



ARISTA WHITE PAPER

Deploying IP Storage Infrastructures

The cost of deploying and maintaining traditional storage networks is growing at an exponential rate. New requirements for compliance and new applications such as analytics mean that ever-increasing volumes of unstructured data are collected and archived.

Legacy storage networks cannot meet the need to scale-out capacity while reducing capital and operational expenditures. In response to this challenge, new storage architectures based on IP/Ethernet have evolved, and are being adopted at an ever-increasing rate.

While technologies such as NAS and iSCSI are widely deployed, Fibre Channel SANs have maintained a dwindling but still strong presence in enterprise architectures. This paradigm is beginning to change as scale-out storage systems enabled by Software Defined Storage (SDS) are increasingly mature. The ability to reclaim stranded Direct Attached Storage (DAS) assets in a server infrastructure combined with the efficiencies gained by running storage traffic over the same IP/Ethernet network as standard data traffic provides an undeniably strong opportunity to reduce both the capex and opex required to deploy and run a storage infrastructure. According to some estimates, as much as a 60% capex reduction is achievable by deploying an SDS architecture which leverages server DAS to create a VSAN. Even without such radical savings, scale-out storage relying on IP/Ethernet offers dramatic savings over traditional Fibre Channel based storage. Most future IT asset deployments will leverage 10 Gigabit Ethernet (10GbE) and now 40 Gigabit Ethernet [40GbE] for the underlying storage interconnect for newer applications.

Traditional datacenter networks however are not designed to support the traffic patterns and loss characteristics to reliably deploy an IP/Ethernet storage infrastructure. Instead, a new class of switches and designs is required to ensure success. Arista Networks has responded to the need for a new IP/Ethernet storage fabric. Arista provides industry-leading products and solutions that deliver operational and infrastructure efficiencies that have been previously unavailable.

ARISTA



CHALLENGES IN IMPLEMENTING AN IP/ETHERNET STORAGE NETWORK

The distributed nature of scale-out IP/Ethernet storage architectures is driving a need to reevaluate the way networks are deployed today. Scale-out storage has a very different set of challenges from legacy storage transport. Figure 1 shows these three common challenges:

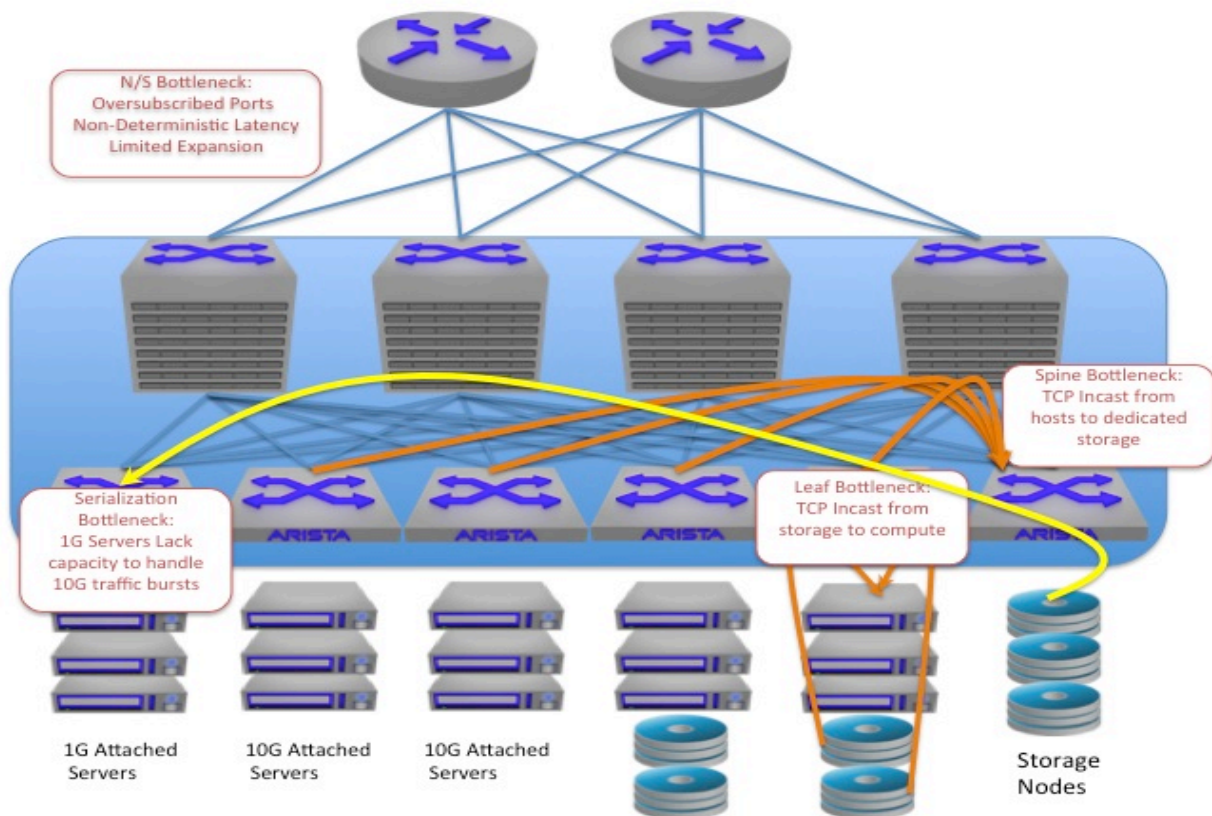
- Scale-out Storage networks require massive east to west performance and scalability.
- TCP incast occurs when a distributed storage and compute frameworks such as Hadoop or scale-out is implemented.
- Storage and compute devices interconnect at different speeds.

SCALE-OUT STORAGE TECHNOLOGIES

The term scale-out storage can be used to refer to multiple different types of storage technology. Generically, the term refers to storage systems

whereby a software layer presents an aggregate view of storage resources across multiple nodes. This enables storage resources to be added or removed from the system as a whole to provide at or near linear growth to the system by adding new resources. Storage resources are treated as building blocks rather than monolithic entities. This approach is contrary to the scale up approach used in legacy storage arrays where the size and breadth of the storage systems themselves was a boundary requiring new, higher scale systems when more capacity was needed.

