



Third-party information brought to you courtesy of Dell®

NIC Partitioning (NPAR) FAQs

Revision History

<i>Revision</i>	<i>Date</i>	<i>Change Description</i>
2CS57712-K-FAQ101-D10	12/13/11	Updated: <ul style="list-style-type: none">• “Purpose” on page 6• “Audience” on page 6• T6.4 software release by adding VMWare ESX 4.1, FCoE, DCB, and BACS4 information.• “Q6: How many iSCSI offload protocol HBA device partitions are allowed?”• “Q9: What does a partition’s Relative Bandwidth Weight value mean?”• “Q10: What is a partition’s Maximum Bandwidth value range?”• Table 2: “SF Versus NPAR Mode,” on page 14• “Q22: How are Windows BASP Teaming, Linux/VMWare Bonding affected by NPAR mode?”• “Q24: What criteria is used to generate IEEE 802.3x link-level flow control pause frames? ” on page 16 Added: <ul style="list-style-type: none">• “Q1: What protocols does the 57712-k support?”• “Q2: What is Broadcom's NIC Partitioning (NPAR)?”• Table 1: “Protocols Available in Operation Systems Versus SF and NPAR Mode,” on page 9• “Q12: How is the receive direction traffic flow rate affected by different NPAR settings?”• “Q13: Does the 57712-k support DCB Enhanced iSCSI (iSCSI TLV over DCB)?”• “Q29: Is NPAR mode supported at 1 GbE negotiated link speeds on the 57712-k?” on page 18• “Q30: What Dell M1000e Blade Chassis 10 GbE IO Modules support the 57712-k NDC?” on page 18
2CS57712-K-FAQ100-R	04/20/11	Initial release

Broadcom Corporation
5300 California Avenue
Irvine, CA 92617

© 2011 by Broadcom Corporation
All rights reserved
Printed in the U.S.A.

Broadcom®, the pulse logo, Connecting everything®, and the Connecting everything logo are among the trademarks of Broadcom Corporation and/or its affiliates in the United States, certain other countries and/or the EU. Any other trademarks or trade names mentioned are the property of their respective owners.

Table of Contents

About This Document	6
Purpose.....	6
Audience.....	6
Acronyms and Abbreviations.....	7
Technical Support	7
Frequently Asked Questions (FAQ)	8

List of Figures

Figure 1: Broadcom 57712-k Converged NDC..... 6

Figure 2: Non-DCB Mode Zero Sum Relative Bandwidth Weight TX Traffic Flow 11

Figure 3: Non-DCB Mode 100% Sum Relative Bandwidth Weight TX Traffic Flow 12

Figure 4: DCB Mode TX Traffic Flow 13

List of Tables

Table 1: Protocols Available in Operation Systems Versus SF and NPAR Mode9

Table 2: SF Versus NPAR Mode14

About This Document

Purpose

This document provides a list of frequently asked questions (FAQs) pertaining to the NIC Partitioning (NPAR) on a Dell® PowerEdge® M710HD and M915 Blade Servers installed with the Broadcom® 57712-k Converged Network Daughter Card (NDC) Dual Port 10 GbE option (see [Figure 1](#)).

