Technical Report: iSCSI multipathing possibilities on Windows with Data ONTAP

Richard Jooss | 12/2006 | TR-3441

TECHNICAL REPORT

Network Appliance, a pioneer and industry leader in data storage technology, helps organizations understand and meet complex technical challenges with advanced storage solutions and global data management strategies.

Abstract

This document discussions the various multipathing options that are available for iSCSI on Microsoft Windows in conjunction with Data ONTAP 7.1 or later. The pros and cons of each solution will be discussed with the intention of helping the reader determine which solution is best for their particular environment.

Network Appliance Inc.

Table of Contents

1. Introduction

This document discusses the various multipathing options that are available for iSCSI on Microsoft Windows in conjunction with Data ONTAP 7.1 or later. The pros and cons of each solution will be discussed with the intention of helping the reader determine which solution is best for their particular environment.

2. Intended Audience

This paper is intended for system and storage architects designing iSCSI solutions with NetApp storage appliances. It is assumed:

- The reader has as a minimum general knowledge of Network Appliance hardware and software solutions, particularly in the area of block access
- The reader has some experience with block-access protocols such as Fibre Channel or iSCSI
- This paper is based on the functionality of **Data ONTAP™ release 7.1 and later** and the **Microsoft iSCSI software Initiator 2.0 and later**.

3. Multipathing Definition and Algorithms

Multipathing is the ability to have multiple paths from a server to a storage array. Multipathing provides protection against hardware failures (cable cuts, switches, HBAs, etc…) 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. Multipathing is often split into two categories, active-active and active-passive. Although these two categories might be too simplified for today's midrange arrays which can have multiple active and multiple secondary paths. A multipathing solution is generally considered to be active-active when I/O for a single LUN is going down multiple paths simultaneously. There are multiple algorithms used for deciding how to spread the I/O across the available paths such as:

- Active-passive – A single path is used and the other, passive, paths are only used when the active path is not available
- Round robin – I/O is spread evenly across the available paths using a round robin algorithm.
- Least queue depth – The path to use for an I/O is decided by looking at which available path has the smallest number of outstanding I/Os.
- Fewest bytes – The path to use for an I/O is selected by choosing the path with the lowest number of bytes outstanding.
- Weighted path – Each path is given a weight which is used to determine which path should be used for an IO.

4. Multipathing Possibilities

When an application writes data, the data follows through the host side storage stack and out through its storage connect (i.e. SCSI, iSCSI/Ethernet, Fibre Channel) to the storage array where the data will be stored. Figure 1 below shows the storage stack for an iSCSI configuration without any multipathing

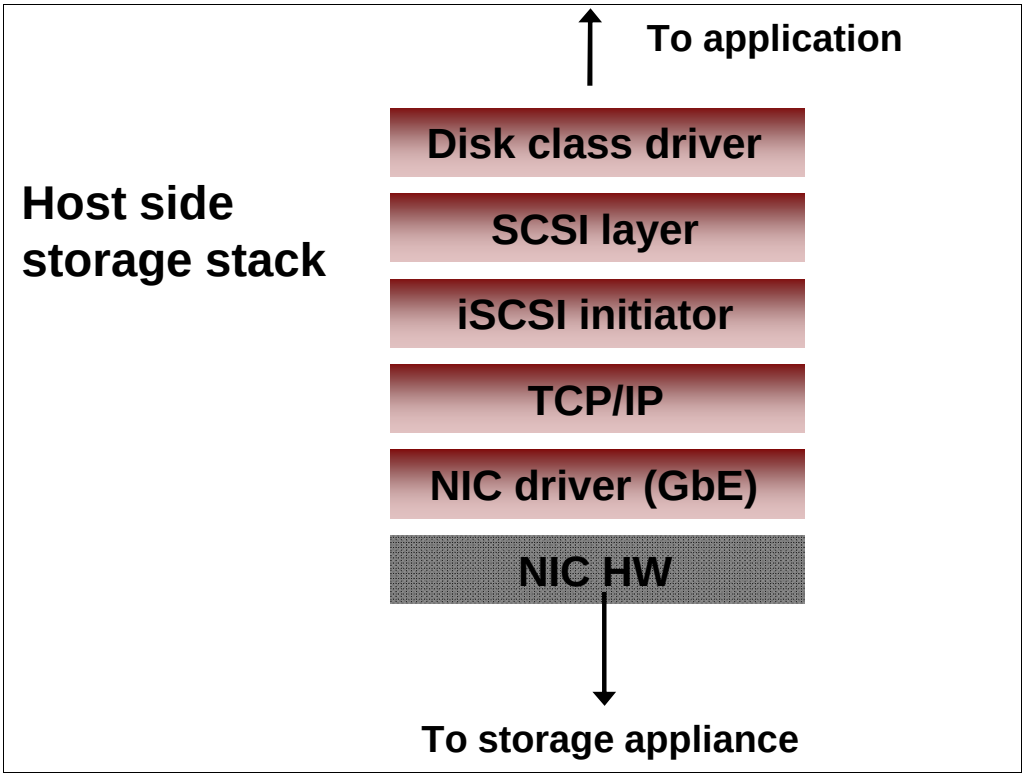

Figure 1 iSCSI storage stack with no multipathing

Multipathing is achieved by means of some sophistication, at some layer in the storage stack. The application writes to a single filesystem or 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 in the stack, both above and below the multipathing-capable layer.

