Windows Server 2008 R2 NIC Optimization and Best Practices with EqualLogic SAN

A Dell EqualLogic Reference Architecture

Dell Storage Engineering
September 2013

Revisions

Date	Description
September 2013	Initial release

Table of contents

Acknowledgements

This best practice white paper was produced by the following members of the Dell Storage team:

Engineering: Clay Cooper

Technical Marketing: Guy Westbrook

Editing: Camille Daily

Additional contributors: Mike Kosacek, Steve Williamson, and Fred Spreeuwers

Feedback

We encourage readers of this publication to provide feedback on the quality and usefulness of this information by sending an email to SISfeedback@Dell.com.

SISfeedback@Dell.com

Executive summary

This reference architecture explores the configuration options available for improving Dell™ EqualLogic™ PS Series SAN performance using the Broadcom® BCM57810 or Intel® X520 10 GbE network adapters and Windows Server® 2008 R2 on a Dell™ PowerEdge™ 12th generation server. Recommended OS and NIC configurations are given based on the results of SAN performance testing.

1 Introduction

Dell EqualLogic PS Series arrays provide a storage solution that delivers the benefits of consolidated networked storage in a self-managing iSCSI storage area network (SAN) that is affordable and easy to use, regardless of scale.

In every iSCSI SAN environment, there are numerous configuration options at the storage host which can have an effect on overall SAN performance. These effects can vary based on the size and available bandwidth of the SAN, the host/storage port ratio, the amount of network congestion, the I/O workload profile and the overall utilization of system resources at the storage host. One setting might greatly improve SAN performance for a large block sequential workload yet have an insignificant or slightly negative effect on a small block random workload. Another setting might improve SAN performance at the expense of host CPU utilization.

This technical paper quantifies the effect on iSCSI throughput and IOPS of several configuration options within the Broadcom and Intel 10 GbE adapter properties and the Windows Server 2008 R2 TCP stack during three common SAN workloads. It also takes into account the value of certain settings in congested network environments and when host CPU resources are at premium. From the results, recommended configurations for an EqualLogic PS Series SAN are given for each tested NIC type.

In order to focus on the pure SAN performance benefits of the tested configuration options, Data Center Bridging (DCB) and Broadcom Switch Independent Partitioning, also known as NIC Partitioning (NPAR) were excluded from testing.

Note: The performance data in this paper is presented relative to baseline configurations and is not intended to express maximum performance or benchmark results. Actual workload, host to array port ratios, and other factors may also affect performance.

1.1 Audience

This technical white paper is for storage administrators, SAN system designers, storage consultants, or anyone who is tasked with configuring a host server as an iSCSI initiator to EqualLogic PS Series storage for use in a production SAN. It is assumed that all readers have experience in designing and/or administering a shared storage solution. Also, there are some assumptions made in terms of familiarity with all current Ethernet standards as defined by the Institute of Electrical and Electronic Engineers (IEEE) as well as TCP/IP and iSCSI standards as defined by the Internet Engineering Task Force (IETF).

2 Technical overview

iSCSI SAN traffic takes place over an Ethernet network and consists of communication between PS Series array member network interfaces and the iSCSI initiator of storage hosts. The Broadcom BCM57810 NetXtreme® II and the Intel X520 10 GbE network adapters were used as the iSCSI initiators during this project.

The Broadcom BCM57810 network adapter features iSCSI Offload Engine (iSOE) technology which offloads processing of the iSCSI stack to the network adapter. When using iSOE mode, the network adapter becomes a host bus adapter (HBA) and a host-based, software iSCSI initiator is not utilized. This is as opposed to non-iSOE mode in which the network adapter functions as a traditional NIC and works with a software iSCSI initiator. This paper refers to the non-iSOE mode of operation as NDIS mode. In Windows, NDIS refers to the Network Driver Interface Specification, a standard application programming interface (API) for NICs.

The following three initiator modes of operation were tested:

1. Broadcom BCM57810 NDIS mode
2. Broadcom BCM57810 iSOE mode
3. Intel X520

Appendix B provides a detailed list of the tested configuration options and default values for each NIC type as well as for the Windows Server 2008 R2 TCP stack.

3 Test configurations and methodology

The following section addresses the reasoning behind SAN design decisions and details the SAN configurations. Performance testing methodology, test case sequence, and results analysis are also explained.

3.1 Simplified SAN

Every effort was made to simplify and optimize the test configurations so that the performance effects of each option could be isolated. The following configuration and design elements helped to achieve this goal.

- All unused protocols disabled for each network adapter
- Eight volumes within a single storage pool, evenly distributed across array members
- An isolated SAN with no LAN traffic
- Load balancing (volume page movement) disabled on the array members
- DCB was not used
- The NIC bandwidth was not partitioned

Load balancing is recommended for production environments because it can improve SAN performance over time by optimizing volume data location based on I/O patterns. It was disabled for performance testing to maintain consistent test results. It is enabled by default.

Base and congested SAN designs were chosen and are described in the following sections. See Appendix A for more detail about the hardware and software infrastructure.

3.1.1 Base SAN configuration

The first SAN design chosen was a basic SAN with a redundant SAN fabric and an equal number of host and storage ports. Having a 1:1 host/storage port ratio is ideal from a bandwidth perspective. This helped to ensure that optimal I/O rates were achieved during lab testing. Figure 1 shows only the active ports of the PS Series array members.

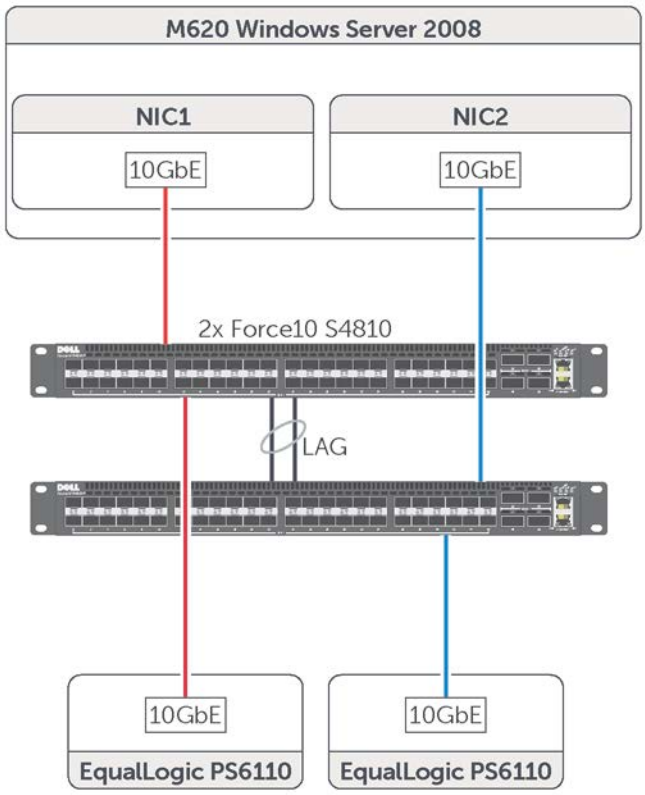

Figure 1 Physical diagram of the base SAN configuration

- Two switches
- Two array members each with a single port
- A single host with two 10 GbE NIC ports
- A 1:1 storage/host port ratio

3.1.2 Congested SAN configuration

The second SAN design was constructed to mimic a host port experiencing network congestion. In this SAN design, a single host port was oversubscribed by four storage ports for a 4:1 storage/host port ratio. Since only one host port existed, the SAN fabric was reduced to a single switch. Figure 2 shows only the active ports of the PS Series array members.

A non-redundant SAN fabric is not recommended for a production SAN environment.

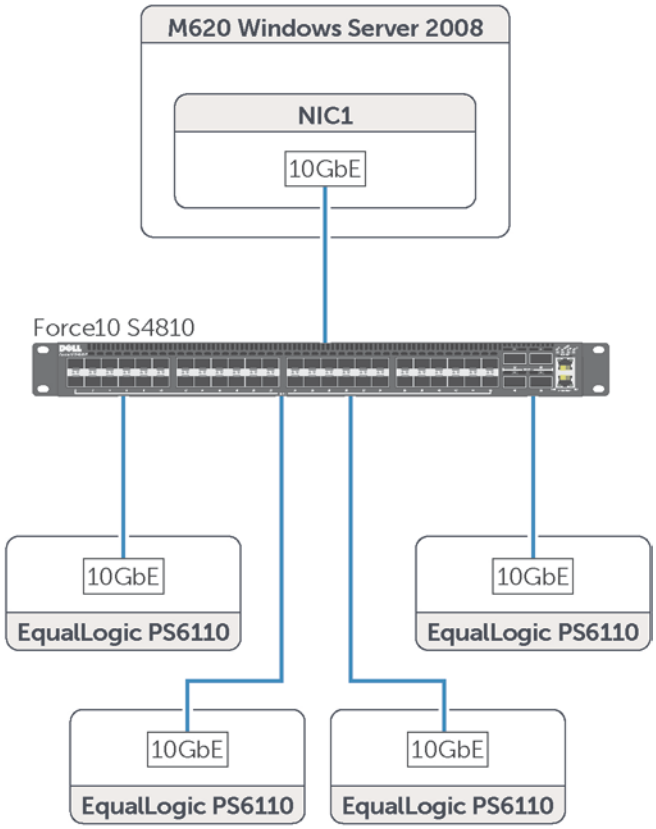

Figure 2 Physical diagram of the congested SAN configuration

- One switch
- Four array members each with one port
- A single host with one 10 GbE NIC
- A 4:1 storage/host port ratio

3.2 I/O execution and evaluation

Prior to each test run, the host was restarted to confirm configuration changes were in effect. After boot, the even distribution of iSCSI connections across host and storage ports and of active array member ports across SAN switches was confirmed.

The following three vdbench workloads were run:

- 8 KB transfer size, random I/O, 67% read
- 256 KB transfer size, sequential I/O, 100% read
- 256 KB transfer size, sequential I/O, 100% write

For every test case, each vdbench workload was run three times for twenty minute durations and the results were averaged.

Vdbench IOPS results were used to evaluate 8K random workload performance. Vdbench throughput results were used to evaluate 256K sequential workload performance. Host and array member retransmission rates and CPU utilization were also examined.

See Appendix C for a list of vdbench parameters.

3.3 Statistical analysis

In order to ensure stable results, the relative standard deviation among the three performance results for each test case workload was calculated. When the relative standard deviation was greater than 1% for a given test case workload, that particular workload performance test was repeated.

3.4 Test case sequence

The performance and effect on system resources of adapter and Windows Server 2008 R2 TCP stack options were evaluated using the test cases listed below.

