Lenovo

Lenovo FCoE Converged Networking Architecture Best Practices

Introduces converged networking and Fibre Channel over Ethernet

Provides FCoE sample configurations and topologies

Explains redundancy and capacity considerations

Recommends preferred techniques for implementing converged networks

Scott Lorditch

LENOVO PRESS

Abstract

This paper presents recommended topologies for use with converged networking, focusing on the use of Fibre Channel over Ethernet (FCoE) with Lenovo Networking switches. It includes sample configurations for several Lenovo Networking products which can be customized for use in a production environment. There are samples using the most current Lenovo Networking firmware (CNOS) as well as the older firmware (ENOS).

This paper is intended for use by system and networking engineers working on those technologies, and who have a working understanding of Ethernet networking.

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

Do you have the latest version? We update our papers from time to time, so check whether you have the latest version of this document by clicking the **Check for Updates** button on the front page of the PDF. Pressing this button will take you to a web page that will tell you if you are reading the latest version of the document and give you a link to the latest if needed. While you're there, you can also sign up to get notified via email whenever we make an update.

Contents

Introduction

This document shows our preferred topology for use with converged networking, including mostly FCoE because of its complexity.

This document presents key considerations for using FCoE, including recommended topologies for FCoE deployment and examples of switch configurations for FCoE. We also provide a brief summary of other Ethernet-based SAN technologies. Some of the configurations will be based upon embedded switches used in the Lenovo Flex System chassis but all of them can be readily adapted to Lenovo top-of-rack switches.

The paper does not include discussion of Hyperconverged architectures, where storage and compute resources reside in the same set of physical devices. For discussion of the Nutanix hyperconverged product implemented on Lenovo appliances and networking, see the following two documents:

- *Networking Guide for Lenovo Converged HX Series Nutanix Appliances*

  http://lenovopress.com/lp0546
- *Networking Guide for Lenovo Converged HX Series Nutanix Appliances (CNOS Switch Firmware)*

  http://lenovopress.com/lp0595

Much of the content of this document is updated from, and refers to, content from Chapter 2 and Chapter 8 of the Lenovo Press publication *Lenovo Networking Best Practices for Configuration and Installation*, SG24-8245, which is available from:

http://lenovopress.com/sg248245

FCoE considerations and prerequisites

This section describes preferred practices when Fibre Channel over Ethernet (FCoE) is used in the network.

General considerations

Switch firmware must be flashed with FCoE actively configured to ensure that the Fibre Channel ASIC onboard the RackSwitch™ G8264CS and Flex System CN4093 switches is properly updated.

Ensure that the server that is sending FCoE data has the current firmware and drivers installed on the Converged Network Adapter (CNA). Adapter vendors made significant changes in FCoE implementation over the past few years. Running older FCoE drivers and firmware is a significant cause of unexpected behavior and poor performance.

FCoE and FCFs

The FCoE protocol allows servers to use a single network connection to enable access to TCP/IP and storage resources. Access to storage is accomplished by encapsulating the Fibre Channel protocol inside an Ethernet frame.

When hosts send FCoE traffic to a switch, the primary task for the switch is to forward the FCoE packets to the nearest FCF. When the packet arrives at an FCF, the Fibre Channel protocol frame is de-encapsulated and sent to a Fibre Channel Switch.

Switches, such as the Flex System EN4093 and the RackSwitch G8124 and G8264 provide FCoE transit functionality only. The Flex System CN4093 and RackSwitch G8264CS provide FCoE and can function as an FCF; they have ports which can be configured as true Fibre Channel ports. Other vendors provide switches with similar capabilities.

Recent offerings from storage vendors provide FCF functionality inside storage arrays. When such a storage product is used, the traffic is never forwarded over a true Fibre Channel link and no FC ports are involved.

NPV versus full fabric modes

When a switch is functioning as the FCF, the FCF must be configured to operate in one of two modes: NPV (sometimes called “tributary” mode) or Full Fabric.

If storage is connected directly to an G8264CS or CN4093, the switch must be configured for its FCF to operate in Full Fabric mode. Zoning and all other Fibre Channel administrative tasks must be performed on the G8264CS or CN4093.

When the G8264CS or CN4093 switch is connecting to a Fibre Channel Fabric, the FCF should be configured in NPV mode. NPV mode is the preferred configuration because it allows all Fibre Channel administration to be performed on the existing Fibre Channel switches.