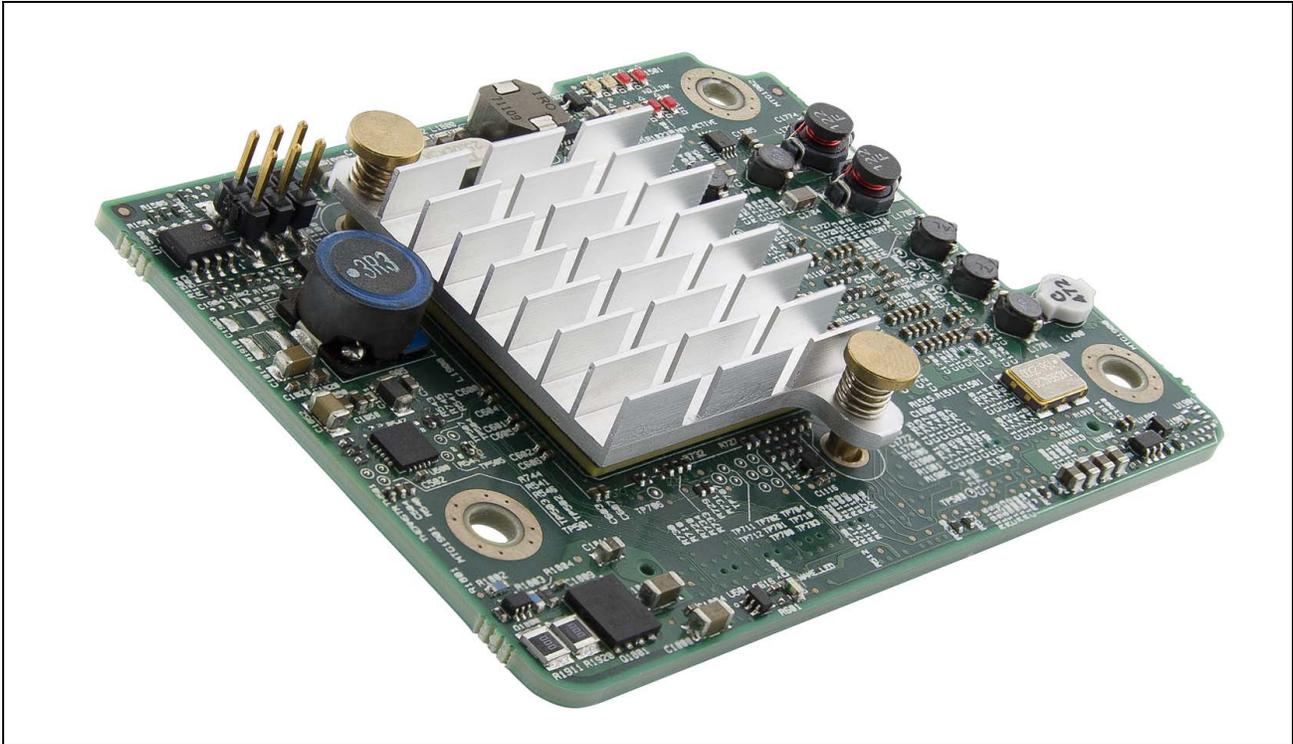


Figure 1: Broadcom 57712-k Converged NDC

Audience

This document is written for the network administrator who wishes to partition the Broadcom network controller on a Microsoft® Windows Server 2008 R2, VMWare® ESX/ESXi 4.1, SUSE Linux Enterprise Server (SLES) and Red Hat® Enterprise Linux® (RHEL) system with up to eight functions (four per port) over Ethernet, and/or four functions (two per port) over iSCSI HBA, or two functions (one per port) over FCoE HBA, plus two functions (one per port) over iSCSI HBA in operating systems where the specific HBA can be enabled.

Acronyms and Abbreviations

In most cases, acronyms and abbreviations are defined on first use. For a comprehensive list of acronyms and other terms used in Broadcom documents, go to:

<http://www.broadcom.com/press/glossary.php>.

Technical Support

Broadcom provides customer access to a wide range of information, including technical documentation, schematic diagrams, product bill of materials, PCB layout information, and software updates through its customer support portal (<https://support.broadcom.com>). For a CSP account, contact your Sales or Engineering support representative.

In addition, Broadcom provides other product support through its Downloads and Support site (<http://www.broadcom.com/support/>).

Frequently Asked Questions (FAQ)

Q1: What protocols does the 57712-k support?

- 10 Gigabit Ethernet (10 GbE)
- TCP Chimney Offload (TOE)
- Data Center Bridging (DCB) (See [Table 1 on page 9](#))
 - Priority-based Flow Control (PFC), IEEE 802.1Qbb
 - Enhanced Transmission Selection (ETS), IEEE 802.1Qaz
 - Data Center Bridging Exchange (DCBx), IEEE 802.1Qaz
- Fibre Channel over Ethernet (FCoE) Hardware Offload
 - Over 1,700,000 IOPs per dual-ported device
- Internet Small Computer System Interface (iSCSI) Hardware Offload
 - iSCSI Type-Length Value (TLV) for DCB
 - Over 1,000,000 IOPs per dual ported device
- Single-Root I/O Virtualization (SR-IOV) Ready
- Switch Independent NIC Partitioning (NPAR)
- Pre-boot eXecution Environment (PXE) Boot
- FCoE Boot from Storage Area Networking (SAN)
- iSCSI Boot

Q2: What is Broadcom's NIC Partitioning (NPAR)?

- Broadcom NPAR allows a single physical adapter to be partitioned into multiple virtual Ethernet/iSCSI/FCoE adapters, which present additional PCI Functions that appear as Networking or Storage devices to the operating system. Broadcom NPAR simultaneously supports up to eight virtual Ethernet adapters (four per port) and four virtual Host Bus Adapter (HBAs) (two per port), with user configurable relative weight and maximum bandwidth allocation for traffic shaping and Quality of Service (QoS) control.
- Broadcom's NPAR technology enables the user to both reduce the number of adapters needed in virtual server environments by uniting multiple lower rate 1 GbE NICs into a single adapter AND increasing overall bandwidth available on limited PCIe slot rack and blade servers; all the while maintaining flexible traffic ports (i.e., Ethernet and HBA offload being available on the same partition at the same time), traffic isolation, service quality, and bandwidth tuning for your management/backup/migration/production networks.

Q3: How many partitions are created per port?

- A maximum of four partitions are created per port. There are two ports on the device, so there are a total of eight partitions made available in NPAR mode.

Q4: What are the supported operating systems?

- The NPAR mode supported OSs are Microsoft Windows Server 2008 (and R2), Microsoft Windows Server HyperV R2, RHEL (v5.5 or better and v6.x), Solaris® 10u9, SUSE Linux Enterprise Server 11 (SLES11) SP1, and VMWare ESX/ESXi 4.1 as shown in [Table 1 on page 9](#). The drivers may not be inbox.

Table 1: Protocols Available in Operation Systems Versus SF and NPAR Mode

Operating System	SF Mode			NPAR Mode		
	Ethernet	iSCSI Offload	FCoE Offload	Ethernet	iSCSI Offload	FCoE Offload
Windows 2008 OS ^a	YES	YES	YES	YES	YES	YES
Windows 2008 R2 OS ^a	YES	YES	YES	YES	YES	YES
Windows 2008 R2 Hyper-V OS	YES	YES	YES	YES	YES	YES
Oracle Solaris 10u9 OS	YES	NO	NO	YES	NO	NO
Linux OSs ^b	YES	YES	YES ^c	YES	YES	YES ^c
VMWare ESX/ESXi 4.0 OS	YES	NO	NO	YES	NO	NO
VMWare ESX/ESXi 4.1 OS	YES	YES	NO	YES	Not certified	NO

a. DCB (DCBX/PFC/ETS) Supported.

b. DCB (DCBX/PFC/ETS) supported in RHEL v6.x and SLES11 SP1 only.

c. FCoE Offload supported in RHEL v6.x and SLES11 SP1 only.

Q5: How many Ethernet protocol device partitions are allowed?