Initially, tests were run to compare a totally default configuration to the baseline configuration chosen for testing. The baseline configuration included the following non-default settings:

- Jumbo frames enabled
- Flow control enabled (if not already enabled by default)
- Maximum receive and transmit buffers (if applicable)

Each subsequent test case consisted of a single option being toggled from the default setting to show its effect relative to the baseline configuration defined above. In test cases where a technology option had an adapter and a corresponding OS setting, for example Receive Side Scaling, both settings were changed simultaneously prior to the test case execution. In the case of TCP/IP offload engine (TOE) / chimney support, chimney was disabled for baseline testing and enabled for the TOE test case. This was done

because chimney support is set to *automatic* by default in Windows Server 2008 R2, which dynamically offloads TCP connections based on a set of criteria including total throughput, making it difficult to control the iSCSI connection offload state during testing.

3.4.1 Broadcom BCM57810 NDIS mode test case sequence

The following tables show the test case sequence used to evaluate the effect of tested configuration options for the Broadcom BCM57810 in NDIS mode. **Bold text** indicates the changed value for each test scenario.

Table 1 Baseline test case sequence for Broadcom BCM57810 NDIS mode

Test case	Frame size	Flow Control	Rx / Tx buffers	Other adapter settings	Windows Server TCP stack setting	Comments
1	Standard	Auto	Default	Default	Default	Default configuration
2	**Jumbo**	Auto	Default	Default	Default	Jumbo performance effect
3	Jumbo	**On**	**Maximum**	Default	Default	Baseline configuration to evaluate all subsequent settings

Table 2 Test case sequence for Broadcom BCM57810 NDIS mode to evaluate other options

Test case	Frame size	Flow Control	Rx / Tx buffers	Other adapter settings	Windows Server TCP stack setting	Comments
4	Jumbo	On	Maximum	**Interrupt moderation disabled**	Default	
5	Jumbo	On	Maximum	**Receive Side Scaling (RSS) disabled**	**RSS disabled**	
6	Jumbo	On	Maximum	**RSS queues of 16**	Default	RSS enabled by default at adapter and in Windows Server TCP stack.
7	Jumbo	On	Maximum	**TCP Connection Offload enabled**	**Chimney enabled**	Both settings required to activate TCP Offload Engine (TOE)
8	Jumbo	On	Maximum	Default	**Receive Window Auto-tuning disabled**	

Test case	Frame size	Flow Control	Rx / Tx buffers	Other adapter settings	Windows Server TCP stack setting	Comments
9	Jumbo	On	Maximum	Default	**Delayed ACK algorithm disabled**	
10	Jumbo	On	Maximum	Default	**Nagle's algorithm disabled**	
11	Jumbo	On	Maximum	**Large Send Offload (LSO) disabled**	Default	

3.4.2 Broadcom BCM57810 iSOE mode test case sequence

The following table shows the test case sequence used to evaluate the effect of tested configuration options for the Broadcom BCM57810 in iSOE mode. **Bold text** indicates the changed value for each test scenario.

As can be seen in the table below, when in iSOE mode the Broadcom 57810 has a much more limited set of adapter options. Also, in iSOE mode the Windows Server 2008 R2 TCP stack options have no effect since the entire iSCSI and TCP stack are offloaded to the adapter by design.

Table 3 Test case sequence for Broadcom BCM57810 iSOE mode

Test case	Frame size	Flow Control	Rx / Tx buffers	Other adapter settings	Windows Server TCP stack	Comments
1	Standard	Auto	N/A	N/A	N/A	Default configuration
2	**Jumbo**	Auto	N/A	N/A	N/A	Jumbo performance effect
3	Jumbo	**On**	N/A	N/A	N/A	Recommended settings

Bold text indicates the changed value for each test scenario.

3.4.3 Intel X520 test case sequence

The following table shows the test case sequence used to evaluate the effect of tested configuration options for the Intel X520. **Bold text** indicates the changed value for each test scenario.

Table 4 Baseline test case sequence for Intel X520

Test case	Frame size	Flow Control	Rx / Tx buffers	Other adapter settings	Windows Server TCP stack	Comments
1	Standard	On	Default	Default	Default	Default configuration
2	**Jumbo**	On	Default	Default	Default	Jumbo performance effect
3	Jumbo	On	**Maximum**	Default	Default	Baseline configuration to evaluate all subsequent settings

Table 5 Test case sequence for Intel X520 to evaluate other configuration options

Test case	Frame size	Flow Control	Rx / Tx buffers	Other adapter settings	Windows Server TCP stack	Comments
4	Jumbo	On	Maximum	**Interrupt moderation disabled**	Default	
5	Jumbo	On	Maximum	**Receive Side Scaling (RSS) disabled**	**RSS disabled**	
6	Jumbo	On	Maximum	**RSS queues of 16**	Default	RSS enabled by default at adapter and in Windows Server TCP stack.
7	Jumbo	On	Maximum	Default	**Receive Window Auto-tuning disabled**	
8	Jumbo	On	Maximum	Default	**Delayed ACK algorithm disabled**	
9	Jumbo	On	Maximum	Default	**Nagle's algorithm disabled**	
10	Jumbo	On	Maximum	**Large Send Offload (LSO) disabled**	Default	

4 Results and analysis

All test case performance results for each NIC mode and workload combination are presented in this section. For the sake of analysis, a 5% margin of error is assumed and only performance differences greater than this were acknowledged as significant.

Based on the results, recommended configurations for each NIC mode will be given in Section 5.

4.1 Baseline testing

The following non-default settings were used as a baseline configuration for further testing and evaluation.

4.1.1 Jumbo Frames

Jumbo frames enable Ethernet frames with payloads greater than 1500 bytes. In situations where large packets make up the majority of traffic and additional latency can be tolerated, jumbo packets can reduce CPU utilization and improve wire efficiency.

Figures 3 - 5 illustrate the dramatic effect that enabling jumbo frames can have on large block workloads. Significant throughput increases were observed for both read and write large block workloads on both the Broadcom and Intel network adapters in all supported operating modes (i.e., NDIS, iSOE).

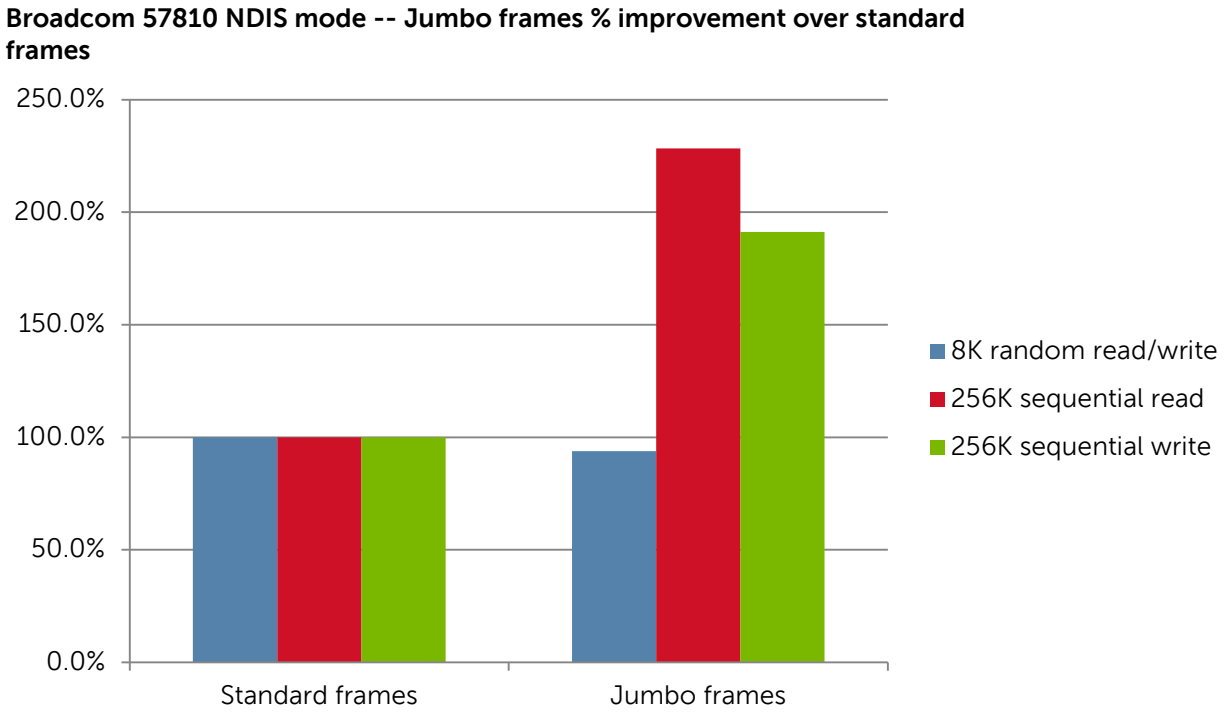

Figure 3 The performance effects of enabling jumbo frames when using Broadcom 57810 in NDIS mode

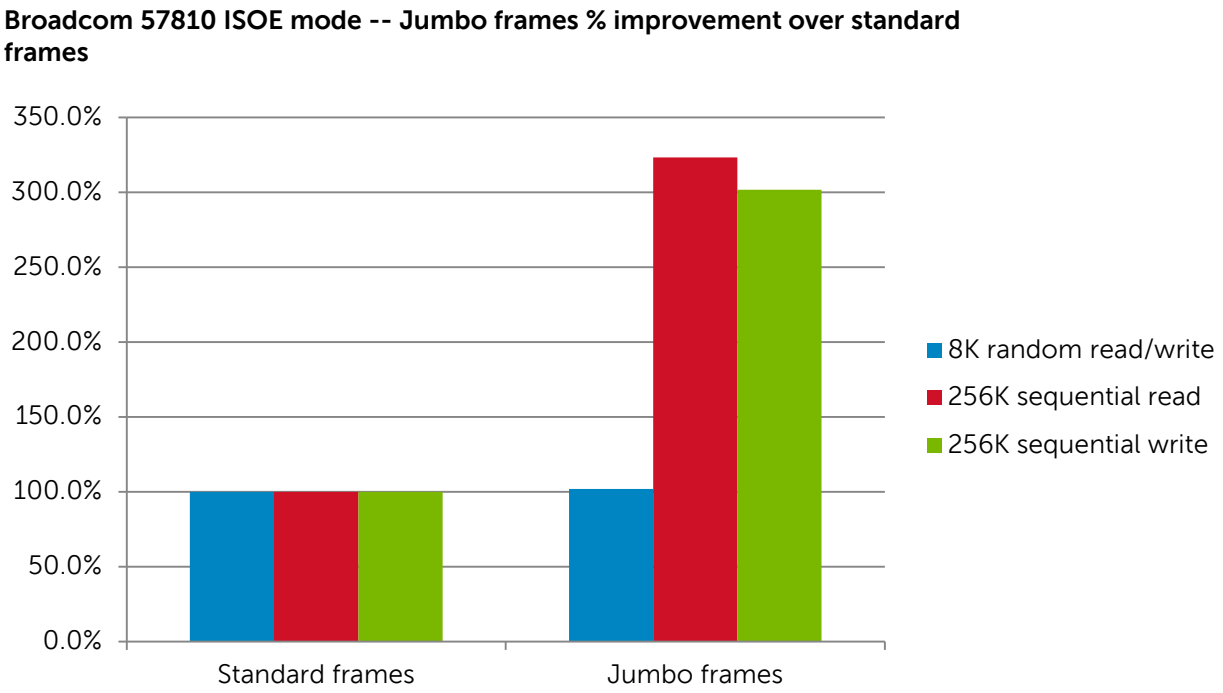

Figure 4 The performance effects of enabling jumbo frames when using Broadcom 57810 in iSOE mode