Scale-out technologies are available for NAS, SAN, and hybrid environments from most storage vendors. Additionally, newer technologies, such as SDS, harvest disk resources across servers within a compute infrastructure, and can often include external dedicated storage as well. The DAS on the compute nodes is presented to the compute infrastructure as a pool of storage resources, enabling any to any connectivity for storage traffic rather than



compartmentalized storage infrastructures as in a traditional i.e. non-scale-out NAS or iSCSI environment. As some of these systems employ the concept of locality so as to minimize traffic across the network, a new layer of east-west traffic needs to be considered, namely, that within an individual cabinet.

EAST/WEST SCALABILITY

At the very core, storage is simply data. The more data that can be moved from host to target, the more efficiently the applications will run on the hosts. This data often traverses multiple links, and many networks are not designed to provide a non-blocking solution with deterministic latency. Traditional three-tier designs can be massively oversubscribed in the data center because they were designed primarily for north to south traffic.

Storage data I/O must be resilient and ensure reliable transmission to prevent loss or corruption. Packet loss is undesirable and can dramatically affect performance of storage networks. Handling large volumes of east/west traffic with minimal packet loss is table stakes for delivering a high performance storage network. Given the issues which can arise from packet loss in storage networks, Arista recently performed a set of tests which measured the number of TCP retransmits in an oversubscribed east-west network with deep buffers vs. one with shallow buffers. To validate the importance of deep buffers at both the leaf and spine layers, one leaf switch is an Arista 7280SE switch which has 128MB of buffers per 10G port, while the other is the relatively shallow buffered Arista 7050 with a shared buffer of 9MB for all ports. At the spine layer, an Arista 7504 is used configured with 48 MB of buffer per 10G port. The topology is shown in figure 2.

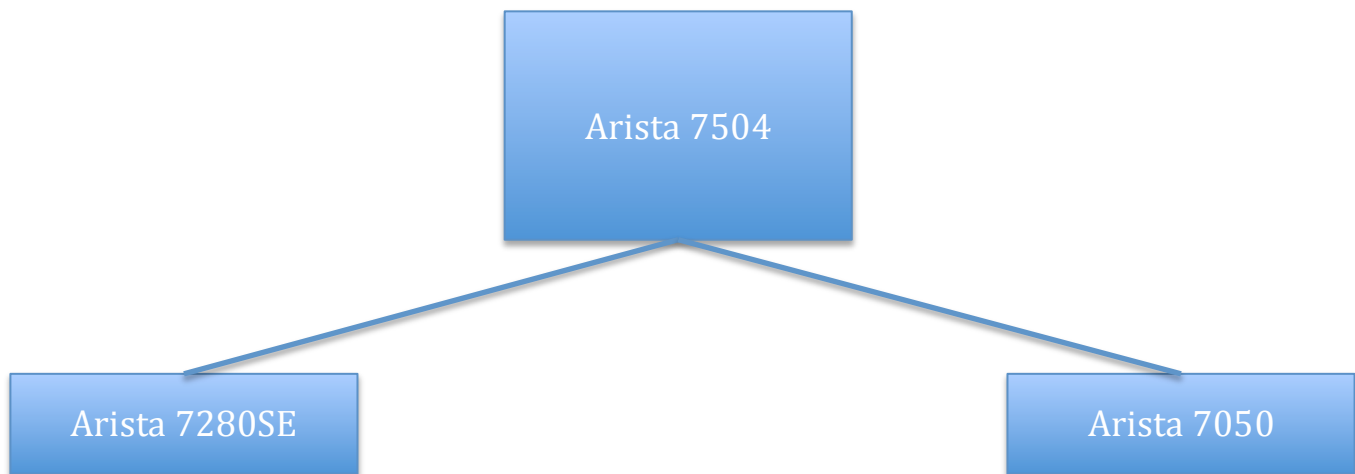


Figure 2: Simple Leaf/Spine Network

To simulate shallow buffering across the board, the 7280SE and 7504 were manually configured so as to have 1 MB/10G port of buffering available. The deep-buffered network was oversubscribed at a ratio of 5.33:1, while the shallow-buffered network saw a lower oversubscription ratio of only 4:1. Even the shallow buffered switch setup used in this test uses deeper buffers than most datacenter switches on the market today, which typically use dynamic buffering against a shared pool of 10-15MB for the entire switch.

Running traffic at maximum capacity, TCP retransmits were measured. Looking at the retransmits in this simple 2 leaf east/west network, even with the very conservative approaches to buffer sizing, the results are striking. Figure 3 shows the deep buffered system.

