# Table of contents

Overview........................................................................................................................................... 3  
Introduction....................................................................................................................................... 3  
Storage node ....................................................................................................................................... 4  
  10GbE connectivity .............................................................................................................................. 6  
  Interface speed and duplex settings ................................................................................................... 6  
  Flow control ...................................................................................................................................... 6  
  Jumbo frames .................................................................................................................................... 7  
  Network bonding ............................................................................................................................... 7  
  Separate management networks ......................................................................................................... 10  
  IP addressing .................................................................................................................................. 12  
Network switching ............................................................................................................................... 12  
  Port buffering cache ......................................................................................................................... 12  
  Switching methodology .................................................................................................................... 13  
  Flow control .................................................................................................................................... 13  
  Jumbo frames .................................................................................................................................... 13  
  VLAN ................................................................................................................................................ 14  
  Spanning Tree Protocol ..................................................................................................................... 14  
  Unicast storm control ....................................................................................................................... 15  
  Recommended network switches for HP StoreVirtual Storage......................................................... 15  
Host configuration ............................................................................................................................... 18  
  Jumbo frames .................................................................................................................................... 18  
  Flow control .................................................................................................................................... 18  
  Multipathing I/O................................................................................................................................ 19  
  MPIO for VMware vSphere ................................................................................................................ 19  
  MPIO for Microsoft Windows Server .................................................................................................. 19  
  Microsoft Windows Server 2012 TCP Software and Hardware NIC settings.................................... 20
Overview

This paper provides network design recommendations for deploying HP StoreVirtual Storage. Design practices described within this paper serve as a blueprint to guide network configurations for the current and previous generations of HP StoreVirtual Storage, HP P4000 G2 SAN Solutions, and HP StoreVirtual 4000 Storage and will be referred to that way in this paper.

As shown in Figure 1. Architecture Overview, the principle components of an HP StoreVirtual installation are:

- **Target**—Storage systems and storage clusters
- **Network**—Networking infrastructure
- **Initiators**—Servers and virtual machines accessing shared storage

Guidelines provided in this paper have been tested and refined over time to guide network planning, architect design, and implementation of HP StoreVirtual Storage. Network considerations included in this document take into account standard and multi-site clusters as well as typical disaster recovery (DR) configurations with Remote Copy in the context of networking. This document serves as a recommendation guide and should not be considered an implementation guide for all customer deployments of HP StoreVirtual.

Readers of this white paper have basic experience designing and deploying network infrastructures for iSCSI storage, and are knowledgeable of the OSI model in general, and Layers 2 and 3 networking in particular.

Figure 1. Architecture Overview

---

Introduction

HP StoreVirtual leverages the industry-standard iSCSI protocol over Ethernet to provide block-based storage to application servers. HP StoreVirtual uses a distributed scale-out architecture and clusters two or more storage nodes into a flexible pool of storage, a storage cluster. The iSCSI protocol enables the block level data transfers over Ethernet between an initiator (application server) and a target (storage system).

From a networking perspective, HP StoreVirtual Storage provides the following key advantages:

- Adheres to standard Ethernet and IP network design principles
- Leverages standard Ethernet switching and routing infrastructure
- Supports any Ethernet infrastructure
- Supports synchronous replication of volumes between two or three data centers (Network RAID in Multi-site Clusters)
- Supports asynchronous replication of volumes (Remote Copy) over long distances using IP networks
- Simplifies management through a single pane of glass, the Centralized Management Console
Leveraging IP networking and the iSCSI protocol, HP StoreVirtual provides a scalable, flexible, and resilient architecture—allowing you to expand storage capacity incrementally and online by attaching additional storage nodes to existing clusters on the network.

This paper will look at the networking aspects of the various components of an environment with HP StoreVirtual Storage (see Figure 1. Architectural Overview). As a rule of thumb, the network for HP StoreVirtual should provide a high level of resiliency, high bandwidth, and low latency. Typically, a dedicated infrastructure for HP StoreVirtual is considered best practice.

Figure 2. Storage and networking technologies for increased resiliency

The technologies shown in Figure 2, Storage and networking technologies for increased resiliency, are the common practices in HP StoreVirtual installations. While network level technologies such as trunking and bonding are used between switches and on the storage nodes, multipathing drivers on the application servers are used to manage additional resiliency of redundant paths. This paper will provide pointers on how to configure these technologies.

Storage node

Network interfaces on HP StoreVirtual Storage are used for communication between storage nodes in a cluster for replication of data, between clusters for asynchronous replication remote snapshots, management traffic as well as storage access via iSCSI. Depending on which HP StoreVirtual platform is deployed, there are various connectivity options available:

- Rack-mount HP StoreVirtual Storage models come with four 1GbE and some come with two additional 10GbE interfaces. HP StoreVirtual 4330, HP StoreVirtual 4530, HP StoreVirtual 4730 Storage systems come with four 1GbE and two 10GbE interfaces.
- Previous generations came with two 1GbE ports with optional upgrades for an additional two 10GbE ports (HP LeftHand P4000; HP P4000 G2 Storage Systems).
- HP StoreVirtual VSA comes with one or more virtual network interfaces.
HP StoreVirtual 4330 Storage and HP P4000 G2 Storage Systems may be upgraded with two additional 10GbE ports. Fibre Channel-capable storage nodes, such as the HP StoreVirtual 4330 FC Storage, come with two 10GbE ports and four 1GbE ports.

Storage nodes require minimal user configuration to bring them online as available systems within the Centralized Management Console except for cabling and IP addressing. Our guidance for configuring network adapter settings such as network bonding, jumbo frames, flow control, interface duplex, and speed are described in the sections to follow. IP addressing and network interface settings on storage nodes can be configured using the Centralized Management Console or through the LeftHand OS (LHOS) Configuration Interface on the local console. Note that bonding may only be configured in the CMC. HP iLO provides an Integrated Remote Console for remote administration so you can control your server from any location through HP iLO’s web interface. This interface provides access to the remote computer’s keyboard and video. For the StoreVirtual VSA 2014, the hypervisor’s management utilities provide access to this interface.

**Figure 3.** Network settings on storage node in the Centralized Management Console

![Network settings on storage node in the Centralized Management Console](image)

Figure 3, Network settings on storage node in the Centralized Management Console, shows the speed, duplex, and frame size settings in the Centralized Management Console for an interface which is not part of a bond. It is recommended to configure network settings on each storage node before joining these nodes to a new or existing management group. When configuring network settings on a storage node after it has been added to the management group and cluster, the storage node may become temporarily unavailable. All network configuration changes should be carefully planned; specifically if quorum or data availability depends on a single storage node. For more information on quorum in management groups, refer to the HP StoreVirtual User Guide.
10GbE connectivity

10GbE network interfaces provide higher bandwidth than 1GbE adapters. To use 10GbE interfaces on HP StoreVirtual, the LeftHand OS interface needs to be configured to enable the 10GbE interface on the storage node with best practices recommending bonding. All storage nodes in the same HP StoreVirtual cluster should have the same connectivity to the network; i.e., mixing 10GbE and 1GbE interfaces as the primary LeftHand OS interface on the same cluster is not considered a best practice configuration.

10GbE connectivity for iSCSI on HP StoreVirtual Storage makes sense for large sequential workloads, like streaming, video editing, and large file operations (network-bound). Essentially, all workloads, which require high sequential throughput per volume, benefit from 10GbE. In other cases where high number of I/O operations are required (disk-bound), 1GbE is typically sufficient.

HP StoreVirtual Storage 10GbE upgrade options include 10GbE SFP+ SR optics, which work with standard OM3 cabling. Alternatively, direct-attach cables (DAC) may be used to connect HP StoreVirtual Storage to the storage network.