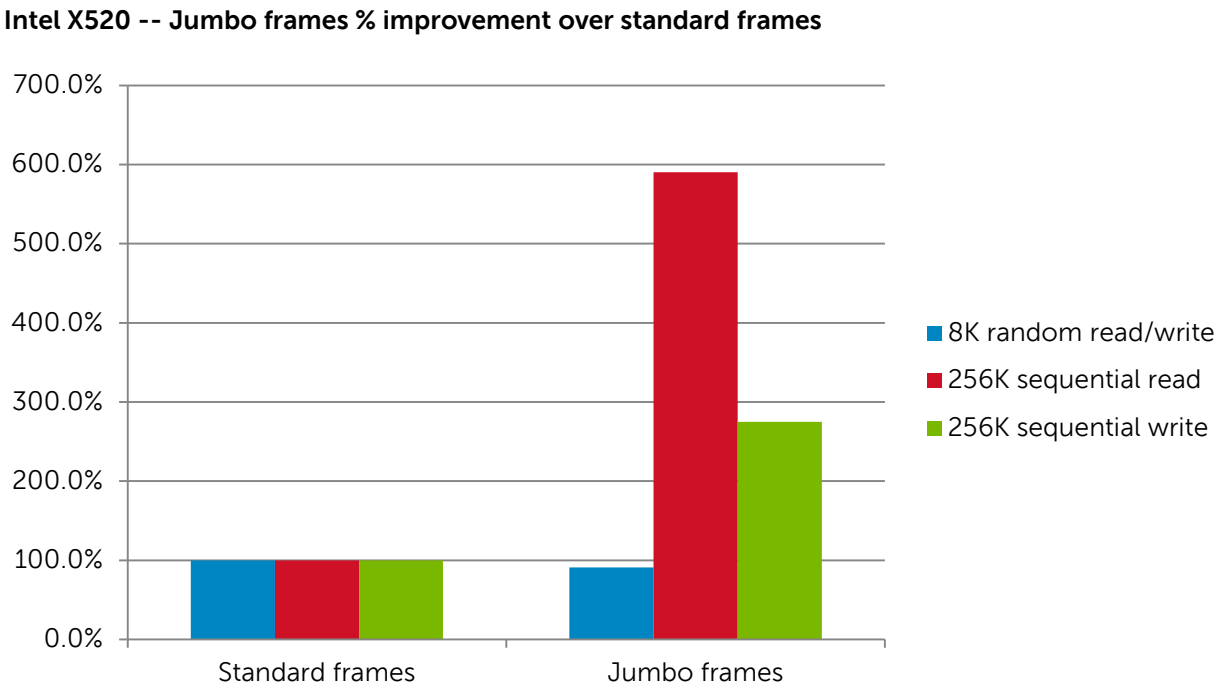

Figure 5 The performance effects of enabling jumbo frames when using Intel X520

4.1.2 Flow control

Flow control is a link-level mechanism that enables the adapter to respond to or to generate flow control (PAUSE) frames, helping to regulate network traffic. It is recommended that flow control be enabled as it is well-known to be of benefit in a congested network environment.

4.1.3 Receive and Transmit buffers

Rx / Tx buffers are used by the adapter when copying data to memory. Increasing this value can enhance performance with only a slight increase in system memory utilization. Maximizing buffer allocation is particularly important on a server with heavy CPU utilization and can also be beneficial during times of network congestion.

4.2 Other available performance tuning options

For both the network adapter and the Windows Server 2008 R2 TCP stack, there are other options available which can have an effect on SAN performance under the right circumstances. Section 4.2 defines these options and discusses the results of the performance testing. It is important to understand that the performance results described below may not translate to all EqualLogic PS Series SAN environments. The material presented below identifies setting and workload combinations that have a clearly positive or negative impact on iSCSI SAN performance. It is recommended that each potential configuration change be evaluated in the environment prior to implementation.

For more information on performance tuning options for the networking subsystem see *Performance Tuning Guidelines for Windows Server 2008 R2* at http://msdn.microsoft.com/en-us/library/windows/hardware/gg463392.aspx

The tuning guidelines and recommendations in *Performance Tuning Guidelines for Windows Server 2008 R2* are for TCP/IP network interfaces in general and are not particular to iSCSI networking.

4.2.1 Interrupt moderation

With interrupt moderation, a network adapter attempts to reduce the number of interrupts required by sending a single interrupt for multiple events rather than an interrupt for each event. While this can help lower CPU utilization it can also increase SAN latency. It is enabled by default on both the Broadcom and Intel adapters. Disabling interrupt moderation increased 256K sequential read throughput by 7% when using the Broadcom 57810 adapter in NDIS mode, though no significant performance effect was observed on any other workload or with the Intel X520.

4.2.2 Receive side scaling

RSS balances incoming traffic across multiple CPU cores, up to one logical process per CPU core. This can improve the processing of receive-intensive workloads when the number of available logical processors outnumbers network adapters. It is enabled by default on both the Broadcom and Intel adapters and globally in Windows Server 2008 R2.

In addition to disabling RSS to test against the baseline (RSS enabled), maximizing the number of RSS queues was also tested. A larger number of queues increases network throughput at the expense of CPU utilization.

Disabling of RSS reduced 256K sequential read throughput by 13% on the Intel adapter, while on the Broadcom adapter 256K sequential write throughput increased by 13%.

Increasing the number of RSS queues did not have a significant effect on any workload performance for either the Broadcom or Intel adapter.

4.2.3 TCP/IP offload engine

TOE offloads the processing of the entire TCP/IP stack to the network adapter and is available on the Broadcom 57810 network adapter. For TOE to be active, TCP connection offload must be enabled on the network adapter and chimney support must be enabled in Windows Server 2008 R2. By default, TCP connection offload is enabled on the Broadcom 57810, while Windows Server 2008 R2 chimney support is set to automatic. In order to ensure that the iSCSI sessions are immediately offloaded, it is recommended that chimney support be set to *enabled* rather than *automatic*.

TOE improved throughput by 7% during the 256K sequential read workload and by 10% during the 256K sequential write workload. In the congested SAN configuration, TOE eliminated storage array retransmissions and reduced CPU utilization by 35% during the 256K sequential read workload as shown in Figure 6 below.

4.2.4 Large send offload

LSO enables the adapter to offload the task of segmenting TCP messages into valid Ethernet frames. Because the adapter hardware is able to complete data segmentation much faster than operating system software, this feature may improve transmission performance. In addition, the adapter uses fewer CPU resources. It is enabled by default on both Broadcom and Intel adapters. No significant performance effects were observed during testing.

4.2.5 TCP checksum offload

Enables the adapter to verify received packet checksums and compute transmitted packet checksums. This can improve TCP performance and reduce CPU utilization and is enabled by default. With TCP checksum disabled, iSCSI connection instability and increased array packet retransmission were observed during testing and so the results were omitted from the performance diagrams.

4.2.6 TCP receive window auto-tuning

TCP window auto tuning enables Windows Server 2008 R2 to monitor TCP connection transmission rates and increase the size of the TCP receive window if necessary. It is enabled by default. Disabling auto-tuning reduced 256K sequential read workload throughput by 20% on Broadcom adapters and by over 30% on Intel adapters.

4.2.7 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. Though disabling this algorithm had no effect on the performance of any tested workload, there are cases where disabling TCP delayed ACK for an iSCSI interface may improve performance. For example, poor read performance and unreliable failover have been observed during periods of network congestion on Microsoft iSCSI cluster nodes. In certain cases, disabling TCP delayed ACK on iSCSI interfaces might be recommended by Dell Enterprise Support.

4.2.8 Nagle's algorithm

Nagle's algorithm is a technique to improve TCP performance by buffering output in the absence of an ACK response until a packet's worth of output has been reached. This algorithm has the potential to interact negatively with a TCP receiver utilizing the delayed ACK algorithm, since the delayed ACK algorithm may delay sending an ACK under certain conditions up to 500 milliseconds.

While disabling Nagle's algorithm did not demonstrate an effect on the performance results of the sustained workloads, there may be cases where disabling Nagle's algorithm on the storage host improves iSCSI SAN performance. Bursty SAN I/O or a high frequency of iSCSI commands can trigger ACK delays and increase write latency. As with disabling TCP delayed ACK, Dell Enterprise Support might recommend disabling Nagle's algorithm on iSCSI interfaces in certain cases.

4.2.9 iSCSI Offload Engine

iSOE offloads the processing of the entire iSCSI stack to the network adapter and is available on the Broadcom 57810 network adapter. For iSOE to be active it must be enabled in the Broadcom Advanced Control Suite (BACS). By default it is disabled.

Once the iSOE function is enabled for the adapter (and the NDIS function disabled), it disappears from the native Windows Server networking administration and monitoring tools and appears as a storage controller in Windows device manager. It must be configured and managed through BACS.

In the base SAN configuration, iSOE increased the IOPS of the 8K random read/write workload and the throughput of the 256K sequential read workload by 10% relative to the baseline NDIS configuration. 256K sequential write throughput was increased by an impressive 70%.

Like TOE, iSOE had a positive effect on storage array member retransmission and CPU utilization. In the congested SAN configuration during the 256K sequential read workload, the array member retransmission rate was reduced from .11% when disabled to zero when enabled. The CPU utilization decreased significantly from ~6% to ~1%. See Figure 6 below for an illustration.

