, administrative costs, and additional power and cooling from multiple components.

The concept of I/O consolidation and unification or convergence enables the ability of network adapters, switches, and storage systems to use the same shared Ethernet physical infrastructure to transport different types of protocol traffic. For the IT network manager, I/O consolidation equates to installing, operating and maintaining a single network infrastructure instead of three, while still having the ability to differentiate between traffic types. The data center manager can purchase fewer host bus adapters and NICs, cables, switches, and storage systems, reducing power, equipment, and administrative costs.

Lossless 10-Gibabit Ethernet

I/O consolidation and unification promise to support both storage and TCP/IP network traffic on a single network infrastructure. One of the primary enablers of I/O consolidation is lossless 10-Gigabit Ethernet, a technology with bandwidth, data integrity, and latency characteristics sufficient to support multiple traffic flows on the same link. The following factors are driving adoption and the eventual ubiquity of 10GbE as a shared I/O transport:

- Server virtualization enables application workload consolidation, which contributes to higher network throughput demands and higher bandwidth utilization
- Virtualization aggregates multiple applications and OS instances on a single physical server with each application and OS instance generating significant I/O traffic. This places an overwhelming demand on existing multiport 1GbE infrastructures.
- Multi-socket, multi-core server technology supports higher workload levels, which demand greater throughput from IP networking and Fibre Channel Storage Area Network (SAN) interconnections.
- Increasing use of network storage requires higher bandwidth links between servers and storage.

Ethernet enhancements – Standards overview

For 10GbE to be a strong viable option for server I/O consolidation and storage networking, enhancements must be added to Ethernet to perform functions of other transports and open the door for converging multiple fabrics onto a single shared lossless Ethernet networking transport. The diverse nature of the technologies needed to enable convergence requires the development of several new industry standards that cover Fibre Channel, Ethernet,

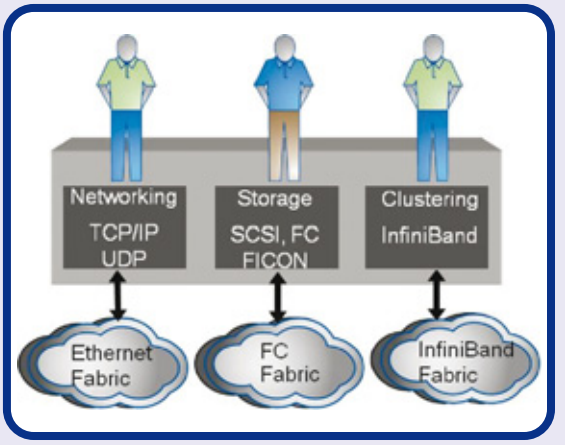

Figure 1

and Link Layer (Layer 2) routing. The FCoE, L2 routing, and Data Center Bridging (DCB) protocols are being developed by three different industry standards bodies, each focusing on technology areas that fall under a specific domain of expertise:

INCITS Technical Committee T11: FCoE

- FCoE (figure 2) and FIP are defined in FC-BB-5, which describes how Fibre Channel protocol is encapsulated, transported and mapped over lossless Ethernet. The T11 committee completed its technical work for FC-BB-5 in June 2009, and forwarded the draft standardto INCITS, where it was approved and will be published soon.

IEEE: Data Center Bridging

- Data Center Bridging development by an IEEE 802.1 work group is aimed at adding new extensions to bridging and Ethernet so that it becomes capable of converging LAN and storage traffic on a single link. DCB is designed to make Ethernet more like Fibre Channel, because the new features being added to Ethernet are solving issues that Fibre Channel faced in the past and successfully resolved. IEEE is expected to complete its work on the various components of DCB in the second half of 2010. The new enhancements are:

802.1Qbb: Priority-based Flow Control (PFC)

- Establishes eight priorities for flow control based on the priority code point field in the IEEE 802.1Q tags. This enables control over individual data flows on shared lossless links. PFC allows Fibre Channel storage traffic encapsulated in FCoE frames to receive lossless service from a link that is being shared with traditional LAN traffic, which is loss-tolerant.
- PFC provides link level congestion control that alleviates flow control difficulties common in TCP/IP environments.

802.1Qaz: Enhanced Transmission Selection (ETS)

- ETS provides the capability to group each type of data flow, such as storage or networking, and assigns an identification number to each of the traffic class groups. The value of this new feature lies in the ability to manage bandwidth on the Ethernet link by allocating portions (percentages) of the available bandwidth to each of the groups. Bandwidth allocation allows traffic from the different groups to receive their target service rate (such as 8 Gbps for storage and 2 Gbps for LAN). Bandwidth allocation provides quality of service to applications sharing a common transport medium.
- ETS incorporates Data Center Bridging Exchange (DCBX), a discovery and initialization protocol that discovers the resources connected to the Enhanced Ethernet cloud and establishes its limits. DCBX distributes the local configuration and detects the misconfiguration of ETS and PFC between peers. It also provides the capability for configuring a remote peer with PFC, ETS, and application parameters. The application parameter is used for informing the end station which priority to use for a given application type (e.g. FCoE, iSCSI). DCBX leverages the capabilities of IEEE 802.1AB Link Layer Discovery Protocol (LLDP).