There are various layers in the storage stack where this split from a single to multiple paths can occur. Each option has possible advantages and limitations.

4.1 Link Aggregation

One possible split point would be 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 and 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", "teaming", and "trunking". The term "trunking" is technically incorrect, as this term is used to describe VLAN packet tagging as specified in IEEE 802.1q. Link aggregation is not storage specific and all network traffic benefits from the multiple connections. Figure 2 shows where link aggregation fits into the storage stack.

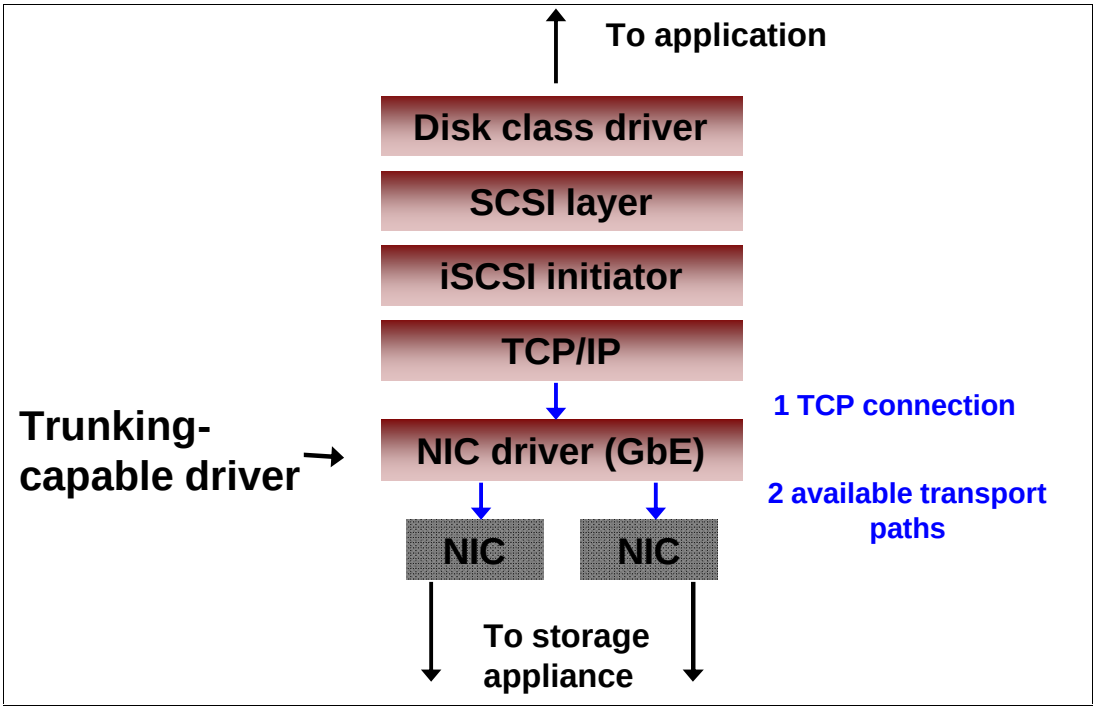

Figure 2 iSCSI storage stack with link aggregation

Trunking implementations typically support active-passive and active-active path selection algorithms. However, it should be noted that for active-active configurations where both (or multiple) paths are used simultaneously, the same path is always selected for a given destination endpoint to avoid out of order delivery. This fact means that the link aggregation does not increase the throughput above the capability of a single path for a connection between a single server and a single storage appliance. Starting with the Microsoft iSCSI software initiator v2.0 Microsoft explicitly does not support link aggregation on the host. Link aggregation (vif) on the storage side is supported.

Pros:

- Transparent to all network protocols – The advantages of link aggregation is shared not just with iSCSI but other network traffic such as NFS, CIFS etc….
- Well known mature technique
- Available on both host and storage appliance (VIFs)

Limitations:

- Not supported with the Microsoft iSCSI software initiator
- Aggregated interfaces must be connected to the same network, often the same switch or card within a switch limiting the physical isolation of the multiple paths
- Dependent on having aggregation capable drivers and switches
- Not possible to get aggregated throughput for a single server to single storage appliance pair
- Not currently supported by hardware iSCSI initiators

4.2 iSCSI multi-connection sessions (MCS)

Multi-connection sessions are an optional part of the iSCSI specification and create multiple paths starting at the iSCSI session layer of the storage stack. Both the iSCSI initiator (host) and iSCSI target (storage system) need to support multi-connection sessions in order to configure sessions with multiple connections. Figure 3 shows where iSCSI multi-connection sessions fit into the storage stack.

