Considerations When Selecting a Fabric for Storage Area Networks:

A G2M Research White Paper

White Paper Key Points

1. The choice of a SAN fabric and protocol has a significant impact on application performance, availability, and cost.

2. A dedicated storage fabric such as Fibre Channel provides the highest throughput and lowest latency/jitter, albeit at a higher CapEx compared to existing Ethernet LANs.

3. When a dedicated SAN isn't an option, Ethernet-based storage protocols offer lower CapEx and OpEx but at the expense of performance.

4. Where it is a close call, deploy higher speeds of the SAN fabric that you are currently using, as it represents the lowest risk and potential to disrupt your operations.

Considerations When Selecting a Fabric for Storage Area Networks: Executive Summary

Today, there are a variety of architectures, fabrics, and protocols that can be utilized for Storage Area Networks (SANs). Because of this, the challenges in choosing the right approach, fabric, and protocol to optimize application performance are not trivial, nor are all the tradeoffs obvious when making such as choice. The choice of a storage architecture (and the fabric and protocols utilized) is generally driven by four factors:

1) The "scale" of the SAN; e.g., how many nodes need to access the shared storage pool.

2) The performance required, as measured by throughput (either bandwidth or packets per second for a given packet size) and latency (the time it takes to complete a storage command).

3) The consistency of the performance that applications require before their availability is compromised.

4) Whether a dedicated SAN can be utilized, or whether a converged network is required, either due to economics or other constraints.

For applications that can afford a dedicated SAN and require thousands of nodes sharing the same storage namespaces and/or significant throughput per initiator node with low jitter and low latency, Fibre Channel (FC) has several advantages over Ethernet-based fabrics. While converged Ethernet networks (one that operates as both a SAN and a LAN) are more economical and can boast clock speeds of up to 100Gb/second today, network congestion can be a significant issue for converged networks, impairing the usable bandwidth delivered and significantly adding to jitter. While utilizing a dedicated Ethernet-based SAN would reduce these effects, it also significantly increases SAN costs.

For use cases where a dedicated SAN is not desirable or economical, Ethernet converged networks utilizing NVMe over Fabric protocols for the SAN generally offer the best performance and latency, though storage architects must pay attention to the effects of congestion on SAN performance. Understanding where the boundary lies between these two problem spaces, and what to do when in this "gray area", is critical to ensure that your choice meets the shared storage needs of your application.

The Evolution of Storage Area Networks (SANs)

Local Area Networks (LANs), largely based on Ethernet technology, were deployed in both workspaces and data centers in the 1980s. With LANs came file servers, which provided shared storage to workstations over the LAN. As the number of file servers increase in organizations, the need to centralize their contents on high-performance storage systems became apparent. These centralized storage systems connected to servers by LAN. Storage traffic began to consume increasing amounts of bandwidth, often interfering with other LAN operations. This drove the development of a serialized optical networking technology for storage systems, standardized in 1994 by ANSI as Fibre Channel (FC).

SANs utilize block-level storage protocols, providing remote access to either physical or virtual blocks on the shared storage devices. The block-level protocols first used in SANs were based on the Small Computer System Interface, or SCSI specification. Since FC's inception, several Ethernet-based SAN protocols have been proposed as replacements for FC, with the goal of building converged LAN/SAN networks and reducing costs. These include SCSI-based protocols such as Internet SCSI (iSCSI) for Ethernet and Fibre Channel over Ethernet (FCoE).

