Switch Independent Partitioning FAQs
## Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Change Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2CS57712-K-FAQ103-R</td>
<td>04/09/13</td>
<td>Updated:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q1: What protocols does the 57712-k/578xxS controllers support?” on page 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q2: What is Broadcom's Switch Independent Partitioning?” on page 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q3: How many partitions are created per port?” on page 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q4: What are the supported operating systems?” on page 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Table 1: “Protocols Available in Operating Systems vs. SF and Switch Independent Partitioning Mode,” on page 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q5: How many Ethernet protocol device partitions are allowed?” on page 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q6: How many iSCSI offload protocol HBA device partitions are allowed?” on page 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q7: How many FCoE offload protocol HBA device partitions are allowed?” on page 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q17: How are PXE/iSCSI/FCoE remote boots affected by Switch Independent Partitioning mode?” on page 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q16: What are the similarities and differences between SF and Switch Independent Partitioning modes for advanced features such as stateless offload, teaming, VLAN, etc.?” on page 18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Table 2: “SF Versus Switch Independent Partitioning Mode,” on page 18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q21: How are the Switch Independent Partitioning settings modified in the operating system?” on page 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q26: Does Switch Independent Partitioning support Jumbo frames?” on page 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q27: How can I view the current Switch Independent Partitioning settings in the operating system?” on page 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q29: Is Switch Independent Partitioning mode supported at 1GbE negotiated link speeds on the 57712-k/578xxS?” on page 23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q30: What Dell M1000e Blade Chassis 10GbE IO Modules support the 57712-k/578xxS devices?” on page 23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q31: How do I enable DCB to work with Switch Independent Partitioning mode?” on page 23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Figure 3: “DCB Mode TX Traffic Flow with Lossless FCoE, Lossy iSCSI, and Lossy Ethernet,” on page 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Added:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Figure 4: “DCB Mode TX Traffic Flow with Lossless iSCSI and Lossy Ethernet,” on page 16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Figure 5: “DCB Mode TX Traffic Flow with Lossless FCoE and Lossless iSCSI and Lossy Ethernet,” on page 17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- “Q32: Does the 57712/578xx support Windows RSS BaseProcessorNumber and MaxProcessorNumber in Windows Server 2008 R2 and 2012?” on page 25</td>
</tr>
<tr>
<td>Revision</td>
<td>Date</td>
<td>Change Description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Removed:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Figures one through seven.</td>
</tr>
<tr>
<td>2CS57712-K-FAQ102-R</td>
<td>05/02/12</td>
<td><strong>Updated:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• T7.2 software release by adding the 578xxS device family and VMWare ESXi 5.0 information.</td>
</tr>
<tr>
<td>2CS57712-K-FAQ101-R</td>
<td>12/14/11</td>
<td><strong>Updated:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Purpose” on page 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Audience” on page 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• T6.4 software release by adding VMWare ESX 4.1, FCoE, DCB, and BACS4 information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q6: How many iSCSI offload protocol HBA device partitions are allowed?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q9: What does a partition’s Relative Bandwidth Weight value mean?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q10: What is a partition’s Maximum Bandwidth value range?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Table 2: “SF Versus NPAR Mode,” on page 14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q22: How are Windows BASP Teaming, Linux/VMWare Bonding affected by NPAR mode?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q24: What criteria is used to generate IEEE 802.3x link-level flow control pause frames?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Added:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q1: What protocols does the 57712-k support?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q2: What is Broadcom’s NIC Partitioning (NPAR)?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Table 1: “Protocols Available in Operation Systems Versus SF and NPAR Mode,” on page 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q12: How is the receive direction traffic flow rate affected by different NPAR settings?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q13: Does the 57712-k support DCB Enhanced iSCSI (iSCSI TLV over DCB)?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q29: Is NPAR mode supported at 1GbE negotiated link speeds on the 57712-k?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Q30: What Dell M1000e Blade Chassis 10GbE IO Modules support the 57712-k NDC?”</td>
</tr>
<tr>
<td>2CS57712-K-FAQ100-R</td>
<td>04/20/11</td>
<td>Initial release</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Non-DCB Mode Zero Sum Relative Bandwidth Weight TX Traffic Flow ............................................13
Figure 2: Non-DCB Mode 100% Sum Relative Bandwidth Weight TX Traffic Flow ..............................................14
Figure 3: DCB Mode TX Traffic Flow with Lossless FCoE, Lossy iSCSI, and Lossy Ethernet ..............................15
Figure 4: DCB Mode TX Traffic Flow with Lossless iSCSI and Lossy Ethernet ...................................................16
Figure 5: DCB Mode TX Traffic Flow with Lossless FCoE and Lossless iSCSI and Lossy Ethernet .....................17
Figure 6: USC DCB Settings ...............................................................................................................................24
Figure 7: BACS4 DCB Settings ..........................................................................................................................24
Figure 8: CCM CDB Settings .............................................................................................................................25
List of Tables

