



Alex Osuna
Rick Jooss
Tarik Jose Maluf
Patrick Strick

Windows Multipathing Options with IBM System Storage N series

Introduction

This IBM® Redpaper™ provides a description of the various multipathing options available for iSCSI and Fibre Channel SANs on Microsoft® Windows® in conjunction with Data ONTAP®. The pros and cons of each solution are discussed with the intention of helping the reader determine the best solution for the reader's particular environment. In this paper, we also introduce the concepts of multipathing, and we briefly outline the installation steps and describe the management interface for the Windows operating system (OS).

Overview

To have a highly available storage area network (SAN), steps must be taken so that no single failure results in an outage. In this paper we look at redundancy in the links connecting hosts to storage systems and the options available to achieve a highly available SAN infrastructure.

Multipathing is the ability to have multiple data paths from a server to a storage array. Multipathing protects against hardware failures (cable cuts, switches, HBAs, and so on) and can provide higher performance limits by utilizing the aggregate performance of multiple connections. When one path or connection becomes unavailable, the multipathing software automatically shifts the load to one of the other available paths. Active/active and active-passive are usually used to categorize multipathing. However, current midrange storage solutions support multiple active and multiple passive paths simultaneously. A multipathing solution is generally considered to be active/active when an I/O for a single LUN travels multiple paths simultaneously to its destination.

Figure 1 shows an example of a single multipath. The server has two HBA cards connected to two SAN switches, providing connectivity redundancy to the storage subsystem.

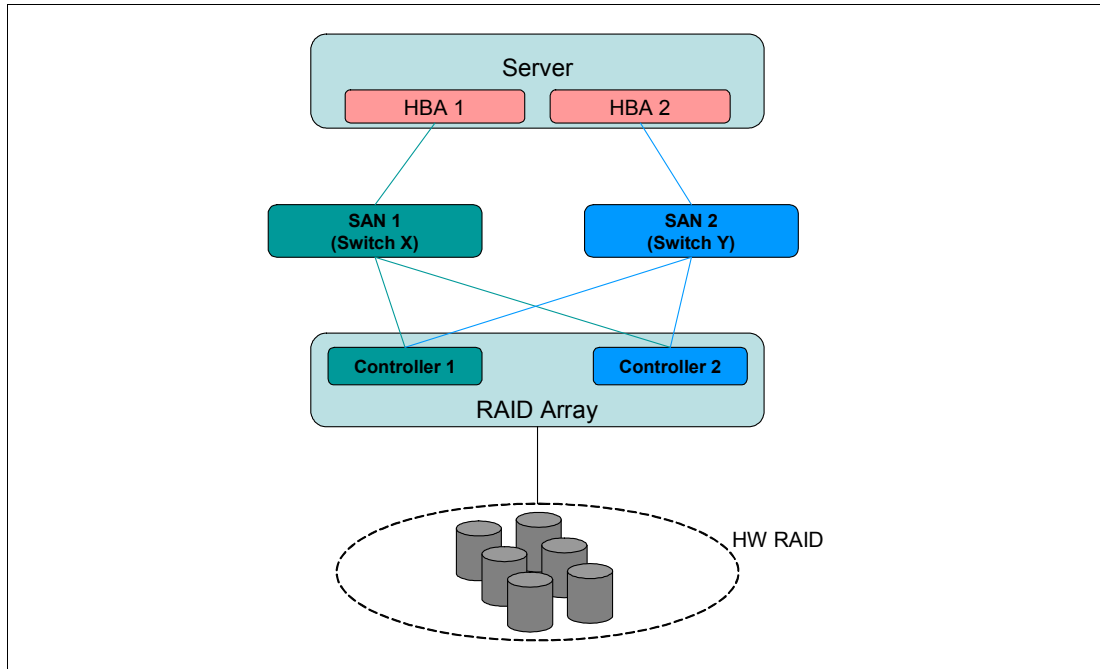


Figure 1 Multipath configuration with one RAID device and additional paths

Eliminating single points of failure

Electrical and mechanical components can always fail. The way to achieve high availability is by eliminating single points of failure in the environment. That way, when an individual component fails, the overall system continues to be available to users. Any multipathing solution should utilize separate adapters, cables, and switches to avoid a single point of failure.

Windows Storage Stack

When an application writes data to disk, that data flows through the host-side storage stack and through its storage interconnect (for example, parallel SCSI, Fibre Channel, iSCSI, and so on) to the storage array.

Figure 2 illustrates the storage stack for Microsoft Windows 2008.

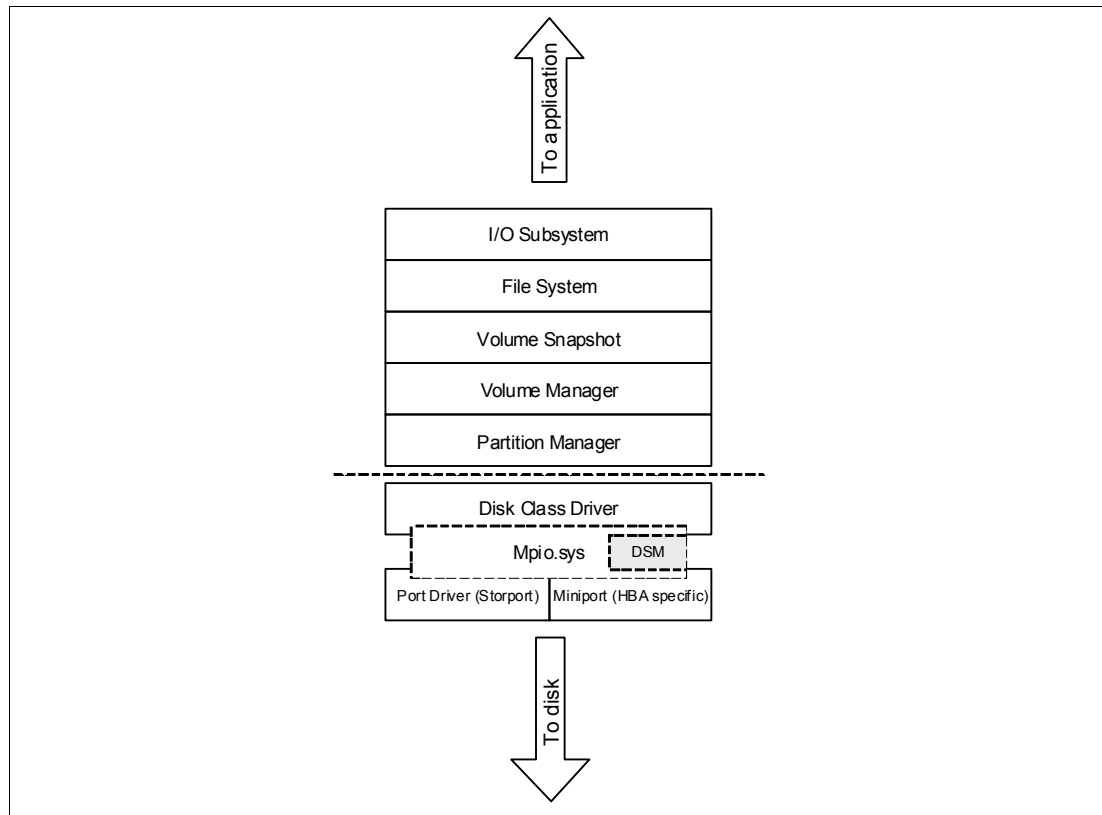


Figure 2 Microsoft Windows Storage Stack

Multipathing is achieved by sophistication at some layer of the storage stack. The application writes to a single file system of a raw device. The multipathing-capable layer receives the request and routes it to one of the underlying data paths. This routing is performed transparently to the other layers of the stack, both above and below the multipathing layer. There are various layers in which this split from a single path to multiple paths can occur. Each option has its advantages and limitations.

Link aggregation

One possible split point is at the NIC driver layer using TCP/IP link aggregation. Link aggregation is the technique of taking several distinct Ethernet links and making them appear as a single link. It is specified by the 802.3ad IEEE specification. Traffic is directed to one of the links in the group using a distribution algorithm. This technology is referred to by many names, including *channel bonding* and *teaming*. Link aggregation is not storage specific, and all types of network traffic can benefit from the multiple connections.

Note: The Microsoft iSCSI software initiator does *not* support link aggregation on the host. Link aggregation on the storage side (a VIF in Data ONTAP) is supported by both Microsoft and N series.

Figure 3 shows four separate network interfaces (e0a, e0b, e0c, and e0d) before they are grouped into a vif (virtual interface) link aggregation.