One new entrant to the storage interface arena is the Non-Volatile Memory Express (NVMe) interface. NVMe, which was standardized early in this decade, is a block-level protocol which significantly reduces latency compared to SCSI protocols. Several remote access versions of NVMe, known as NVMe over Fabric (NVMe-oF) have been developed, including NVMe over RDMA (NVMe-oRDMA) for Ethernet, NVMe over TCP (NVMe-oTCP) for Ethernet, and NVMe over Fibre Channel (NVMe-oFC). The first NVMe-oF protocols were standardized in 2016, and the first NVMe-oF arrays from major storage vendors started to appear in mid-late 2017. The first major deployments of NVMe-oF based storage is expected to occur in late 2018 or 2019, but are not expected to reach significant levels of deployment (>10% of all new deployments) until the next decade.
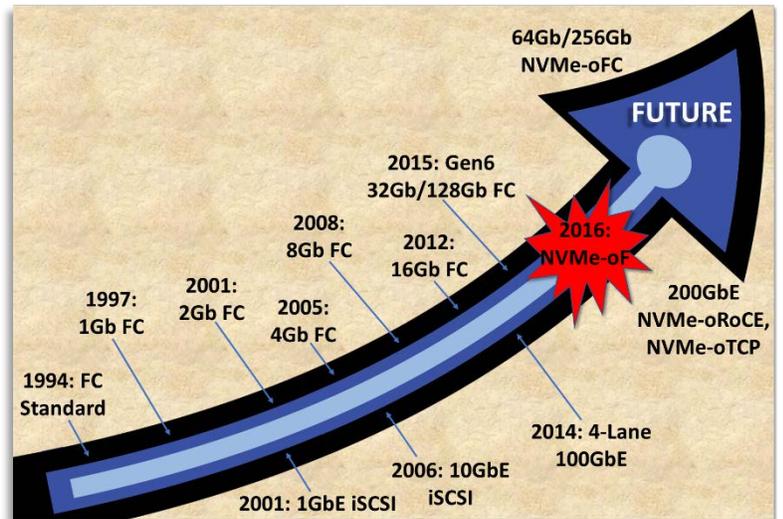
Criteria for Choosing the Right SAN Fabric and Protocol

The choice between using Fibre Channel or Ethernet is largely dictated by whether a dedicated SAN is required and/or desirable. Ethernet based SAN implementations are nearly always done in the context of a converged network architecture; FC networks are always utilized as dedicated SANs. For many use cases such as telecom central offices and embedded applications, a dedicated SAN is not an option. In other cases, the mission critical nature of the application alone justifies a dedicated SAN. Examples of this include billing systems and enterprise resource planning (ERP) systems for Fortune 500 companies, and raw content editing and post-production workflows in the media and entertainment (M&E) industry. However, there are many cases where the line between whether a dedicated SAN is justified can be unclear.

Since most SANs are built for no more than a few specific applications, the needs of these applications generally drive the requirements of the SAN. The criteria which drive the selection of the SAN protocol and fabric include the following (see Table 1):

1) **The Scale of Storage Connectivity:** How many application instances need to be able to simultaneously access the shared storage (what is the scale of the connectivity)?

Fabric/ Protocol	Dedicated SAN?	Scalability	Ease of Expansion	Through -put	Latency	Jitter	Delivery Reliability	Cost	Adoption
FC/FCP	Yes	5	5	4	4	5	5	2	5 ↔
FC/NVMe-oFC	Yes	5	5	4	5	5	5	2	1 ↑
Ethernet/iSCSI	No	3	2	3	2	2	2	5	3 ↓
Ethernet/NVMe-oF	No	4	2-3	5	3-4	3	3	5	1 ↑

Table 1: Rating of SAN Protocols Across Various Criteria (*Ratings: 1=worst to 5=best*)

2) **Storage Throughput Needs for Each Application Instance**: What are the storage performance needs of each application instance? This includes the throughput requirements (usually measured in GB/second) and the size of the packets (large which typically means 4KB per packet or more, or small which generally means less than 1KB per packet). Throughput can also be measured in packets per second (PPS), which is the GB/second divided by the packet size.

3) **Latency Needs for the Application**: Latency at the application level is the time delay from when a read/write is issued by the application to when the read/write is completed at the application level. Latency is typically measured in microseconds (us).

4) **Throughput and Latency Jitter Requirements**: Throughput jitter is the throughput's variance over time, quantified as the average difference from the mean throughput that the application instance experiences. For example, if the mean throughput to the application instance is 100K PPS with a jitter of 10K PPS, the application would see throughput between 90-110K PPS fifty percent of the time. Latency jitter is measured in a similar manner.