- Four Ethernet protocol devices can be enabled per port for a maximum of eight on the dual-ported 57712-k device.
- For Windows® OSs, the Ethernet protocol can be enabled with or without any other offload protocols being enabled on the same partition.
- For Linux® OSs, the Ethernet protocol for all partitions is always enabled.
- For VMWare® ESX/ESXi 4.1 OSs, the Ethernet protocol for all partitions should always be enabled.
- Windows TOE enabling requires that selected partition also have the Ethernet protocol enabled.

Q6: How many iSCSI offload protocol HBA device partitions are allowed?

- Two iSCSI offload protocol HBA devices can be enabled per port on any two of the four partitions for a total maximum of four on the dual-ported 57712-k device.
- In Windows OSs, the iSCSI offload HBA device can be enabled with or without the Ethernet protocol being enabled on the same partition.
- In Linux OSs, the iSCSI offload protocol must be enabled with the Ethernet protocol being enabled on the same partition.
- The iSCSI offload protocol HBA is not certified by VMWare for ESX/ESXi 4.1 OSs in NPAR mode. The iSCSI non-offload initiator mode is available for all partitions on every port.

Q7: How many FCoE offload protocol HBA device partitions are allowed?

- One FCoE offload protocol HBA devices can be enabled per port on any one of the four partitions, for a total maximum of two on the dual-ported 57712-k device.
- In Windows OSs, the FCoE offload HBA device can be enabled with or without the Ethernet protocol being enabled on the same partition.
- In Linux OSs, the FCoE offload protocol HBA must be enabled with the Ethernet protocol being enabled on the same partition.

- Additionally, one iSCSI offload protocol HBA device can be enabled per port on any of the three remaining partitions.
- A maximum of four HBA offload protocols can be enabled on the dual-ported 57712-k device.
- The FCoE protocol is not supported on VMWare ESX/ESXi 4.1 OSs.

Q8: How should the available traffic type protocols be configured for the various modes?

- For NPAR mode of operation using the iSCSI non-offload pathway initiator mode on Windows, enable the Ethernet protocol for the specific port's partition. On Linux and VMWare ESX/ESXi 4.1, the Ethernet (i.e., L2) protocol should always be enabled.
- For NPAR mode of operation with iSCSI hardware offload HBA initiator mode on Windows/Linux, enable the iSCSI offload protocol for the specific port's partition.
- For NPAR mode of operation with FCoE hardware offload HBA initiator mode on Windows/Linux, enable the FCoE offload protocol for the specific port's partition.
- For Single Function (SF) mode of operation (NPAR disabled), all three protocols (Ethernet, iSCSI, and FCoE offload) are available by default, on both ports for Windows and Linux OSs. The VMWare ESX/ESXi 4.1 OS's support both Ethernet and iSCSI offload protocols only in SF mode.
- Note: For Windows, the offload protocol resource may need to be enabled for the specific offload device to be detected/enumerated when the device is first installed.

Q9: What does a partition's Relative Bandwidth Weight value mean?

- Each individual partition's Relative Bandwidth Weight value can be from 0% to 100%, but the sum of the Relative Bandwidth Weights for all four partitions on the same port **must** equal either 0% or 100%.
- The Relative Bandwidth Weight value is more than a simple minimum bandwidth setting. It is used to dynamically determine the bandwidth ratio of a partition's send traffic verses the other partition's sending traffic. The send traffic is approximately that partition's Relative Bandwidth Weight value divided by the sum of all the other current actively sending partition's Relative Bandwidth Weight values. This assumes each individual actively sending partition is able to reach that amount via their individual sending applications and that partition's Maximum Bandwidth value is not restricting the flow to below that amount. If an individual partition's send traffic flow can not fill the level allocated to it, the other actively sending partitions will attempt to use the surplus bandwidth within the constraints of their Relative Bandwidth Weight and Maximum Bandwidth settings.
- Setting the individual partition's Relative Bandwidth Weights (in USC) to equal a sum other than 0% or 100% will result in undefined behavior.
- Setting the Relative Bandwidth Weight value to all 0%'s is not exactly the same as setting them all to 25%. All 0%'s means each individual traffic flow gets equal weight with respect to the other traffic flows while setting them all to 25% means each partition's aggregated traffic flow gets equal weight with respect to each of the other partition's aggregated traffic flows.
- If the SUM of the four partitions on the same port is not 0%, and one or more of the partitions has a 0% Relative Bandwidth Weight value, then the Relative Bandwidth Weight value used by the sending bandwidth allocation logic is actually 1%. This means that the send traffic percent allocated to a partition with 0% is never zero.
- The reset default Relative Bandwidth Weight value is 0% for each partition.

Q10: What is a partition's Maximum Bandwidth value range?

- A partition's Maximum Bandwidth value range is 1 to 100 in increments of 1% of the port's current Link Speed (in 100 Mbps increments for a 10 GbE link and in 10 Mbps increments for a 1 GbE link when the

57712-k's KR interface is connected to an IO Module with a 1 GbE link). The sum total of the four partition's Maximum Bandwidth values on the same port do not have to equal 100. The allowable sum-total value range is from four (undersubscribed – all four partitions set to 1) to 400 (oversubscribed – all four partitions set to 100).

- If the four partitions were set to undersubscribe the available bandwidth, then all unassigned transmit/send direction bandwidth would be unavailable for use by any of the four partitions – i.e., if the Maximum Bandwidth sum of the four partitions on a port was 4%, then the unassigned 96% of the available bandwidth would be unavailable and could never be used.
- The reset default Maximum Bandwidth value is 25% for each partition.

Q11: How is the NPAR transmit direction traffic flow rate affected by different mode settings?