Table 1: Protocols Available in Operating Systems vs. SF and Switch Independent Partitioning Mode ..........10
Table 2: SF Versus Switch Independent Partitioning Mode ........................................................................ .....18
About This Document

Purpose

This document provides a list of frequently asked questions (FAQs) pertaining to the Switch Independent Partitioning (also referred to as NPAR in other documentation) on Servers installed with the Broadcom® 57712, 57800S, 57810S and 57840S Converged Network Interface Cards (cNICs), LAN On Motherboard (LOMs), blade mezzanine, and rack riser cards.

Audience

This document is written for the network administrator who wishes to partition the Broadcom network controller on a Microsoft® Windows Server, VMWare® ESX/ESXi, Oracle® Solaris®, Citrix® XenServer™, SUSE Linux Enterprise Server (SLES) and Red Hat® Enterprise Linux® (RHEL) system. There may be up to eight functions (four per port) over Ethernet, four functions (two per port) over iSCSI HBA, 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/578xxS controllers support?
- 10-Gigabit Ethernet (10GbE)
  - 10GBASE-T
  - SFP+
  - KR
- TCP Chimney Offload (TOE)
- Windows Server 2012 Receive Segment Coalescing (RSC)
- Data Center Bridging (DCB) (See Table 1 on page 10)
  - 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
  - 57712-k: Over 1,600,000 IOPS per dual-ported device
  - 57800S/57810S: Over 2,500,000 IOPS per dual-ported device
  - 57840S: Over 2,640,000 IOPS per quad-ported device
- Internet Small Computer System Interface (iSCSI) Hardware Offload
  - iSCSI Type-Length Value (TLV) for DCB
  - 57712-k: Over 1,000,000 IOPS per dual-ported device
  - 57800S/57810S: Over 1,500,000 IOPS per dual-ported device
  - 57840S: Over 1,670,000 IOPS per quad-ported device
- Single-Root I/O Virtualization (SR-IOV) on Windows Server 2012
  - 64 SR-IOV Virtual Functions (VFs) per Single Function (SF) port
  - Total of 64 SR-IOV VFs per Switch Independent Partitioned port
- Switch Independent Partitioning
- Preboot eXecution Environment (PXE) Boot
- FCoE Boot from Storage Area Networking (SAN) on Windows, Linux and VMWare ESXi 5.1
- iSCSI Boot on Windows, Linux and VMWare
- Virtual Extensible LAN (VXLAN) on VMWare ESXi 5.1
- RX/TX multiqueue
  - Windows Virtual Machine Queues (VMQs)
  - VMWare NetQueues
- Receive-Side Scaling (RSS)/Transmit-Side Scaling (TSS)
- Check Sum Offload (CO)
- TCP Segmentation Offload (TSO)
- Giant Send Offload (GSO)
- Large Send Offload (LSO)/Large Receive Offload (LRO)
Q2: What is Broadcom’s Switch Independent Partitioning?

- Broadcom Switch Independent Partitioning 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 Switch Independent Partitioning simultaneously supports up to:
  - eight (57712-k/57810S/57840S) or four (57800S) virtual Ethernet adapters, and
  - four (57712-k/57800S/57810S) or eight (57840) virtual Host Bus Adapters (HBAs)
  - all with user-configurable relative weight and maximum bandwidth allocation for traffic shaping and Quality of Service (QoS) control.

Q3: How many partitions are created per port?

- A maximum of four partitions are created per port on the 57712-k/57810S based adapters. There are two ports on the device, so there are a total of eight partitions made available in Switch Independent Partitioning mode.
- A maximum of two partitions are created per 10G port on the 57800S based adapters. There are two 10G ports on the 57800S device, so there are a total of four partitions made available in Switch Independent Partitioning mode.
- A maximum of two partitions are created per 10G port on the 57840S based adapters. There are four 10G ports on the 57840S device, so there are a total of eight partitions made available in Switch Independent Partitioning mode.