Other important criteria include delivery reliability (the likelihood that a packet reaches its destination quickly enough to avoid a storage stack exception), cost (both CapEx and OpEx), and adoption (both absolute adoption, and whether adoption is increasing, flat, or decreasing), which is an indicator of the long-term viability of a SAN technology.

Comparing the Various Fabric and Protocol Choices

While comparing the various protocols in an absolute sense is highly dependent on the properties of all of the network components, some general statements can be made. Since FC networks are dedicated SANs, FC generally does better from a performance, latency, and jitter standpoint than Ethernet-based networks. That is not to say there aren't instances where Ethernet SANs can provide better performance metrics than FC networks. 100GbE NVMe-oF networks that are used nearly exclusively for storage can provide higher throughput and similar latencies, though scalability of such networks beyond the rack level today is problematic.

Table 1 below compares these various protocols across seven criteria. The following sections will discuss each set of fabrics and protocols, the primary use cases for these today, and what drives each protocols' success (or lack thereof) for the primary use cases. For this study, we will not consider protocols that have not achieved/are not expected to achieve widespread adoption such as iSER (which is primarily limited to high-performance computing), FCoE (which never achieved widespread adoption and whose install base is shrinking), and NVMe-oiWARP (which is the "dark horse" in the race among NVMe over Ethernet protocols).

Fibre Channel

History, Evolution, and Adoption
Since its inception over two decades ago, Fibre Channel (FC) has been the fabric of choice for SANs. Since 2001, over 100M FC ports have been shipped,

with an estimated 46M FC ports in operation today[1], most of which are Gen 5 (16Gb FC) or Gen6 (32Gb FC). While FC unit sales have been declining by roughly 4% per year since early in this decade[2,3,4], this decline has been primarily due to the growth of cloud storage, which today utilizes either Ethernet storage networking or direct attach storage[5]. FC is the predominant fabric utilized by leading storage array vendors such as Dell/EMC, NetApp, and Pure Storage. FC switches are available from Broadcom and Cisco, and FC Host Bus Adapters (HBAs) are available from ATTO, Broadcom (Emulex), Marvell (QLogic), and others. The primary protocols supported on FC fabrics is Fibre Channel Protocol (FCP), which is a SCSI-based protocol.

The current maximum network speed for FC is 32Gb/second in a single lane, while inter-switch links (ISLs) using four lanes can operate at 128Gb/second. In the next few years, FC speeds are expected to double to 64Gb/second on a single lane, and up to 256Gb/second on quad-lane ISLs. The bulk of FC SANs deployed today are 16GbFC, followed by 8GbFC and 32GbFC.

Use Cases and Considerations

As a dedicated storage network, FC SANs provided significantly better performance in most "real world" use cases than Ethernet-based alternatives such as iSCSI, albeit with higher CapEx and OpEx costs than Ethernet alternatives due to the cost of managing two networks. However, FC SANs are highly scalable and are immune to network congestion due to the use of a credit-based flow control mechanism, enabling FC SANs to be built that contain thousands of nodes without impacting throughput, latency, or jitter. FC SANs also support multi-pathing, improving performance and reliability. These factors simplify both the performance optimization and the addition of storage to FC SANs, tasks that are significantly more difficult to accomplish on converged Ethernet networks due to the need to constantly retune performance on converged networks as workload change and/or storage is added to the SAN.

Because of these capabilities, FC has been widely deployed in Fortune 500 (F500) enterprises, as well as in high-performance workflows. Use cases for FC SANs include:

- Large database clusters such as Enterprise Resource Planning (ERP) and billing systems.
- Banking financial management systems
- Airline and travel booking systems
- Large retail point of sale systems
- Persistent storage for large in-memory database solutions such as SAP HANA.
- Large media and entertainment workflow clusters such as raw content post-production and video editing systems.
- Oil and Gas analytics clusters