Interface speed and duplex settings

The default interface speed and duplex mode for 1GbE interfaces on StoreVirtual is set to auto for both the speed and mode of duplexing on storage systems. With auto/auto set, the speed and duplex setting are automatically negotiated between interfaces on the storage node and switch ports. This setting should remain unchanged in most environments. If auto-negotiation is not resulting in 1GbE on the storage nodes, this may be an indication that the switch ports are not configured correctly or that the cabling is faulty. The default interface speed and duplex mode on HP StoreVirtual Storage with 10GbE SFP+ network adapters are hard set to 10GbE and full duplex. HP StoreVirtual Storage with 10BASE-T connectors will auto-negotiate with 1GbE switches.

Best practices

- Set interfaces on HP StoreVirtual Storage and switch ports to auto-negotiate, thus allowing the devices to auto-negotiate speed and duplex settings
- Verify that the speed and duplex of storage interfaces and switch ports match

Flow control

Flow control allows the receiver to instruct the sender to pause transmission of Ethernet traffic when the receiver senses congestion or buffer cache pressure. The receiver accomplishes this by sending pause control frames to the sender, which causes the sender to cease transmission for a brief period. This allows the receiving network interface time to process buffered cache. To take advantage of this performance increase, all network enabled devices in the iSCSI network path between the initiator and target including switches, routers, HBAs, and NICs must allow flow control enabled for receive (Rx) and transmit (Tx).

It is recommended to enable send and receive flow control on all edge ports handling iSCSI traffic. Some switches will allow to configure flow control for receive (Rx), transmit (Tx), or both (Rx/Tx).

IP storage networks are unique in the amount of sustained bandwidth that is required to maintain adequate performance levels under heavy workload. For HP StoreVirtual, flow control should be enabled on all appropriate servers, storage systems, and switch ports. The default flow control setting on storage systems’ network interfaces is auto-negotiate, which is recommended by HP.

Best practices

- Set flow-control setting to auto-negotiate on storage systems which is the default (see Figure 3).
- Enable flow control on all switch ports, which are used for iSCSI to achieve optimal performance.
- Set the desired flow control setting on each interface before you create the bond; modifying the flow control settings on bonded interfaces is not possible without deleting the bond first.
**Jumbo frames**

Configuring jumbo frames within a StoreVirtual installation should not be taken lightly and should be employed by storage and network administrators that are knowledgeable of storage I/O characteristics as well as jumbo frame support for Layers 2 and Layer 3 network devices within their network. A key feature of any storage network design is the ability to move data traffic as fast as possible incurring the least amount of latency. HP internal testing has shown an increase in performance in the order of 10–20 percent for large sequential read and write workloads when using jumbo frames with 9,000 bytes. To take advantage of this performance increase, all network enabled devices in the iSCSI network path between the initiator and target including switches, routers, HBAs, and NICs must have jumbo frames enabled with exactly the same frame size.

The default maximum transmission unit (MTU) in Ethernet is 1,500 bytes; the resulting frame with Ethernet header and preamble is 1,518 bytes. Any Ethernet frame larger than the default is considered a “jumbo frame.” The maximum allowed frame size supported by HP StoreVirtual Storage is 9,000 bytes. The setting in HP StoreVirtual is called “Frame Size,” see Figure 3.

The actual impact on storage throughput when using jumbo frames depends on application workload and I/O characteristics. To take advantage of jumbo frames and avoid problems, all devices on the same network segment (the path between the servers and the storage cluster), including the network switches, must have the same frame size setting (i.e. same MTU).

**Best practices**

- For large sequential workloads, jumbo frames can improve overall throughput.
- Cut-through switching is a method for packet switching systems, wherein the switch starts forwarding a frame (or packet) before the whole frame has been received, normally as soon as the destination address is processed. Engineering noted that cut-through switch enablement in some switch configurations improved performance even more. Additional Layer 2 and Layer 3 switch queuing mechanisms may also further improve overall system performance.
- When using jumbo frames, all devices on the same network segment need to have the same frame size settings (typically this means all switches, all storage nodes interfaces in the same management group, and all network adapters of application servers).
- Enable jumbo frames before adding storage nodes to a new or existing management group.

**Network bonding**

*Figure 4.* Creating a new interface bond in the Centralized Management Console
Designing best practice iSCSI solutions start with ensuring resilient network communication and traffic distribution. Storage nodes are designed to address network fault tolerance and bandwidth aggregation by bonding network interfaces. Depending on storage system hardware, network infrastructure design, and Ethernet switch capabilities, network interfaces can be configured to support one of three bonded techniques. As changing the default speed, duplex, and flow control settings of interfaces can only be performed on interfaces that are not bonded, these settings should be configured prior to bonding network interface.

Network interfaces are bonded by joining two or more physical interfaces into a single logical interface (a so-called bond) with one IP address. The logical interface utilizes two or more physical interfaces. A bond is primarily used to increase fault tolerance at the network level, thus tolerating the failure of a network adapter, cable, switch port, and or a single switch in some cases. Bandwidth aggregation is another reason to create bonds.

Bonds may be used for cluster and iSCSI connectivity, as well as management traffic; for example, on HP StoreVirtual 4330 Storage with four 1GbE ports: Two interfaces in one bond for inter-node cluster/iSCSI communication and two interfaces in another bond for CMC management operations on a different network segment.

Table 1 highlights the bonding techniques supported on HP StoreVirtual Storage. Ensure that the physical path mapping to switches are known to ease configuration of switch bonds. Alternatively, known MAC address to switch port mapping may be investigated with some managed switches by examining port ARP tables and matching the HP StoreVirtual port MAC addresses. Some switches support creating a single logical switch across multiple switches in a highly available IRF or trunk configured backplane, thereby extending link aggregation benefits across multiple switch configurations. Link aggregation (LACP; 802.3ad) requires configuration of the port group trunks on the physical switch to match the bonds on the storage.

<table>
<thead>
<tr>
<th>NIC bonding support</th>
<th>Description</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive load balancing (ALB)</td>
<td>Uses interfaces simultaneously for data transfer (receive and transmit load balancing). If one connection fails, the remaining connection continues without disruption to the storage nodes operation or iSCSI sessions. It offers the benefits of the increased transmission rates without additional switch configuration. Interfaces in an ALB bond may be connected to different switches and therefore supports switch failover.</td>
<td>Preferred</td>
</tr>
<tr>
<td>Active/Passive (A/P)</td>
<td>Interfaces are bonded with one active (preferred) and the other passive (standby). The preferred NIC is selected for the logical bond interface to use. If the preferred NIC fails, then the logical interface begins using the standby NIC in the bond until the preferred NIC resumes operation. When the preferred NIC resumes operation, data transfer resumes on the preferred NIC. No additional switch configuration is required to support this bonding method.</td>
<td>Simplest</td>
</tr>
<tr>
<td>Link aggregation (LACP; 802.3ad)</td>
<td>Both interfaces are bonded into a link aggregation interface to deliver both bandwidth aggregation and redundancy. The bonded interface communicates to the network at the bandwidth of the sum of the interface. If one NIC fails, the remaining connections continue with no disruption to the iSCSI clients. To use link aggregation, the edge switch must support IEEE 802.3ad thus requiring switch configuration.</td>
<td>Most complex</td>
</tr>
</tbody>
</table>
Table 2 provides a quick glance at supported and unsupported NIC bonding configuration for StoreVirtual systems installed with 1 and/or 10 Gigabit NICs. Note: Bonding of bonded interfaces is not supported.