Q4: What are the supported operating systems?

- The Switch Independent Partitioning mode-supported OSs are Microsoft Windows Server 2008 (and R2), Microsoft Windows Server HyperV R2, Windows Server 2012, Windows Server 2012 HyperV, RHEL (v5.5 or later and v6.1 or later), Oracle Solaris® 10u9/10u10 or later, SUSE Linux Enterprise Server 10 (SLES10)/11 (SLES11) SP1 or later, Citrix XenServer v6.0 or later, VMware ESXi 5.x or later, and VMWare ESX/ESXi 4.1 as shown in Table 1 on page 10. The fully featured drivers might not be included in box.
Q5: How many Ethernet protocol device partitions are allowed?

- Four Ethernet protocols can be enabled per port for a maximum of eight on the dual-ported 57712-k/57810S devices. On the 57800S, two Ethernet protocol devices can be enabled per 10G port for a maximum of four on the two 10G ports. The two 57800 1GbE ports stay in Single Function mode. On the 57840 4x10GbE device, two Ethernet protocol devices can be enabled per 10G port for a maximum of eight on the four 10G ports.

- For Windows® OSs, the Ethernet protocol can be enabled with or without any other offload protocols being enabled on the same partition.

- For Linux®, Solaris, Citrix XenServer, and VMWare ESX OSs, the Ethernet protocol for all partitions is always enabled.

- Windows TOE enabling requires the selected partition to have the Ethernet protocol enabled.
Q6: How many iSCSI offload protocol HBA device partitions are allowed?

- Two iSCSI offload protocol HBAs can be enabled per port on any two of the partitions for a total maximum of four on the 57712-k/57800S/57810S devices and a total maximum of eight on the 57840S device.
- In Windows OSs, the iSCSI offload HBAs can be enabled with or without the Ethernet protocol being enabled on the same partition.
- In Linux and VMWare OSs, the iSCSI offload protocol is enabled with the Ethernet protocol being enabled on the same partition.

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

- One FCoE offload protocol HBA can be enabled per port on any one of the partitions for a total maximum of two on the 57712-k/57800S/57810S devices and a total maximum of four on the 57840 device.
- FCoE Offload protocol is not supported over the 10GBASE-T interfaces.
- In Windows OSs, the FCoE offload protocol HBA can be enabled with or without the Ethernet protocol being enabled on the same partition.
- In Linux, Solaris, and VMWare ESX OSs, the FCoE offload protocol HBA is enabled with the Ethernet protocol being enabled on the same partition.
- Additionally, one iSCSI offload protocol HBA can be enabled per port on any of the remaining partitions.
- A maximum of four HBA offload protocols can be enabled on the 57712-k/578xxS devices.
- The FCoE protocol is not supported on the VMWare ESX/ESXi 4.1, RHEL 5.x, SLES10, and Citrix XenServer OSs.

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

- For Switch Independent Partitioning 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, Solaris, Citrix XenServer, and VMWare ESX 4.1, the Ethernet (i.e., L2) protocol should always be enabled.
- For Switch Independent Partitioning mode of operation with iSCSI hardware offload HBA initiator mode on Windows/Linux/VMWare, enable the iSCSI offload protocol for the specific port’s partition.
- For Switch Independent Partitioning mode of operation with FCoE hardware offload HBA initiator mode on Windows/Linux/Solaris/VMWare, enable the FCoE offload protocol for the specific port’s partition.
- For Single Function (SF) mode of operation, all three protocols (Ethernet, iSCSI, and FCoE offload) are available by default on both 10G ports for Windows, VMWare ESXi 5.0, and Linux OSs. The VMWare ESX/ESXi 4.1 OSs support both Ethernet and iSCSI offload protocols in both SF and Switch Independent Partitioning mode. The Solaris OS support both Ethernet and FCoE in both SF and Switch Independent Partitioning modes. The Citrix XenServer OS supports Ethernet mode in both SF and Switch Independent Partitioning modes.

Note: For Windows, the offload protocol resource (in BACS) might need to be enabled for the specific offload device to be detected/enumerated when the device is first installed. Refer to the Switch Independent Partitioning Setup Guide for details.
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 recommended Relative Bandwidth Weight value range is either all 0% OR no lower than 10% on a single partition.