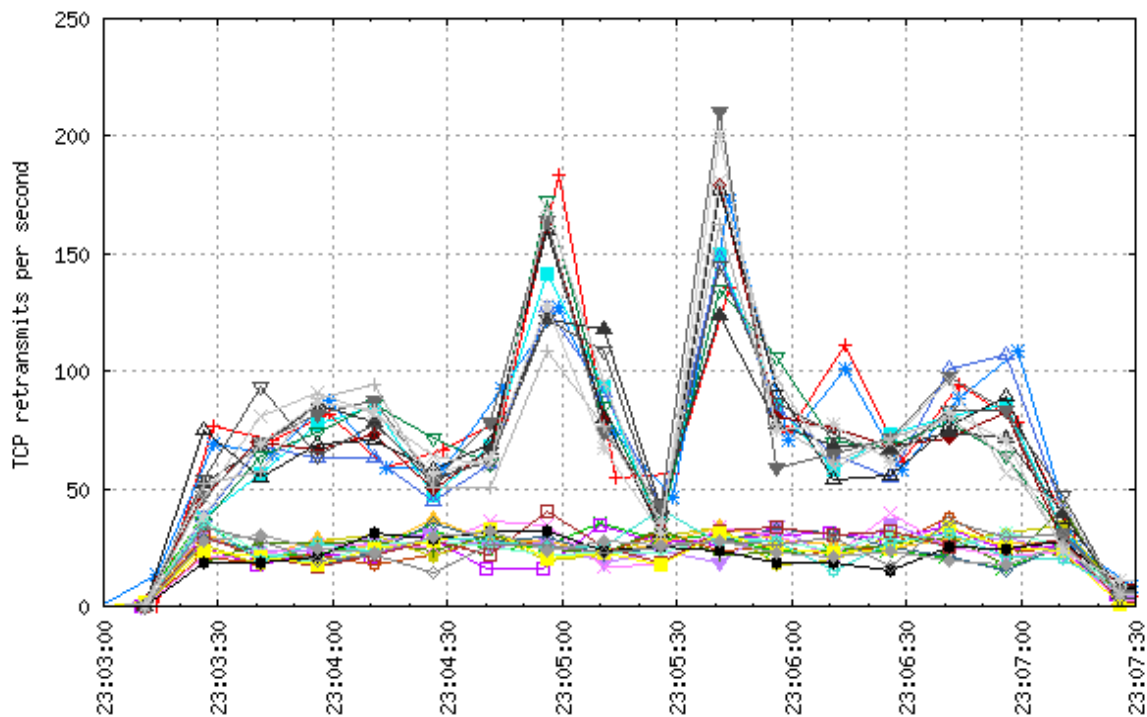


Figure 3: TCP retransmits in 5.3:1 oversubscribed network with deep buffers

While retransmits are seen in the deep buffered network, as expected given the level of oversubscription, the hosts connected to the 7280SE leaf switch have below 50 retransmits per second consistently. Hosts connected to the shallow buffered 7050 switch experience higher levels of retransmits across the board with spikes greater than 200 retransmits per second. This demonstrates the benefit of large buffers at the leaf to handle large amounts of east/west traffic.

How does buffering in the spine affect performance? When simulating a shallow buffered system at the spine, and re-running the test and measuring the retransmits, a different story is apparent. In figure 4, we see the results on the shallow-buffered end-to-end network.

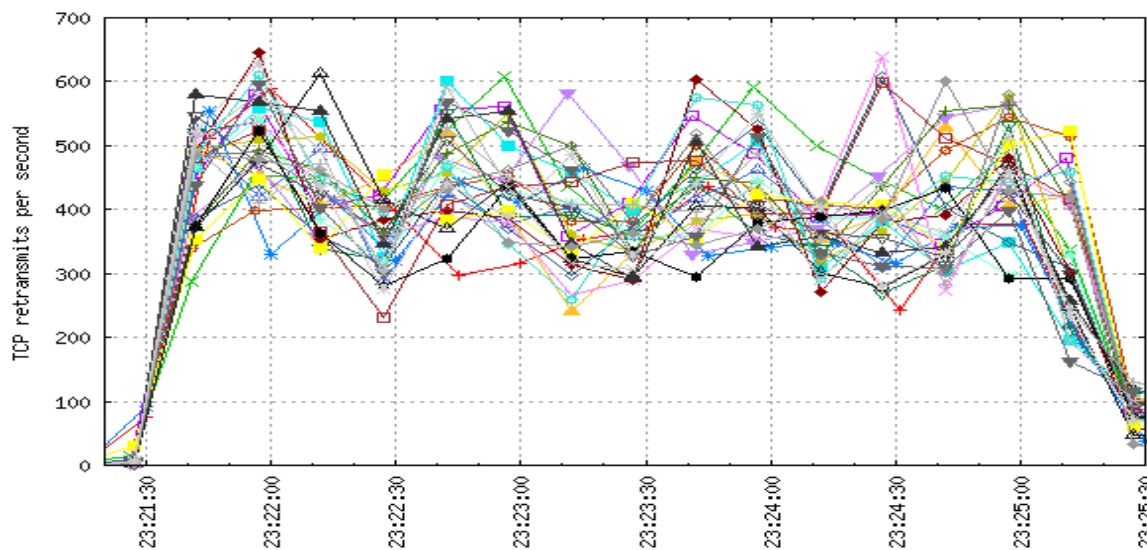


Figure 4: TCP retransmits in a 4:1 oversubscribed network with shallow buffers

As can be seen in Figure 4, the retransmits increase dramatically when lowering the buffer size to a level far greater than many datacenter switches can support today. This data shows that supporting high rates of east-west traffic requires large buffers at both the leaf and spine layers.

TCP INCAST

TCP incast is a many-to-one communication problem that can occur in data networks. As distributed storage frameworks have been deployed, a single host request for data can result in responses from many storage nodes simultaneously. This response oversaturates the connection to the host. The move toward virtualization and maximizing efficiency has driven storage devices to service many hundreds or even thousands of hosts. When many hosts access the same storage device simultaneously, a massive incast condition happens at the storage device and on the intermediate network links. As the number of hosts and leaf switches in the network increases, the likelihood of incast does as well, both on the leaf->storage links as well as the spine->leaf links.

Similarly, if a host attempts to access data that is stored across multiple nodes as is seen in many SDS solutions, the storage nodes will respond simultaneously to maximize parallelism and cause an incast pattern towards the requestor. At the end of the day, regardless of direction, in an incast communication pattern multiple senders concurrently transmit a large amount of data to a single receiver (host, switch, etc). The data from all senders traverses a bottleneck link in a many-to-one fashion. This condition often results in a microburst that overwhelms a single host or switch port. Traffic is dropped, which causes TCP timeouts and retransmits which ultimately degrades application performance.

Incast is a prevalent issue in IP storage deployments. To deal with incast, buffering is the most powerful tool in a network architect's toolbelt. Ability to buffer large bursts of data neutralizes the effect of a microburst by queuing the traffic above the link bandwidth rather than dropping it. Arista has the only solutions in the industry capable of providing the necessary buffers to deal with TCP incast at both the spine and leaf layers of the network.

STORAGE AND COMPUTE DEVICES INTERCONNECTING AT DIFFERENT SPEEDS

Another area where bottlenecks can occur is in networks where devices are connected at different speeds; Today many networks are a mixture of 1Gbps, 10Gbps, 40Gbps, and 100Gbps. This speed mismatch can cause a condition where a server is connected at 1Gbps but the storage system is connected at 10Gbps. In this scenario, a single request is sent at 1Gbps, but the responses come back at 10Gbps and must be serialized to 1Gbps. Even in situations where both hosts at the end of a connection are connected at the same speed can experience this. This massive speed mismatch can easily lead to buffer exhaustion in a switch, especially at the leaf layer where speed differentials frequently occur. This results in dropped packets, which degrades the performance of applications. To ensure packets are delivered, deep buffered switches like the Arista 7500E and 7280SE are required where speed differentials occur in the network.