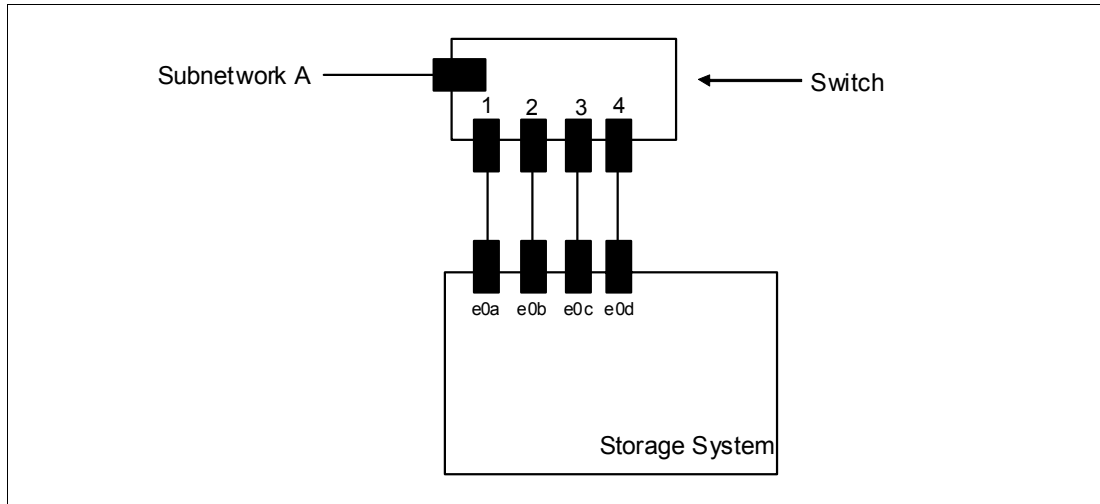


Figure 3 Network interfaces before link aggregation

Figure 4 shows the four network interfaces grouped into a single vif called Trunk1.

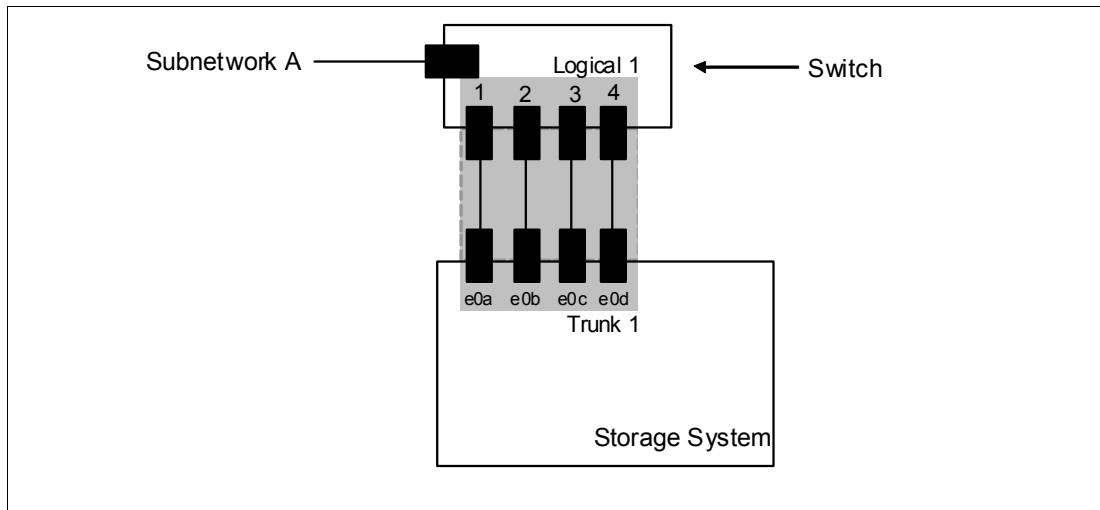


Figure 4 Network Interfaces after Link Aggregation

Advantages

There are some advantages of using link aggregation. Some of the main ones are:

- ▶ Transparent to all network protocols. The advantages of link aggregation are shared not just with iSCSI, but also with other network traffic such as NFS and CIFS.
- ▶ Well-known, mature technique.
- ▶ Available on both host and storage systems.

Disadvantages

Link aggregation has limitations and disadvantages, as listed below. Make sure that none of them will affect your environment before implementing it.

- ▶ It is not supported on the host with the Microsoft iSCSI software initiator.
- ▶ It is not possible to get aggregated throughput for a single host to a single storage system.
- ▶ It does not work with hardware iSCSI initiators (HBAs).
- ▶ Aggregated interfaces must be connected to the same network, often the same switch or card within a switch, thus limiting the physical isolation of the multiple paths
- ▶ It is dependent on aggregation-capable drivers and switches.

Multiconnection sessions

Multiconnection sessions (MCS) are part of the iSCSI specification. They create multiple paths within a single iSCSI session using separate TCP connections. Both the iSCSI initiator (host) and iSCSI target (storage system) need to support multiconnection sessions for them to be used. Current versions of Data ONTAP and Microsoft Windows support multiconnection sessions. As of Data ONTAP 7.3, the default maximum number of connections per session is 32. Refer to the *N series Interoperability Matrix* on the following website for the most up-to-date information regarding supported Data ONTAP and initiator releases.

<http://www-03.ibm.com/systems/storage/network/interophome.html>

Figure 5 illustrates where iSCSI multiconnection sessions fit into the storage stack.

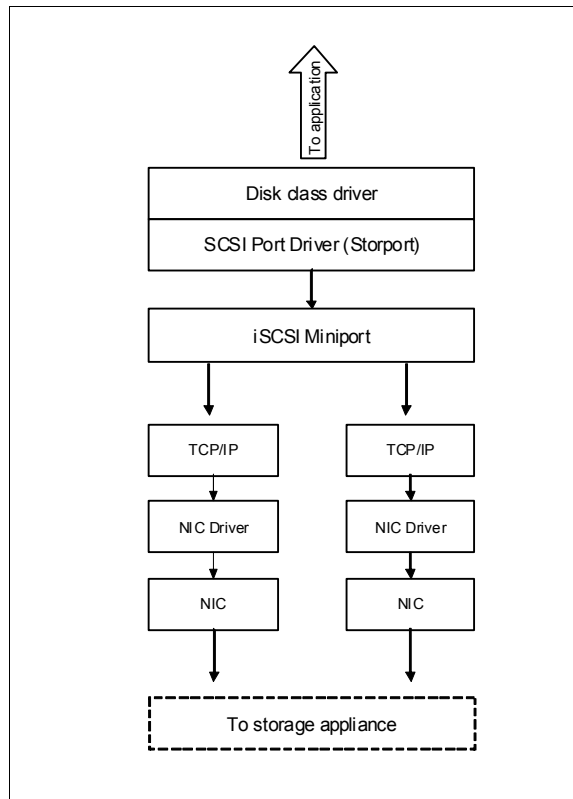


Figure 5 Microsoft Windows Storage Stack with multiconnection sessions

iSCSI HBAs do not currently support multiconnection sessions. Given that the iSCSI initiator portion of the stack resides on the HBA, implementing multiconnection sessions across HBAs poses challenges. The Microsoft iSCSI initiator supports multiconnection sessions across a single path or multiple paths. SnapDrive® for Windows will work with preexisting iSCSI connections that have multiconnection sessions enabled, but will not create a multiconnection session-enabled connection and has no knowledge of those created manually.

iSCSI multiconnection sessions can be performed over a single target or initiator port or can utilize multiple ports on either end. If multiple target ports are used, all target interfaces for the connection must be in the same target portal group. By default, each interface is in its own target portal group.

Attention: When used with multi-connection sessions, the Windows iSCSI software initiator creates a persistent association between the target portal group tag value and the target interfaces. If the tag value changes while an iSCSI session is active, the initiator will be able to recover only one connection for a session. To recover the remaining connections, you must refresh the initiator's target information.

You can view a list of the current target portal groups using the `iscsi tpgroup show` command (Example 1). For each target portal group, the list includes the name, tag, and interfaces that belong to the group.

Example 1 Target portal groups

```
itsotuc*> iscsi tpgroup show
TPGTag  Name                Member Interfaces
1000    e0a_default         e0a
1001    e0b_default         e0b
1002    e0c_default         e0c
1003    e0d_default         e0d
```

Although technically possible, mixing iSCSI multiconnection sessions and MPIO multipathing styles to the same LUN is not supported by Microsoft or IBM.