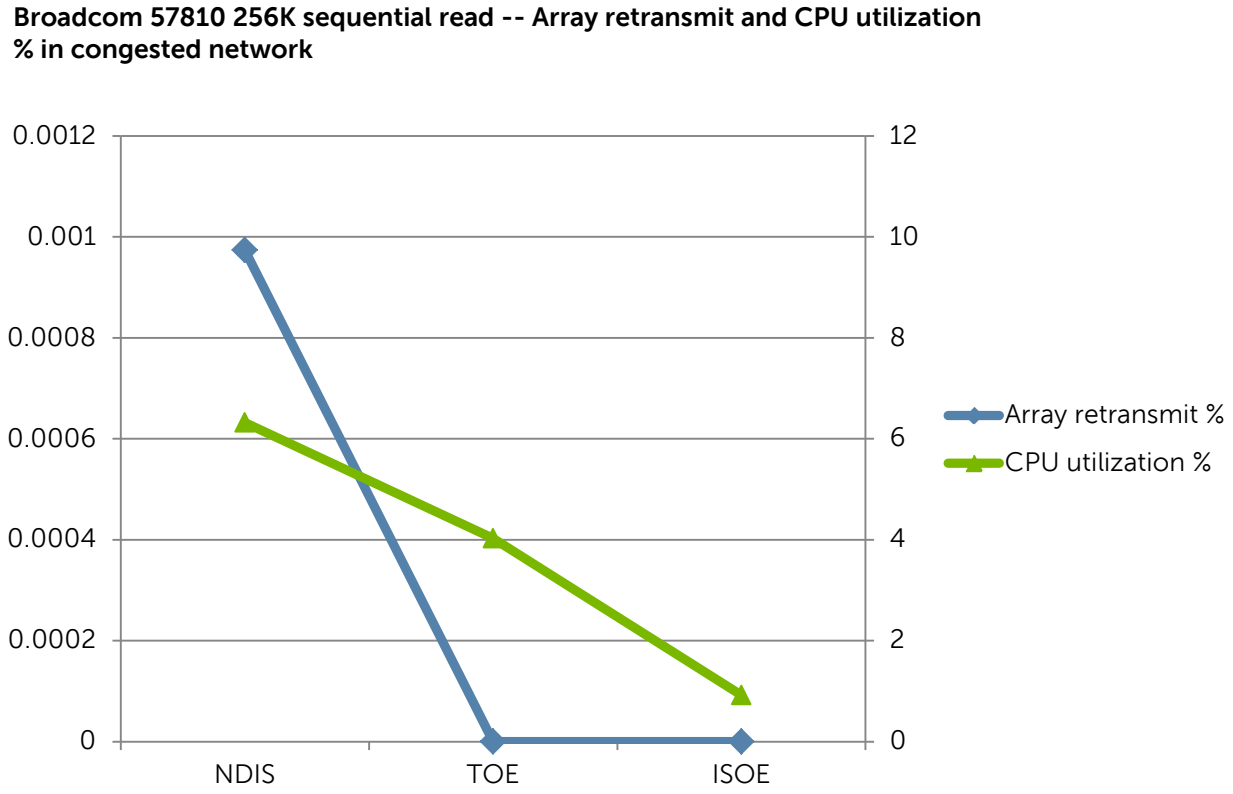

Figure 6 The effects on array retransmissions and CPU utilization of TCP and iSCSI offload engines when using Broadcom 57810 as compared to NDIS mode during a 256K sequential read workload in the congested SAN configuration.

4.3 Broadcom BCM57810 NDIS mode performance results

Section 4.3 shows the performance results when using the different configuration options during the three tested workloads using the Broadcom BCM57810 network adapter in NDIS mode.

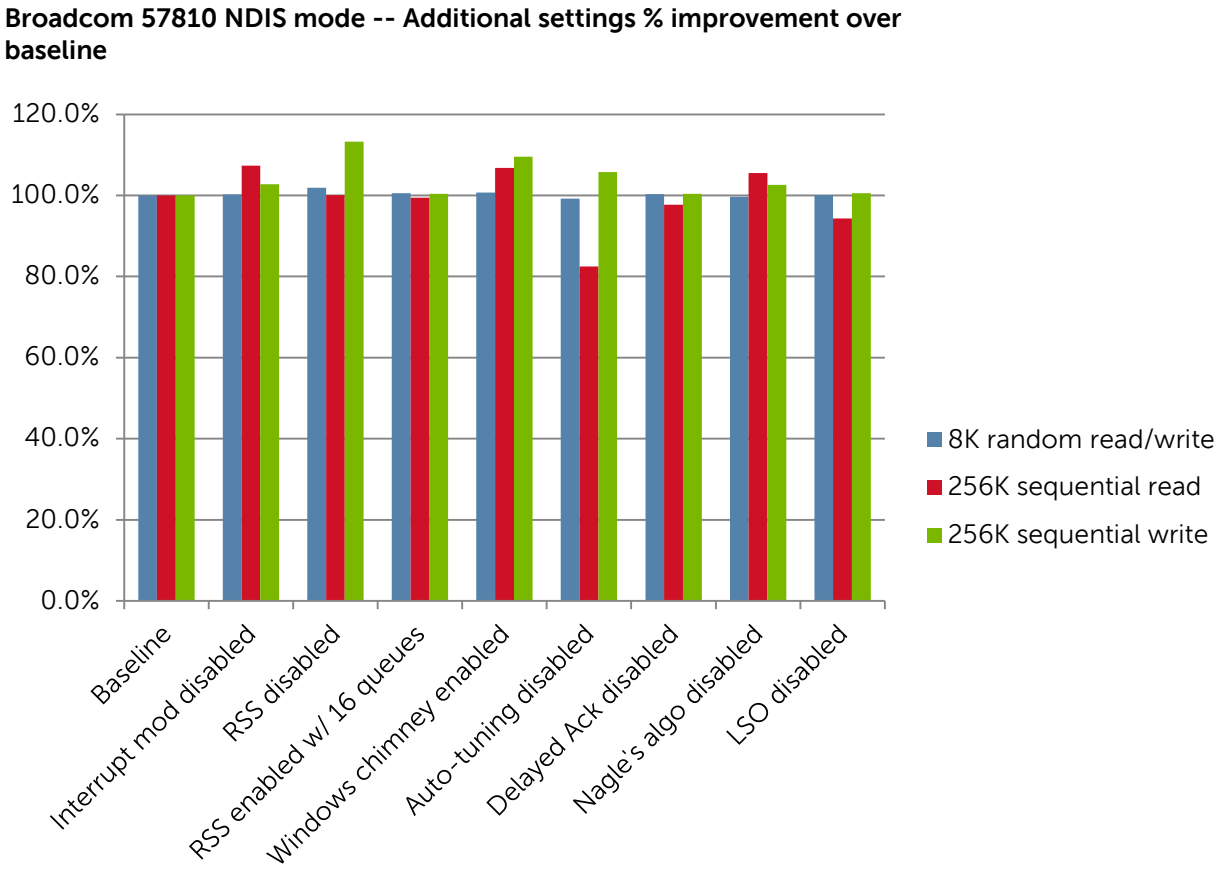

Figure 7 Performance effects of individual configuration changes relative to the baseline configuration on a Broadcom 57810 in NDIS mode

4.4 Broadcom BCM57810 iSOE mode performance results

Section 4.4 compares the performance results with ISOE mode to those of NDIS mode and NDIS mode with TOE during the three tested workloads using the Broadcom BCM57810 network adapter.

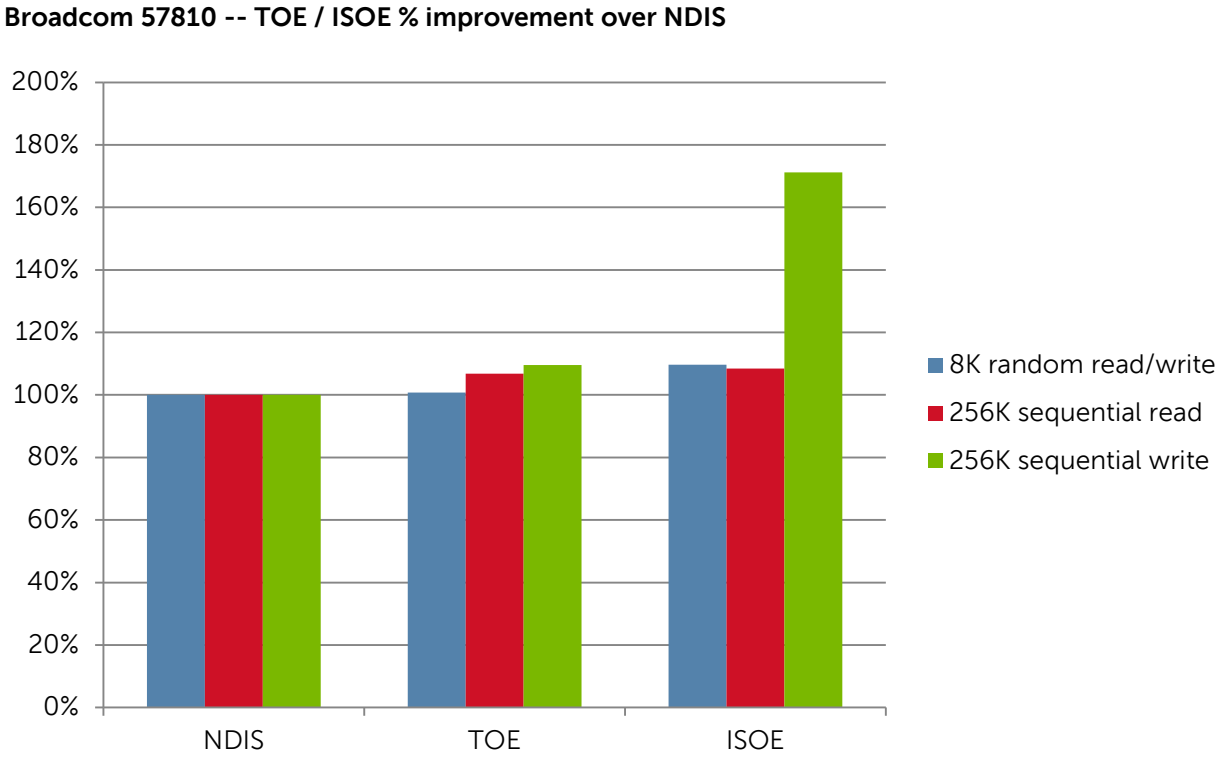

Figure 8 The performance effects of the iSCSI offload engine when using Broadcom 57810 as compared to NDIS mode and TCP offload

4.5 Intel X520 performance results

Section 4.5 shows the performance results with the different configuration options during the three tested workloads using the Intel X520 network adapter.