- 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 versus 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 below that amount. If an individual partition's sent traffic flow cannot 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% is not exactly the same as setting all of the 57712-k/57810S partition values to 25%. All 0% 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. The 57800S/57840's two partitions (per physical port) would be set to 50% each for the same effect.

- If the sum of the four partitions on the same port is not 0%, and one or more of the partitions have 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 10GbE link and in 10 Mbps increments for a 1GbE link when the 57712-k/578xxS 10G interface is connected to a 1GbE link). The sum total of the partitions’ Maximum Bandwidth values on the same port does not have to equal 100. The allowable sum-total value range on the 57712-k/57810S is from 4 (undersubscribed – all four partitions set to 1) to 400 (oversubscribed – all four partitions set to 100). The 57800S/57840 value range can be 2 (both partitions set to 1) to 200 (both partitions set to 100).

- If all partitions were set to undersubscribe the available bandwidth, then all unassigned send/transmit direction bandwidth would be unavailable for use by any of the 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 not be used.

- The reset default Maximum Bandwidth value is 25% (57712-k/57810S) and 50% (57800S/57840S) for each partition.
Q11: How is the Switch Independent Partitioning 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 1. 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 partitions’ 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 1: Non-DCB Mode Zero Sum Relative Bandwidth Weight TX Traffic Flow](image-url)
In non-DCB mode, where the sum of the partition’s Relative Bandwidth Weights equals 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 2. This means a specific partition’s Relative Bandwidth Weight value will restrict the traffic flows sharing that partition’s bandwidth allocation, as 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. The recommended Relative Bandwidth Weight value range in this case is no lower than 10% on a single partition.

Figure 2: 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 has both an FCoE-enabled and Ethernet-enabled protocols transmitting to two different Priority Groups. Each partition’s Maximum Bandwidth setting and the overall connection link speed as shown in Figure 3. The FCoE traffic type is in PG1, and all of the other traffic types (iSCSI Offload and Ethernet) are in the PG0. Each Priority Group has its own ETS value. Similar 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 and are in the same Priority Group. Refer to the User Guide's examples for details.

Figure 3: DCB Mode TX Traffic Flow with Lossless FCoE, Lossy iSCSI, and Lossy Ethernet
This second example of DCB mode shows the 1st and 4th Partitions having both an iSCSI-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. In this example, the lossless iSCSI Offload over DCB traffic (also known as iSCSI-TLV) type is in PG1, and all of the other traffic types (specifically Ethernet) are in the default PG0. Each Priority Group has its own ETS value. The two partition’s lossless iSCSI Offload go through its assigned Priority Group while the Lossy Ethernet traffic goes through the other.

Figure 4: DCB Mode TX Traffic Flow with Lossless iSCSI and Lossy Ethernet
This third example of DCB mode shows the 1st Partition again having both an FCoE-enabled and Ethernet-enabled protocol transmitting to two different Priority Groups while the 4th Partition has the iSCSI-enabled and Ethernet-enabled protocols also transmitting to two different Priority Groups and each partition’s Maximum Bandwidth setting and the overall connections link speed as shown in Figure 5. In this example, the lossless FCoE Offload over DCB traffic type (from Partition 1) is in PG1, the lossless iSCSI Offload over DCB traffic (also known as iSCSI-TLV) type (from Partition 4) is in PG2, and all of the other traffic types (specifically the Ethernet from Partitions 1 through 4) are in the default PG0. Each Priority Group has its own ETS value. The two different lossless Offloads (FCoE and iSCSI) go through their assigned Priority Groups while the Lossy Ethernet traffic goes through another.

Figure 5: DCB Mode TX Traffic Flow with Lossless FCoE and Lossless iSCSI and Lossy Ethernet

Note: VMWare ESX/ESXi 4.1, Citrix XenServer, RHEL 5.x, and SLES 10 do not support DCB mode of operation on the 57712-k/578xxS.

Q12: How is the receive direction traffic flow rate affected by different Switch Independent Partitioning settings?