<table>
<thead>
<tr>
<th>Supported NIC bonded configurations</th>
<th>Number of ports x NIC type</th>
<th>Active-passive</th>
<th>802.3ad (LACP)</th>
<th>ALB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HP LeftHand P4000, HP LeftHand P4000 G2 Storage</strong></td>
<td>2 x 1GbE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>2 x 10GbE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>1 x 1GbE + 1 x 10GbE in single bond</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Multiple bonds of same type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Multiple bonds of different type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>HP StoreVirtual 4330, HP StoreVirtual 4530, HP StoreVirtual 4730, HP StoreVirtual 4335 Storage</strong></td>
<td>2 x 1GbE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>3 x 1 GbE</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>4 x 1GbE</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>2 x 10GbE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>1 x 1GbE + 1 x 10GbE in single bond</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Multiple bonds of same type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Multiple bonds of different type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>HP StoreVirtual 4630, HP P4800 G2 for HP BladeSystem, HP LeftHand P4800 for HP BladeSystem</strong></td>
<td>2 x 10GbE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 3. Comparison of active-passive, link aggregation, and adaptive load balancing characteristics

<table>
<thead>
<tr>
<th>Feature</th>
<th>Active-passive</th>
<th>Link aggregation</th>
<th>Adaptive load balancing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>Use of 1 active NIC at a time provides single interface bandwidth. The stand-by NIC is passively awaiting the active NIC’s failure before becoming active.</td>
<td>Simultaneous use of both NICs aggregated bandwidth. All interfaces are active.</td>
<td>Simultaneous use of both NICs increases bandwidth. All interfaces are active however both interfaces send but only one port receives.</td>
</tr>
<tr>
<td>NIC failover protection</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Protection during switch failure in high-availability configuration</td>
<td>Yes, when NICs are connected to different switches.</td>
<td>Yes, when connected to multiple switches virtualized as a single managed switch port Trunk Group. Virtualized switches typically interconnect via dedicated trunks, IRF or dedicated backplanes and manage across both as a single entity.</td>
<td>Yes, when NICs are connected to different switches.</td>
</tr>
<tr>
<td>Link aggregation (802.3ad/LACP) Trunk group on switch and bonded interface</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Nice to know

In an active/passive bond configuration, the preferred NIC is the “active” and the other NICs are “passive.” When the active link fails, then the status of the NIC becomes passive (failed) and one of the other NIC in the bond becomes active. This configuration remains until the failed preferred interface is brought back online. When the failed preferred interface is brought back online, it becomes “active” again and the other NIC returns to the passive state.

Separate management networks

On platforms with more than two network adapters, which are usually bonded for iSCSI and cluster communication, it is possible to connect storage nodes to a separate management network. This allows the storage nodes to access network services, which are not available on the iSCSI network, such as NTP, DNS, or SMTP servers and SNMP trap receivers (HP Remote Support, HP Systems Insight Manager, HP OneView, and the HP Helion OpenStack). The additional network can also be used by administrators to perform storage management operations using the HP StoreVirtual Storage Centralized Management Console, HP StoreVirtual Storage Command-Line Interface (CLIQ) or the HP StoreVirtual REST Application Programming Interfaces (API). The connection to this management network segment can be established using a single network adapter or an additional bond on the storage nodes. Figure 5, Environment with a separate management network, shows a management network which is separate from the storage infrastructure components.
To enable a separate management network, configure the network interface or interface bond in the HP StoreVirtual Centralized Management Console. Make sure that the bond which is used for iSCSI and cluster communication is “LeftHand OS Interface” on the Communication tab of the storage nodes’ network properties (see Figure 6, LeftHand OS interface designation).

**Figure 6.** LeftHand OS interface designation in the Centralized Management Console
**IP addressing**

Using static IP addresses for HP StoreVirtual Storage is best practice. In cases where DHCP is used, reserve IP addresses for all storage systems. The Virtual IP and subnet representing the iSCSI Target Portal must be statically assigned. Address bonds cannot be configured with DHCP, therefore, manually assign the IP address to the bond. When using separate networks for iSCSI and management (on models with more than two network interfaces), the IP addresses should be on a separate subnet.

**Network switching**

Network interfaces on HP StoreVirtual Storage (physical storage nodes as well as HP StoreVirtual VSA) can connect to industry-standard network switches (adhering to the IEEE 802.3 Ethernet standard). While any compliant Ethernet switch works, the capabilities and feature sets of the switch are key to designing a robust and resilient network for storage.

Key design criteria for iSCSI networks should address:

- Performance
- Reliability
- High availability
- Low latency

To address these design criteria, HP recommends selecting switches that supports the following software and hardware feature:

**Software features:**
- Jumbo frame support
- Flow control
- Loop detection
- Quality of services
- Link aggregation protocol
- Traffic filtering (access controls)
- Virtual local area networks (VLANs)
- Routing (optional)
- Switch virtualization (optional)

**Hardware features:**
- Non-blocking backplane
- Sufficient port buffering
- Wire-speed throughput

**Port buffering cache**

For sequential workloads with larger block sizes, it is recommended that the switch has sufficient buffer cache. In these environments, it is recommended to allocate at least 512 KB of buffer cache per port. For example, if the switch has 48 ports, which are all used for iSCSI, the recommendation is to have at least 24 MB of switch buffer cache for those ports.
Switching methodology

The advantage of cut-through switching is that the switch can start transmitting an Ethernet frame on a destination port before it was received in its entirety on a source port—this is in contrast to store and forward packet switching in which the switch has to receive the entire frame first before transmitting it to its destination.

When examining only the first few bytes of a frame, which carries the destination media access control (MAC) address, it saves a considerable amount of latency time to forward the entire frame to its destination port over examining the frame. Different switch technologies aim to reduce this latency.

Flow control

Flow control should be enabled on all switch ports that handle iSCSI traffic to prevent buffer overflow. HP Networking switches recommended in table 5 support both send and receive pause frames.

Nice-to-know

- Flow control is a temporary means of congestion management and is not a resolution for switch backplane oversubscription. Flow control utilizes the buffer of the sending switch for a brief period to manage the switch buffer of the receiving switch during period of congestion.
- Some switch manufacturers do not recommend configuring flow control when using jumbo frames, or vice versa. Consult the switch manufacturer documentation for guidance on this issue.
- Some manufacturer switch models may have limited flow control support, the ability to receive and respond to pause frames, but do not support the ability to originate the sending of pause frames. HP Networking series of switches listed in table 5 can respond to pause frames generated by targets and initiators, allowing them to control the overall data flow between servers and storage volumes.

Best practices

- HP recommends enabling flow control. Flow control should be enabled on all switch ports.

Jumbo frames

When using jumbo frames for the iSCSI network, all switch ports on the network segment need to be configured for jumbo frames. If the ports are not properly configured, the switch may discard frames. This may cause connectivity and performance issues, which are difficult to troubleshoot.

Nice-to-know

- Jumbo frame sizes are not standardized across all switch manufacturers.
- On HP Networking switches, jumbo frames can be enabled globally, per-VLAN and or on a per port interface basis.
- Switches that don’t support jumbo frames, drop jumbo frames.

Best practices

- For a jumbo frame to be transmitted in whole from end to end, every network device within the network path must support that frame size.
- When employing jumbo frames in iSCSI networks, remember to set the MTU on initiator access, trunk and bonds for all appropriate ports within the iSCSI data path.
- HP doesn’t recommend the use of jumbo frames in all customer deployments. However, depending on the characteristics of the iSCSI workloads, HP test results have shown performance gains of 10–20 percent when designed and configured properly for large sequential read and write workloads.
- Set the switch infrastructure to a frame size 4 bytes higher than initiators and HP StoreVirtual Storage highest setting if CRC checksum is enabled plus an additional 4 bytes if the switching infrastructure enables VLANs.
VLAN

A VLAN is a logical grouping of devices that share the same broadcast domain with a common set of network requirements. For HP StoreVirtual Storage, the network should be isolated from other networks, either by using VLANs or separate network equipment.

VLANs allow for a logical segmentation of a single physical switch. A port which is a member of one VLAN cannot communicate with ports which are on another VLAN; this creates separate networks segments on a single switch. By segmenting traffic, broadcasts from one segmentation do not impact any other. A single port is typically a member of one “native” VLAN; however, ports can also carry traffic from multiple VLANs. This means that all frames which are not on the native VLAN of this port are tagged with the VLAN ID. This is true for all incoming and outgoing frames on that port.

HP StoreVirtual Storage does not natively support tagged VLANs. If HP StoreVirtual Storage is on a separate iSCSI VLAN, the appropriate switch ports must be in the iSCSI VLAN (untagged for iSCSI VLAN; primary VLAN ID must be set to the VLAN ID of the iSCSI VLAN).