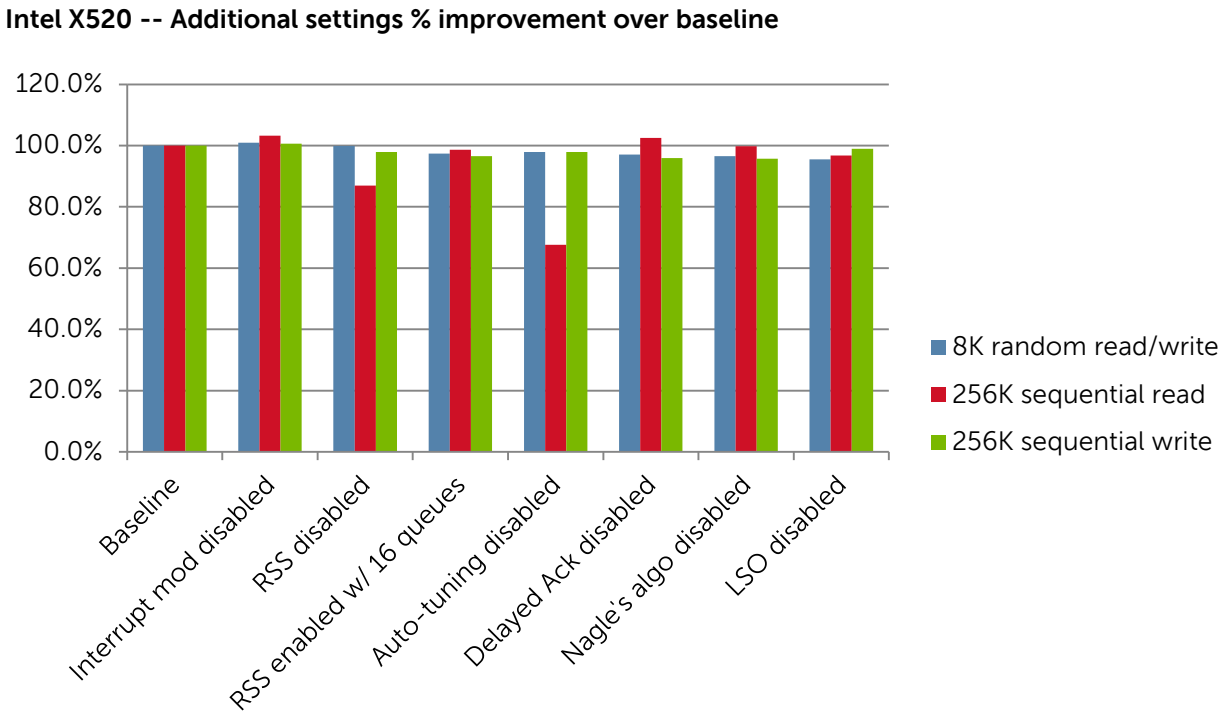

Figure 9 Performance effects of individual configuration changes relative to the baseline configuration on an Intel X520

5 Best practice recommendations

In this section, recommended configurations will be given based on the performance results and analysis detailed in Section 4. Non-default settings are recommended only when a compelling difference in performance for one or more workloads was observed, or when a setting is known to provide benefit during network congestion or heavy CPU utilization. Only the non-default settings are listed.

For a complete list of tested options and default values as well as instructions on making configuration changes to the storage host, see Appendix B.

5.1 Broadcom BCM57810 NDIS mode recommended configuration

Based on the performance results and analysis for each workload, the following NIC and OS configuration changes are recommended.

Table 6 Broadcom BCM57810 NDIS mode recommended adapter configuration

Setting	Default value	Recommended value
Flow Control	Auto	Rx & Tx Enabled
Jumbo packet	1514	9614
Receive buffers	0=Auto	3000
TCP Connection Offload	Disabled	Enabled
Transmit Buffers	0=Auto	5000

Table 7 Broadcom BCM57810 NDIS mode recommended Windows Server 2008 R2 TCP configuration

Setting	Default value	Recommended value
Chimney Offload State	Automatic	Enabled

5.2 Broadcom BCM57810 ISOE mode recommended configuration

Based on the performance results and analysis for each workload, the following NIC configuration changes are recommended.

Since all TCP functions for the SAN interfaces are offloaded in iSOE mode, the Windows Server 2008 R2 TCP configuration has no effect and no changes are required.

Table 8 Broadcom BCM57810 iSOE mode recommended adapter configuration

Setting	Default value	Recommended value
Flow control	Auto	Rx & Tx enabled
MTU	1500	9600

5.3 Intel X520 recommended configuration

Based on the performance results and analysis for each workload, the following NIC configuration changes are recommended.

Table 9 Intel X520 recommended adapter configuration

Setting	Default value	Recommended value
Jumbo packet	Disabled	Enabled
Receive Buffers	512	4096
Transmit Buffers	512	16384

6 Conclusion

For an EqualLogic PS Series SAN, enabling jumbo frames and flow control for both the Broadcom 57810 and the Intel X520 is recommended. If not using the Broadcom iSCSI Offload Engine, receive and transmit buffers should also be maximized.

When using the Broadcom BCM57810, TCP Offload Engine should be enabled for its ability to decrease CPU utilization and also to lower retransmission rates in congested networks.

The Broadcom BCM57810 iSCSI Offload Engine is another compelling option. Not only did it decrease retransmission rates in a congested network and lower CPU utilization by an even greater amount than did TOE, it also exhibited performance benefits during every workload.

One thing to consider when using iSOE is the difference in administration procedures. Once iSOE is enabled, the network adapter disappears from the native Windows Server network management and monitoring tools and appears as a storage controller in Windows Server device manager. It must be configured and monitored in the Broadcom Advanced Control Suite.

A Test configuration details

Hardware	Description
Blade enclosure	Dell PowerEdge M1000e chassis: • CMC firmware: 4.31

Blade server	Dell PowerEdge M620 server:
	• Windows Server 2008 R2 Datacenter SP1
	• BIOS version: 1.6.1
	• iDRAC firmware: 1.37.35
	• (2) Intel® Xeon® E5-2650
	• 64GB RAM
	• Dual Broadcom 57810S-k 10GbE CNA
	  • NDIS driver: 7.4.23.0
	  • ISOE driver: 7.4.3.0
	  • Firmware: 7.4.8
	• Dual Intel x520-k 10GbE CNA
	  • Driver: 2.11.114.0
	  • Firmware: 14.0.12
	• Dell EqualLogic Host Integration Toolkit v4.5.0

Blade I/O modules	(2) Dell 10Gb Ethernet Pass-through module
SAN switches	(2) Dell Force10 s4810 • Firmware: 8.3.12.1
SAN array members	(2) Dell EqualLogic PS6110XV • (2) 10GbE controllers • Firmware: 6.0.4

B Network adapter and TCP stack configuration details

This section provides more detail about the configuration options and default settings of the network adapter properties and the Windows Server 2008 R2 TCP stack.

B.1 Broadcom BCM57810 NDIS mode adapter options

The following table lists the tested adapter options for the Broadcom BCM57810 NetXtreme II 10 GigE NIC in NDIS mode along with the default value.

Table 10 Broadcom BCM57810 NDIS mode adapter options

Setting	Default value
Flow control	Auto
Interrupt moderation	Enabled
Jumbo packet	1514
Large Send Offload V2	Enabled
Number of RSS queues	8
Receive buffers	0=Auto
Receive Side Scaling	Enabled
TCP Connection Offload*	Enabled
TCP/UDP Checksum Offload	Rx & Tx Enabled
Transmit buffers	0=Auto

* To be enabled, this option must be enabled in the NIC adapter settings and Windows Server 2008 R2 TCP Chimney offload must be enabled.

B.2 Configuring Broadcom BCM57810 adapter properties in NDIS mode

Adapter properties for the Broadcom BCM57810 NDIS adapter can be set in the traditional Windows Server adapter properties dialog box in the **Advanced** tab or with the Broadcom Advanced Control Suite (BACS), a separate application from the native Windows management tools.

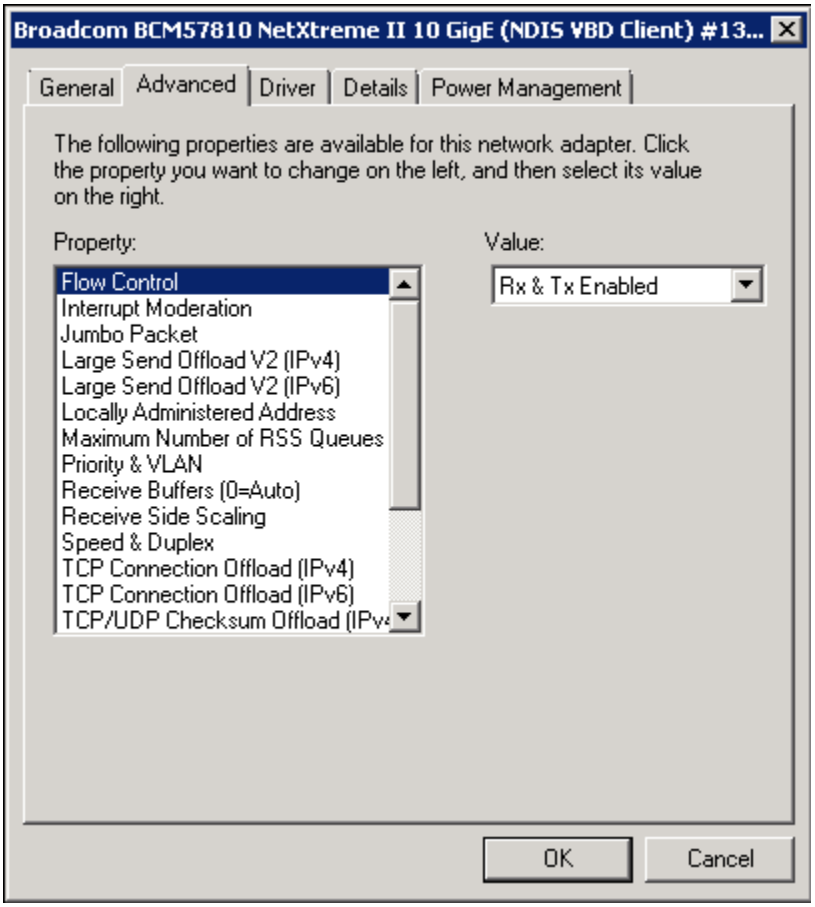

Figure 10 Windows 2008 R2 adapter properties window for the Broadcom NDIS adapter