Zoning

No issues or special considerations exist relative to zoning and Lenovo networking. Hard or soft zoning can be implemented on the Fibre Channel switch.

When zones are created, the preferred configuration is to have a single initiator and single target per zone. When a device is added or removed from a fabric, the fibre switch sends RSCN events to all devices in the zone. Active hosts in the zone that are not entering or leaving can have latency effects when this situation occurs. Creating a zone of one initiator and one target prevents this situation from occurring.

FCoE topologies

This section will present the preferred topology for use with FCoE and an alternative. There are other variations which can also be used with FCoE which are less preferred and are not discussed here.

Full mesh topology with Virtual Link Aggregation

This topology is preferred (except for when conditions prevent its use, such as those conditions that described later in this section). It provides connectivity with the following qualities:

- High availability, which enables the environment to survive the failure of one of two access switches, which connect directly to the servers, or one of two upstream switches (or the links that connect to them), or both.

- All of the links between an access switch and an upstream switch are active and can carry production traffic. None of the links are blocked by Spanning Tree to prevent network loops.
- The server-facing (INT*x or other*) ports on the embedded switches can be channeled together. This configuration works with the high-availability features. Aggregation-based Active/Active NIC teaming modes are available if the server's ports are channeled.

Do not implement this topology if any of the following conditions are true:

- The upstream switches do not support some types of cross-switch link aggregation, such as vPC, VSS, Virtual Link Aggregation (vLAG), MCLAG, or a stacking feature. In this case, other designs should be used, and can enable most of the availability and usage features for this topology. An example of one such design is included under "Spanning-Tree Topology" below.
- The customer does not have two switches that they plan to use to connect to the configuration. A single upstream switch design is not recommended because it contains a single point of failure and can isolate the servers from the remainder of the network - including the storage arrays - if that single switch fails.

A diagram of an example of this topology is shown in Figure 1. Equivalent designs can be deployed with a pair of embedded switches or top-of-rack switches with servers that are dual-homed.

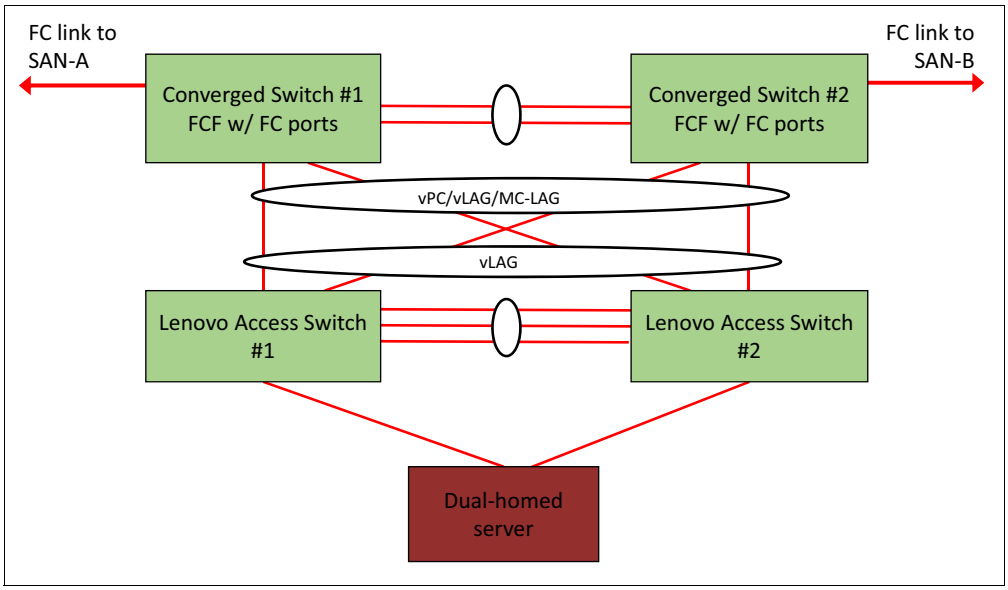

Figure 1 Full mesh network design

FCoE with vLAG Examples

This section presents examples of preferred practices for enabling FCoE within a vLAG environment, including the following:

- Flex Switch CN4093 switch
- Flex Switch EN4093 to G8264CS switches
- Flex Switch EN4093 to Nexus 5548 switches

Flex Switch CN4093 switch

The Flex Switch CN4093 switch provides connectivity to a standard Ethernet network and a Fibre Channel environment as an FCoE Gateway. (FCF)

The top-of-rack equivalent of the CN4093 is the RackSwitch G8264CS; the key difference between them for purposes of this example is the port naming convention: the G8264CS identified ports only by numbers and does not name them with INTxx or EXTxx. Omni ports on the G8264CS, which can be configured to function as 10Gb Ethernet ports or as Fibre Channel ports, are numbered 53-64.