- In non-DCB mode, where the sum of the partition's Relative Bandwidth Weights equals zero (i.e., each partition's Relative Bandwidth Weight is set to zero), each individual traffic flow – in this example, Partition 1's iSCSI (P1i) and Ethernet (P1e) through Partition 4's Ethernet (P4e) and iSCSI (P4i) – is equally scheduled to transmit within the limitations of the partition's Maximum Bandwidth and the overall connection's link speed as shown in [Figure 2](#). This means if the Maximum Bandwidth of a specific partition is set to less than 100%, then the traffic flows sharing that partition will be further restricted to where their combined traffic flow bandwidth will be capped by that per partition setting. The actual inter-partition ratio of the two sharing traffic flows is controlled by the host OS. Think of the Maximum Bandwidth as a fixed sized funnel with the OS determining how the two sharing traffic types are pouring into that funnel. If all four partition's individual Maximum Bandwidths are set to 100% (i.e., they are all unrestricted), then each actively sending traffic flow (without regard to which partition they are on) will equally share the transmit directions total bandwidth (i.e., TX link speed).

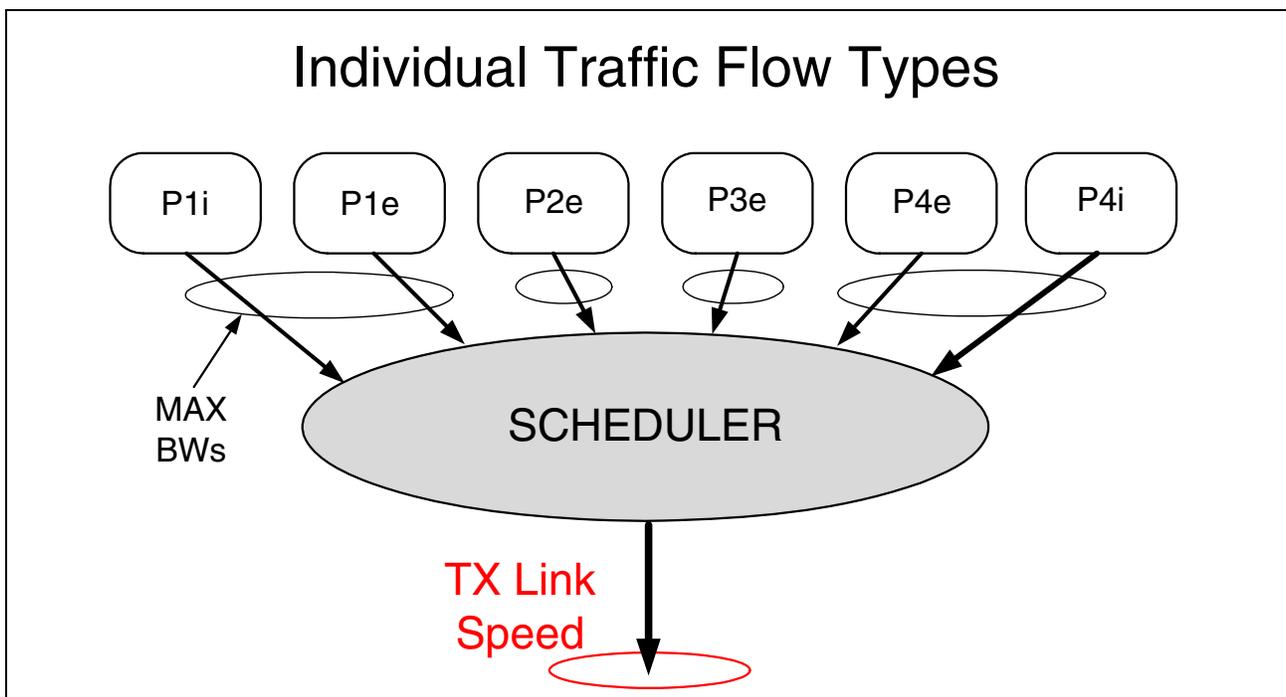


Figure 2: Non-DCB Mode Zero Sum Relative Bandwidth Weight TX Traffic Flow

- In non-DCB mode, where the sum of the partition's Relative Bandwidth Weights equal 100, each Partition's combined traffic flow is equally scheduled to transmit within the limitations of the partition's Relative Bandwidth Weight and Maximum Bandwidth settings and the overall connection's link speed as shown in [Figure 3](#). This means a specific partition's Relative Bandwidth Weight value will restrict the traffic flows sharing that partition's bandwidth allocation, as if one combined traffic flow with respect to the other actively sending partitions. The partition's send flow rate is based on the ratio of that partition's individual weight versus the aggregated weights of all the other actively sending partitions. Furthermore, each partition's combined traffic flow will be capped by that partition's Maximum Weight setting. The actual inter-partition ratio of the two sharing traffic flows is controlled by the host OS. Think of the dynamic weight ratio as a variable sized funnel that could be further restricted by the Maximum Bandwidth fixed sized funnel with the OS determining how the sharing traffic types are pouring into the combined funnels.