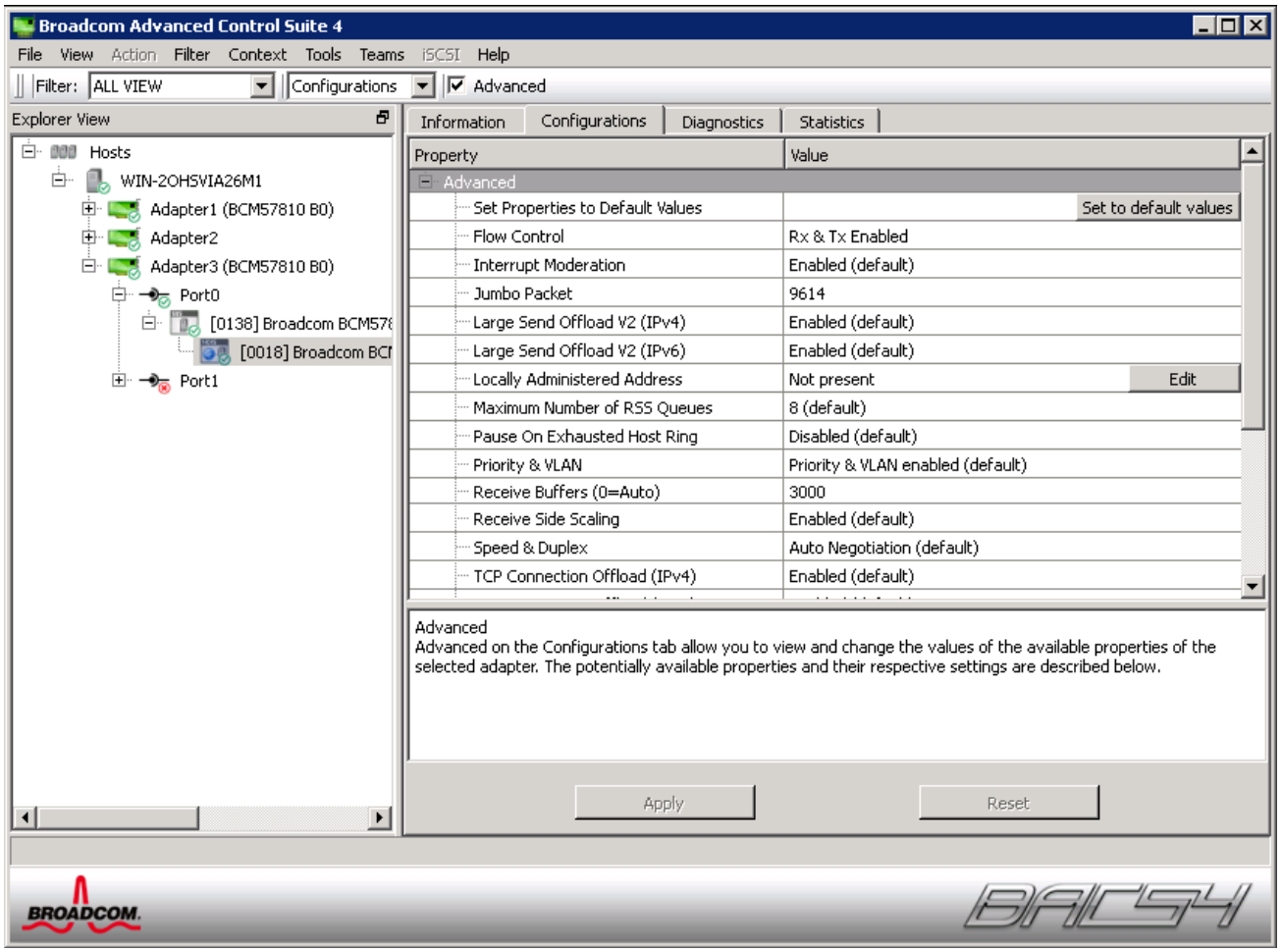

Figure 11 Adapter properties for the Broadcom NDIS adapter using the Broadcom Advanced Control Suite

B.3 Broadcom BCM57810 iSOE mode adapter options

The following table lists the tested adapter options for the Broadcom BCM57810 NetXtreme II 10 GigE NIC in iSOE mode along with the default value.

Table 11 Broadcom BCM57810 iSOE mode adapter options

Setting	Default value
Flow control	Auto
MTU*	1500

* Equivalent to jumbo packet option above.

Once iSOE is enabled, only the Flow control and MTU options are available for configuration.

B.4 Configuring Broadcom BCM57810 adapter properties in iSOE mode

Adapter properties for the Broadcom BCM57810 iSOE adapter must be set in BACS. After enabling iSOE mode with BACS, jumbo frames and flow control settings can be established.