Best practices

- Isolate iSCSI traffic on its own network segment (separate VLAN or dedicated network infrastructure), thus separating management and general production traffic on separate VLANs.
- The preferred protocol for combining two or more physical VLAN tagged uplinks between switches to aggregate bandwidth while providing fault tolerance is the link aggregation control protocol (LACP/802.3ad).
- Plan for redundant network paths exist between all interconnected switches allowing an alternative path for traffic to traverse in the case of a primary path failure.
- Note that VLANs may not be enabled on HP StoreVirtual Storage networks originating from the storage, but may be enabled and managed from the switch infrastructure and initiator host. The HP StoreVirtual VSA and HP StoreVirtual VSA 2014 may have their virtual NICs VLAN tagged from the parent hypervisor.

Spanning Tree Protocol

Spanning Tree Protocol (STP) is a link management protocol that detects and prevents loops in an Ethernet network by helping ensure that only one active path exists between two or more interconnected switches. There are several enhanced versions of the common STP protocol that offer faster convergence of the STP protocol. HP recommends using either Rapid Spanning Tree Protocol (RSTP) or Multiple Streaming Tree Protocol (MSTP) for loop prevention and path fail-over to support a fault-tolerant IP storage network.

RSTP is simplest of the two protocols to configure. RSTP operates independent of network VLAN configuration, and employs a single spanning tree instance throughout the Layer 2 bridged network. RSTP protocol blocks physical redundant links within a switched network. Blocked links are available as failover paths.

MSTP allows for the design of multiple STP instances over 802.1q links (tagged links) and enables VLANs to be mapped to its spanning-tree instance. MSTP provides multiple forwarding paths for data traffic and enables load balancing. It improves the fault tolerance of the network because a failure in one STP instance, or forwarding path, may not affect other STP instances forwarding for the other VLANs. MSTP offers advanced STP features and is more complex to configure and design than RSTP.

Nice-to-know

- RSTP and MSTP achieve faster convergence on uplink or Inter-Switch Link (ISL) ports than the traditional STP protocol.
- MSTP and RSTP are industry-standard IEEE spanning tree protocols; and are the default protocols supported by switch manufacturers adhering to open and industry-standard network protocols.
- A switch port connected to the network interface of a server or storage node is commonly considered an edge port. This type of port should not be configured to participate in the STP protocol process.
**Best practices**

- Do not enable RSTP/MSTP on a switch access port connected to end stations (i.e., iSCSI initiator or storage node); if RSTP/MSTP is enabled, configure the port as an edge port.
- Configure BPDU guard on ports configured as edge ports to prevent ports configured as edge port from sending or receiving BPDU.
- Minimize the dependence on RSTP/MSTP (i.e., use Stacking, Virtual Switching Technology, IRF).
- Enable RSTP/MSTP on all ports between switches.
- Set the root bridge manually rather than allowing the switches to go through the root bridge election process.
- Configure RSTP/MSTP in all cases where two or more switches are interconnected via more than one physical link to help minimize the chance of a network loop caused by improper configurations (i.e., IRF, link aggregation group, and stacking of switches).

**Unicast storm control**

A traffic storm occurs when an excessive amount of packets create network traffic that is transmitted by a single port and sensed by the switch as a traffic storm (i.e., denial of service). iSCSI traffic is bursty in nature and may sometimes utilize a very high link percentage that sometimes may be incorrectly sensed as a traffic storm by the switch.

**Best practices**

HP recommends disabling unicast storm control on iSCSI ports.

**Recommended network switches for HP StoreVirtual Storage**

HP Networking has an extensive portfolio of Ethernet switches that are ideally suited for iSCSI deployments. These switches are designed to deliver high-performance, low-latency, and wire-speed Layer 2 to 4 network services. HP Networking top of rack series of switches offer unmatched combination of 1GbE and 10GbE ports density, high-availability architecture, Intelligent Resilient Framework (IRF) technology as well as a full range of Layer 2 and or Layer 3 features depending on switch selection. These switches are built to support open and industry-standards networking protocol.

These switches can be deployed within the network core or as a high-performance switch in the convergence layer or network edge of SMB and enterprise networks.

<table>
<thead>
<tr>
<th>HP Switch Model</th>
<th>10GbE</th>
<th>Uplink</th>
<th>Intelligent Resilient Framework (IRF)</th>
<th>Port Buffer Cache</th>
<th>IPv6</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5900 CP</td>
<td>48 SFP+</td>
<td>4 QSFP+40-GbE</td>
<td>Yes. Up to 9</td>
<td>95 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td>5900AF-48XG-4QSF+</td>
<td>48 SFP+</td>
<td>4 QSFP+40-GbE</td>
<td>Yes. Up to 9</td>
<td>95 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td>5820X-24XG-SFP+</td>
<td>24 SFP+</td>
<td></td>
<td>Yes. Up to 9</td>
<td>2 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td>5406R zl2</td>
<td>48 SFP+</td>
<td></td>
<td>Yes</td>
<td>36 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td>5412R zl2</td>
<td>96 SFP+</td>
<td></td>
<td>Yes</td>
<td>36 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td>8206zl</td>
<td>48 SFP+</td>
<td></td>
<td>Yes</td>
<td>144 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
<tr>
<td>8212zl</td>
<td>96 SFP+</td>
<td></td>
<td>Yes</td>
<td>144 MB shared</td>
<td>Yes</td>
<td>Full</td>
<td>Full</td>
</tr>
</tbody>
</table>

Recommended HP Networking pocket guide and Individual product specification data sheets
h20195.www2.hp.com/v2/getpdf.aspx/4AA1-6237ENU.pdf?ver=7.1
Table 4. List of recommended HP Networking Ethernet switches for HP StoreVirtual Storage systems

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3800-24G</td>
<td>The HP 3800 Switch Series is a family of fully managed Gigabit Ethernet switches available in 24-port and 48-port models with either SFP+ or 10GBASE-T uplinks. The 3800 Switch Series utilizes the latest HP ProVision ASIC technology and advances in hardware engineering. Meshed stacking technology is implemented in the HP 3800 Switch Series to deliver chassis-like resiliency in a flexible, stackable form factor. This 24-port model operates at wire-speed and is a non-blocking switch. This model delivers 88 Gb/s of switching capacity and 65.4 million packets per second. This switch offers extensive Layer 2 and Layer 3 switching and routing features. This switch has redundant hot-swappable power supplies and designed to be co-located at the top of rack. This switch incorporates 18 MB of packet buffer.</td>
</tr>
<tr>
<td>5830AF-48G</td>
<td>The HP 5830AF-48G model offers 48 1GbE ports, up to four 10GbE ports, high-availability architecture and full Layer 2 and Layer 3 switching/routing capabilities. In addition to wire-speed performance on all ports, the switches include HP IRF technology; which support local or geographically distributed HP 5830 switches to be interconnected for higher resiliency and performance. This versatile switch is ideal for use in the network core or as a high-performance network edge switch for virtualized infrastructures. It delivers 160 Gb/s of routing and switching capabilities (wire-speed). The platform delivers exceptional hardware reliability offering redundant power supplies. This is a store and forward switch with a port buffer size of 1 GB.</td>
</tr>
<tr>
<td>5920</td>
<td>The HP 5920 Switch Series is made up of high-density 10GbE, ultra-deep packet buffering, and Top-of-rack (ToR) switch. This switch is part of the HP FlexFabric solution module of the HP FlexNetwork architecture. This switch is ideally suited for deployments at the server access layer of large enterprise data centers. They are also designed for content delivery networks, especially when used to eliminate network congestion at the I/O that is associated with the heavy use of server virtualization, as well as bursty storage applications. This switch offers industry leading advance Layer 2/3 switching and routing protocol. This switch fully supports HP IRF technology; up to four 5920 switches can be grouped together in an IRF configuration that allows them to be configured and managed as a single switch. This is a store and forward switch; it provides up to a 3.6 GB of packet buffer to eliminate network congestion.</td>
</tr>
<tr>
<td>6600-24XG</td>
<td>The HP 6600 Switch Series offers both 1GbE and 10GbE models. The 6000 Switch Series consists of the most advanced data center server edge switches in the HP Networking product line. This 24-port 10GbE model is a wire-speed and non-blocking switch. This model delivers 322 Gb/s of routing and switching capacity and 240 Mb/s packet forwarding rate. This switch offers extensive Layer 2 and Layer 3 advanced switching and routing features. This switch has front-to-back cooling, redundant hot-swappable power and redundant hot-swappable fans, and designed to be co-located at the top of a server rack and ideally suited for the enterprise data center. This switch incorporates 108 MB of packet buffer shared across all 10GbE ports. This is a store and forward capable switch.</td>
</tr>
</tbody>
</table>