ARISTA IP/ETHERNET STORAGE NETWORK SOLUTIONS

Arista Networks is the leader in building scalable high-performance data center networks. To meet the challenges of transporting storage on an IP/Ethernet network, the Arista solution includes these features and technology to assist with the IP storage challenges, as shown in Figure 5:

- Massively scalable Leaf/Spine non-blocking topology
- Deep buffering technology and lossless delivery
- Operational visibility into the network to provide an early warning of congestion

MASSIVELY SCALABLE LEAF/SPINE TOPOLOGY

Arista provides industry-leading solutions that offer a superior price to performance ratio when designing networks to support IP/Ethernet storage. The Arista solution ensures effective use of all available bandwidth in a

non-blocking mode and provides failover and resiliency when any individual chassis or port has an outage condition. Multi-chassis Link Aggregation (MLAG) and Equal Cost Multipath Routing (ECMP) provide standards based, non-proprietary, multipath technologies at Layer 2 and Layer 3. These technologies currently scale linearly to more than 50,000 compute and storage nodes, both physical and virtual.

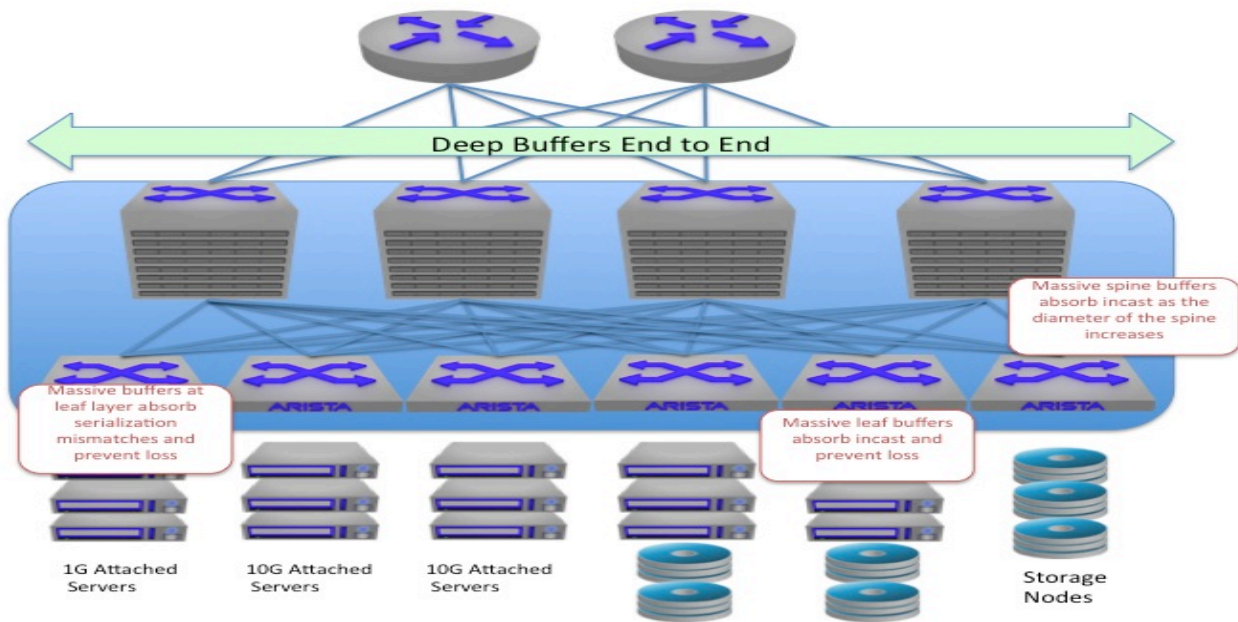


Figure 5: Arista Deep Buffered IP Storage Network Design

Next-generation multicore server CPUs with dense virtual machines (VMs) and storage make the use of this type of uncompromised Leaf/Spine topology critical. In addition uplink, downlink, and peer ports that are not oversubscribed and are all switched at wire speed are very important.

The Arista Leaf/Spine topology effectively solves the IP/Ethernet storage challenge of massive east to west traffic by delivering a solution with deterministic latency characteristics, any-to-any non-blocking host communication, and deep buffers capable of absorbing the largest of bursts and TCP incast traffic patterns.

BUFFERING TECHNOLOGY AND LOSSLESS DELIVERY

Switch port buffering is a very important consideration as mentioned previously. The Arista 7048, 7280, and 7500E switches are designed to appropriately allocate bandwidth to all traffic flows in the data center. The Arista 7048 switches have 768MB of packet buffer memory, which allows the architecture to scale in congestion scenarios. All ports can simultaneously buffer up to 50ms of traffic without any drops. The deep buffer architecture of the Arista 7048 switch also provides a platform for lossless asymmetric connections when 10GbE nodes, such as storage devices, communicate with 1GbE nodes, such as servers, in the data center.



Figure 6: Arista 7280E Series Switches

For the 10G leaf node, Arista has introduced the 7280SE series (Figure 6) switch that provides unparalleled performance for IP/Ethernet storage networks. Built on the same technology as the industry leading 7500E chassis products, the 7280SE delivers an unprecedented 9GB of buffer capacity in a 1U form factor. The 7280SE switches employ a Virtual Output Queue (VoQ) architecture to ensure fairness and lossless port-to-port forwarding in the most extreme conditions. The 7280SE provides industry best solutions dealing with massive intra-cabinet IP storage flows, incast towards the spine, and ability to support emerging SDS architectures. Beyond the VoQ architecture and deep buffering, the 7280SE is the world's first 1U switch that supports 100G Ethernet links to the spine. Simply put, no other 1U switch can compare for enabling scale-out storage architectures.

Buffering at the spine is also a key consideration. Table 1 contains real world buffer utilization information from spine switches on a per-port basis across multiple market segments. Without sufficient buffering in the spine, drops will be experienced.