- A partition’s Relative Bandwidth Weight and Maximum Bandwidth values affect only its send/transmit/outgoing/egress traffic.
- The receive direction traffic of the four partitions of a single port is not weighted or restricted by the 57712-k/578xxS device Switch Independent Partitioning settings. The receive direction acts in a similar way when in SF mode: the incoming traffic on the port is controlled by the switch port to which it is connected.
Q13: Does the 57712-k/578xxS support DCB Enhanced iSCSI (iSCSI Offload TLV over DCB)?
   – Yes, the 57712-k/578xxS can assign iSCSI offload traffic to be Lossless and to its own dedicated priority group independent of the other TCP/IP networking or FCoE Offload traffic, in both SF and Switch Independent Partitioning 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 Switch Independent Partitioning mode) or port (in SF mode).
   – Alternately, TOE can be enabled or disabled for the entire system (all ports) using BACS4 (Host - Configurations - Property - System Management - Chimney Offload State) or 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/adapter link speed. This value represents the specific partition’s send/transmit/outgoing maximum rate and not the receive/incoming link speed rate, which is always the port’s current link speed.

Q16: What are the similarities and differences between SF and Switch Independent Partitioning modes for advanced features such as stateless offload, teaming, VLAN, etc.?
   – Refer to Table 2.

Table 2: SF Versus Switch Independent Partitioning Mode

<table>
<thead>
<tr>
<th>Feature</th>
<th>SF</th>
<th>Switch Independent Partitioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethernet Adapters</td>
<td>1 per port</td>
<td>Up to 4 per port; see FAQ entry</td>
</tr>
<tr>
<td>See “Q5: How many Ethernet protocol device partitions are allowed?” on page 10.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iSCSI Hardware Offload HBAs</td>
<td>1 per port</td>
<td>Up to 2 per port; see FAQ entry</td>
</tr>
<tr>
<td>See “Q6: How many iSCSI offload protocol HBA device partitions are allowed?” on page 11.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCoE Hardware Offload HBAs</td>
<td>1 per port</td>
<td>Up to 1 per port; see FAQ entry</td>
</tr>
<tr>
<td>See “Q7: How many FCoE offload protocol HBA device partitions are allowed?” on page 11.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows TCP Chimney Offload (TOE)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TCP Segmentation Offload (TSO)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Large Send Offload (LSO) and Large Receive Offload (LRO)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Giant Send Offload (GSO)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IPv4/IPv6 TCP-UDP-IP Checksum Offload (CO)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IP Address Multihoming</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
### Table 2: SF Versus Switch Independent Partitioning Mode (Cont.)

<table>
<thead>
<tr>
<th>Feature</th>
<th>SF</th>
<th>Switch Independent Partitioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>PXE/ISCSI/FCoE Boot</td>
<td>Yes</td>
<td>Yes – first partition of each port; see FAQ entry See “Q17: How are PXE/ISCSI/FCoE remote boots affected by Switch Independent Partitioning mode?” on page 20.</td>
</tr>
<tr>
<td>Wake-on-LAN (WOL)</td>
<td>Yes</td>
<td>Yes – see FAQ entry See “Q18: How is Wake On LAN (WOL) Out Of Band (OOB) affected by Switch Independent Partitioning mode?” on page 20.</td>
</tr>
<tr>
<td>PAUSE Flow Control</td>
<td>Yes</td>
<td>Yes – see FAQ entry See “Q23: Is IEEE 802.3x link-level flow control supported in SF and Switch Independent Partitioning mode?” on page 21.</td>
</tr>
<tr>
<td>Jumbo Frames (up to 9600 Bytes)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>iSCSI Offloads</td>
<td>128 per port</td>
<td>128 per partition – for up to two iSCSI Offload Protocol enabled partitions</td>
</tr>
<tr>
<td>RX/TX L2 Ethernet Queues (RSS/NetQueue/VMQ)</td>
<td>16 per port</td>
<td>4 per partition</td>
</tr>
<tr>
<td>MSI</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>MSI-X</td>
<td>1024</td>
<td>1024</td>
</tr>
<tr>
<td>OS specific Multipath I/O (MPIO)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VLAN (insertion/striping/filtering)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Q17: How are PXE/iSCSI/FCoE remote boots affected by Switch Independent Partitioning 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, VMWare ESXi 5.1, and Linux all support PXE/iSCSI/FCoE remote boots.
- Solaris 10u10 supports FCoE remote boots.
- VMWare ESXi 4.1/5.0 support PXE/iSCSI remote boots.
- VMWare ESX/ESXi 4.1/5.0 do not support FCoE remote boots.
- VMWare ESX 4.1 (the non "i" version) supports PXE but not iSCSI remote boots.
- In SF or Switch Independent Partitioning modes, PXE booting of Windows Hyper-V VMs over a Smart Load Balanced and Failover (SLB) Team is supported. The HyperV Mode must be enabled when creating the SLB team.