The table below summarizes the switch capabilities of HP Networking switches to meet the demands of small, medium, and enterprise iSCSI network. To learn more visit hp.com/go/networking.
### Table 5. HPN switch features by model small, medium, and enterprise

<table>
<thead>
<tr>
<th>Feature</th>
<th>2920-24G/2920-48G</th>
<th>5920-24G</th>
<th>5900 CP</th>
<th>3800-24G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access port</td>
<td>1GbE</td>
<td>10GbE</td>
<td>1GbE</td>
<td>1GbE</td>
</tr>
<tr>
<td>Uplink port</td>
<td>10GbE</td>
<td>10GbE</td>
<td>10GbE</td>
<td>10GbE</td>
</tr>
<tr>
<td>IRF</td>
<td>X</td>
<td>√</td>
<td>√</td>
<td>X</td>
</tr>
<tr>
<td>Flow control</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Jumbo frame</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>802.3x (Flow Control)/802.3ad (LACP)</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Store-and-forward</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Packet buffering</td>
<td>Shared</td>
<td>Shared</td>
<td>Shared</td>
<td>Shared</td>
</tr>
<tr>
<td>Port buffer size</td>
<td>18 MB/1.125 MB</td>
<td>3.6 GB</td>
<td>95 MB</td>
<td>18 MB</td>
</tr>
<tr>
<td>Routing</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>6600-24XG</th>
<th>5406R/5412R zl2</th>
<th>5820X-24G/5830AF-48G</th>
<th>8206/8212 zl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access port</td>
<td>10GbE</td>
<td>10GbE</td>
<td>1GbE</td>
<td>10GbE</td>
</tr>
<tr>
<td>Uplink port</td>
<td>10GbE</td>
<td>10GbE</td>
<td>10GbE</td>
<td>10GbE</td>
</tr>
<tr>
<td>IRF</td>
<td>X</td>
<td>√</td>
<td>√</td>
<td>X</td>
</tr>
<tr>
<td>Flow control</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Jumbo frame</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>802.3x (Flow control)/802.3ad (LACP)</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Store-and-forward</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Packet buffering</td>
<td>Shared</td>
<td>Shared</td>
<td>Dedicated/Shared</td>
<td>Shared</td>
</tr>
<tr>
<td>Port buffer size</td>
<td>108 MB</td>
<td>36 MB</td>
<td>2 MB per port/1 GB</td>
<td>144 MB</td>
</tr>
<tr>
<td>Routing</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>
**Host configuration**

The server is typically configured in one of three ways to communicate to iSCSI targets presented by HP StoreVirtual Storage over the network. The three initiator configurations are:

1. iSCSI Software Initiator with a standard network adapter
2. iSCSI Software Initiator combined with an accelerated network adapter. Referred to as a TCP/IP offload engine (TOE)
3. Hardware iSCSI host bus adapters (HBAs) or converged network adapters

From the server, network adapters which are used by iSCSI Software Initiators show up as network adapters and iSCSI HBAs are classified as storage controllers. HP recommends:

- Two NICs or two iSCSI HBAs installed per server, dedicated to iSCSI storage traffic
- Plus additional NICs for other traffic types (server cluster network, virtual machine networks, etc.)
- Disable any service or protocol which is not actively used on network adapters dedicated to iSCSI traffic (e.g. file and print sharing protocols)

For a current list of supported iSCSI HBA initiators and operating systems, examine the HP Single Point of Connectivity Knowledge (SPOCK) website, hp.com/storage/spock, to obtain detailed information about supported HP product configurations. HP does not recommend any particular 10GbE adapter for third-party servers.

**Jumbo frames**

Like the other components on the network segment, hosts need to be configured with the same MTU size as storage systems and network switches. When using jumbo frames, make sure that all network adapters used for iSCSI are properly configured.

**Best practices**

- For VMware® vSphere hosts—set to the same MTU value on Virtual Switches (i.e., VMware vNetwork Standard Switch and vNetwork Distributed Switch), VMkernel ports, physical network adapters, and virtual NICs of the guest operating system.
- For Microsoft® Windows Server hosts—enable jumbo frames on network adapters to take advantage of larger MTU frame sizes. The process of enabling jumbo frames varies based on network adapter and driver. To set jumbo frames in Windows go to the device manager, select the network interface, then go to the advanced tab, select the jumbo frame setting and then set the correct MTU size. If virtual machines on Hyper-V should also use jumbo frames for iSCSI, the virtual switch(es) and network adapter of the virtual machines need to be configured for jumbo frames too.

**Flow control**

Ideally, flow control on hosts is configured to auto-negotiate the flow control settings with the switch ports (like storage nodes). If flow control auto-negotiation is not available, enable flow control for sending and receiving.

**Best practices**

- Hardware-based flow control is recommended on all server network interfaces.
- VMware vSphere hosts have flow control enabled by default; the physical adapters (vmnics) are set to auto-negotiate.
- For Microsoft Windows Server® hosts, enable flow control on all server network adapters attached to iSCSI segments. Flow control is enabled by selecting the appropriate network adapter and then by selecting the “Advance Tab.” HP recommends setting flow control settings to Rx/Tx enabled; this enables network adapters to generate and respond to flow control frames. Make certain all server network adapters attached to iSCSI segments are configured the same; with flow control enabled.
**Multipathing I/O**

For servers, multipathing makes use of redundant physical adapters to create logical paths between the initiator and target. Each network interface card or HBA should be connected to the resilient storage network to provide continued access to storage in the event of a component failure.

### MPIO for VMware vSphere

HP recommends native multipathing using round-robin load balancing policy. Round-robin provides performance advantage by load balancing host I/O requests across all available network paths. Each path to the iSCSI SAN establishes a unique iSCSI session to the target; thus providing high availability, aggregate bandwidth, and optimal network performance. Configuring round-robin multipathing requires at least two network ports and two VMkernel ports connected to a virtual switch.

To guide VMware vSphere and HP StoreVirtual multipathing integration, refer to the technical white paper titled *“HP StoreVirtual (formerly LeftHand) Storage with VMware vSphere: Design considerations and Best Practices”.* This paper provides detailed information on how to integrate VMware vSphere 5.0 with HP StoreVirtual Storage to include configuration steps to implement multipathing.

**Best practices**

- Use at least two network adapters and round-robin multipathing for improved performance and failover of the iSCSI connections.

- VMkernel ports with iSCSI bindings should be separate from host management and virtual networks traffic used by virtual machines. Best practice is to separate VMkernel port traffic by functionality such as iSCSI, vMotion, Fault Tolerance (FT), and virtual machines, requiring a minimum of eight network adapters.

- HP recommends using network adapters capable of offloading TCP/IP processes from the host CPU to specialized hardware on the network adapter. Refer to VMware Compatibility Guide or HP SPOCK for a list of compatible network adapters.

### MPIO for Microsoft Windows Server

For Windows Server, the built-in MPIO framework is used to manage redundant paths to HP StoreVirtual Storage. In addition to supporting the generic Microsoft Device Specific Module (DSM) for iSCSI, HP StoreVirtual Storage also supports the HP StoreVirtual DSM for Microsoft MPIO which can improve performance and reduce latency.