Ethernet/iSCSI

History, Evolution, and Adoption

iSCSI was once seen as a potential competitor to/replacement for FC, but it never achieved the scale of adoption and deployment that FC had achieved, as it was limited to the throughput of 1 Gigabit Ethernet (1GbE) until 10GbE emerged. However, the vast bulk of Ethernet SANs today are iSCSI-based, and have extremely low costs per port because iSCSI utilizes standard Ethernet network interface cards (NICs) and switches. iSCSI is supported by all Ethernet NICs (with the right software driver) and Ethernet switches. This cost factor and universal interoperability allowed iSCSI to achieve a significant installed base[6] and support from most major storage array manufacturers. However, this cost advantage is offset by iSCSI's increased latency (relative to FC) due to its more complex protocol stack, and Ethernet's susceptibility to congestion as shown in Figure 1.
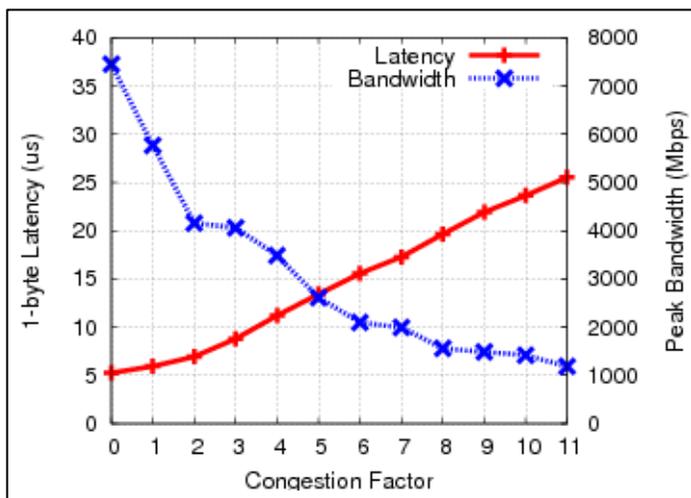

Figure 1: Ethernet Congestion vs Latency

The current maximum network speed for Ethernet is 25Gb/second in a single lane (25GbE), with quad-lane Ethernet reaching 100Gb/second (100GbE). The bulk of Ethernet converged LAN/SAN networks deployed today are based on 10GbE. While 25GbE is expected to supplant 10GbE in the market, that point is probably 3 years away. 100GbE is currently limited to applications where its high cost per port is justified by the performance required. Network congestion management schemes are also critical in 100GbE, and add complexity to its deployment, management, and optimization, especially when used as a SAN.

Use Cases and Considerations

iSCSI has found application in use cases where low cost for shared storage is the driving factor, and where performance and data delivery reliability are not critical factors. This has historically included small to medium business (SMB) applications, and low-performance workflows and applications that require shared storage such as file sharing, email, or web servers. iSCSI's high latency is its largest challenge.

Typical throughputs are 40%-50% of line rate. Throughput jitter and latency jitter are also significantly higher for iSCSI networks than other SANs for all but the simplest network topologies. This is due to the potential for out-of-order packet delivery and packet loss. While the theoretical scale of Ethernet networks is unlimited, iSCSI over Ethernet

networks start to experience congestion and "noisy neighbor" issues well before they reach even half of their theoretical throughput. These issues negatively impact the actual network throughput by causing retries that reduce network capacity even on 100GbE networks, as well as network reliability for storage applications. Ethernet's susceptibility to network congestion also negatively impacts its throughput, as shown in Figure 2.

NVMe-oF Protocols:

One new set of emerging protocols to watch for converged networks are the various flavors of NVMe-oF. These protocols include NVMe over Fibre Channel (NVMe-oFC), and NVMe over RDMA over Converged Ethernet (NVMe-oRoCE) and NVMe over TCP (NVMe-oTCP) for Ethernet. NVMe-oFC runs on standard FC networks and has all of the advantages inherent in a dedicated SAN, but also enjoys rhe reduced latency resulting from the elimination of the SCSI protocol. NVMe-oRoCE utilizes a flow control methodology like that of FC's credit-based flow control to deliver packets in-order and without loss. NVMe over TCP (NVMe-oTCP) takes a different approach by utilizing standard TCP flow control, allowing interoperability with L2 Ethernet switches and management applications. While TCP-based NVMe adapters have roughly 10% higher latency