Advantages

Advantages of multiconnection sessions are:

- ▶ Part of the iSCSI specification
- ▶ No extra vendor multipathing software required
- ▶ No dependency on aggregation-capable Ethernet infrastructure

Disadvantages

The MCS disadvantages listed below can affect your environment if not previously mapped:

- ▶ Not manageable by SnapDrive iSCSI connection wizard.
- ▶ Not currently supported by iSCSI initiator HBAs.

- ▶ Not supported for MS software initiator boot. For iSCSI software booted hosts, Microsoft does not support MCS. The iSCSI boot leading connection cannot be protected, which could lead to disconnection from the boot LUN, causing a kernel crash. Due to this MCS limitation, MPIO must be used as the high-availability solution. Mixing MPIO and MCS is not supported.
- ▶ Load balance policy is set on a per-session basis. All LUNs in an iSCSI session share the same load balance policy.

Multipath I/O (MPIO)

The classic way to do multipathing is to insert a separate multipathing layer into the storage stack. This method is not specific to any underlying transport and is the standard way to achieve multipath access to iSCSI, Fibre Channel, and even parallel and serial SCSI targets. There are multiple implementations of this type of multipathing on various operating systems. With Microsoft Windows, each storage vendor supplies a device-specific module (DSM) for its storage array. In addition, Microsoft provides its own DSM (iSCSI only for Windows 2000 Server and Windows 2003 Server, both FCP and iSCSI for Windows Server 2008).

Figure 6 illustrates how MPIO fits into the storage stack.

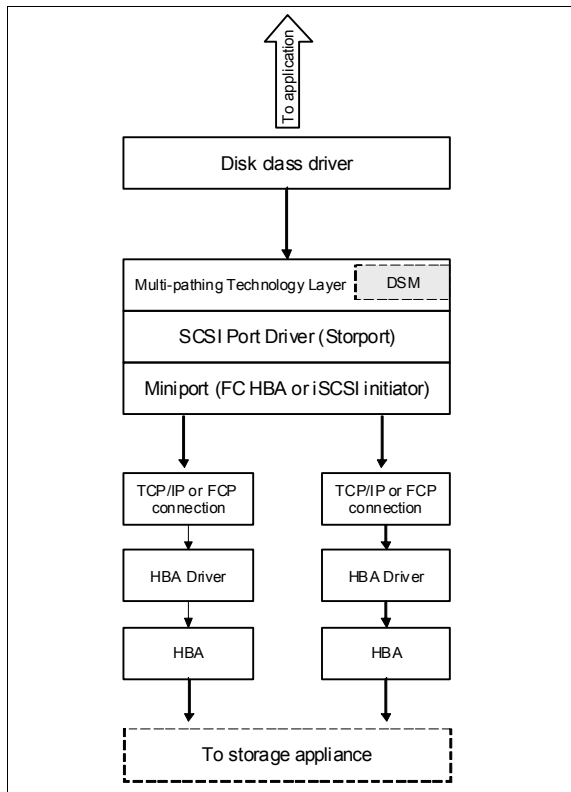


Figure 6 Microsoft Windows storage stack with MPIO

Since MPIO occurs above the miniport driver layer, the MPIO driver only sees SCSI devices and does not know about the transport protocol. This allows Fibre Channel and iSCSI paths to the same LUN to be mixed. Since the protocols have different access characteristics and performance, we recommend that, if they are mixed, they be used in an active-passive configuration in which one takes over if the other fails.

Note: The Data ONTAP DSM claims all LUNs that it discovers on IBM N series storage systems. These LUNs have the vendor identifier and product identifier (VID/PID) pair. Other DSMs can be used on the same Windows host to claim LUNs from other storage systems with other VID/PID values.

The Microsoft iSCSI Initiator for Windows Server 2003 includes a DSM (msiscsidsm) that can manage the iSCSI paths that use the software initiator. The Data ONTAP DSM and msiscsidsm can co-exist as long as both versions are listed on the appropriate support matrixes.

Microsoft Windows Server 2008 includes a DSM (msdsm) that can manage FC and iSCSI paths. The msdsm claims any devices not claimed by other DSMs. The Data ONTAP DSM and the msdsm can co-exist as long as the configuration is listed on the appropriate support matrixes.

Advantages

Some of the main advantages of MPIO are:

- ▶ No dependency on aggregation-capable Ethernet infrastructure.
- ▶ Very mature implementation.
- ▶ Supports software and hardware iSCSI initiators (HBAs).
- ▶ Can mix protocols between paths (for example, iSCSI and FCP).
- ▶ Each LUN can have its own load balance policy.

One of the disadvantages is that an extra multipathing technology layer is required.

Asymmetric logical unit access (ALUA)

Not all paths available to a LUN necessarily have equal access. In a N series high-availability cluster pair, one node owns the LUN, but, in the case of Fibre Channel, ports on both nodes provide access. Paths using the cluster partner cross the cluster interconnect to reach the LUN. Paths to ports on the owning node are called *primary* or *optimized*. Paths to ports on the cluster partner are commonly called *unoptimized*, *partner*, *proxy*, or *secondary*. Figure 7 illustrates a primary and partner path. The blue-shaded lines are the paths over which data traffic flows.

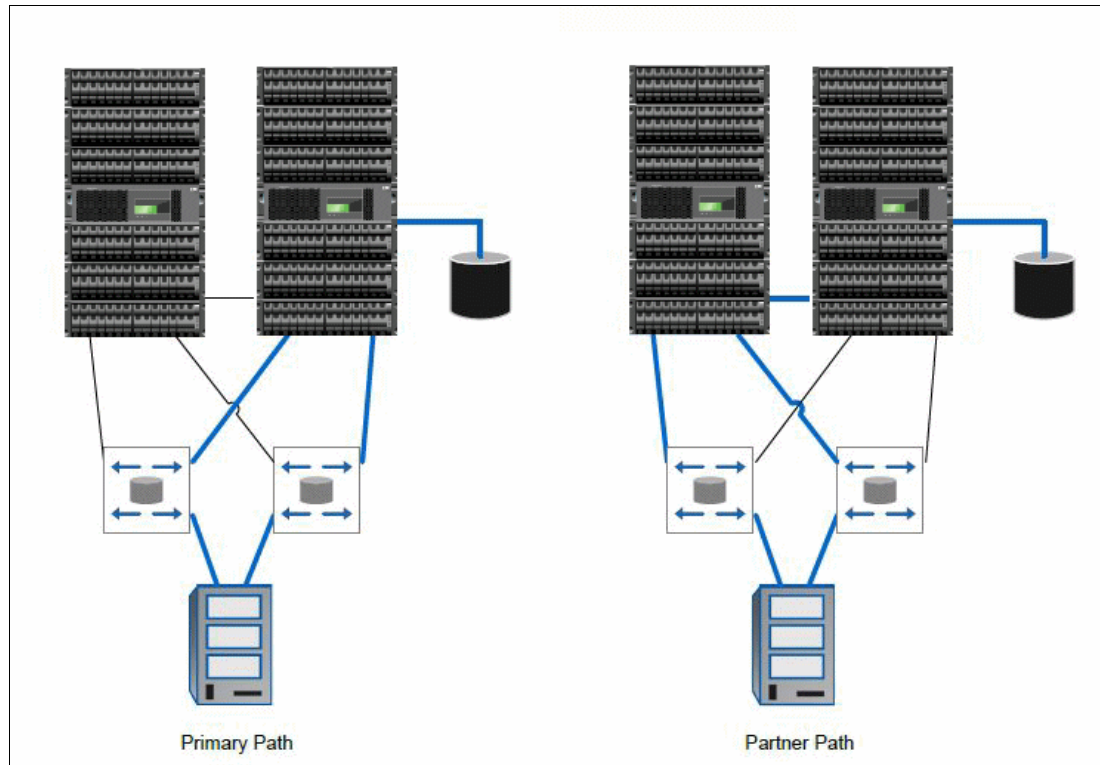


Figure 7 Fibre Channel path failover

So that data traverses the primary paths only, the host must communicate with the storage controller to determine which paths are primary and which are proxy. This has traditionally been done with vendor-specific multipathing software. A standardized method was added to the SCSI specification called Asymmetric Logical Unit Access (ALUA) and was implemented in Data ONTAP 7.2 and Windows 2008. ALUA allows the initiator to query the target about path attributes, such as which paths are primary or secondary. As a result, multipathing software can be developed to support any array that uses ALUA.

Because with iSCSI there is no proxy path and because link failover operates differently from Fibre Channel, ALUA is not supported on iSCSI connections.