HP StoreVirtual DSM for MPIO can increase the throughput for sequential workloads by optimizing the data path. In addition, host and target gain efficiencies of decreased CPU utilization and latencies. In cases of multi-site, the HP StoreVirtual DSM for MPIO also provides site awareness, ensuring local host to local storage replicated to remote storage, thereby minimizing data traversing between synchronously replicated sites. Once installed, MPIO works automatically in the background for all iSCSI sessions established by Microsoft’s iSCSI Software Initiator for volumes on HP StoreVirtual Storage. iSCSI HBAs do not support the HP StoreVirtual DSM for MPIO.

For more information, refer to the *“Configuring HP StoreVirtual (formerly LeftHand)”* white paper which covers multipathing for Microsoft Windows Server in more detail.

**Nice-to-know**

- Microsoft does not support network adapter teaming with the iSCSI Software Initiator

- HP StoreVirtual DSM for Microsoft MPIO can be found at [hp.com/go/p4000downloads/](http://hp.com/go/p4000downloads/)

**Best practices**

- Use multiple network adapters to establish iSCSI sessions for multipathing

- The preferred load balancing options for HP StoreVirtual DSM for MPIO and Microsoft DSM for iSCSI is round-robin
Microsoft Windows Server 2012 TCP Software and Hardware NIC settings

Windows Server 2012 supports enhanced TCP settings increasing throughput and support may vary depending upon the network interface installed. Examine table 6 for a summary of the default settings.

Delayed ACK algorithm
The delayed ACK algorithm is a technique to improve TCP performance by combining multiple ACK responses into a single response. This algorithm has the potential to interact negatively with a TCP sender utilizing Nagle’s algorithm, since Nagle’s algorithm delays data transmission until a TCP ACK is received.

Nagle’s algorithm
Nagle’s algorithm, named after John Nagle, is a means of improving the efficiency of TCP/IP networks by reducing the number of packets that need to be sent over the network. Nagle’s algorithm works by combining a number of small outgoing messages, and sending them all at once.

Receive Segment Coalescing (RSC)
RSC is a stateless offload technology that helps reduce CPU utilization for network processing on the receive side by offloading tasks from the CPU to an RSC-capable network adapter.

Receive Side Scaling (RSS)
RSS enables network adapters to distribute the kernel-mode network processing load across multiple processor cores in multi-core computers. The distribution of this processing makes it possible to support higher network traffic loads than would be possible if only a single core were to be used.

Receive Window Auto-Tuning
Receive Window Auto-Tuning feature which improves performance for programs that receive TCP data over a network. Note that slower data transfers or connectivity loss may occur if the network infrastructure uses older router technology not supporting this feature.

TCP Chimney Offload
TCP Chimney Offload is a networking technology that helps transfer the networking processing workload from the host CPU to a network adapter’s processor during network data transfer.

<table>
<thead>
<tr>
<th>Table 6. Default TCP setting for Windows Server 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2012 default TCP settings</td>
</tr>
<tr>
<td>Delayed ACK algorithm</td>
</tr>
<tr>
<td>Nagle’s algorithm</td>
</tr>
<tr>
<td>Receive Segment Coalescing (RSC)</td>
</tr>
<tr>
<td>Receive Side Scaling (RSS)</td>
</tr>
<tr>
<td>Receive Window Auto-Tuning Level</td>
</tr>
<tr>
<td>TCP Chimney Offload</td>
</tr>
</tbody>
</table>
Multi-site cluster networking requirements

Multi-site clusters in HP StoreVirtual enable block level synchronous data replication between sites using Network RAID. Site-awareness for server and storage systems is built in to multi-site clusters. A multi-site cluster can have up to three sites. As volumes in a multi-site cluster span all sites, each site has the same number of storage nodes and volumes in these environments should always be protected using Network RAID (Network RAID 10 for clusters with two sites; Network RAID 10+1 for clusters stretching three sites).

Figure 7. Typical multi-site cluster environment

The configuration in Figure 7, Typical multi-site cluster environment, is for customers who want to replicate data synchronously between two data centers. HP StoreVirtual Fail-Over Manager, which is ideally installed in a third site, provides quorum in the event of a site failure. For more information, please refer to the Specialized Manager section in the HP StoreVirtual User Guide.

From a networking perspective, there are various options to achieve resiliency. Multiple, aggregated links between data centers are recommended.

The two primary factors that contribute to the health of a multi-site cluster are network latency and bandwidth. High network latency is a primary cause of slow I/O performance, or worse case iSCSI target disconnects. When designing a network for multi-site clusters, latency should be kept to a minimum; HP recommends a round trip latency of less than two milliseconds (2ms). Focus on reducing latency throughout the design from end-to-end, while taking into consideration the following:

- Distance between sites—make certain distance between all clustered sites meets the 2ms latency requirement. Increasing distance increases latency. Some technologies exist for latency reduction which may increase distances without impacting latency.
- Number of routed hops between sites—each router or switch in the data path adds delay as the packet is received, processed, and then forwarded.
- Advanced network features—firewalls, NAT devices, QoS, or access control lists add latency.
**Best practices**

- Use a single IP subnet with Layer 2 switching.
- Configure redundant network paths between sites to help minimize link failure in a two-site configuration. In multi-site clusters with three sites, HP recommends a triangle network design, and not a V-design to ensure the highest availability.
- For 1GbE installations allocate 50 MB/s per storage node of bandwidth in the sites between data centers; for environments with 10GbE storage connectivity, allocate 250 MB/s per storage node. For campus cluster deployments with Fibre Channel connectivity with HP StoreVirtual Storage, plan for 500 MB/s per storage node (only two sites supported).
- In the example (see Figure 7) above with 1GbE storage connectivity and two storage nodes in each site, designate 50 MB/s x 2 = 100 MB/s of bandwidth between the sites with at least one 1GbE link; two for redundancy.
- HP StoreVirtual Fail-Over Manager (FOM) is not in the data path and has to be connected using at least Fast Ethernet connections (100 Mb/s) with a latency not exceeding 50ms. No iSCSI data traverses between storage nodes and the FOM.

**Remote Copy**

Remote Copy is an asynchronous data replication feature within HP StoreVirtual Storage. This feature addresses requirements for high levels of data protection in the context of disaster recovery, business continuity, and backup and recovery. Remote Copy uses HP StoreVirtual volume snapshot and thin provisioning features. Using Remote Copy, you can replicate data from and to StoreVirtual clusters. Distances can be local (within the same data center or on the same campus), Metro (in the same city) or WAN (long distances). Remote Copy takes point-in-time snapshots of volumes and replicates the data to the remote site. Snapshots provide a means to implement secure offsite tape backup at recovery sites. Use cases also include replication between remote/branch offices and data centers.

**Figure 8.** Two management group sites in a Remote Copy environment

One of the most important aspects of designing a successful disaster recovery solution using Remote Copy is properly designing and sizing the network bandwidth and usage requirements between primary and recovery sites. The sections to follow provide an overview of WAN connectivity options, bandwidth planning, and remote bandwidth tuning recommendations for using Remote Copy in an IP network. Figure 8, Two management group sites in a Remote Copy environment, shows a typical setup for Remote Copy. Two individual HP StoreVirtual management groups can use an IP network (private lines or encrypted site-to-site VPNs over the Internet) to replicate data. For more information on which ports should be opened on firewalls, please consult the “HP StoreVirtual Storage User Guide - Chapter 22 LeftHand OS TCP and UDP port usage (page 271)” document.
Asynchronous replication

With Remote Copy, there are virtually no IP distance limitations for replicated solutions, the primary and recovery sites may span hundreds or even thousands of kilometers, using different types of network access connectivity between sites. With asynchronous replication, it is understood that the bandwidth between the two sites is limited. Writes to the primary volume are acknowledged immediately back to the client, and then replicated to the remote volume as part of a snapshot. The replication time required to copy data between sites is determined by the amount of data that needs to be copied and a function of the bandwidth availability. The first snapshot must copy the entire thinly provisioned data representation of the volume whereas subsequent remote snapshots only copy the delta change of data between the last remote snapshot version. The consideration of data change and bandwidth should be considered carefully prior to creating a Remote Snapshot Schedule.