Figure 2: Per-Node Performance With Congestion
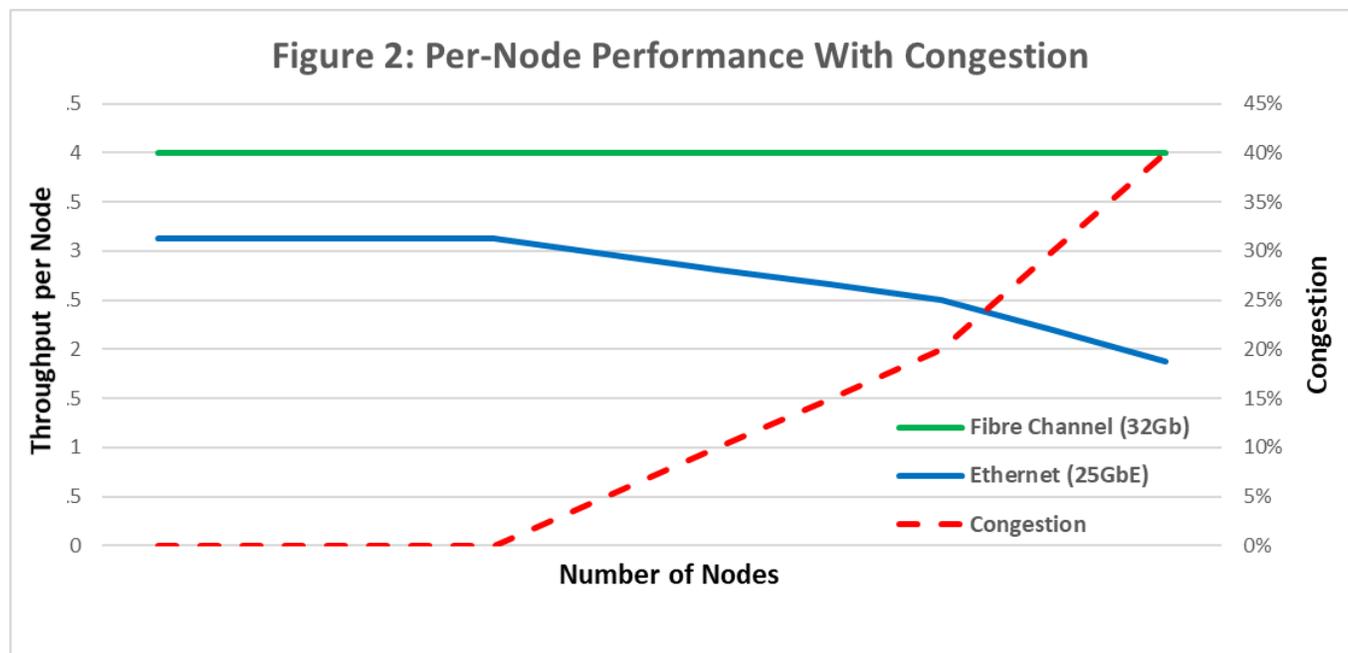
than NVMe-oRoCE in laboratory conditions, under realistic conditions the difference is minimal.

The primary issue with all NVMe-oF protocols is the relative immaturity of the larger storage ecosystem, particularly the availability of storage arrays that support NVMe-oF. A secondary consideration (especially with NVMe-oRoCE) is the susceptibility to network congestion and congestion spreading. Under network congestion conditions in lossless networks such as DCB networks, initiators are expected to limit the rate of packet injection into the network, in effect "capping" the performance application instances in NVMe-oRoCE networks. This effect can impact both network scale and throughput/latency jitter.