Windows hosts support ALUA with certain combinations of Windows, Data ONTAP, Host Utilities, and MPIO software.

Table 1 lists configurations that support ALUA. Use the interoperability matrix to determine a supported combination of Windows, Data ONTAP, host utilities, and MPIO software. Then enable or disable ALUA based on the information in Table 1.

Table 1 Windows-supported ALUA configurations

Windows version	MPIO software	Minimum data ONTAP	Supported?
Server 2008 Server 2008 R2	Microsoft DSM (msdsm)	7.3.0	Yes
Server 2008 SP2 Server 2008 R2	Data ONTAP DSM 3.4 and later	7.3.2	Yes
Server 2008 Server 2008 R2	Data ONTAP DSM 3.3.1 and earlier	None	No
Server 2008 Server 2008 R2	Veritas DSM from Storage Foundation for Windows 5.1 and earlier	None	No
Server 2003 SP2 Server 2003 R2	Data ONTAP DSM 3.4 and later	7.3.2	Yes
Server 2003	Data ONTAP DSM 3.3.1 and earlier	None	No
Server 2003	Veritas DSM from Storage Foundation for Windows 5.1 and earlier	None	No

Attention: Ensure that your host supports ALUA before enabling it. Enabling ALUA for a host that does not support it can cause host failures during cluster failover.

If ALUA is not automatically enabled for an igroup, you can manually set the `alua` option to `yes`.

To check and enable ALUA:

1. Check whether ALUA is enabled by entering the `igroup show -v igroup_name` command (Example 2).

Example 2 Checking whether ALUA is enabled in a specific igroup

```
itsotuc*> igroup show -v lnxgrp
  lnxgrp (FCP):
  OS Type: linux
  Member: 10:00:00:00:c9:6b:76:49 (logged in on: vtic, 0a)
  ALUA: No
```

2. If ALUA is not enabled, enter the following command to enable it:

```
igroup set igroup_name alua yes
```

Data ONTAP DSM and Microsoft DSM (Windows 2008)

The Data ONTAP DSM for Windows MPIO is a device-specific module that works with Microsoft Windows MPIO drivers (`mpdev.sys`, `mpio.sys`, and `mpspfltr.sys`) to manage multiple paths between IBM N series storage systems and Windows host computers.

Two primary multipathing options are available for a Windows 2008 or newer host:

- ▶ The built-in Microsoft MPIO feature
- ▶ The Data ONTAP Device Specific Module

Windows Server 2008 introduced a native MPIO feature that utilizes ALUA for path selection. It is enabled as a feature in Windows 2008 and supports both Fibre Channel and iSCSI. The standard set of load balance policies is available, including failover only, failback, round-robin, round-robin with subset, least queue depth, and weighted paths. Windows Server 2008 R2 also adds the least blocks policy. The default policy for FCP connections is round-robin with subset, and the default for iSCSI is failover.

Microsoft Windows Server 2008 includes a DSM (msdsm) that can manage FC and iSCSI paths. The msdsm claims any devices not claimed by other DSMs. The Data ONTAP DSM and the msdsm can co-exist as long as the configuration is listed on the appropriate support matrixes.

If the Microsoft MPIO feature is used for LUNs connected over Fibre Channel, ALUA must be enabled on the Data ONTAP igroup to which its initiator connects. Example 3 shows ALUA being enabled in igroup ThinkCentre. Enablement should be done prior to the LUN being discovered by the host.

Example 3 Enabling ALUA in specific igroup

```
itsonas*> igroup set ThinkCentre alua yes
itsonas*> igroup show -v ThinkCentre
  ThinkCentre (FCP):
    OS Type: windows
    Member: 21:00:00:e0:8b:14:f6:85 (logged in on: vtic, 0b)
    ALUA: Yes
```

The Data ONTAP DSM provides standard load balance policies and adds an easy-to-use interface (both GUI and CLI).

Attention: You can have a maximum of 32 paths to a LUN. This maximum applies to any mix of FC and iSCSI paths. This is a limitation of the Windows MPIO layer. Additional paths can be created, but are not claimed by the DSM. Exceeding 32 paths leads to unpredictable behavior of the Windows MPIO layer and possible data loss.

Load balance policies

When multiple paths to a LUN are available, a consistent method of utilizing those paths should be determined. This is called the load balance policy. There are six standard policies in Windows Server 2008, and they apply to multiconnection sessions and MPIO:

- ▶ **Failover:** No load balancing is performed. The application specifies a primary path and a set of standby paths. The primary path is used for processing device requests. If the primary path fails, one of the standby paths is used. Standby paths must be listed in decreasing order of preference (the most preferred path first).
- ▶ **Failback:** Failback is the ability to dedicate I/O to a preferred path whenever it is functioning. If the preferred path fails, I/O is directed to an alternate path until function is restored to the preferred path, but I/O automatically switches back to the preferred path when function is restored.
- ▶ **Round robin:** The DSM uses all available paths for I/O in a balanced, round-robin fashion.

- ▶ Round-robin with a subset of paths: The application specifies a set of paths to be used in a round-robin fashion and a set of standby paths. The DSM uses paths from the primary pool of paths for processing requests, as long as at least one of the paths is available.

The DSM uses a standby path only when all primary paths fail. Standby paths must be listed in decreasing order of preference (most preferred path first). If one or more of the primary paths become available, DSM uses the standby paths in their order of preference.

Figure 8 shows an example of four paths (A, B, C, and D) connected to a host. Paths A, B, and C are listed as primary paths, and D is a standby path. The DSM chooses a path from A, B, and C in round-robin fashion as long as at least one of them is available. If all three fail, the DSM uses D, the standby path. If A, B, or C become available, DSM stops using D and switches to available paths among A, B, and C.

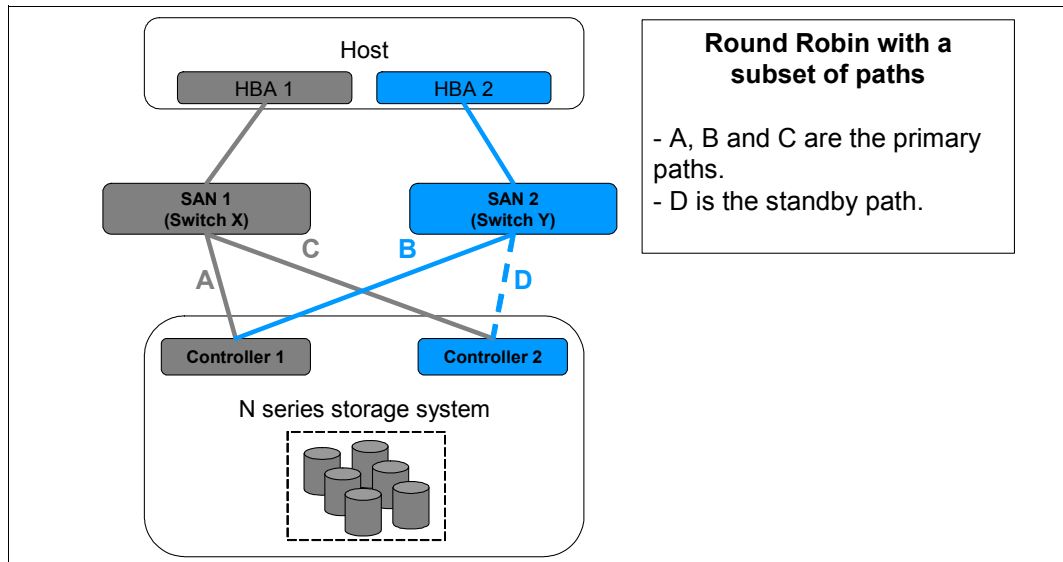


Figure 8 Round Robin with a subset of paths with no degradation

If the three primary paths fails, the standby becomes available (Figure 9).