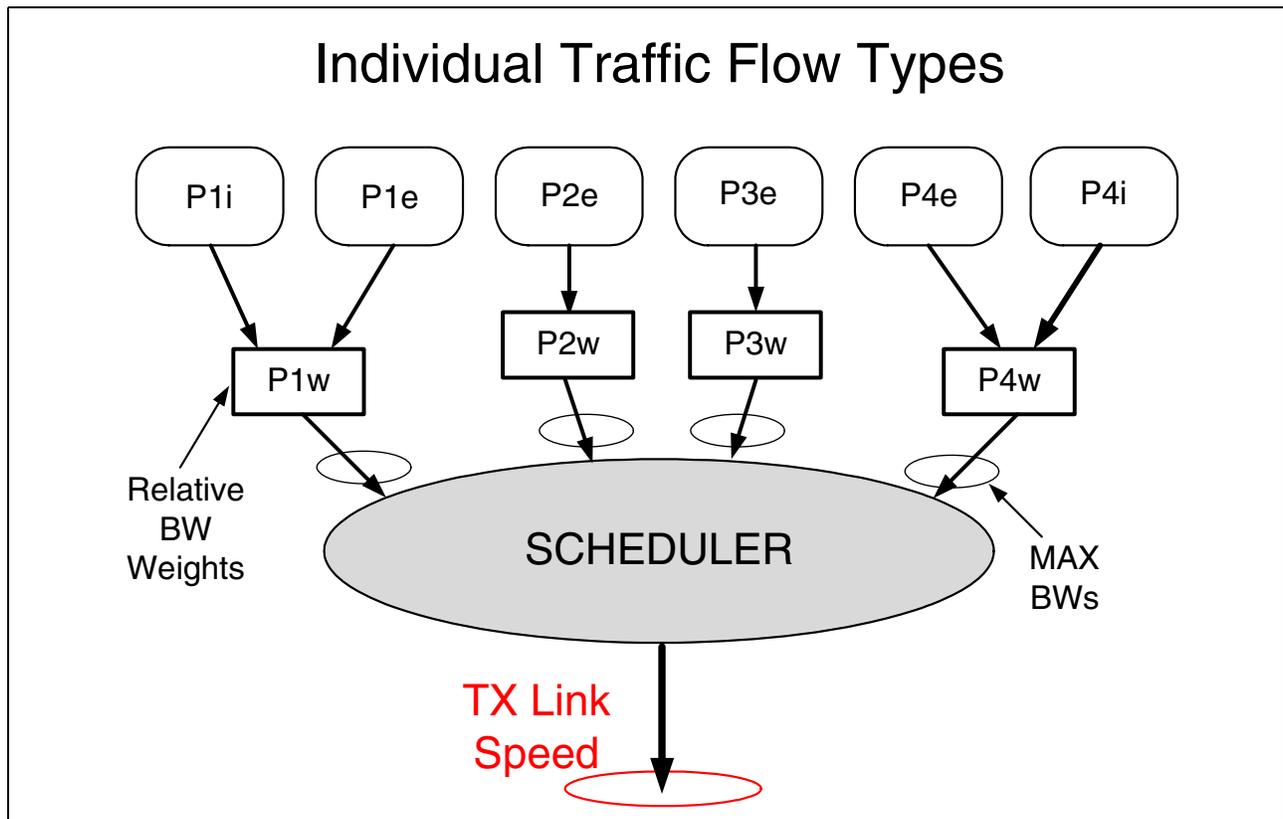


Figure 3: Non-DCB Mode 100% Sum Relative Bandwidth Weight TX Traffic Flow

- In DCB mode, all of the Partition's Relative Bandwidth Weights are disregarded and the individual traffic flows are scheduled to transmit within the limitations of the Priority Group's ETS value (determined by its Traffic Type – this example shows the 1st Partition have both an FCoE enabled and Ethernet enabled protocol transmitting to two different Priority Groups) and each partition's Maximum Bandwidth setting and the overall connections link speed as shown in [Figure 4 on page 13](#). In this example, the FCoE traffic type is in PG1 and all of the other traffic types (iSCSI and Ethernet) are in the PGO. Each Priority Group has its own ETS value. DCB Lossless iSCSI (iSCSI-TLV) could be used in place of FCoE in this example for a similar effect where the Lossless iSCSI Offloaded traffic would go through its assigned Priority Group while the Lossy Ethernet traffic would go through the other. Similarly to the

other two rate controlling modes, the host OS determines the actual inter-partition traffic ratio for the cases where two traffic types share the same partition. See the User Guide's examples for more details.



Note: VMWare ESX/ESXi 4.1 does not support DCB mode of operation on the 57712-k.