Decision-Making Outside the Clear FC/Ethernet Use Cases

How do you decide what SAN fabric to use when your application doesn't clearly fit into either the FC or Ethernet use cases? Assuming a dedicated SAN is an option, Table 2 provides some decision criteria to utilize when deciding on a SAN fabric and protocol, both for greenfield deployments and for improvements to existing SANs. If you are deploying a new SAN, start with what you know. For instance, deploying an FC SAN in an organization that has not previously done so would be a high risk; the same is true for large NVMe-oRoCE deployments. That is not

to say that such a decision might not be the best answer; rather such an approach should only be undertaken with design and deployment support (e.g., professional services) from your storage/server vendor.

Picking the Right SAN

The choice of a fabric and protocol for a storage area network (SAN) can have a significant effect on application performance and availability. For use cases where a dedicated SAN is not desirable or economical, NVMe-based Ethernet protocols offer performance that approaches FC with lower CapEx and OpEx costs, but storage architects must also consider the effect that non-storage traffic can have on network congestion which can impact storage performance and reliability, as well as the availability and maturity of NVMe-oF storage arrays and storage appliances. For enterprise applications where the size of the SAN is in the thousands of ports, or for use cases where application instances require both high throughput and low latency, FC SANs continue to be the lowest risk choice for storage architects. The ability to scale and grow FC SANs without impacting performance (which increases the OpEx of converged Ethernet SANs) offsets the higher CapEx costs of an FC SAN. Because FC SANs are dedicated to storage traffic, they can consistently provide high throughput, low latency, and extremely low jitter.

Network Type		HDD or SAS/SATA Flash Storage System	NVMe-oF Based Storage System
Deploying on an Existing Network	10GbE	Ethernet/iSCSI	Ethernet/NVMe-oTCP
	16GbFC	FC/FCP	FC/NVMe-oFC
Deploying a New Network	Large Workgroup (>10)	16GbFC/FCP	16GbFC/NVMe-oFC
	Small Workgroup (<6)	10-25GbE/iSCSI	25GbE/NVMe-oRoCE
	Single Rack	Ethernet/iSCSI	NVMe-oRoCE
	Multiple Racks, Replacing 10Gb iSCSI	25GbE/iSCSI	25GbE/NVMe-oTCP
	Multiple Racks, Replacing 8Gb FC	16/32Gb FC/FCP; 25GbE/iSCSI	16/32Gb NVMe-oFC; 25GbE/NVMe-oTCP

Table 2: Decision-Making Criteria for Selecting a SAN

About G2M Research

G2M Research provides targeted industry research for emerging enterprise technology markets, including market sizing, ecosystem mapping and market analysis. In addition to providing our standardized analysis in markets such as NVMe and endpoint security, G2M Research also performs customized research for companies. G2M Research is a part of G2M Communications, Inc. For more information, visit www.g2minc.com/research.

About ATTO Technology

For 30 years, ATTO Technology, Inc., has been a global leader across the IT and media & entertainment markets, specializing in storage and network connectivity and infrastructure solutions for the most data-intensive computing environments. ATTO works with partners to deliver end-to-end solutions to better store, manage and deliver data. Working as an extension of customer's design teams, ATTO manufactures host and RAID adapters, network adapters, storage controllers, Thunderbolt enabled adapters and software. ATTO solutions provide a high level of connectivity to all storage interfaces, including Fibre Channel, SAS, SATA, iSCSI, Ethernet, NVMe over Fabric, and Thunderbolt. ATTO is the Power Behind the Storage.

References

1 – FCIA, "Fibre Channel's Growth Trend", Mark Jones (2017).

2 – NETWORKComputing, "The Sinking Fibre Channel SAN", Jim O'Reilly (January 4, 2016).

3 – NETWORKComputing, "Fibre Channel Really Is Dead", Kevin Deierling (April 23, 2015).

4 – BusinessWire, "Technavio's Top Three Trends to Impact the Fibre Channel Adapter Market" (April 4, 2017).

5 – Dell Oro Group Blog, "Storage Systems Market in Transition" (July 21, 2017).

6 - SearchStorage, "Essential Guide: How to choose the best hardware for virtualization; iSCSI", Margaret Rouse (2014).