STORAGE SWITZERLAND REPORT

STORAGE EVOLUTION: FC SAN, IP SAN, ETHERNET SAN

George Crump, Senior Analyst

Yes, there's a difference between IP and Ethernet

The demand for shared storage is greater than ever, as are the options to connect servers to a shared pool of storage resources. The primary goals of each new storage connectivity method is to do one of the following: reduce costs, reduce complexity, increase performance or increase flexibility. Seldom are all four goals addressed simultaneously. A new breed of Ethernet SAN solutions, running on the lightweight ATA over Ethernet (AoE) protocol, may actually address each of these four goals. AoE is storage over Ethernet but it's not IP SAN (iSCSI) or NAS. What is AoE and how might it address the growing need for shared storage?

The need for shared storage is traditionally driven by a desire to improve storage economics by leveraging capacity across multiple servers or to improve data protection by centralizing the data repository. Over the past few years a new driver has emerged and it's now the dominant justification for shared storage - server virtualization. Server Virtualization needs shared storage so that multiple host servers can transport virtual machines between each other, and it needs high performance.

This new justification, server virtualization, has accelerated the adoption of shared storage in newer environments and has broadened its use in established ones. It has also changed the way that data is being accessed, creating an increasingly dynamic workload on the storage infrastructure. In the past each server was individually making requests of the storage subsystem. In the virtualized environment each host can have 20, 30 or more servers inside of it, all making storage requests. As a result, customers are facing scaling and performance limitations caused by bottlenecks in the storage infrastructure. This may result in more complexity as workarounds are developed to aggregate the ports accessing this infrastructure or to manage access priorities to critical virtual machines.

Mainframe environments used network protocols like FICON and ESCON to connect to storage resources, and these connection-based mainframe protocols evolved into Fibre Channel storage area networks. Fibre Channel, backed by the top storage vendors in the 1990's, grew into the dominant technology for enterprise storage networks. While other legacy networks like token ring and InfiniBand declined in popularity, the partner ecosystem and vendor backing around Fibre Channel allowed it to carve out a substantial market. What started out as small 8-port SANs grew over time into sprawling network topologies with hundreds of ports, complete with costly switches, HBAs, and specialized multi-pathing software for high availability. These static network topologies were designed primarily for large servers running database workloads, and it's unlikely the early designers every envisioned the chaos of modern virtualized server environments with technologies like VMware's VMotion.

As Fibre Channel began to establish itself as the shared storage protocol of choice, other protocols were developed to address its limitations, mostly focused on cost and complexity. Two of the more common protocols, especially in the virtualized environment, were iSCSI and NAS via NFS. Both had the advantage of using standard Ethernet cabling to connect the servers to the shared storage. In most cases this allowed a data center to reduce the amount of specialized switches and HBAs that needed to be bought. Each though had its challenges.

iSCSI, although Ethernet based, essentially encapsulated SCSI storage commands in a higher-level TCP/IP wrapper to provide reliable delivery. TCP/IP protocols, designed for routing Internet traffic across multiple router hops, also introduced a heavy overhead for transporting data from servers to storage. Some of this performance penalty was addressed with specialized iSCSI HBAs, but performance still lagged behind Fibre Channel, and many environments did not deploy the more costly HBAs for acceleration. While iSCSI did have the apparent simplicity of running on the same IP network that the administrator was used to, it also mimicked the connection-based topology of Fibre Channel networks. As a result, iSCSI retained much of the complexity of Fibre Channel, requiring configuration of features such as multi-pathing software and port bonding.

Similar to iSCSI, NFS environments also leverage the existing Ethernet infrastructure, but are reliant on IP for protocol transmission. While NAS is often cited as an easier use case for shared access to data, its traditional application has been for serving file data to connected user clients, not data to servers. Improvements in NAS capabilities have increased its adoption in the data center for server data sharing, but it has not seen widespread adoption, in comparison to traditional Fibre Channel or even iSCSI. NAS, whether it be CIFS or NFS, is still essentially a file system, and as a result, it also has the overhead of translating from one file system to another.

The limitation in both of these environments is that while they drive out much of the cost of a comparable Fibre Channel installation, the simplicity gains are often only temporary - until the environment needs to scale. When performance becomes an issue, and in virtualized environments it often does, complexity has to be added to these configurations to increase performance from the host, through the switch and to the storage devices. Customers are looking for a solution that can address both the cost issues of Fibre Channel and the performance scaling issues of iSCSI/NAS. New Ethernet SAN solutions based on ATA over Ethernet (AoE) may offer a compelling solution to these challenges.

AoE leverages Ethernet cabling similar to how iSCSI and NAS do; but that's the extent of the similarity. As drive technology developed, the intelligence for managing the drive moved away from the controller and onto the drive itself. The drive was then directly connected to the motherboard by way of a ribbon cable instead of needing an onboard HBA. At its most basic level AoE moves that drive out of the server so it can be shared, and replaces the ribbon cable with an Ethernet cable. The communication across that Ethernet connection uses raw Ethernet frames at Layer 2, rather than the heavier TCP/IP layer, to transport reads and writes. Considering the increasing ubiquity of 10 Gb Ethernet, with 40 Gb and 100 Gb on the horizon, Ethernet SAN provides a compelling "bare metal" performance story relative to traditional protocols.

Without this IP limitation, AoE can scale to meet the growing demands being placed on shared storage. AoE has been supported in the Linux kernel since 2005, and servers can install standard Ethernet NICs with AoE driver software (essentially AoE HBAs) to accelerate performance. Because AoE runs on Layer 2, Ethernet SANs eliminate the need for multi-pathing software, switch configuration, port bonding, and other similar features of connection-based SANs. Server I/O is automatically load balanced over all available ports, so SAN connectivity can be scaled linearly by adding inexpensive 1 GbE or 10 GbE HBAs.

The AoE protocol was invented and released to the open source community by Brantley Coile, CTO of storage vendor CORAID. Companies like CORAID build on this platform to offer storage virtualized disk shelves that allow capacity to be discretely shared with specific hosts or made available to multiple hosts in the environment. Each shelf has its own controller and virtualization engine so performance can be scaled out without bottlenecks as capacity is added. Each additional shelf leads to more performance while any server in the environment is able to access any of the storage systems in the environment via its Ethernet connectivity.

From a management standpoint, AoE storage looks like a direct-attached SCSI disk. Storage can be managed from within the hypervisor layer, providing "self-service" control to application owners. For the massive wave of new virtualization environments and users, eliminating complex SAN topologies is very attractive.

With AoE we may have a protocol that fulfills all four of the goals of shared storage: reduce costs, reduce complexity, increase performance and increase flexibility. It's a protocol that delivers potentially great simplicity and cost savings compared to iSCSI and NAS. And, with CORAID's implementation, it may be capable of going head-to-head with Fibre Channel on performance.