Q18: How is Wake On LAN (WOL) Out Of Band (OOB) affected by Switch Independent Partitioning 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 Switch Independent Partitioning 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 supported for all partitions on both ports at 1 Gbps link speed only.

Q20: How is Switch Independent Partitioning enabled through USC?
- Reboot to Dell’s Unified Server Configurator (USC) using the Dell® UEFI bootup option and follow the setup guide instructions. Detailed information on USC and the UEFI boot option can be found on the Dell support site.

Q21: How are the Switch Independent Partitioning settings modified in the operating system?
- In Windows and Linux OSs, you can change and enable/disable the partitioning mode or settings in the operating system using BACS4. With BACS4, changes to Maximum Bandwidth and Relative Bandwidth Weight settings do not require a reboot, but traffic may be interrupted on the affected ports when the changes are applied.
- In VMWare/Solaris/XenServer, you can only change and enable/disable Switch Independent Partitioning settings in USC/CCM. USC/CCM can also be used to modify the Switch Independent Partitioning settings in Linux/Windows.
- In newer versions of EthTool, the "ethtool -s ethX speed NNNNN" command supports setting the 57712-k/578xxS partition’s Maximum Bandwidth value without a system reboot, where ethX is the partition and NNNNN is the speed which is in 1 Mbps increments (i.e., it is NOT a percentage):
  - For a 10GbE link connection you could set it from 100 (which is equivalent to 100 Mbps or 1% of the 10GbE link speed) to 10000 (which is equivalent to 10 Gbps or 100% of the 10GbE link speed) – the last two zero positions are not used.
  - For a 1GbE link connection you could set it from 10 (which is equivalent to 10 Mbps or 1% of the 1GbE link speed) to 1000 (which is equivalent to 1 Gbps or 100% of the 1GbE link speed) – the last zero position is not used.
- In some cases, a system reboot is required before any changes take effect.
**Q22: How are Windows Broadcom Advanced Server Program (BASP) Teaming, Windows Server 2012 in-OS NIC Teaming, and Linux/VMWare Bonding affected by Switch Independent Partitioning mode?**

- A team cannot contain more than one partition from the same port.
- Switch Independent Partitioning adapters can be teamed with other non-Switch Independent Partitioning adapters, such as the 5709 and 57711 or with other vendor Ethernet adapters.
- In Windows OS, Smart BASP Load Balanced and Failover (SLB with or without Auto-Fallback) switch independent teaming mode is available for teams using Switch Independent Partitioning adapters.
- Switch dependent teaming (IEEE 802.3ad LACP and Generic/Static Link Aggregation (Trunking)) cannot use a Switch Independent Partitioning 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.
- FCoE and iSCSI traffic connections use Multipath I/O (MPIO) to provide Load Balancing and Failover protection at the OS’s storage level when there is more than one connection to the same storage LUN.
- FCoE and iSCSI Offload should not be enabled on Switch Dependent teamed ports; except for FCoE Offload on LACP teams on Virtual Port Channel (vPC) configured Cisco switches.
- FCoE and iSCSI Offload can be enabled on SLB teamed ports but is not regulated by the SLB teaming software.

**Q23: Is IEEE 802.3x link-level flow control supported in SF and Switch Independent Partitioning mode?**

- Yes. Link-level flow control is available in both SF and Switch Independent Partitioning 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 Switch Independent Partitioning mode, link-level flow control is configured on a per-port basis through USC/CCM 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 Switch Independent Partitioning 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 NICs, networking (L2 Ethernet) host buffers and storage (iSCSI/FCoE hardware offloaded) host buffers are maintained separately. If networking host buffers are too few and are allowed to generate link-level pause frames, 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 Switch Independent Partitioning 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, whereas 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 Switch Independent Partitioning modes?

Yes. Priority Flow Control (PFC) is negotiated as part of DCB which is controlled on a per-port basis. The Data Center Bridging eXchange (DCBx) negotiated lossless 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/578xxS DCB supports both lossless FCoE Offload and lossless iSCSI Offload TLV modes of operation.

If PFC is not enabled during the DCBx exchange, IEEE 802.3x link-level flow controls settings will apply.

Q26: Does Switch Independent Partitioning support Jumbo frames?