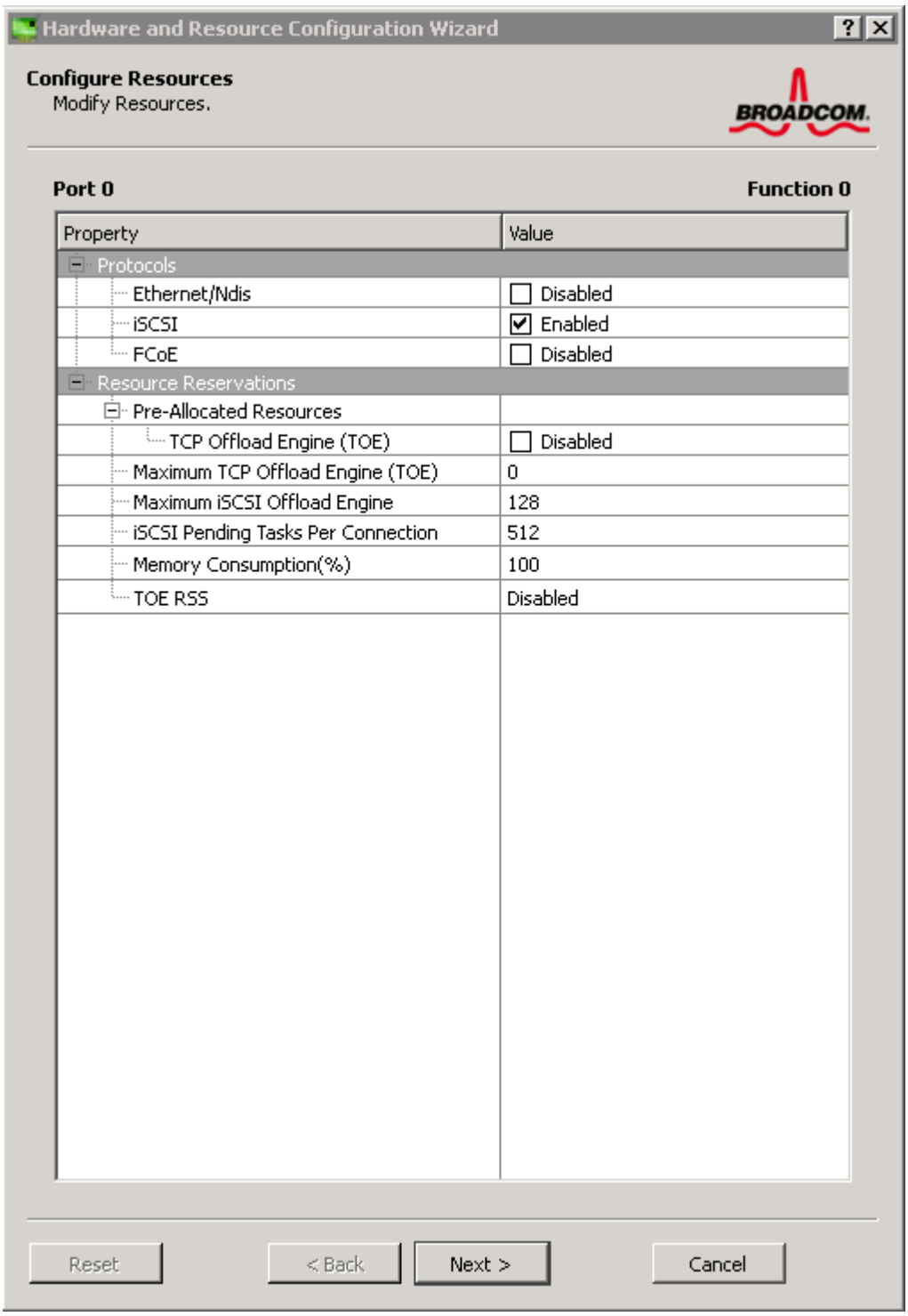

Figure 12 Enabling iSOE mode in the Broadcom Advanced Control Suite

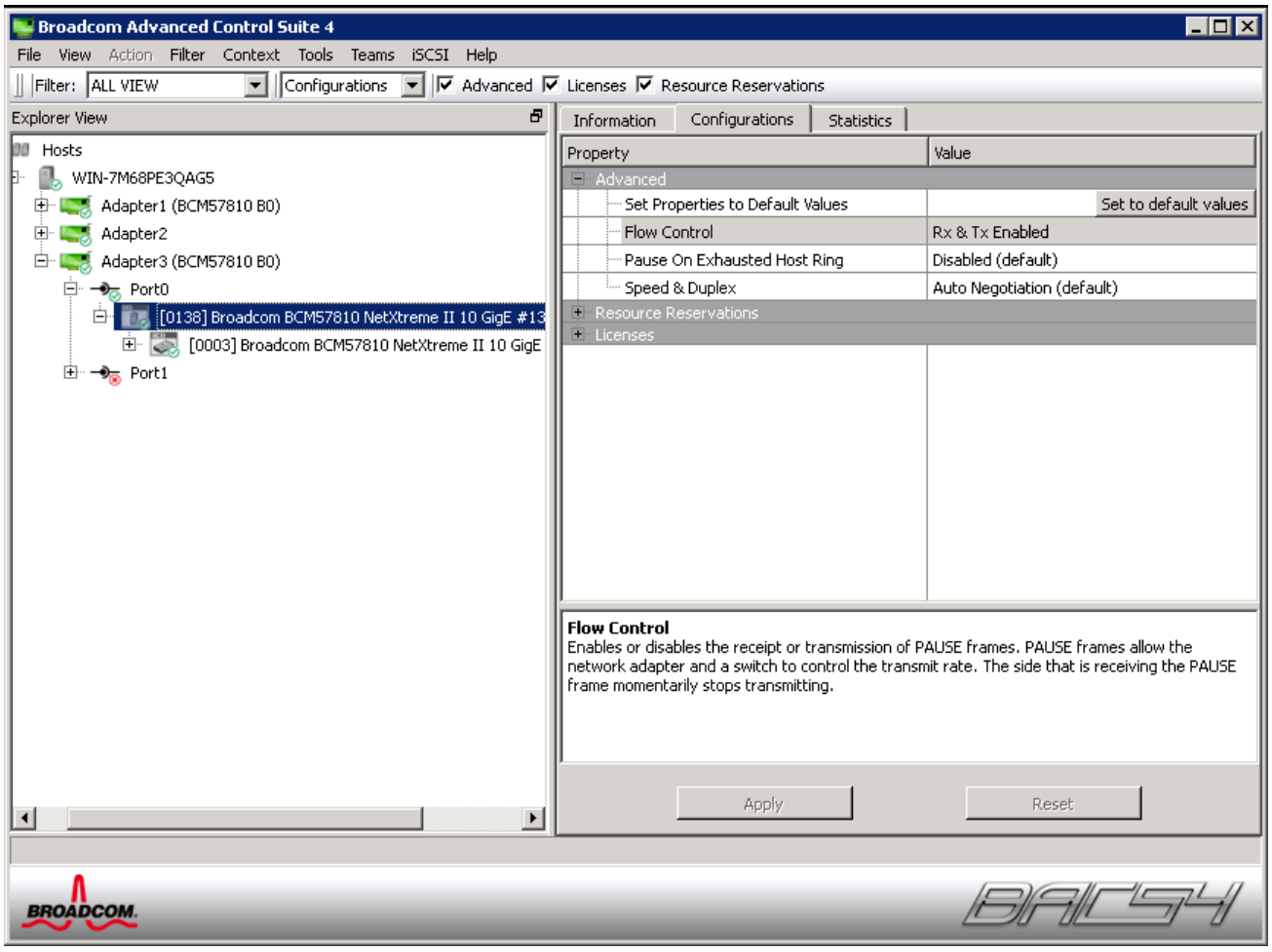

Figure 13 Configuring flow control in the Broadcom Advanced Control Suite

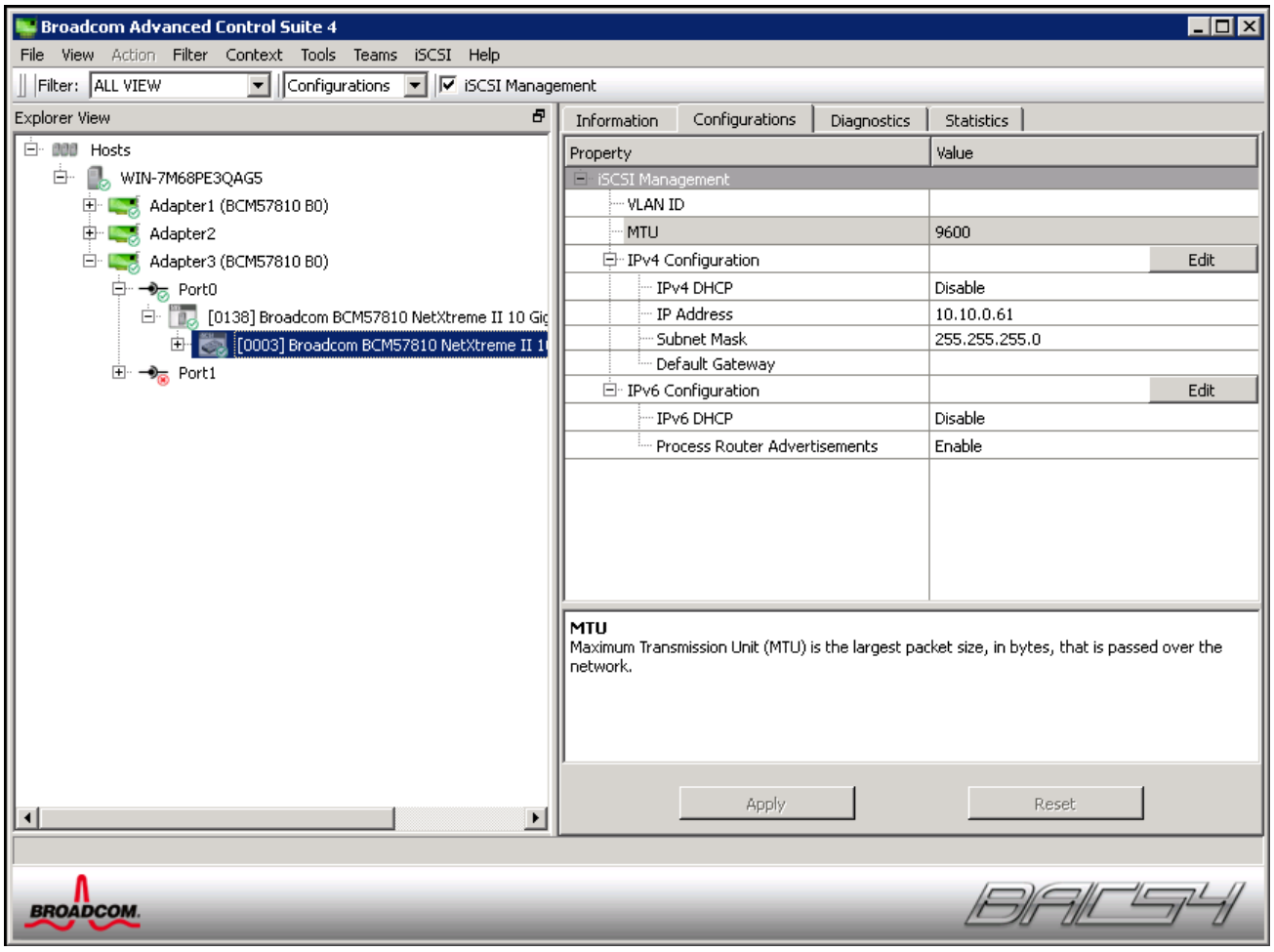

Figure 14 Configuring jumbo frames (MTU) in the Broadcom Advanced Control Suite

B.5 Intel X520 adapter options

The following table lists the tested adapter options for the Intel X520 10 GigE NIC along with the default value.

Table 12 Intel X520 adapter options

Setting	Default value
Interrupt Moderation	Enabled
Jumbo packet	Disabled
Large Send Offload V2	Enabled
Maximum Number of RSS Queues	2
Flow Control	Rx & Tx Enabled
Receive Buffers	512
Transmit Buffers	512
Receive Side Scaling	Enabled
IPv4 Checksum Offload	Enabled
TCP Checksum Offload	Enabled

B.6 Configuring Intel X520 adapter properties

Adapter properties for the Intel X520 NDIS adapter can be set in the traditional Windows Server adapter properties window in the **Advanced** tab.

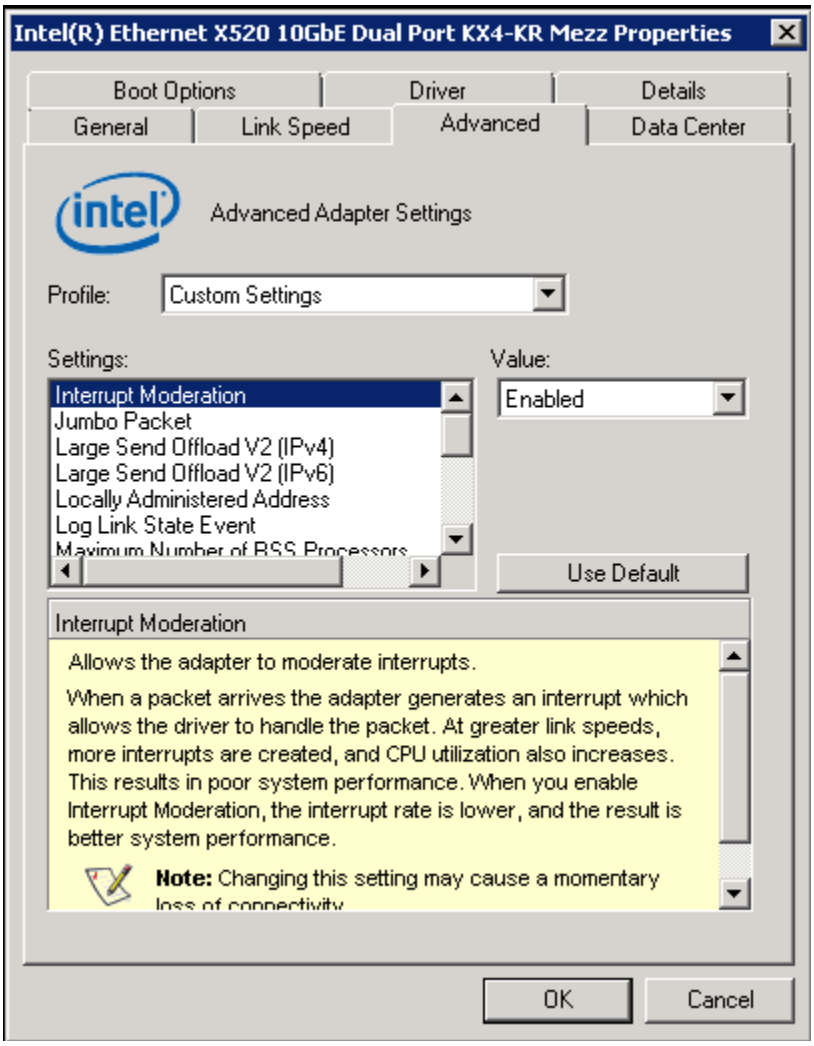

Figure 15 Windows 2008 R2 adapter properties window for the Intel X520 adapter

B.7 Windows Server 2008 R2 TCP stack options

The following table lists the tested TCP stack options for Windows Server 2008 R2 along with the default value.

Table 13 Windows Server 2008 R2 TCP stack options

Setting	Default value
Receive-Side Scaling State*	Enabled
Chimney Offload State*	Automatic
Receive Window Auto-Tuning Level	Normal
Delayed ACK algorithm	Enabled
Nagle's algorithm	Enabled

* There are network adapter options which correspond to these OS options. During test cases, the option's test case value was achieved by changing the option setting at both the network adapter and the OS.

B.8 Configuring the Windows Server 2008 R2 TCP stack

Windows Server 2008 R2 TCP stack options can be configured using the *netsh* command in PowerShell.

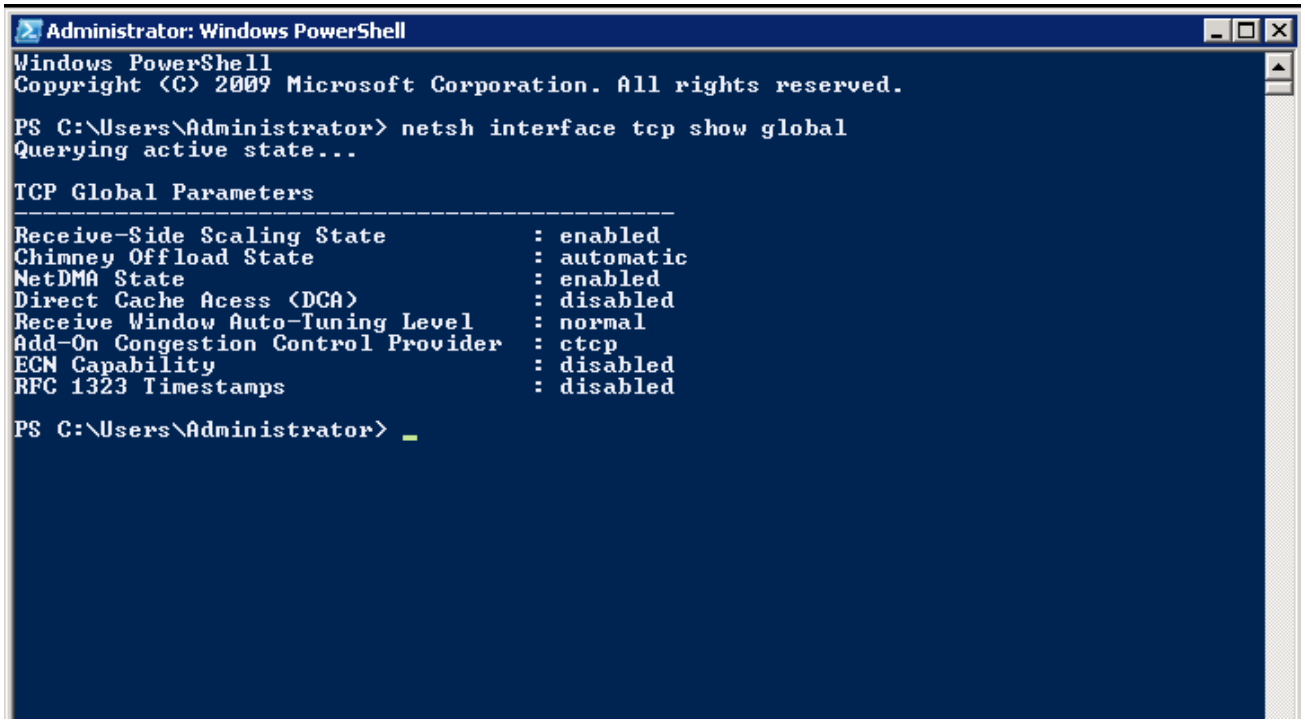

Figure 16 Using the *netsh* command in Windows Server 2008 R2 PowerShell.