Table 1: Real World Spine Switch Buffer Utilization

User Segment	Environment	Max Buffer Utilization Observed
Software Vendor	Engineering Build Servers	14.9 MB
Oil & Gas	Storage Cluster – Medium	33.0 MB
Online Shopping	Big Data – Hadoop 2K Servers	52.3 MB
HPC – Research	HPC Supercomputer Cluster	52.4 MB
Animation	Storage Filer (NFS)	74.0 MB

A quick note on buffer bloat; this is a term that has been used for close to 20 years which indicates a condition whereby the amount of buffers in a device can cause additional latency and jitter within the network that is ultimately detrimental to application performance. While it's undeniable that buffering induces latency, the amount of latency introduced by 50MB of buffering on a 10G interface is equivalent to 40ms. For storage applications, insertion of a 40ms delay in extreme conditions is vastly preferable to dropping packets and going into TCP slow start. QoS mechanisms can easily be deployed which ensure that latency sensitive applications are transmitted ahead of non-latency sensitive applications to ensure that no adverse effects are encountered in even the deepest buffered switch.

DCB AND PFC

In addition to well-engineered buffering technology, Arista supports other mechanisms that help to ensure the lossless delivery for IP/Ethernet storage by utilizing the IEEE standardized set of features called Data Center Bridging (DCB). DCB works to make Ethernet a lossless medium and thus is well suited for storage networks that run directly over Ethernet such as FCoE or AoE. Protocols that rely on the network transport to provide the guaranteed in-order delivery native to TCP based protocols.

DCB has two important features: Data Center Bridging Exchange (DCBx) and Priority Flow Control (PFC). These features enable Arista switches to carry storage data more reliably. DCBx is an exchange protocol that capable

switches can use to exchange data center bridging capabilities to ensure consistent configuration and support for lossless transport.

PFC enables switches to implement flow-control measures for multiple classes of traffic. Switches and edge devices slow down traffic that causes congestion and allow other traffic on the same port to pass without restriction. Arista switches can drop less important traffic and tell other switches to pause specific traffic classes so that critical data is not dropped. This Quality of Service (QoS) capability eases congestion by ensuring that critical storage I/O is not disrupted or corrupted and that other non-storage traffic that is tolerant of loss can be dropped.

It is important to note however, that for protocols such as iSCSI, the underlying TCP transport can render the lossless, in-order deliver characteristics provided by DCB redundant. TCP's innate congestion control mechanisms provide a per-flow lossless delivery of traffic, which adjusts to network congestion on a per-flow basis. When used in conjunction with DCB, the lossless nature of DCB can mask network issues and the TCP flows riding the connection will not naturally back off. This can cause impact across multiple flows and applications, or even entire network segments, which can all be blocked while the storage traffic takes priority.

The architecture of a switch must be taken into consideration when deciding whether or not to deploy PFC. While a VoQ architecture such as those present in Arista's 7500E or 7280SE switches can operate well in a PFC enabled infrastructure, switches which employ an internal Clos architecture for a fabric are not well suited to a PFC enabled deployment. The issue is a direct function of leveraging "switch on a chip" architectures to create a multi-stage fabric within a chassis. Consider the implications of congestion: When such a switch experiences congestion on a PFC enabled port, the port ASIC must generate a pause frame towards its upstream neighbor. In the case of an internal Clos fabric, this neighbor can quite likely be a fabric chip. This causes all traffic in the lossless class on the fabric chip to be paused. As the fabric is paused, being a natural convergence point for flows, the fabric buffers will fill up very rapidly causing drops. At this point, the fabric, experiencing congestion will send a pause to all connected interfaces, which typically translates to all ingress ports. The net effect is a PFC initiated head of line blocking across all ports on the switch. For this reason PFC is not recommended on a Clos fabric based switch.

Arista recommends that the ramifications of TCP over DCB be thoroughly evaluated, and the proposed design lab tested to validate intended behavior prior to deployment. To simplify the architectural decision process, eliminate switches without deep buffers from consideration for an IP storage network. If PFC is required, ensure that the switch does not employ an internal Clos architecture, instead, choose switches with either a VoQ architecture such as the Arista 7500E or 7280SE, or which leverage fabric-less architectures distributed among multiple switches to avoid a cascading pause.

ARISTA LANZ PROVIDES OPERATIONAL VISIBILITY TO REDUCE CONGESTION

IP/Ethernet storage network environments require deterministic characteristics with regard to throughput, performance, and latency. These characteristics are theoretical only if the network infrastructure cannot provide real time monitoring of changing traffic patterns and their effect on the design characteristics.

Traditional tools that are used to monitor network traffic patterns, such as Remote Monitoring (RMON) and Simple Network Management Protocol (SNMP), are based on a polling model where data typically is collected at one second or longer intervals. An increasing number of data center hosts are connected at 10GbE. Within a one-second interval, a 10GbE interface can go from idle to forwarding over 28 million packets and back to idle again. In a traditional polling model, this 28 million-packet burst can become invisible.

The Arista Latency Analyzer (LANZ) is a pro-active, event-driven solution that is designed to provide real-time visibility of congestion hot spots and their effect on application performance and latency at a nanosecond

resolution. The level of granularity that LANZ delivers is made possible by its unique event-driven architecture. The traditional polling model provides visibility only at discrete intervals. LANZ reports on congestion events as they occur. LANZ provides visibility into the network where IP/Ethernet storage is attached and it can be used to ensure that necessary interconnects are available to ensure a lossless transport.

OTHER ARISTA OPERATIONAL ADVANTAGES

In addition to the features already described, Arista solutions also provide many other operational advantages, including:

- Industry-leading power efficiency per 10GbE port at less than 3.5W
- Single binary image across all platforms
- Open APIs for third-party integration
- Advanced Event Management for customizable actions based upon specific events
- Zero-touch provisioning (ZTP) for rapid deployment and expansion of storage clusters
- VM Tracer for dynamic and automated provisioning of VLANs in virtualized environments
- Hardware virtual extensible LAN (VXLAN) support to allow for seamless integration with virtualized compute infrastructure

Visit www.arista.com more information on each of these features and functionality.

REFERENCE ARCHITECTURES

When choosing an architecture for an IP/Ethernet storage network, consideration of the goals of the project must be taken into account when selecting the design. For example, if cost and space saving are of paramount importance, minimization of network elements follows naturally and the approach is typically to deploy a collapsed or shared network infrastructure. Likewise if performance, maintenance of fully segmented networks, and data protection and data segmentation are key, then an IP/Ethernet SAN is the likely architecture choice. While both collapsed and dedicated approaches are sometimes taken, most often a hybrid or semi-collapsed approach is used. Dedicated leaf switches and NICs connect storage for mission critical applications to the spine while general purpose applications leverage a shared infrastructure.