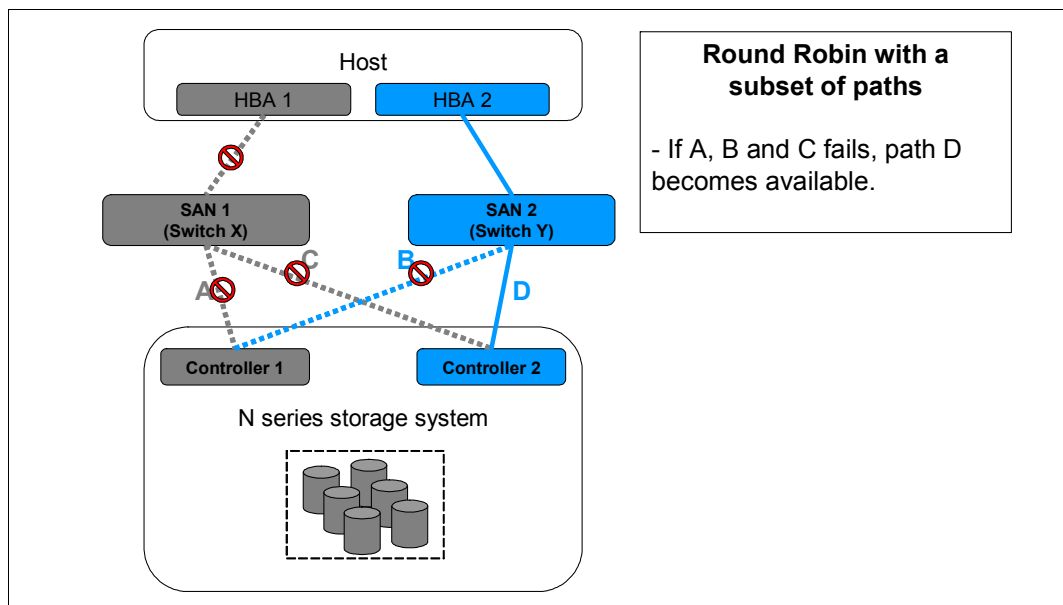


Figure 9 Round-robin with standby path enabled

- ▶ Dynamic least queue depth: The DSM routes I/O to the path with the least number of outstanding requests.
- ▶ Weighted path: The application assigns weights to each path. The weight indicates the relative priority of a given path. The larger the number, the lower the priority. The DSM chooses the path that has the least weight from among the available paths.

Note: The Microsoft DSM preserves load balance settings even after the computer is restarted. When no policy has been set by a management application, the default policy that is used by the DSM is either round-robin, when the storage controller follows the true active/active model, or simple failover in the case of storage controllers that support the SPC-3 ALUA model. With simple failover, any one of the available paths can be used as the primary path, and remaining paths are used as standby paths.

iSCSI network design suggestions

iSCSI is a standard-based transport protocol for connecting storage devices over a TCP/IP network. To the operating system or application that uses the storage, an iSCSI device appears to be a local SCSI disk drive. The iSCSI protocol defines the way that standard SCSI commands and responses are encapsulated into TCP/IP packets and transmitted over an Ethernet network.

The user of iSCSI storage is an initiator. The provider of iSCSI storage is a target. For N series systems, each host has one or more initiators, and the N series storage system is the target. The set of initiators, targets, and the connecting network makes up an iSCSI storage area network (SAN).

The simplest iSCSI system consists of a single host connected to a single storage system through a direct Ethernet crossover cable (Figure 10). A complex iSCSI SAN could contain many hosts and storage systems connected by one or more switched Ethernet networks.

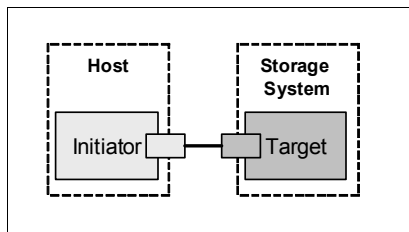


Figure 10 iSCSI direct-attached connection with a crossover cable

iSCSI network topologies

The iSCSI protocol is defined by RFC 3270, published by the Internet Engineering Task Force. A copy of the standard can be obtained from the URL below:

<http://www.ietf.org/rfc/rfc3270.txt>

The first decision that a customer needs to make is whether to run iSCSI traffic over a physically separate dedicated network. A dedicated iSCSI Ethernet infrastructure can include its own switches or VLANs. For smaller configurations, hosts can connect directly to non-clustered storage systems using crossover cables.

Note: We strongly suggest that if using multiple paths or sessions with iSCSI, that each path be isolated on its own subnet.

On the storage system, the iSCSI service should be disabled on network interfaces that will not be used for iSCSI sessions. Once disabled, the service rejects subsequent attempts to establish new iSCSI sessions over that interface. This increases security by only allowing iSCSI connections on predetermined ports.

Starting in Data ONTAP 7.3, iSCSI access lists have been implemented to give more granular control. Access can be granted to specific initiators over specific storage interfaces. Use the **iscsi interface access list add** command to create access lists (Example 4).

Example 4 Creating and showing iSCSI access lists.

```
itsotuc*> iscsi interface accesslist add iqn.1991-05.com.microsoft:win.itso.com -a
Adding interface e0a to the accesslist for iqn.1991-05.com.microsoft:win.itso.com
Adding interface e0b to the accesslist for iqn.1991-05.com.microsoft:win.itso.com
Adding interface e0c to the accesslist for iqn.1991-05.com.microsoft:win.itso.com
Adding interface e0d to the accesslist for iqn.1991-05.com.microsoft:win.itso.com

itsotuc*> iscsi interface accesslist show
Initiator Name          Access List
iqn.1991-05.com.microsoft:win-oz15boz5iej.itso.com    e0a, e0b, e0c,e0d
```

We suggest a network topology that minimizes the risk of unauthorized access to or modification of data as it traverses the network. You can limit access to data through the use of direct cabling, switched network environments, virtual LANs (VLANs), and dedicated storage network interfaces where appropriate.

Three topologies can be used to design iSCSI networks. Each has advantages and disadvantages.

Shared switched iSCSI network

Shared configurations run both iSCSI and other Ethernet traffic over the same physical network. There might be performance and throughput issues, especially with 10 and 100 megabit cards and switches. Because it is shared with other traffic and hosts, this option is less secure than a dedicated network. Implement available security features to reduce exposure.

Use VLANs to segregate iSCSI from other network traffic in shared configurations. The VLAN provides additional security and simplifies network troubleshooting. A N series storage system can be configured as a VLAN-aware device that processes VLAN tags, or the VLAN can be managed at the switch port level and be transparent to the storage system.

Figure 11 shows a shared switch network with iSCSI connections.

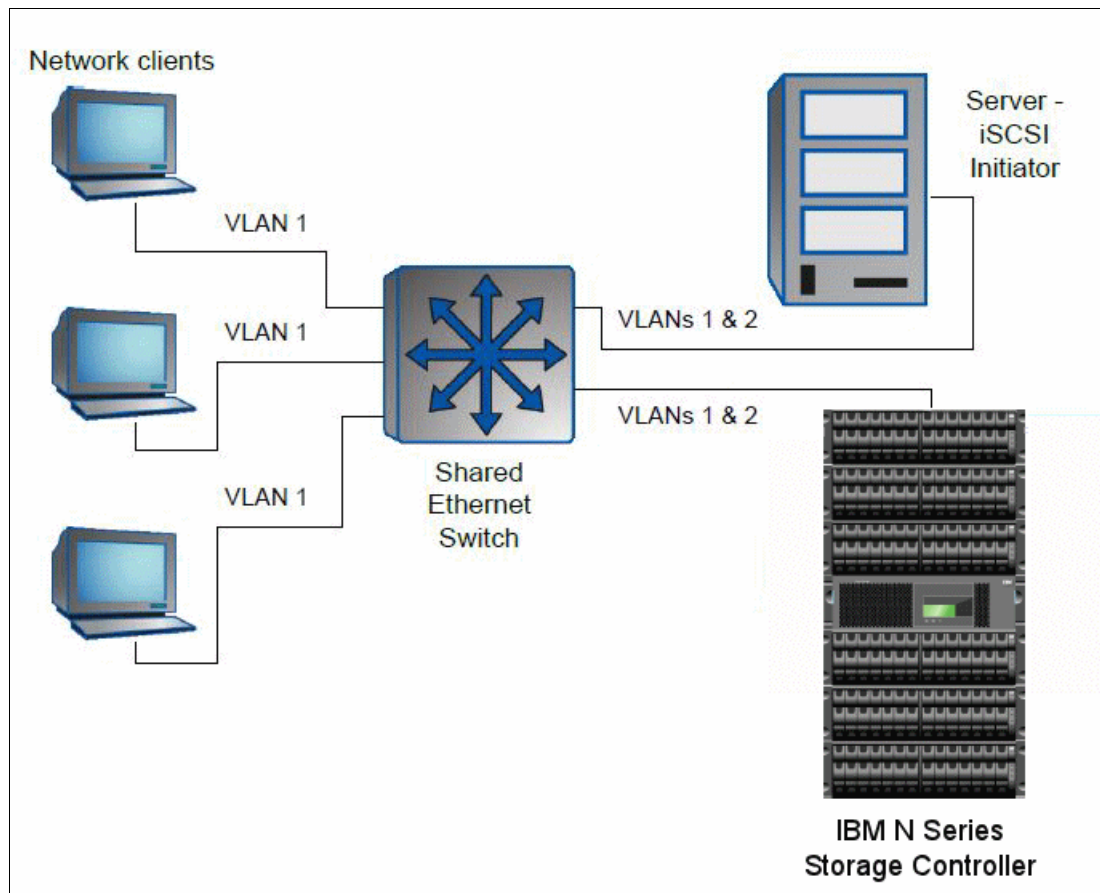


Figure 11 Shared switched network

Advantages

Shared switched network advantages are:

- ▶ Link aggregation is possible if supported by the switch.
- ▶ Multiple switches can be used for redundancy.
- ▶ The number of hosts and storage systems is limited only by the available switch ports.
- ▶ The existing Ethernet switch infrastructure is utilized, saving money.