Figure 2

802.1Qau: Quantized Congestion Notification (QCN)

- This end-to-end congestion management mechanism enables the throttling of traffic at the edge nodes of the network in the event of traffic congestion.
- QCN is an independent protocol and PFC and ETS do not require QCN to deliver FCoE over lossless DCB links.

IETF: TRILL

- Internet Engineering Task Force (IETF) is developing a new shortest path frame routing protocol in multi hop environments. The new protocol is called Transparent Interconnection of Lots of Links (TRILL) and is expected to be completed in the second half of 2010:
- TRILL provides a Layer 2 multipath alternative to the single-path and network bandwidth-limiting Spanning Tree Protocol (STP) currently deployed in data center networks.
- TRILL also provides Layer 2 multi-hop routing capabilities that are essential for expanding the deployment of DCB/FCoE solutions beyond access layer server I/O consolidation and into larger data center networks.

Why TRILL

The ever increasing adoption of virtual environments in data centers necessitates a more resilient L2 networking infrastructure. Efficient and reliable L2 infrastructure is needed to support the I/O demands of virtual applications especially when applications are migrated across servers or even different data centers. Today's STP-based networks limit the available network bandwidth and fail to maintain reliable, complex network architectures. Although there may be many Equal Cost Multiple Paths (ECMPs) or physical paths through the network at any given time, all traffic will flow along the path that has been defined by a spanning tree that includes all network devices and nodes. By restricting traffic to this tree, loops in the logical topology are prevented at the expense of blocking alternative network paths.

While STP solves the problem of traffic loops, it prevents network capacity from being fully used. Algorithms that calculate this spanning tree may take a considerable amount of time to converge following changes in the status of the configurations. During that time, the regular flow of traffic must be halted to prevent the type of network saturation described above. Even if multiple simultaneous spanning trees are used for separate VLANs to better distribute traffic flows, the traffic in any one VLAN will still suffer from the same disadvantage of not being able to use all of the available capacity in the network. TRILL will enable multipathing for L2 networks and remove the restrictions placed on data center environments by STP single path networks. Data centers with converged networks will also benefit from the multi-hop capabilities of TRILL Routing Bridges (Rbridges) which enable multi-hop FCoE solutions.

What is TRILL

To eliminate the restriction of STP single path through the network, the IETF formed a working group to study and solve this problem. In summary, the group was charged with developing a solution that:

- Uses shortest path routing protocols as opposed to STP
- Works at Layer 2
- Supports multi-hopping environments
- Works with an arbitrary topology
- Uses an existing link-state routing protocol
- Remain compatible with IEEE 802.1 Ethernet networks that use STP

The result was a protocol called TRILL. Although routing is ordinarily done at Layer 3 of the ISO protocol stack, by making Layer 2 a routing layer, protocols other than IP, such as FCoE, can take advantage of this increased functionality. Multi-hopping allows specifying multiple paths through the network. By working in an arbitrary topology, links that otherwise would have been blocked are usable for traffic. Finally, if the network can use an existing link-state protocol, solution providers can use protocols that have already been developed, hardened, and optimized. This reduces the amount of work that must be done to deploy TRILL.

What TRILL does and does not do

Although TRILL can serve as an alternative to STP, it doesn't require that STP be removed from an Ethernet infrastructure. Most networking administrators can't, and will not, just rip and replace their current deployments just for the sake of implementing TRILL. So hybrid solutions that use both STP and TRILL are not only possible but will be the norm for at least the near term future. TRILL will also not automatically eliminate the risk of a single point of failure, especially in hybrid environments. The goals of TRILL are restricted to those listed above. Another area where TRILL is not expected to play a role is the routing of traffic across L3 routers. TRILL is expected to operate within a single subnet. While the IETF draft standard document mentions the potential for tunneling data, it is unlikely that TRILL will evolve in a way that will expand its role to cover cross L3 router traffic. Existing and well established protocols such as Multiprotocol Label Switching (MPLS) and Virtual Private LAN Service (VPLS) cover these areas and are expected to continue to do so.

Simply put, TRILL enables two primary capabilities:

- Multi-pathing for L2 networks
- Multi-hopping that will benefit FCoE configurations

Summary

TRILL is a new draft standard being created by IETF to accommodate FCoE converged networks and is scheduled to be completed later this year. The goal of TRILL is to create an L2 shortest path robust multi-path and multi-hop routing protocol that eventually will replace the limited and restricted L3 STP. The more resilient L2 will fulfill the needs of 10Gb Ethernet networks with virtualized ecosystem and data migration. TRILL will also enable multi-hop capabilities for FCoE that will drive expanded adoption of FCoE in converged network environments.

www.fcia.org