In figure 7, a collapsed IP storage infrastructure is shown. A collapsed architecture represents the simplest topology because all traffic traverses the normal leaf/spine architecture deployed in the datacenter. Leveraging existing resources, collapsed IP Storage networks provide the most cost effective alternative for deploying IP storage, but in this design, care must be taken to ensure the proper buffering is available to handle the load without causing loss among all traffic in the network. To this end, Arista 7500E spine switches and 7280SE leaf switches are prescribed as they offer the most scalable highest performing options in the industry today.

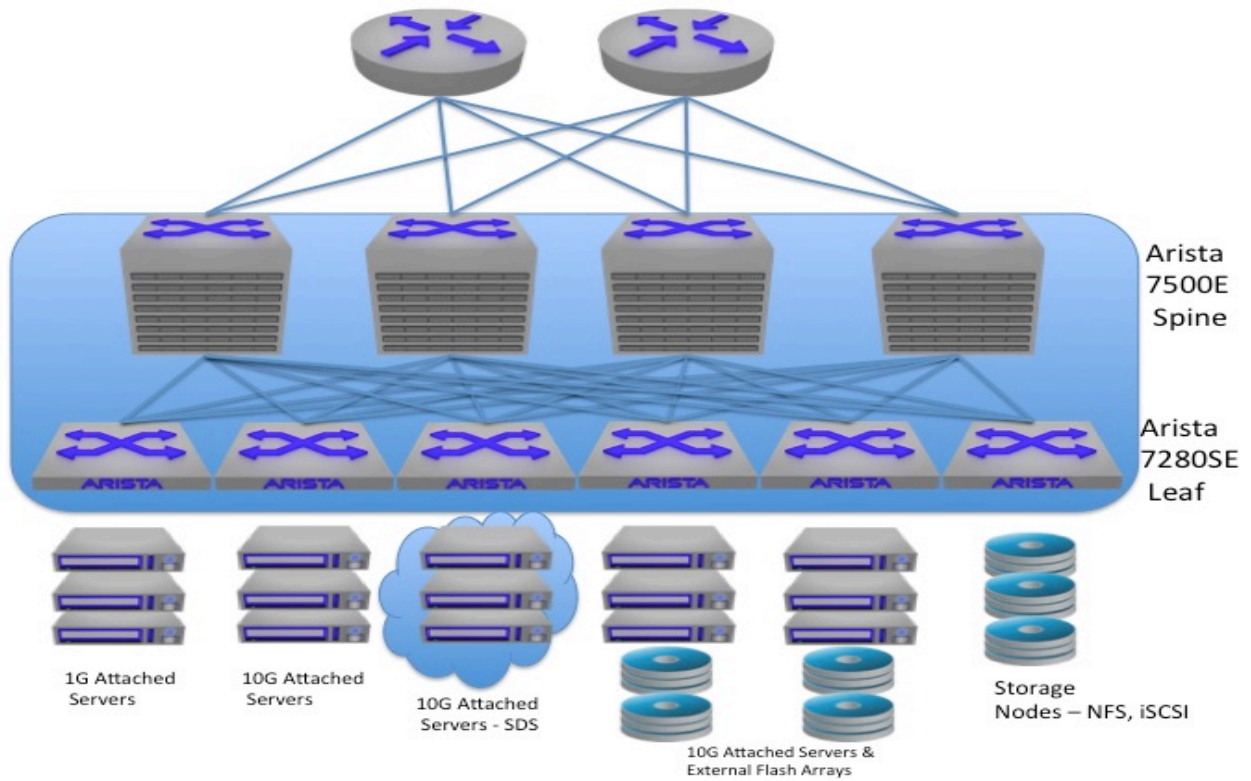


Figure 7: Collapsed IP Storage Infrastructure

Figure 8 shows a semi-collapsed architecture. In this design, a single spine is used for all traffic and dedicated leaf switches are used when the application’s criticality dictates a dedicated infrastructure. For standard network traffic, high performance, low latency switches such as the Arista 7150S can be used, while at the spine and the storage leaf layers, Arista 7500E and 7280SE switches are the best fit due to their unique deep buffer architecture.

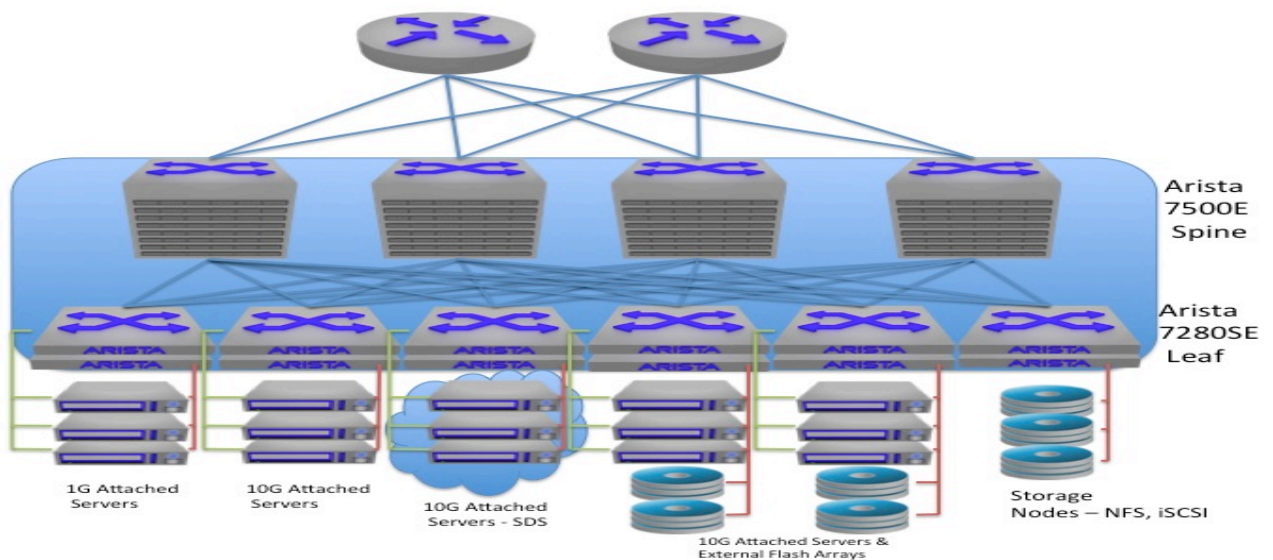


Figure 8: Semi-collapsed IP Storage network

Since the critical applications requiring dedicated storage infrastructure will likely be fewer than those which have less demanding needs, many organizations, rather than choosing dedicated storage leaf switches may choose to deploy a single tier “spline” design. For the same reasons as have been stated previously, when deploying storage, buffers matter in the minimization of packet loss and retransmission. To this end, the 7500E is the go-to platform for an IP storage single tier “spline” network such as that shown in Figure 9.