Disadvantages

The disadvantages of using a shared switched network are:

- ▶ Network bandwidth is shared across public LAN and iSCSI traffic unless initiators and targets are connected to the same switch.
- ▶ It requires switches capable of implementing VLANs.
- ▶ It is less secure than a dedicated switched iSCSI network (if access lists are not implemented).

Dedicated switched iSCSI network

In this configuration, Ethernet switches and cables are dedicated to carrying iSCSI traffic between iSCSI hosts and storage systems. This configuration is similar to a Fibre Channel

fabric in that only iSCSI and related traffic use this dedicated infrastructure. There are additional costs for dedicated Ethernet equipment compared to running iSCSI traffic over the existing Ethernet infrastructure, but you gain security and performance improvements.

Figure 12 shows a dedicated switched iSCSI network.

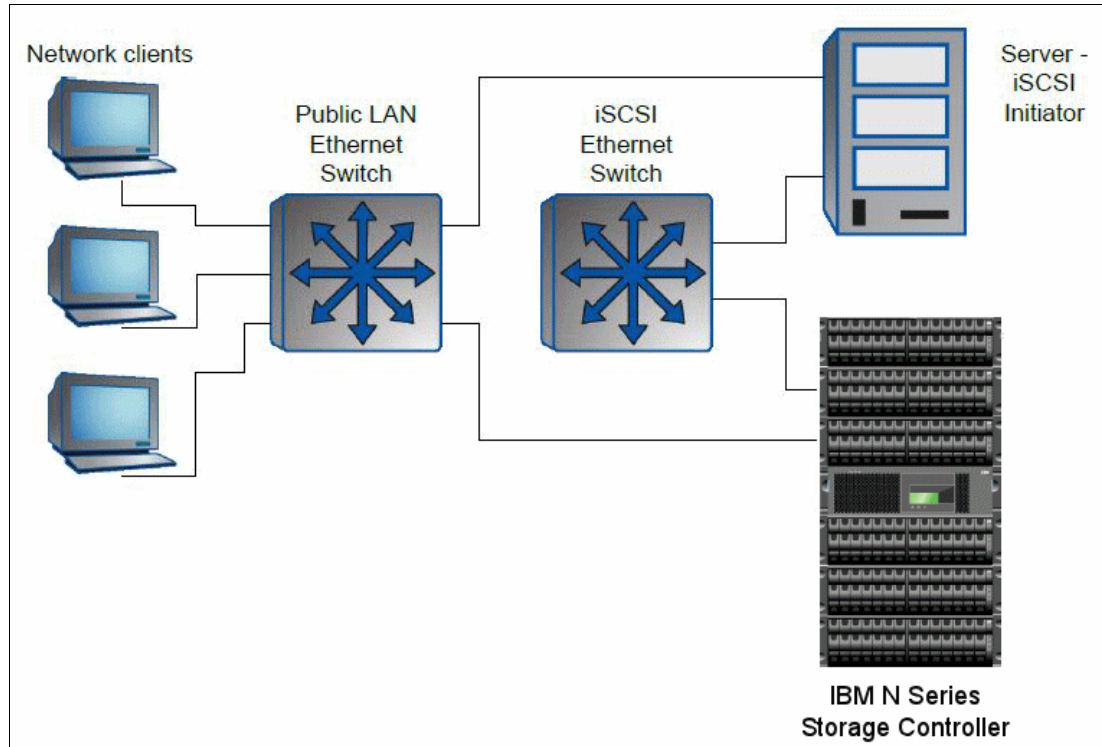


Figure 12 Dedicated switched iSCSI network

Advantages

The advantages are:

- ▶ It is very secure. iSCSI traffic is isolated from public LAN traffic.
- ▶ The full bandwidth of the link is available.
- ▶ Link aggregation is possible if supported by the switch.
- ▶ Multiple switches can be used for redundancy.
- ▶ The number of hosts and storage systems is limited only by available switch ports.
- ▶ You can use less expensive, unmanaged switches because VLANs are not needed.

Disadvantages

There are some disadvantages of using a dedicated switched network:

- ▶ One or more switches must be dedicated to the iSCSI network.
- ▶ Configuration and administration are more complex than direct connection.

Direct-connect iSCSI network

The host is connected to the storage system using crossover cables. No Ethernet switches are involved. This is the most physically secure configuration and allows full bandwidth between the initiator and target.

Figure 13 shows a direct-connect iSCSI network.

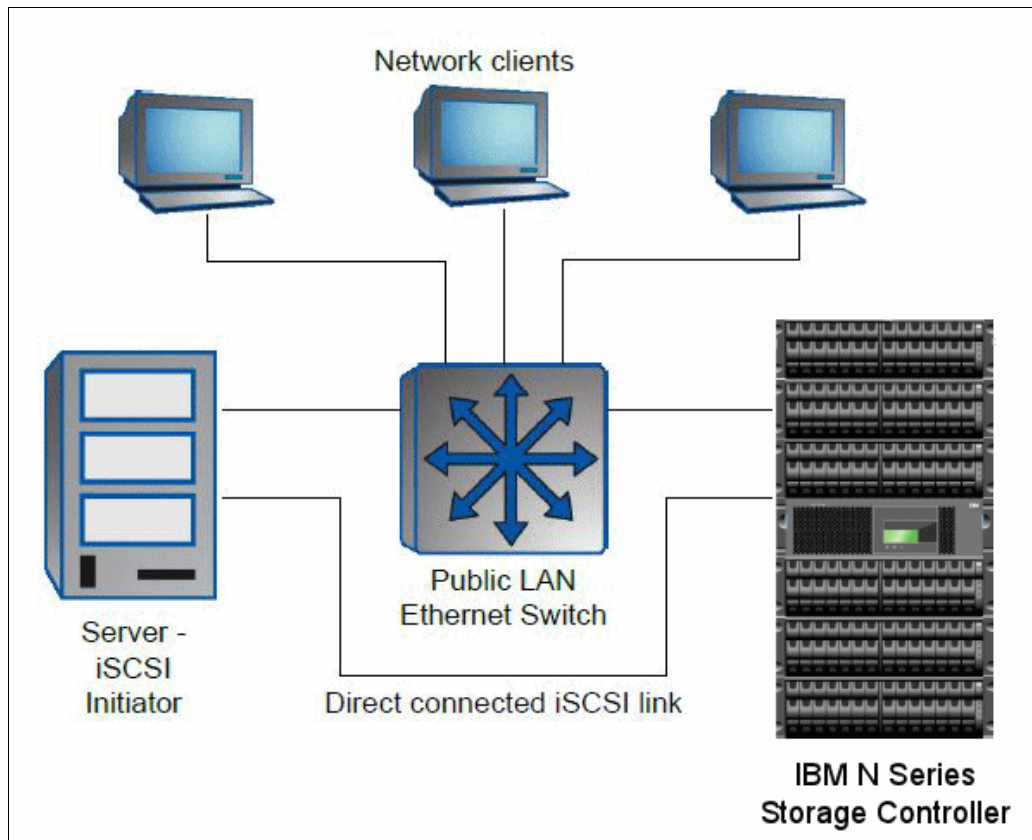


Figure 13 Direct connected iSCSI network

Advantages

Direct-connected iSCSI networks advantages are:

- ▶ It is low cost. No Ethernet switches are required.
- ▶ It is very secure. There is no chance of a man-in-the-middle attack.
- ▶ It is easy to set up and maintain.
- ▶ Full bandwidth of link is available.

Disadvantages

Use of the direct connected iSCSI network can bring some disadvantages as listed below:

- ▶ The number of initiators and/or paths is limited by the number of available network ports.
- ▶ Limited distance between initiator and target.
- ▶ Not supported with storage HA failover.

iSCSI network configuration

The iSCSI protocol is a licensed service on a N series storage system. You must enter the iSCSI license key to enable the protocol and start the service on each storage system.