Yes. Jumbo frames up to 9600 bytes are supported in both Windows and Linux OSs on all partitions, for both Ethernet and iSCSI offload protocol modes in both SF and Switch Independent Partitioning modes.

In SF and Switch Independent Partitioning mode, VMWare ESX/ESXi 4.1 supports standard frames (MTU = 1500 bytes) for iSCSI hardware offload and Jumbo frames for L2 Ethernet (which includes iSCSI non-offload initiator mode).

In SF and Switch Independent Partitioning mode, VMWare ESXi 5.x supports Jumbo frames for both iSCSI hardware offload and L2 Ethernet (which includes iSCSI non-offload initiator mode).

Jumbo Frames are supported on L2 Ethernet in Oracle Solaris and Citrix XenServer in both SF and Switch Independent Partitioning modes.

For Windows, Solaris, VMWare ESXi 5.x and Linux OSs, FCoE protocol’s MTU is always fixed at 2500 bytes.

Q27: How can I view the current Switch Independent Partitioning settings in the operating system?

In Windows OS, you can see what protocols are enabled for a partition in both Windows Device Manager and BACS4. You can 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 OS, you can see what protocols are enabled for a partition in BACS4. You can see the partition’s Relative Bandwidth Weight and Maximum Bandwidth value for all protocol-enabled partitions in BACS4.

In Solaris, XenServer and VMWare OSs, you can see the applicable Ethernet, iSCSI, and FCoE protocol-enabled partitions that 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 detect when the device is in Switch Independent Partitioning mode?

On the 57712-k/578xxS, a switch will see additional MAC addresses (up to four L2 Ethernet MAC addresses), plus two additional iSCSI offload MAC addresses (depending upon the Switch Independent Partitioning configuration), which is slightly more than the 57712-k/578xxS have 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).

A switch port should be set to TRUNK mode if VLANs and/or FCoE protocol mode is enabled on any SF port or Switch Independent Partition.

Q29: Is Switch Independent Partitioning mode supported at 1GbE negotiated link speeds on the 57712-k/578xxS?

- Yes. The attached device and interface (blade KR or SFP+ or 10GBase-T) would have to support 1GbE link speeds.
- For the M1000e blade chassis, the supported 1GbE IO Modules include: 1GbE 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 lowered link speed 1GbE connections.
- Windows TOE and iSCSI Hardware Offload are supported on lowered link speed 1GbE connections on the applicable OSs.
- Switch Independent Partitioning, DCB, FCoE Offload, and iSCSI Hardware Offload are not supported on the two permanent 1GbE ports of the 57800S.

Q30: What Dell M1000e Blade Chassis 10GbE IO Modules support the 57712-k/578xxS devices?

- The currently supported 10GbE IO Modules include: 10GbE Pass Through-k, M8024-k, Force10 I/O Aggregator, Force10 MXL, and M8428-k.
- For FCoE and DCB Lossless operation (to include iSCSI Offload-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 10GbE Pass-Through-k IO module must be externally connected to a switch that supports DCBX.

Q31: How do I enable DCB to work with Switch Independent Partitioning mode?

- DCB is enabled in Dell’s USC (all OS’s) or BACS4 (Windows and Linux OSs) or Broadcom’s Comprehensive Configuration Management (CCM) preboot utility (all OSs).
- A reboot is required after enabling or disabling DCB but DCBX changes sent down from the connected DCB switch are implemented without reboots. Changes in Priority Groups, ETS minimum BW and Losslessness (PFC) are automatically made on the fly.
- DCB works on both Switch Independent Partitioning and normal Single Function mode ports.
- DCB is configurable on a per port granularity. This means all partitions on a port will have DCB enabled or disabled.
- DCB could be enabled on one port of a device and disabled on the other port of the same device.
- DCB requires a CEE-DCBX capable switch connected to the enabled port.
- DCB is required for FCoE and iSCSI Offload TLV.
Figure 6: USC DCB Settings

![USC DCB Settings](image)

Figure 7: BACS4 DCB Settings

![BACS4 DCB Settings](image)
Q32: Does the 57712/578xx support Windows RSS BaseProcessorNumber and MaxProcessorNumber in Windows Server 2008 R2 and 2012?

- Yes, all RSS capable NetXtreme® (5719/5720) and NetXtremeII® (5709/16/5771x/578xx) devices support changes made to the RSS Indirection Table's BaseProcessorNumber and MaxProcessorNumber values using PowerShell version 3.0.
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.