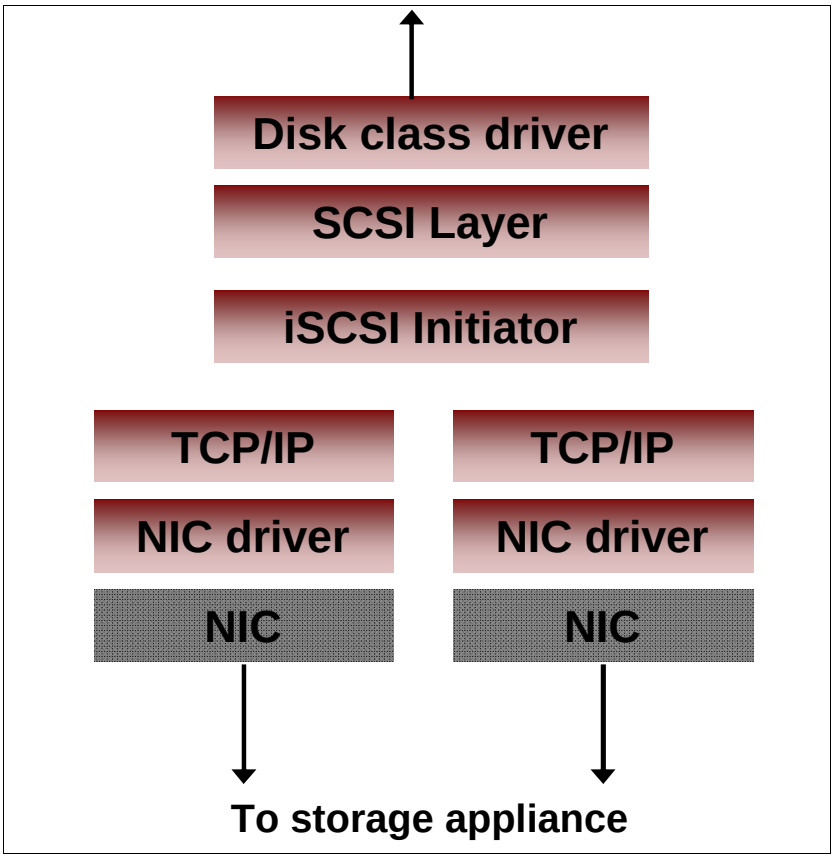

Figure 3 iSCSI storage stack multi-connection sessions

Microsoft added support for multi-connection sessions starting with the 2.0 version of the Microsoft iSCSI software initiator while NetApp added support with Data ONTAP 7.1. Refer to the NetApp iSCSI support matrix on the NOW site for the most up-to-date information regarding supported ONTAP and initiator releases. iSCSI HBAs do not currently support multi-connection sessions. Given that the iSCSI initiator portion of the stack resides on the HBA itself, implementing multi-connection sessions across HBAs will have its challenges. The Microsoft iSCSI software initiator supports both active/passive and active/active algorithms load balancing algorithms. SnapDrive for Windows will work with a pre-existing iSCSI connections which have MCS enabled but will not create a MCS enabled connection and has no knowledge of MSC.

Pros:

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

Limitations:

- Not fully integrated with SnapDrive for Windows
- Not supported by iSCSI initiator HBAs

4.3 Multipathing Technology Layer (MPIO style)

The "classic" way to do multipathing is to insert a separate multipathing layer into the storage stack. This method is not specific to iSCSI or any underlying transport, and is the standard way to achieve multipathed access to Fibre Channel and even parallel SCSI targets.