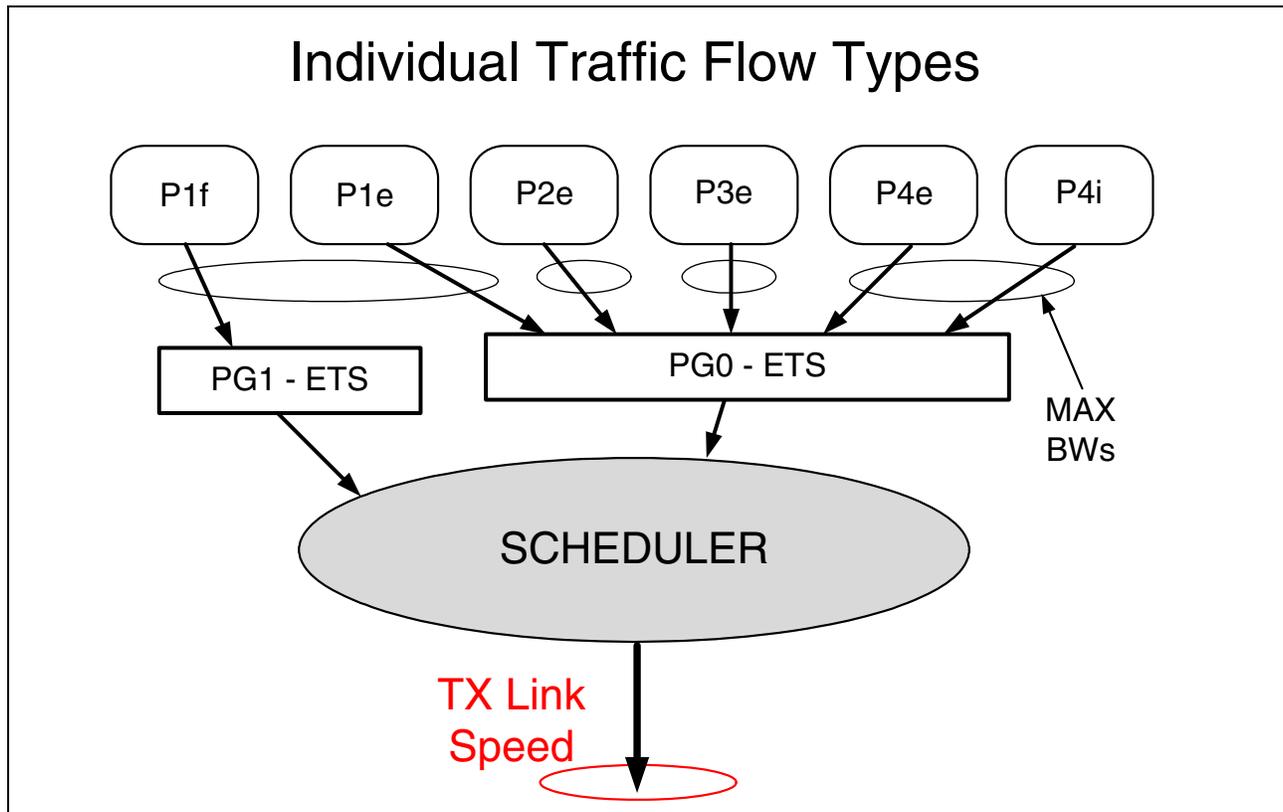


Figure 4: DCB Mode TX Traffic Flow

Q12: How is the receive direction traffic flow rate affected by different NPAR settings?

- A partition's Relative Bandwidth Weight and Maximum Bandwidth values affect only its send/transmit/outgoing traffic.
- The receive direction traffic of the four partitions of a single port are not weighted or restricted by the 57712-k device NPAR settings. The receive direction acts similarly to when in SF mode: the incoming traffic on the port is controlled by the switch port it is connected to.

Q13: Does the 57712-k support DCB Enhanced iSCSI (iSCSI TLV over DCB)?

- Yes, the 57712-k can assign iSCSI offload traffic to be Lossless and to its own dedicated priority group independent of the other TCP/IP networking or FCoE traffic, in both SF and NPAR modes.

Q14: How is Windows OS TOE controlled?

- You can use BACS4 (NDIS Client – Configuration – Advanced Properties) or Windows Device Manager (NDIS Client – Advanced Properties) to enable or disable TOE per partition (in NPAR mode) or port (in SF mode).
- Alternately, TOE can be enabled or disabled for the entire system (i.e., all ports) using the “`netsh int tcp set global chimney=x`” DOS command line instructions where the “x” can be “enabled” or “disabled.”

Q15: How does a partition’s Maximum Bandwidth value affect the displayed Link Speed of the partition adapter?

- A partition’s Maximum Bandwidth value is used by some OS applications to display the partition device/adaptor link speed. This value represents the specific partition's send/transmit/outgoing maximum speed and not the receive/incoming link speed, which is always the port’s current link speed.

Q16: What are the similarities and differences between SF and NPAR modes for advanced features such as stateless offload, teaming, VLAN, etc?

- Refer to [Table 2](#) below.

Table 2: SF Versus NPAR Mode

Feature	SF	NPAR
Ethernet Adapters	2 – 1 per port	Up to 8 – 4 per port; see FAQ entry See “Q5: How many Ethernet protocol device partitions are allowed?” on page 9.
iSCSI Hardware Offload HBAs	2 – 1 per port	Up to 4 – 2 per port; see FAQ entry See “Q6: How many iSCSI offload protocol HBA device partitions are allowed?” on page 9.
FCoE Hardware Offload HBAs	2 – 1 per port	Up to 2 – 1 per port; see FAQ entry See “Q7: How many FCoE offload protocol HBA device partitions are allowed?” on page 9.
Windows TCP Chimney Offload (TOE)	Yes	Yes
TCP Segmentation Offload (TSO)	Yes	Yes
Large Send Offload (LSO) and Large Receive Offload (LRO)	Yes	Yes
Giant Send Offload (GSO)	Yes	Yes
IPv4/IPv6 TCP-UDP-IP Checksum Offload (CO)	Yes	Yes
IP Address Multi-Homing	Yes	Yes
PXE/iSCSI/FCoE Boot	Yes	Yes – first partition of each port; see FAQ entry See “Q17: How are PXE/iSCSI/FCoE remote boots affected by NPAR mode?” on page 15.
Wake On LAN (WOL)	Yes	Yes – see FAQ entry See “Q18: How is Wake On LAN (WOL) Out Of Band (OOB) affected by NPAR mode?” on page 15.

Table 2: SF Versus NPAR Mode (Cont.)