Best practices

• To ensure Remote Copy resiliency, for customer with tight recovery point objectives (RPOs) and recovery time objectives (RTOs), HP recommends considering WAN path redundancy and carrier diversity for WAN links between sites.
• For customers employing site-to-site VPN services, Network Address Translation (NAT), appropriate placement of firewall, public IP addressing, IPsec, authentication, TCP and UDP port forwarding, are top of mind considerations.
• Ensure network time protocol (NTP) is properly configured with correct time zone settings for all storage nodes.
• Use 802.1q tagged switched links to extend VLAN broadcast traffic between primary and secondary sites for non-routed interlinks.
• For routed links between primary and secondary sites, ensure routers are configured with industry standard routing protocols such as Routing Information Protocol (RIP) and Open Shortest Path First (OSPF). HP recommends using OSPF as the routing protocol of choice given it has the richest feature set over RIP.
• When QoS is enabled within your LAN for storage traffic make sure differentiated service code point (DSCP) or class of service (COS) markings are honored by the service provider (across the WAN link) and applied from end to end.

Network bandwidth planning

Sufficient bandwidth between sites can present a challenge for HP StoreVirtual customers to meet their Remote Copy objectives, as high-speed WAN links are typically pricey and constrains most small and medium-size customers to T1 bandwidth speeds for long-distance WAN connectivity. Remote Copy frequency and amount of data change on the primary volume are key factors in calculating network bandwidth.

For example, the amount of data changed divided by the period of time between Remote Copy windows (frequencies) determines the minimum bandwidth required for inter-site connectivity and determines if the RPO objective is met. If our data change rate is 4 GB hourly and our replication window is every four hours, then network bandwidth required to adhere to our Remote Copy schedule is at minimum a 10 Mb/s connection. See table 7 as an example.

To understand network usage, the HP StoreVirtual Centralized Management Console is used to understand snapshot sizing and replication time. For example:

• Set up local snapshot schedule and frequency; let this run for a couple of days to gather snapshot change rate information and transfer times; this information can be reviewed using the remote snapshot tab and remote snapshot details window
• Use information collected to characterize the amount of change on the primary volume for a given period of time

You can calculate Remote Copy transfer rate and times using information gathered using the remote snapshot tab and details window. The remote snapshot tab provides snapshot statistics such as:

• Percent complete—the incremental progress of the Remote Copy operation
• Elapsed time—the incremental time taken by the copy operation
• Data copied—the incremental quantity of data copied
• Current rate—the rate at which data is being copied or when the remote snapshot is complete
• Average rate for the total operation
• State—the status of the operation
The snapshot details window includes the information above plus:

- Start time
- Estimated time remaining
- Completion time
- Data remaining to be copied

Using the start time, completion time, amount of data copied statistic along with Remote Copy frequency are key factors in validating network bandwidth requirement for inter-site links.

### Calculate bandwidth usage

Network bandwidth is commonly measured in megabits per second whereas storage transfer is measured in megabytes per second. To understand the network transfer rate for remote replication sessions, convert megabytes to megabits to determine transfer time in hours. Steps are outlined below:

Steps are outlined below:

1. Convert the size of the volume to bytes: Begin by multiplying the gigabytes by 1,024 to get megabytes.
2. Multiply megabytes by 1,024 to get to kilobytes.
3. Multiply kilobytes by 1,024 to get bytes.
4. Multiply bytes by 8 to yield the number of bits.
5. Divide the number of bits by 1,000 to yield kilobits or again by 1,000 to yield megabits.
6. Divide results by the link speed to get the number of seconds to complete the copy. Convert to the correct order of magnitude, Kb or Mb, before dividing by the corresponding units per second of link speed.

Table 7 provides a working illustration of how to calculate "network transfer rate and time" for a data change rate of 4 GB per hour with an every 4-hour remote snapshot schedule. Meaning the primary volume change rate is 16 GB over a 4-hour period.

<table>
<thead>
<tr>
<th>Conversion description</th>
<th>Conversion factor</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigabytes to megabytes</td>
<td>GB x 1204 = MB</td>
<td>16 GB x 1024 = 16,384 MB</td>
</tr>
<tr>
<td>Megabytes to kilobytes</td>
<td>MB x 1024 = KB</td>
<td>16,384 MB x 1024 = 16,777,216 KB</td>
</tr>
<tr>
<td>Kilobytes to bytes</td>
<td>KB x 1024 = bytes</td>
<td>16,777,216 KB x 1024 = 17,179,869,184 bytes</td>
</tr>
<tr>
<td>Bytes to bits</td>
<td>Bytes x 8 = bits</td>
<td>17,179,869,184 Bytes x 8 = 137,438,953,472 bits</td>
</tr>
<tr>
<td>Bits to kilobits</td>
<td>Bits/1000 = Kb</td>
<td>137,438,953,472 Bits/1000 = 137,438,953 Kb</td>
</tr>
<tr>
<td>Kilobits to megabits</td>
<td>Kb/1000 = Mb</td>
<td>137,438,953 Kb/1000 = 137,438 Mb</td>
</tr>
<tr>
<td>Megabits per second</td>
<td>Mb/network link speed in Mb/s</td>
<td>137,438 Mb/10 Mb/s = 13743 sec</td>
</tr>
<tr>
<td>Seconds to minutes</td>
<td>Second/60 = minutes</td>
<td>13743/60 = 229 min</td>
</tr>
<tr>
<td>Minutes to hours</td>
<td>Minutes/60 = hours</td>
<td>229 min/60 = 3.81 hours &lt; Remote Copy window</td>
</tr>
</tbody>
</table>

Remote Copy schedule is every four hours
### Table 8. Remote Copy transfer rate @ different network link speeds

<table>
<thead>
<tr>
<th>Network media</th>
<th>Network link speed</th>
<th>Remote Copy rate in KB/s</th>
<th>Replication time in hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractional T1</td>
<td>256 Kb/s</td>
<td>32</td>
<td>149.1</td>
</tr>
<tr>
<td>Fraction T1</td>
<td>768 Kb/s</td>
<td>96</td>
<td>49.7</td>
</tr>
<tr>
<td>T1</td>
<td>1.54 Mb/s</td>
<td>193</td>
<td>24.7</td>
</tr>
<tr>
<td>Bonded T1 (2)</td>
<td>3.0 Mb/s</td>
<td>386</td>
<td>12.4</td>
</tr>
<tr>
<td>Bonded T1 (3)</td>
<td>6.0 Mb/s</td>
<td>722</td>
<td>6.2</td>
</tr>
<tr>
<td>Ethernet</td>
<td>10 Mb/s</td>
<td>1,250</td>
<td>3.8</td>
</tr>
<tr>
<td>T3</td>
<td>45.0 Mb/s</td>
<td>5,592</td>
<td>51.2 minutes</td>
</tr>
<tr>
<td>Fast-Ethernet</td>
<td>100.0 Mb/s</td>
<td>12,500</td>
<td>22.9 minutes</td>
</tr>
<tr>
<td>OC-3</td>
<td>155.0 Mb/s</td>
<td>19,375</td>
<td>14.8 minutes</td>
</tr>
<tr>
<td>OC-12</td>
<td>622.0 Mb/s</td>
<td>77,500</td>
<td>3.7 minutes</td>
</tr>
<tr>
<td>Gigabit-Ethernet</td>
<td>1000.0 Mb/s</td>
<td>12,5000</td>
<td>2.3 minutes</td>
</tr>
</tbody>
</table>

### Remote Copy tuning

Remote Copy tuning is used to manage the transfer rate for remote snapshots. Setting the remote bandwidth allows you to define the maximum rate of data transfer between primary and secondary (DR) management groups.

To tune the bandwidth, use the HP StoreVirtual Centralized Management Console to edit the remote bandwidth setting on the DR site by right-clicking on the management group and selecting Edit Management Group. Use Table 8 to guide network bandwidth selection or select Custom from the Edit Remote Bandwidth dialog.