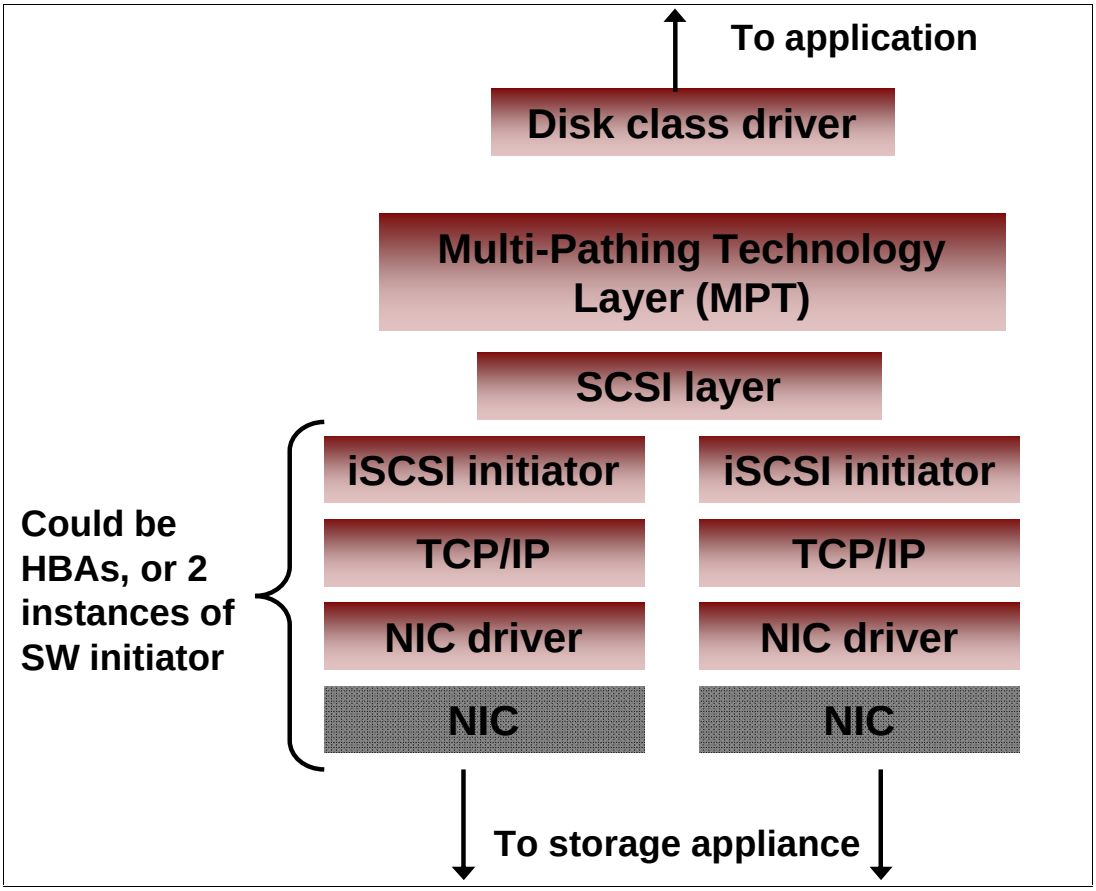

Figure 4 **iSCSI storage stack multipathing technology layer**

There are multiple implementations of this type of multipathing on the various operating systems. The MPIO infrastructure offered by Microsoft is the standard way to do this on Windows Server technologies. With Microsoft's MPIO, each storage vendor supplies a device specific module (DSM) for their storage array. In addition, for iSCSI Microsoft delivers a DSM. Network Appliance supports both the NTAP DSM and the Microsoft iSCSI DSM depending on the version of Data ONTAP and SnapDrive for Windows product is being used. The NTAP DSM uses an active-passive algorithm on a per LUN basis. For a single LUN all IOs will be done across a single path until a failure occurs. However, the LUNs are round robined across the available paths. For example, if there are 2 disks configured, G:\ and H:\, I/O from disk G:\ will utilize one path and disk H:\ will utilize the second path. This allows the full available bandwidth to be utilized as long as multiple LUNs are configured. The Microsoft DSM

supports various load balancing algorithms including active/active algorithms. SnapDrive for Windows fully supports and is integrated with MPIO. The table below shows which version of SnapDrive for Windows is supported with which DSM. For more detailed and exact information about supported Data ONTAP versions, refer to the NetApp iSCSI support matrix on NOW.

	NTAP DSM	Microsoft iSCSI DSM	ONTAP DSM
iSCSI with SnapDrive for Windows v3.2 to v4.1	X		
iSCSI with SnapDrive for Windows v4.2 or later		X	
iSCSI with no SnapDrive		X	
FC with SnapDrive for Windows v4.1 or earlier	X		
FC with SnapDrive for Windows v4.2 or later			X
FC with no SnapDrive			X

Pros:

- Fully supported SnapDrive for Windows
- No dependency on aggregation capable Ethernet infrastructure
- Mature implementation
- Supports software and hardware initiators (HBAs)

Limitations:

- Extra vendor multipathing technology layer required

5. Recommendations

It is not possible to make a single recommendation as to which multipathing solution to use. The following table summaries the capabilities of the possible architectures helping to make a decision as to which is the best solution for a particular environment.

	Link aggregation	MPIO – multipathing technology layer	Multi-connection sessions
Improved bandwidth between a single server and a single storage appliance	No	Yes	Yes
Support of completely physically independent paths	No	Yes	Yes
Support of HW initiators (HBA)	No	Yes	No
Full support with SnapDrive for Windows	Yes	Yes	No