Feature	SF	NPAR
PAUSE Flow Control	Yes	Yes – see FAQ entry See “Q23: Is IEEE 802.3x link-level flow control supported in SF and NPAR mode?” on page 16.
Jumbo Frames (up to 9600 Bytes)	Yes	Yes
iSCSI Offloads	128 per port	128 per partition – for up to two iSCSI Offload Protocol enabled partitions
RX/TX L2 Ethernet Queues (RSS/NetQueue/VMQ)	16 per port	2 per partition
MSI and MSI-X	Yes	Yes
OS specific Multi-Path I/O (MPIO)	Yes	Yes
Teaming – Smart Load Balancing – Switch Independent	Yes	Yes – see FAQ entry See “Q22: How are Windows BASP Teaming, Linux/VMWare Bonding affected by NPAR mode?” on page 16.
Teaming – Link Aggregation – Switch Dependent	Yes	No – see FAQ entry See “Q22: How are Windows BASP Teaming, Linux/VMWare Bonding affected by NPAR mode?” on page 16.
Teaming – Generic Trunking – Switch Dependent	Yes	No – see FAQ entry See “Q22: How are Windows BASP Teaming, Linux/VMWare Bonding affected by NPAR mode?” on page 16.
VLAN (insertion/stripping/filtering)	Yes	Yes

Q17: How are PXE/iSCSI/FCoE remote boots affected by NPAR mode?

- Remote boots will always use the first partition of each port (as if not in partitioned mode) as that boot port’s MAC address.
- Windows and Linux both support PXE/iSCSI/FCoE remote boots.
- VMWare ESXi 4.1 supports PXE/iSCSI remote boots.
- VMWare ESX/ESXi 4.1 does not support FCoE remote boots.
- VMWare ESX 4.1 (the non "i" version) supports PXE but not iSCSI remote boots.
- In SF or NPAR modes, PXE booting of Windows Hyper-V VMs over a Smart Load Balanced and Failover (SLB) Team is not supported; instead use the Switch Dependent Teaming (IEEE 802.3ad Link Aggregation (LACP) and Generic/Static Link Aggregation (Trunking)) modes.

Q18: How is Wake On LAN (WOL) Out Of Band (OOB) affected by NPAR mode?

- Windows/Linux/VMWare ESX/ESXi 4.1 WOL – OOB Advanced Configuration and Power Interface (ACPI) S1/S3/S4/S5 states are supported for function 0 and 1 (i.e., the first partition of each port) and for 1 Gbps link speeds only.

Q19: How is non-OOB WoL affected by NPAR mode?

- Windows WOL ACPI S4/S5 states are supported for all partitions of both ports at 1 Gbps link speed only.
- Linux/VMWare ESX/ESXi 4.1 WOL ACPI S5 state is support for all partitions of both ports at 1 Gbps link speed only.

Q20: How is NPAR enabled through USC?

- Reboot to USC using the Dell® UEFI bootup option and follow the setup guide instructions. More information on USC and the UEFI boot option can be found on the Dell support site.

Q21: How are the NPAR settings modified in the operating system?

- In Windows OS's, you can change and enable/disable the partitioning mode or settings in the operating system using BACS4.
- In Linux/VMWare, you can only change and enable/disable NPAR settings in USC. USC can also be used to modify the NPAR settings in Windows.
- In all cases, a system reboot is required before any changes take effect.

Q22: How are Windows BASP Teaming, Linux/VMWare Bonding affected by NPAR mode?

- A team cannot contain more than one partition from the same port.
- NPAR adapters can be teamed with other non-NPAR adapters, such as for example the BCM5709 and BCM57711.
- In Windows OS, Smart Load Balanced and Failover (SLB with or without Auto-Fallback) switch independent teaming mode is available for teams using NPAR adapters.
- Switch dependent teaming (IEEE 802.3ad LACP and Generic/Static Link Aggregation (Trunking)) cannot use an NPAR virtual adapter. This is due to the way the IEEE standards require Switch Dependent Teaming (IEEE 802.3ad LACP and Generic/Static Link Aggregation (Trunking)) mode to work per entire port instead of MAC address (fraction of a port) granularity.

Q23: Is IEEE 802.3x link-level flow control supported in SF and NPAR mode?

- Yes. Link-level flow control is available in both SF and NPAR modes of operation when DCB's PFC mode is not enabled.
- In non-DCB SF mode, link-level flow control is configured on a per-port basis through the Advanced Properties tab of the NDIS driver of Windows Device Manager, through the Advanced Properties tab of BACS4, or through the Linux/VMWare Ethtool utility.
- In non-DCB NPAR mode, link-level flow control is configured on a per-port basis through USC or BACS4. It is not configured on a per-partition basis since multiple partitions share the same port and link-level flow control operates on a per-port basis. This means that the Advanced Properties (on Windows) or the Ethtool utility (on Linux/VMWare) cannot be used to control link-level flow control in NPAR mode.

Q24: What criteria is used to generate IEEE 802.3x link-level flow control pause frames?

- By default, link-level flow control will only generate pause frames in response to a temporary lack of internal hardware buffering resources.
- Due to the nature of Converged NIC's, networking (L2 Ethernet) host buffers and storage (iSCSI/FCoE hardware offloaded) host buffers are maintained separately. If a lack of networking host buffers were allowed to generate link-level pause frames then storage traffic would also be paused. As a result, Broadcom has chosen not to generate link-level pause frames when the L2 host buffers are limited but instead drop frames destined for the host buffer limited function. This behavior is different from

previous generation products but is intended to provide the best possible performance across both networking and storage functions in the infrequent case where host buffers are low.

- Windows, VMWare, and Linux all provide an option to restore the previous link-level flow control behavior of generating pause frames in response to limited L2 host buffers. In Windows, the option is called “Pause On Exhausted Host Ring” (Windows Device Manager NDIS Advanced Properties or BACS4’s NDIS Configuration Advanced Properties) for each port. In Linux, use the module option parameter named “dropless_fc” (using the “**modprobe bnx2x dropless_fc=1**” or equivalent “insmod” commands). In VMWare, use the command "esxcfg-module -s dropless_fc=1 bnx2x".
- Both Linux and VMWare commands control all partitions of both ports simultaneously. When enabled, link-level pause frames will be generated if any host buffers on the designated partition are low, resulting in paused traffic to all functions on all partitions. In non-DCB NPAR mode, in Windows, all four partitions’ “Pause On Exhausted Host Ring” settings’ must be enabled for it to be ON or disabled for it to be OFF, where as in Linux and VMWare all NetXtremeII bnx2x devices on the host system are enabled (1) or disabled (0) using a single command.
- IEEE 802.3x link-level flow control is disabled when DCB’s PFC mode is enabled.