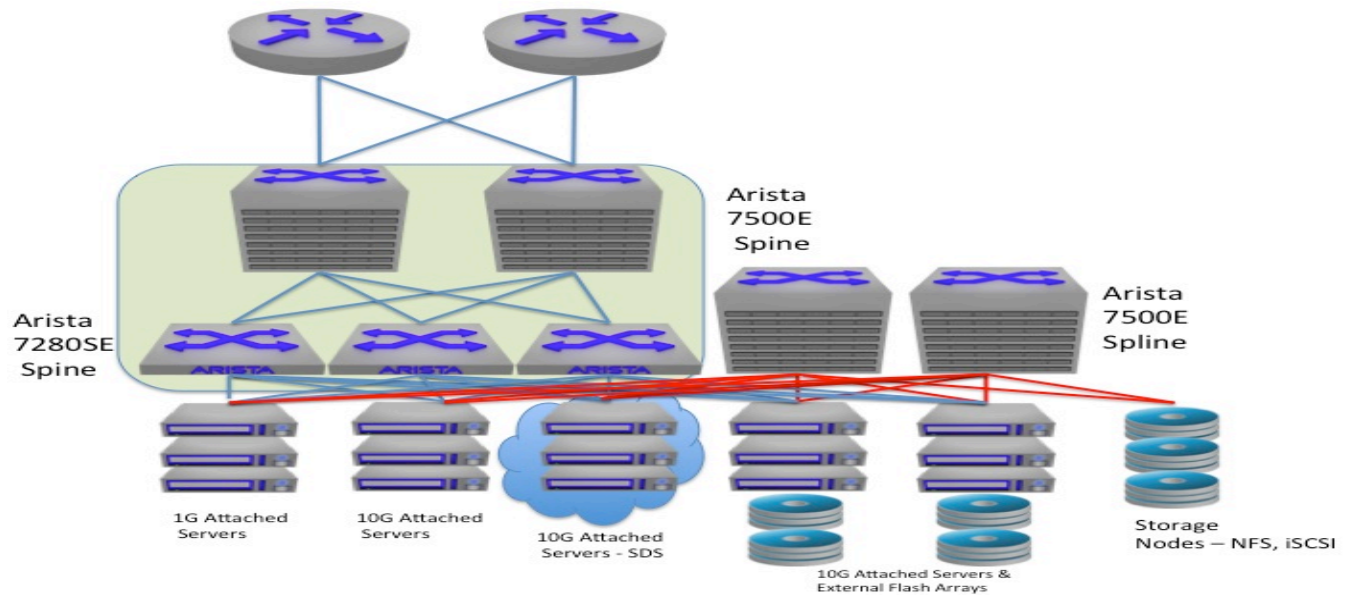


Figure 9: Semi-collapsed IP Storage “Spline”

Finally, for those organizations which require absolutely dedicated networks, this can also be achieved using an IP/Ethernet infrastructure. Less common than most other deployment options, the dedicated network can provide equivalent performance and reliability to a traditional fibre channel SAN with drastically reduced complexity, and minimized cost. An example of this dedicated network design can be found in Figure 10 and utilizes 7500E for the spine and 7280 for the storage leaf layer. The compute leaf switches can be other high performance switches such as the Arista 7150S.