The iSCSI protocol is implemented as a software driver on the storage system. You can configure the iSCSI service to use any or all of the Ethernet ports on the storage system. No

special networking hardware is required, but you should use dedicated gigabit Ethernet ports for iSCSI traffic if possible.

On a host system, the LUN appears as a local SCSI disk. The host is responsible for partitioning and formatting the LUN. The storage system does not understand the contents of a LUN. To the storage system, a LUN is just a set of blocks of arbitrary data.

The iSCSI protocol and the Fibre Channel protocol (FCP) both encapsulate SCSI-3 commands and send them over a network. From the perspective of an application running on a host, the LUNs exposed by both protocols appear as local SCSI disks. Applications should neither know nor care where the storage is actually located and how it is delivered.

Fibre Channel Protocol based SANs always use a dedicated Fibre Channel network. Fibre Channel networks require host bus adapters (HBAs), special copper or optical cables, and Fibre Channel switches for switched configurations. Current Fibre Channel networks operate at 1, 2, or 4 gigabits per second.

iSCSI SANs operate over standard Ethernet networks. The network can be a separate, dedicated network, or iSCSI traffic can be carried over a shared Ethernet network. iSCSI networks typically use 1 gigabit per second Ethernet. It is possible to use 100 megabit or even 10 megabit hardware, but performance will suffer.

Using jumbo frames

By default, Ethernet sends up to 1,500 bytes of data in a single frame. This works well for applications that send small amounts of data, such as client applications. However, for transferring larger blocks of data, as is common in iSCSI, a larger frame size is more efficient.

The term *jumbo frame* typically refers to Ethernet frames with 9000 bytes of data, although it technically applies to any size larger than 1,500 bytes. Unlike the standard frame size, there is no standard size for a jumbo frame. Each network device must typically be configured with the specific maximum transmission unit size that will be used. Therefore, each network device must support the same size for jumbo frames. N series storage systems support jumbo frames on all 1 and 10 gigabit Ethernet interfaces.

Jumbo frames can be used for all Gigabit and 10 Gigabit Ethernet interfaces that are supported on your storage system. The interfaces must be operating at or above 1,000 Mbps.

You can set up jumbo frames on your storage system in the following two ways:

- ▶ During initial setup, the **setup** command prompts you to configure jumbo frames if you have an interface that supports jumbo frames on your storage system.
- ▶ If your system is already running, you can enable jumbo frames by setting the MTU size on an interface. Example 5, shows the e0d card being set the MTU size of 8192 bytes.

Example 5 Changing MTU size in an Ethernet interface

```
itsotuc*> ifconfig e0d mtusize 8192
```

N series host utilities

N series provides a SAN Host Utilities kit for every supported OS. This is a set of data collection applications and configuration scripts. These include SCSI and path timeout values and path retry counts. Also included are tools to improve the supportability of the host in a

N series SAN environment, such as gathering host configuration and logs and viewing the details of all N series presented LUNs.

The Host Utilities is a set of software programs and documentation that enable you to connect host computers to N series storage systems.

The Host Utilities include the following components:

- ▶ An installation program that sets required parameters on the host computer and on certain host bus adapters (HBAs).
- ▶ Diagnostic programs for displaying important information about the host, HBAs, Fibre Channel switches, and storage systems in your network.
- ▶ A program (`san_version.exe`) that reports the Host Utilities version and displays driver version information for Emulex and QLogic HBAs in the host.
- ▶ A program (`mbra1ign.exe`) to detect and correct VHD alignment problems for Hyper-V virtual machines.
- ▶ The Host Utilities support multiple Fibre Channel (FC) paths, multiple iSCSI paths, or a combination of FC and iSCSI paths.

FCP Fabric design recommendations

Of the multipathing topics that were discussed, only MPIO is applicable to Fibre Channel. Three general topologies are available in a Fibre Channel environment:

- ▶ **Direct-attached:** The initiator and target are connected directly by a cable. With IBM N series, there are two possible configurations, single controller and HA controller.
 - Figure 14 shows a direct-attached single controller configuration.

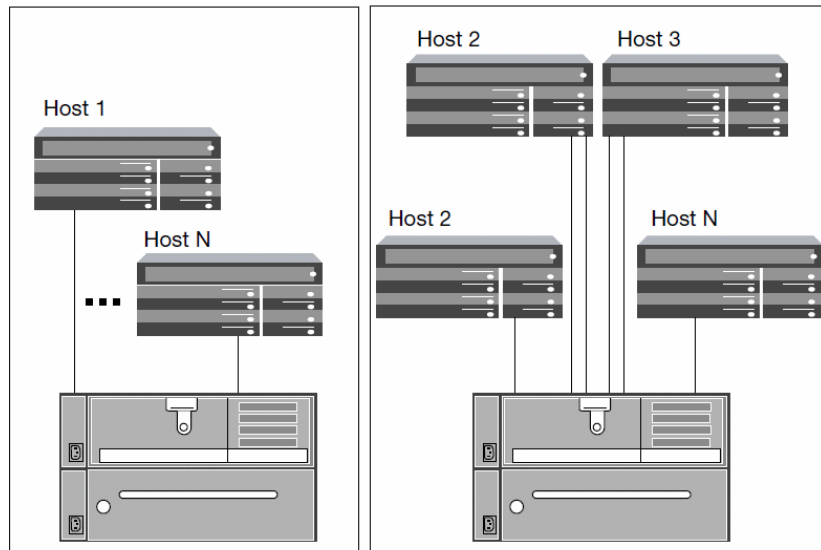


Figure 14 Direct-attached single controller configuration

- Figure 15 shows a direct-attached high availability controller.

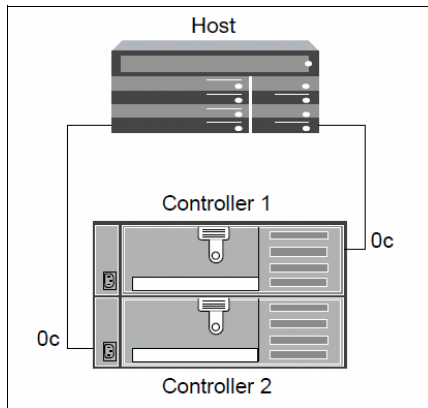


Figure 15 Direct-attached HA configuration

- ▶ Single fabric: All ports of the initiator and target connect to a single switch or fabric of switches. With IBM N series, there are two possible configurations, single controller and HA controller.
 - Figure 16 shows a single fabric with a single controller configuration.

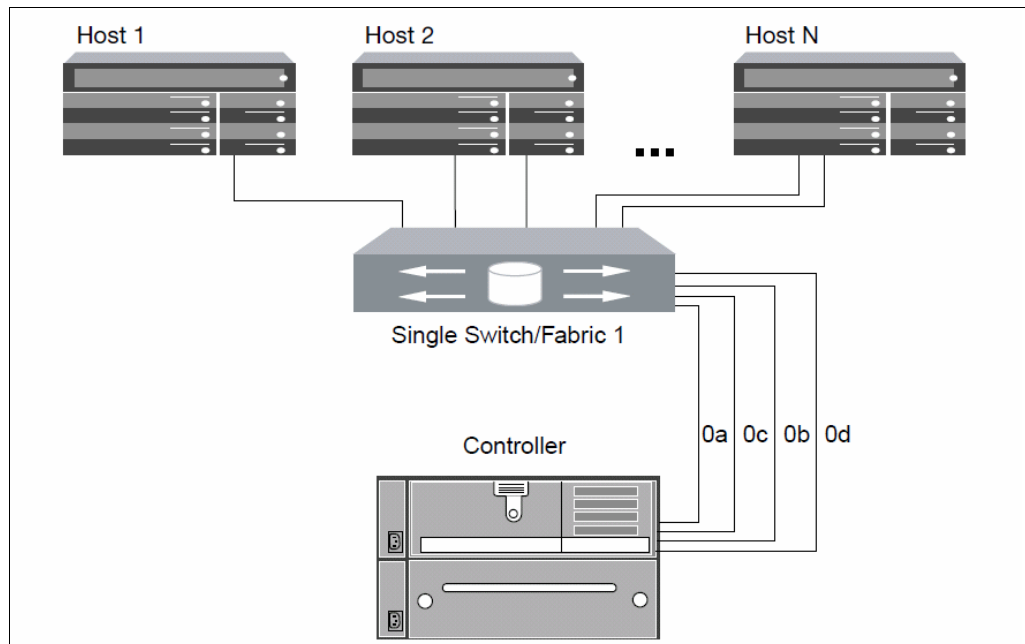


Figure 16 Single fabric single controller configuration

- Figure 17 shows a single fabric with high availability configuration.