**Figure 9.** Bandwidth management for remote management groups
Nice-to-know

• Remote Copy operates at the block level as only changed blocks are captured in snapshots.
• Snapshots are incremental: snapshots save only delta changes associated with the primary volume since the last snapshot was created.
• Remote Copy is robust. If the network link goes offline during the replication process, copying automatically resumes when the link is restored.
• Remote Copy is a pull operation from the remote site.

Best practices

• Test Remote Copy schedules to ensure the frequency of snapshots account for the amount of time required to complete a Remote Copy.
• Use HP StoreVirtual PrimeSync feature for large initial primary volume replications, especially for sites with slow WAN interconnectivity.
• Consider using MPLS VPN services to extend customer’s private network between primary and secondary sites.
• Limit scheduling of Remote Copy to three concurrent tasks.

Frequently asked questions

1. Storage systems can’t ping application servers or other storage systems. HP StoreVirtual Storage is unable to communicate with other services on the network, like the e-mail server. What should I do?

Verify the IP address configuration and ensure that a proper networking IP address, subnet mask, and gateway are configured on all devices. An invalid default gateway or no gateway will interrupt network traffic between segments. Validate physical network connectivity and valid connection lights. Ensure network switch configuration supports a path for all required devices. Validate DNS to IP resolution with reverse pointer lookup records for proper resolution (ping –a hostname). Ensure that firewall rules between testing points allow ICMP/ping traffic.

For validating connectivity from the storage node, the console or text user interface (TUI) supports ping. Validate the response to all devices. Alternatively, the HP StoreVirtual Centralized Management Console also supports issuing a ping command from the specific storage node’s Network TCP/IP tab and selecting ping from the available TCP/IP tasks dropdown.

For validating connectivity from a managed switch, examine the switch capability for issuing a ping command.

For validating connectivity from an iSCSI initiator host or another application server on the network, open a command shell prompt or equivalent. Issue the ping command. A response validates connectivity. Latency and packet loss may also be noted. Note that some operating systems may have system firewall rules enabled restricting ping response.

```plaintext
ing 192.168.2.1
Ping statistics for 192.168.2.1:
   Packets: Sent=4, Received=4, Lost=0 (0% loss),
   Approximate round trip times in milli-seconds:
     Minimum=1ms, Maximum=75ms, Average=20ms
Traceroute is a network diagnostic tool for displaying the route (path) and measuring transit delays of packets across an Internet Protocol (IP) network. It may be used for tracing the intermediate network path and displaying sampled latency.
```
2. HP StoreVirtual Storage provides inconsistent behavior and performance. What should I look at?

Ensure that each storage node network settings match consistently across all storage nodes within the cluster. Inconsistent Best Practices for HP StoreVirtual Storage will be noted within the CMC’s SAN Status Page on the Best Practice Summary view with suggestions for correction. The switch infrastructure of bonded connections must be validated with switch management. Consistency is key. Ensure that the host application server operating system best practices are followed.

3. HP StoreVirtual Storage services or messages indicate Centralized Management communication issues. No error message may suggest a communication issue. “Connection refused” message occurs in the Centralized Management Console.

Error messages may have several causes, however, validating open ports provides a first step guidance if the message context is not applicable.

Ensure that required Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) ports are open between appropriate devices within the network. Different classes of communication exist. Internode storage traffic within a cluster, initiator to target via iSCSI protocol, storage cluster to storage cluster replication and management APIs.

Blocked ports may be implemented within the switching infrastructure. Managed switches may support configuration of per port, per trunk or switch-wide firewall rules specifically enabling TCP/UDP ports or disabling. Network Address Translation may also be implemented mapping specific port translations that could have impact.

Finally, firewall rules may also exist within the host initiator.

For example, to validate firewall rules within Windows

```
netsh firewall show state
```

Lists the blocked ports as well as active listening ports with application associations

```
netsh firewall show config
```

Dumps the Windows firewall configuration detail

The HP StoreVirtual Storage TCP and UDP port usage detail can be found in the HP StoreVirtual Storage User Guide.

Tools that may help troubleshoot include port scanners such as the Open Source nmap and angry IP scanner mapping out network TCP and UDP port capabilities of an infrastructure as seen from the tool. Validate that this troubleshooting is acceptable networking practice with administrators as some networking security products may sense this activity and respond accordingly. (A port scan seems to be a particular network security concern if not intentional with specific white hat purposes).
4. How can communication be verified with the Centralized Management Console (CMC)/CLiQ?

Within the CMC, two ways validate network communication. The first method is implied by the management utility communicating and displaying information about the storage systems. Select <Find -> Clear All Found Items> for the management console to be cleared of any currently found data. After everything is clear, select <Find -> Find Systems> for the Find Systems window. A list of IP addresses to search for may be listed. Validate that the IP address of the storage system is listed. Alternatively, Auto Discover by Broadcast may be selected. If a storage node is not found by Auto Discover, try unselecting Auto Discover by Broadcast and manually adding the IP address of storage node to the IP address list and reselect Find. If the storage system is still not found, validate network configuration.

The second method from within the CMC is the Ping menu option. The advantage of performing the command here is that communication of ping is tested and executed from that storage system. In addition, a specific network interface may be selected, the IP address or hostname may be used to validate name resolution, MTU and a repetition count. Select the storage system (either configured as a cluster member within a management group or as an Available System) and highlight the Network sub-tree option. On the TCP/IP tab, select the TCP/IP Tasks drop down and select Ping.

From within the HP StoreVirtual CMC.

With the CLiQ, the command to run Ping from the storage system is executed. Validation of communication from CLiQ source to storage node is performed with command completion. The result storage system source Ping will indicate additional networking troubleshooting.

cliq utility run=”ping -c 10 10.3.2.1” login=10.1.2.3 userName=admin passWord=secret

5. Ping is validated between storage systems, however, connectivity or specific functionality is impacted. No specific error message suggests a communication issue. What network communication across ports may be subject to networking security and firewall rules?

With HP StoreVirtual communication occurring between storage, client and management on an Ethernet network infrastructure, the transport layer further connects specific HP StoreVirtual services via TCP and UDP with incoming or outgoing requests.

Firewalls are typical network security systems that control incoming and outgoing network traffic based on a set of rules. These rules typically restrict sender and receiver network traffic by limiting the available TCP and UDP ports for communication. A firewall rule implemented from client to target could impact functionality of HP StoreVirtual storage if that rule impedes communication on that TCP or UDP port.

The HP StoreVirtual Storage TCP and UDP port usage detail can be found in the HP StoreVirtual Storage User Guide. Ensure that these ports are open to validate functionality. Specific exclusions within firewall rules may exclude HP StoreVirtual source/destination IPs, source/destination TCP/UDP ports for known services in order to facilitate communication. In some cases, a source/destination ports may be dynamic and random in which case IP source/destination inclusion rules may be easier to apply over specific dynamic ports.
6. **HP StoreVirtual storage is unable to communicate with other network services such as an email server.**

Validate that the networking configuration is valid including a valid gateway configuration. Sometimes, a default gateway is not set to ensure that iSCSI traffic is kept within the local network, however, specific services may depend upon a routable network. Implementations vary depending upon site requirements, however, basic networking rules apply and valid gateway definitions are required for HP StoreVirtual services dependent upon out-of-network communication. Valid routing must also exist within the network. Traceroute may be used to validate the network paths, understanding connection problems including packet loss and high latency by tracing the route from the source to destination.

**Conclusion**

This white paper outlined the important aspects of a network designed for HP StoreVirtual Storage: Resilient network design, low-latency networking, and iSCSI isolation from other services. When leveraging networking features like jumbo frames and flow control, make sure these features are turned on end-to-end on the entire network segment between application servers, switching infrastructure and HP StoreVirtual Storage. On storage systems, make sure bonding is used to increase bandwidth and resiliency. Also, on application servers, use available multipathing options for optimal performance and resiliency. Switch configuration is highly recommended for optimal settings.

**Learn more at**

[hp.com/go/StoreVirtual](http://hp.com/go/StoreVirtual)