Q25: Is IEEE 802.1Qbb PFC supported in SF and NPAR modes?

- Yes. PFC is negotiated as part of DCB which is controlled on a per-port basis. The Data Center Bridging eXchange (DCBx) negotiated loss-less traffic type will have PFC support. Lossy traffic types are not flow control protected beyond what is provided by the upper layers (such as with TCP/IP). The 57712-k DCB supports both loss-less FCoE and loss-less iSCSI (iSCSI TLV) modes of operation.
- If PFC is not enabled during the DCBx exchange, IEEE 802.3x link-level flow control’s settings will apply.

Q26: Does NPAR support Jumbo frames?

- Yes. Jumbo frames up to 9600 bytes are supported in both Windows and Linux OSs on all eight partitions, for both Ethernet and iSCSI offload protocol modes in both SF and NPAR modes.
- In SF mode, VMWare ESX/ESXi 4.1 currently only supports standard frames (MTU = 1500 bytes) for iSCSI hardware offload and Jumbo frames for L2 Ethernet (which includes iSCSI non-offload initiator mode).
- In NPAR mode, VMWare ESX/ESXi 4.1 supports jumbo frames (MTU up to 9600 bytes) for L2 Ethernet (which includes iSCSI non-offload initiator mode).
- For Windows and Linux OS’s, FCoE protocol’s MTU is always fixed at 2500 bytes. The VMWare ESX/ESXi 4.1 OS’s do not support the FCoE protocol on the 57712-k.

Q27: How can I view the current NPAR settings in the operating system?

- In Windows, you can see what protocols are enabled for a partition in both Windows Device Manager and BACS4. You can only see a partition’s Relative Bandwidth Weight settings in BACS4. The Maximum Bandwidth value for an Ethernet protocol-enabled partition can be found in the Network Connections Status window or BACS4. If a partition has only the iSCSI or FCoE offload protocol enabled, you will only see the device and not its Maximum Bandwidth value in the Windows Device Manager but everything is viewable and configurable in BACS4.
- In Linux OSs, you can see the Ethernet, iSCSI and FCoE protocol-enabled partitions are loaded and available for an application’s use. You will not be able to see a partition’s Relative Bandwidth Weight settings. You can see a partition’s Maximum Bandwidth settings.

- In VMWare ESX/ESXi 4.1, you can see the Ethernet protocol-enabled partitions are loaded and available for an application's use. You will not be able to see a partition's Relative Bandwidth Weight settings. You can see a partition's Maximum Bandwidth settings.

Q28: What does a switch see when the 57712-k is in NPAR mode?

- On the 57712-k, a switch will see additional MAC addresses (up to four L2 Ethernet MAC addresses) plus two additional iSCSI offload MAC addresses (depending upon the NPAR configuration) which is slightly more than when the 57712-k is in SF mode.
- In SF mode, the same port will have one L2 Ethernet MAC address and one additional iSCSI offload MAC address. This is no change from the previous generation's 5709, 57710, and 57711 devices which can also present up to two MAC addresses per port (L2 Ethernet and iSCSI hardware offload).

Q29: Is NPAR mode supported at 1 GbE negotiated link speeds on the 57712-k?

- Yes. The M1000e A Fabric IO Module would have to support 1 GbE link. Supported 1 GbE IO Modules include: 1 GbE Pass-Through, M6220, M6348, Cisco® 3032, Cisco 3130G, Cisco 3130X, and M8024. Verify support with Dell.
- The overall rate would be 1 Gbps instead of 10 Gbps.
- DCB and FCoE are not supported on 1 GbE connections.
- Windows TOE and iSCSI Hardware Offload are supported on 1 GbE connections on the applicable OS's.

Q30: What Dell M1000e Blade Chassis 10 GbE IO Modules support the 57712-k NDC?

- The currently supported 10 GbE IO Modules include: 10 GbE Pass Through-k, M8024-k, and M8428-k.
- For FCoE and DCB Lossless operation (to include iSCSI-TLV), the IO Module must support DCB capability exchange (DCBX). See Dell support for the latest IO Module software versions to ensure DCBX support.
- To support DCBX, the 10 GbE Pass Through-k IO module must be externally connected to a switch that supports DCBX.

Dell has tested and certified the Broadcom 57712-k Dual-port 10 GbE Converged Network Daughter Card with NIC Partitioning, TOE, iSCSI, and FCoE ready technology. Dell specifically disclaims knowledge of the accuracy, completeness, or substantiation for all statements and claims made in this document regarding the properties, speeds, or qualifications of the adapter.

Broadcom® Corporation reserves the right to make changes without further notice to any products or data herein to improve reliability, function, or design.

Information furnished by Broadcom Corporation is believed to be accurate and reliable. However, Broadcom Corporation does not assume any liability arising out of the application or use of this information, nor the application or use of any product or circuit described herein, neither does it convey any license under its patent rights nor the rights of others.

Connecting
everything®



BROADCOM CORPORATION

5300 California Avenue

Irvine, CA 92617

© 2011 by BROADCOM CORPORATION. All rights reserved.

Phone: 949-926-5000

Fax: 949-926-5203

E-mail: info@broadcom.com

Web: www.broadcom.com