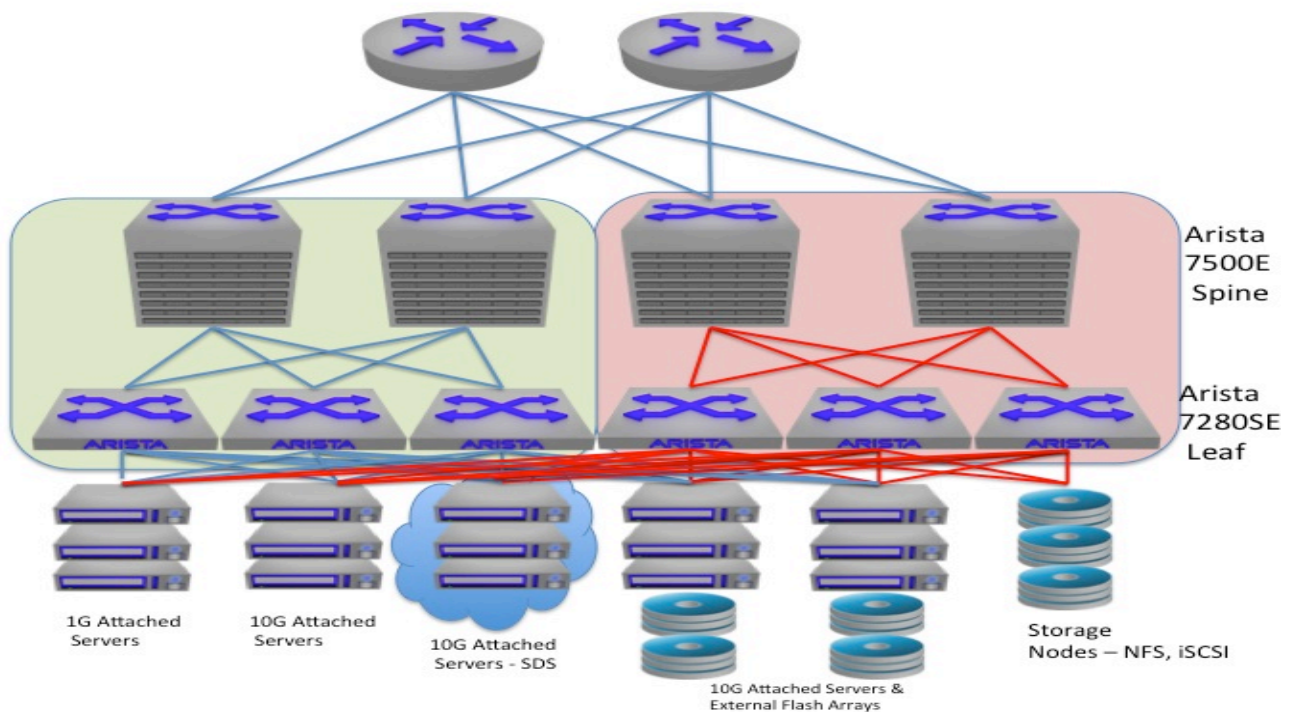


Figure 10: Dedicated IP/Ethernet Storage Network

MIGRATION

Migration of storage from legacy architectures to scale-out IP/Ethernet architectures is a process which needs to be undertaken as a joint effort between the storage teams and network teams so that the expertise of all personnel can best be used. Many storage vendors provide migration solutions native to their offerings. A cross team effort involving the storage team and network team should be undertaken so that the particulars of the storage protocol can be understood from the perspective of loss tolerance, latency sensitivity, and bandwidth utilization. Leveraging LANZ data to provide evidence of congestion events in the network allows the network architect to properly prepare for the real world requirements of a given system.

CONCLUSION

Arista provides the best solution for IP/Ethernet storage networks. Exponential growth in storage needs requires an efficient transport like Ethernet to overcome the expense, interoperability, administrative complexity, and scale-out challenges of traditional approaches. Building an IP/Ethernet storage architecture with high-performance Arista switches maximizes application performance. Arista switches deliver a deep buffers to address TCP incast, operational flexibility and extensibility, and a highly resilient, available and cost effective network to meet the demands of data growth and next-generation storage networks.

APPENDIX A – STORAGE TECHNOLOGIES OVERVIEW

STORAGE AREA NETWORKS

A storage area network (SAN) is an architecture whereby servers access remote disk blocks across a dedicated interconnect. Most SANs use the SCSI protocol to communicate between the servers and the disks. Various interconnect technologies can be used, and each of them requires a specific SCSI mapping protocol.

A SAN is a specialized network that enables fast, reliable access among servers and external or independent storage resources. A SAN is the answer to the increasing amount of data that must be stored in an enterprise network environment. By implementing a SAN, users can offload storage traffic from daily network operations while establishing a direct connection between storage elements and servers. SAN interconnects tie storage interfaces together into many network configurations and across large distances. Interconnects also link SAN interfaces to SAN fabrics.

Building a SAN requires network technologies with high scalability, performance, and reliability to combine the robustness and speed of a traditional storage environment with the connectivity of a network.

FIBRE CHANNEL PROTOCOL

Today, many SANs use the Fibre Channel Protocol (FCP) to map SCSI over a dedicated Fibre Channel. Enterprises that deploy Fibre Channel deploy multiple networks including the LAN network and the dedicated Fibre Channel network. Typically the LAN uses Ethernet technology, which is a basic component of 85% of all networks worldwide, and is one of the most ubiquitous network protocols in existence. Traditional Fibre Channel deployments require the operation of (minimally) two separate network infrastructures, one for data, one for storage.

iSCSI

Internet Small Computer Systems Interface (iSCSI) has gained traction and attention as data centers strive to reduce costs for robust storage. iSCSI rides on IP/Ethernet transport, which alleviates the complexity of a separate traditional Fiber Channel SAN. iSCSI is an ideal solution for many small and medium enterprise organizations. iSCSI relies on TCP/IP protocols, which makes it a natural communication for private and public cloud communications.

The performance advantages of iSCSI are compelling. Storage arrays must keep up with new multicore processors and stack software that can generate a million iSCSI I/O operations per second (IOPS) when connected to a non-blocking storage access switch from Arista. Most modern storage arrays come with iSCSI NICs which are 10Gb Ethernet NICs which employ TCP offload engines for more efficient TCP processing.

NETWORK ATTACHED STORAGE

A longtime popular method for consolidating file based storage. NAS systems differ from a SAN in that instead of block based storage access, file based access is provided. In a NAS, the filesystem is laid out and managed on the NAS itself with remote hosts mounting the system at the directory level typically using NFS or CIFS, depending upon the operating system. While generally lower performance than a block based storage system, NAS is usually good enough for most storage needs. NAS is usually consolidated, lending itself to large probability of incast as multiple hosts across a wide diameter network attempt to write to the file system simultaneously. It is typically more resilient than block storage however when packet loss is encountered.

SOFTWARE DEFINED STORAGE

Software Defined Storage (SDS) is quickly becoming a popular option for next generation storage architectures. As the technology to enable SDS has matured, so has interest. An SDS seeks to reclaim direct attached storage (DAS) which would otherwise be redundant in a system using an external SAN. SDS software creates a pool of resources among the DAS found in an SDS cluster to provide a linearly scalable storage architecture, which does not depend at all on dedicated infrastructure. Because of the massive efficiencies achievable by deploying an SDS architecture,

there are many options to examine in terms of technology vendors.

Network architects should gain a deep understanding of the protocols utilized by these systems as they are often proprietary, and the traffic patterns in such a system do not always follow traditional storage patterns of congestion based on multiple devices fanning into few consolidated network links, or based on differential serialization delays. This said, incast is as likely if not more likely to be experienced when deploying SDS, as the systems rely on massive parallelism to increase IOPs. Such parallelism often results in massive short lived and frequent incasts as hosts requests blocks of data from multiple hosts simultaneously, all of which reply at line rate at the same time.

ARISTA

Santa Clara—Corporate Headquarters
5453 Great America Parkway
Santa Clara, CA 95054
Tel: 408-547-5500
www.arista.com

Ireland—International Headquarters
4325 Atlantic Avenue
Westpark Business Campus
Shannon
Co. Clare, Ireland

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Copyright © 2014 Arista Networks, Inc. All rights reserved. CloudVision, and EOS are registered trademarks and Arista Networks is a trademark of Arista Networks, Inc. All other company names are trademarks of their respective holders. Information in this document is subject to change without notice. Certain features may not yet be available. Arista Networks, Inc. assumes no responsibility for any errors that may appear in this document. 08/14