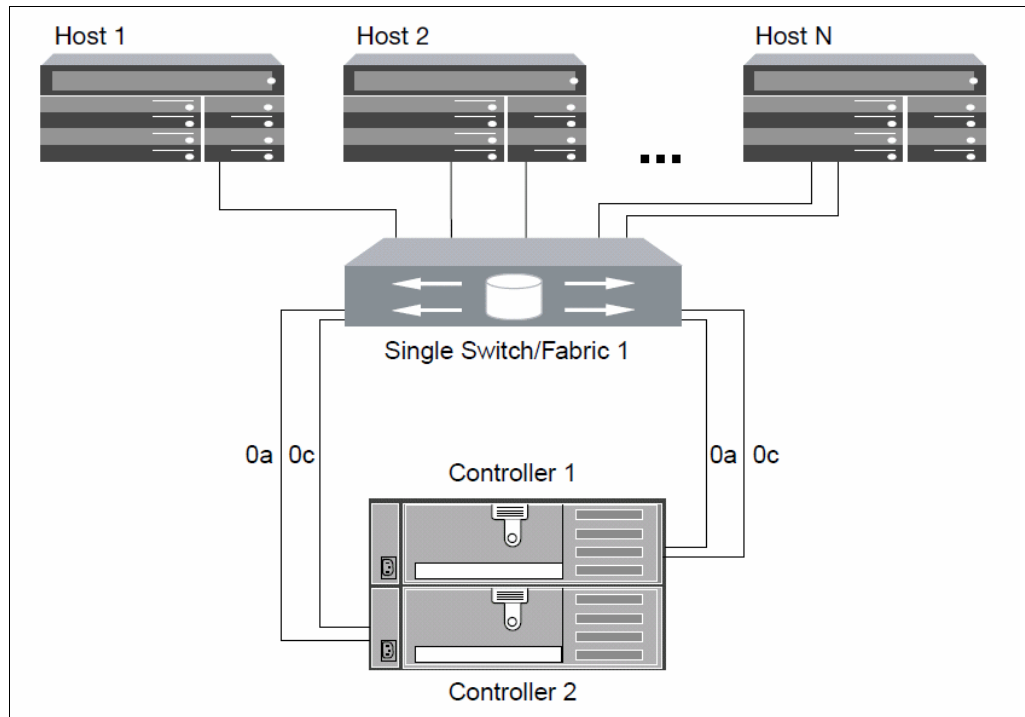


Figure 17 Single fabric HA configuration

- **Multifabric:** Some ports of the initiator or target connect to separate fabrics for redundancy. Figure 18 shows a multifabric HA configuration.

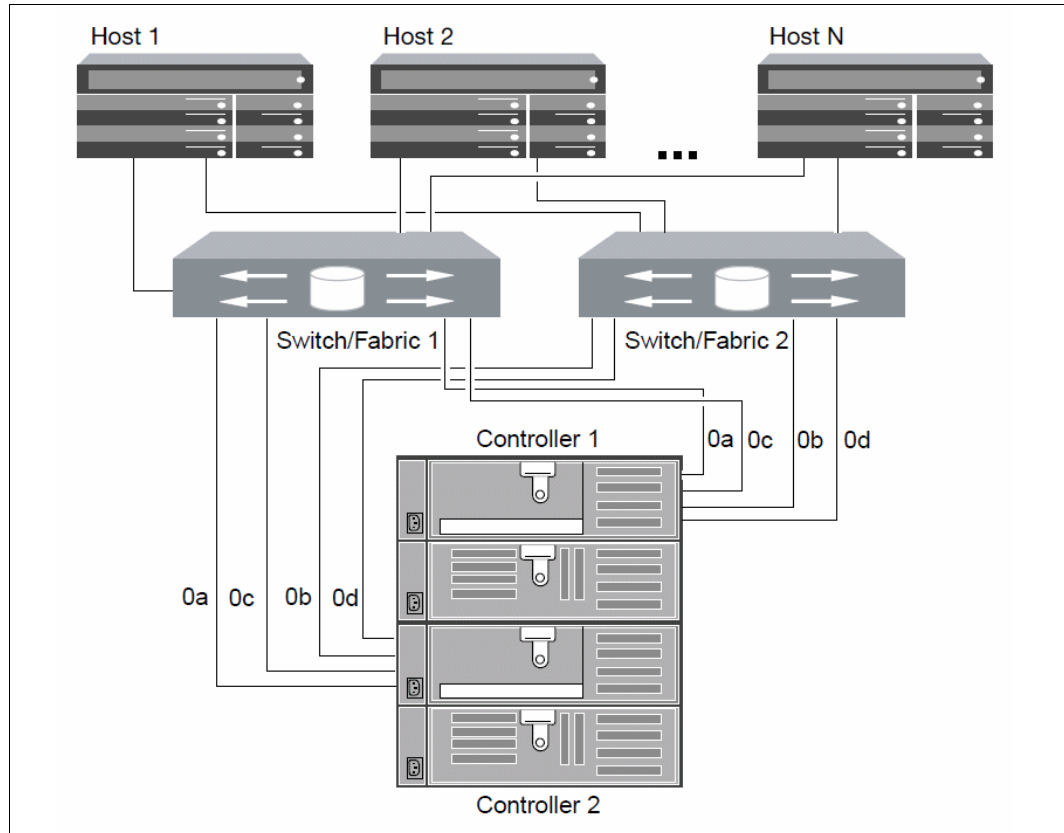


Figure 18 Multifabric HA configurations

We strongly suggest that any SAN solution use redundant components to reduce or eliminate single points of failure. For Fibre Channel this means utilizing multiple HBAs, switches/fabrics, and storage clustering.

The team who wrote this paper

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Tucson Center.

Alex Osuna is a Project Leader at the International Technical Support Organization, Tucson Center. He writes extensively on all areas of IBM storage. Before joining the ITSO six years ago, Alex had been working for over 34 years in the IT industry and holds certifications from IBM, Microsoft, and RedHat.

Rick Jooss has been working within the NetApp Fibre Channel and iSCSI business for the past five years. He was one of NetApp's early proponents of efficient and thin provisioning in NetApp SAN environments. As a Senior Manager of SAN Technical Marketing at NetApp, Rick is responsible for defining technical and business requirements for the SAN ecosystem and has served as a consultant to many of NetApp's largest customers. Before joining NetApp, Rick served as Storage Architect and Delivery Manager on behalf of storage providers for a large global German automaker. He has written and presented on the topic, and is one of the main authors of several best practices in the area.

Tarik Jose Maluf is an IBM Certified IT Specialist in Brazil. Tarik has more than 10 years of experience in the IT industry. He has eight years of experience with Windows Servers Management and related products. Tarik has worked as a Storage Administrator since 2010, and his areas of expertise include enterprise disk and virtualization (IBM Storage Systems DS8000®, DS4000®, SAN Volume Controller, and N series). He holds a degree in information technology from University of Campinas/Brazil (UNICAMP) and certifications from IBM, Microsoft, Citrix, and the Open Group.

Patrick Strick is a Technical Marketing Engineer at NetApp® specializing in SAN (Fibre Channel, FCoE, iSCSI) Storage and Server Virtualization. Previously, he was a Technical Support Engineer at NetApp, and before that he worked in IT in the health care industry.

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ Follow us on Twitter:
<http://twitter.com/ibmredbooks>
- ▶ Look for us on LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore new Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Stay current on recent Redbooks publications with RSS Feeds:
<http://www.redbooks.ibm.com/rss.html>

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document REDP-4753-00 was created or updated on April 29, 2011.

Send us your comments in one of the following ways:

- ▶ Use the online **Contact us** review Redbooks form found at:
ibm.com/redbooks
- ▶ Send your comments in an email to:
redbooks@us.ibm.com
- ▶ Mail your comments to:
IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400 U.S.A.



Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>


The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

The following terms are trademarks of other companies:

SnapDrive, Data ONTAP, NetApp, and the NetApp logo are trademarks or registered trademarks of NetApp, Inc. in the U.S. and other countries.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

DS4000®
DS8000®
IBM®

Redbooks®
Redpaper™
Redbooks (logo) ®

System Storage®

Other company, product, or service names may be trademarks or service marks of others.