To disabled delayed ACK and Nagle's algorithm, create the following entries for each SAN interface subkey in the Windows Server 2008 R2 registry:

Subkey location:

`HKEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ Services \ Tcpip \ Parameters \ Interfaces \ <SAN interface GUID>`

Entries:

`TcpAckFrequency`
`TcpNoDelay`

Value type:

`REG_DWORD, number`

Value to disable:

`1`

B.9 Disabling unused network adapter protocols

Unused protocols can be disabled in the Windows Server 2008 R2 network adapter properties menu.

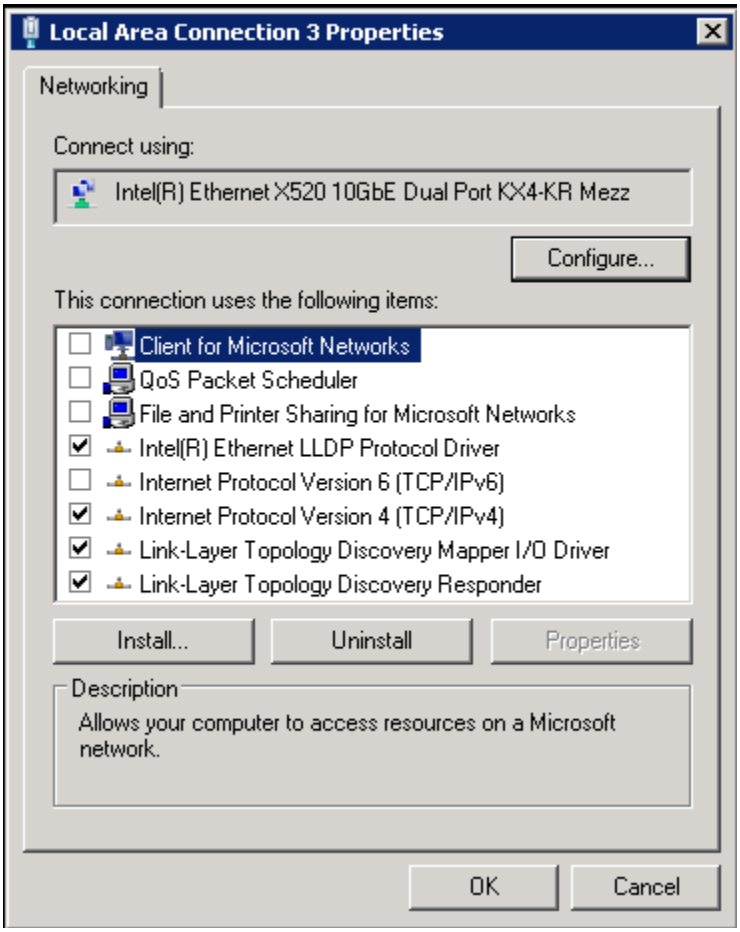

Figure 17 Disabling unused protocols in the Windows Server 2008 R2 network adapter properties menu

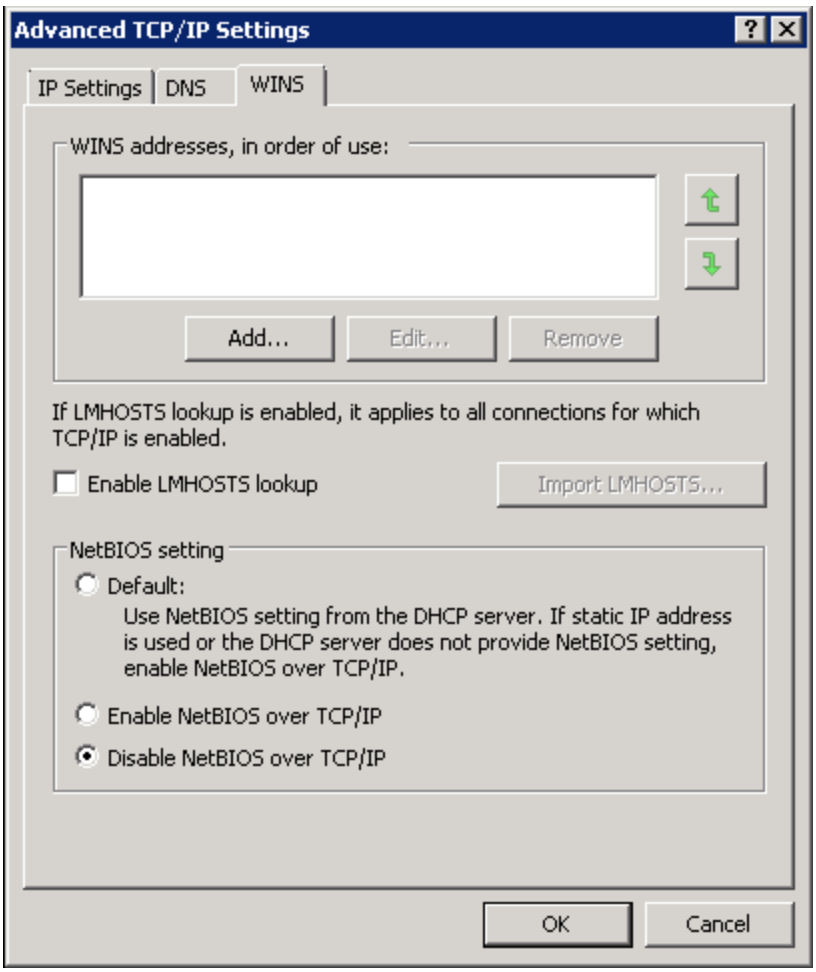

Figure 18 Disabling NetBIOS in the Windows Server 2008 R2 Advanced TCP/IP Settings for the network adapter

C I/O parameters

Vdbench SAN workloads were executed using the following parameters in the parameter file.

Common parameters:

```
hd=default
hd=one,system=localhost
```

iSCSI volumes (random IO):

```
sd=sd1,host=*,lun=\\.\PhysicalDrive1,size=102400m,threads=5
sd=sd2,host=*,lun=\\.\PhysicalDrive2,size=102400m,threads=5
sd=sd3,host=*,lun=\\.\PhysicalDrive3,size=102400m,threads=5
sd=sd4,host=*,lun=\\.\PhysicalDrive4,size=102400m,threads=5
sd=sd5,host=*,lun=\\.\PhysicalDrive5,size=102400m,threads=5
sd=sd6,host=*,lun=\\.\PhysicalDrive6,size=102400m,threads=5
sd=sd7,host=*,lun=\\.\PhysicalDrive7,size=102400m,threads=5
sd=sd8,host=*,lun=\\.\PhysicalDrive8,size=102400m,threads=5
```

iSCSI volumes (sequential IO on two arrays):

```
sd=sd1,host=*,lun=\\.\PhysicalDrive1,size=30m,threads=5
sd=sd2,host=*,lun=\\.\PhysicalDrive2,size=30m,threads=5
sd=sd3,host=*,lun=\\.\PhysicalDrive3,size=30m,threads=5
sd=sd4,host=*,lun=\\.\PhysicalDrive4,size=30m,threads=5
sd=sd5,host=*,lun=\\.\PhysicalDrive5,size=30m,threads=5
sd=sd6,host=*,lun=\\.\PhysicalDrive6,size=30m,threads=5
sd=sd7,host=*,lun=\\.\PhysicalDrive7,size=30m,threads=5
sd=sd8,host=*,lun=\\.\PhysicalDrive8,size=30m,threads=5
```

iSCSI volumes (sequential IO on four arrays):

```
sd=sd1,host=*,lun=\\.\PhysicalDrive1,size=45m,threads=5
sd=sd2,host=*,lun=\\.\PhysicalDrive2,size=45m,threads=5
sd=sd3,host=*,lun=\\.\PhysicalDrive3,size=45m,threads=5
sd=sd4,host=*,lun=\\.\PhysicalDrive4,size=45m,threads=5
sd=sd5,host=*,lun=\\.\PhysicalDrive5,size=45m,threads=5
sd=sd6,host=*,lun=\\.\PhysicalDrive6,size=45m,threads=5
sd=sd7,host=*,lun=\\.\PhysicalDrive7,size=45m,threads=5
sd=sd8,host=*,lun=\\.\PhysicalDrive8,size=45m,threads=5
```

8KB random 67% read workload:

```
wd=wd1,sd=(sd1-sd8),xfersize=8192,rdpct=100,skew=67
wd=wd2,sd=(sd1-sd8),xfersize=8192,rdpct=0,skew=33
```

256KB sequential read workload:

```
wd=wd1,sd=(sd1-sd8),xfersize=262144,rdpct=100,seekpct=sequential
```

256KB sequential write workload:

```
wd=wd1,sd=(sd1-sd8),xfersize=262144,rdpct=0,seekpct=sequential
```

Runtime options:

```
rd=rd1,wd=wd*,iorate=max,elapsed=1200,interval=5
```

Additional resources

Support.dell.com is focused on meeting your needs with proven services and support.

DellTechCenter.com is an IT Community where you can connect with Dell Customers and Dell employees for the purpose of sharing knowledge, best practices, and information about Dell products and your installations.

Referenced or recommended Dell publications:

- EqualLogic Configuration Guide:
  http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19852516/download.aspx
- EqualLogic Compatibility Matrix (ECM):
  http://en.community.dell.com/techcenter/storage/w/wiki/2661.equallogic-compatibility-matrix.aspx
- EqualLogic Switch Configuration Guides:
  http://en.community.dell.com/techcenter/storage/w/wiki/4250.switch-configuration-guides-by-sis.aspx
- The latest EqualLogic firmware updates and documentation (site requires a login):
  http://support.equallogic.com

Force10 Switch documentation:

http://www.force10networks.com/CSPortal20/KnowledgeBase/Documentation.aspx

For EqualLogic best practices white papers, reference architectures, and sizing guidelines for enterprise applications and SANs, refer to Storage Infrastructure and Solutions Team Publications at:

- http://dell.to/sM4hJT

Other recommended publications:

- Performance Tuning Guidelines for Windows Server 2008 R2:
  http://msdn.microsoft.com/en-us/library/windows/hardware/gg463392.aspx