Figure 2 shows SAN A and SAN B for Fibre Channel Multipath isolation. SAN A carries FCoE and Fibre Channel Traffic on the path vHBA-1 → CN4093-1 → SAN Fabric A and SAN B carries FCoE and Fibre Channel Traffic on the path vHBA-2 → CN4093-2 → SAN Fabric B. The vLAG Inter Switch Link (ISL) on the CN4093 Switches carries normal Ethernet traffic only and it is required to prune FCoE VLANs.

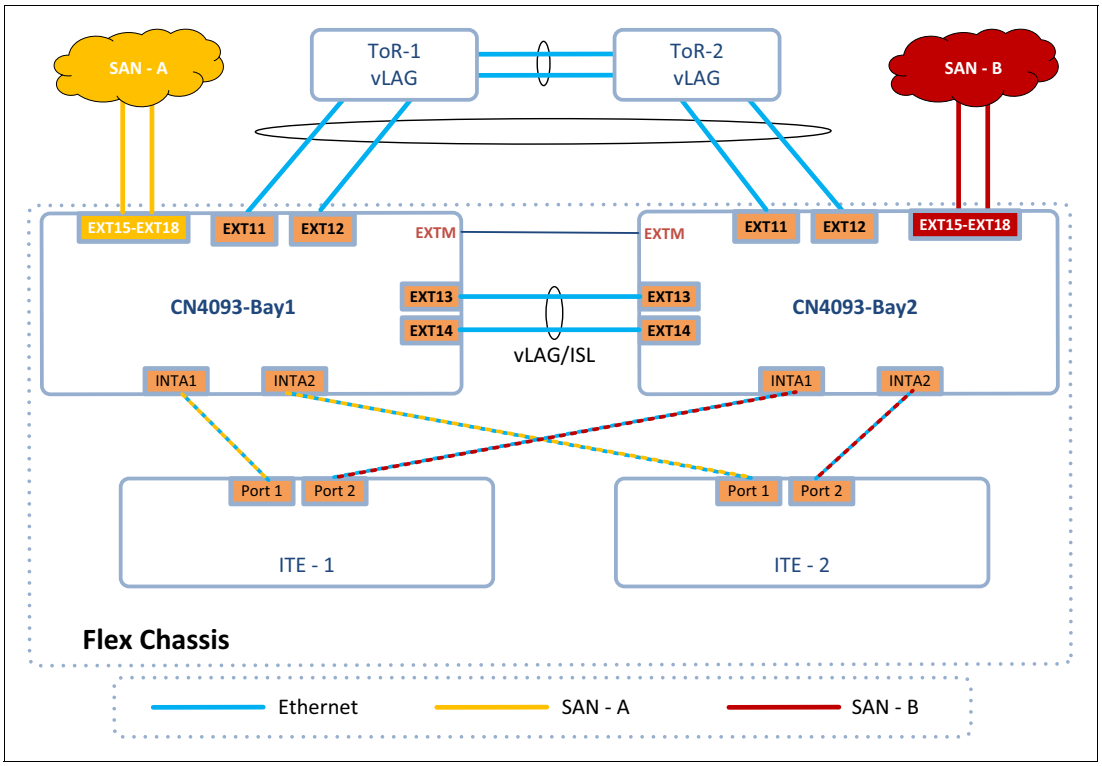

Figure 2 FCoE with vLAG that use CN4093 switches

Consider the following preferred practice guidelines:

- Ensure that vLAG ISL ports contain a non-production VLAN as native VLAN or PVID; for example, VLAN 4090 or VLAN 4094. Disable Spanning Tree for this VLAN to prevent it from participating in Spanning Tree. Do not add this VLAN to any other ports except for the vLAG ISL ports. Doing this will improve recoverability and prevent Spanning Tree issues if one of the CN4093 switches should fail.
- Internal-node-facing ports (for example, INTA ports) on each of the access switches must have VLAN 1 as the native VLAN or PVID VLAN. It is used in the FCoE discovery process.
- Internal-node-facing ports (for example, INTA ports) on each of the access switches must contain the FCoE VLAN. It is commonly VLAN 1001 (SAN A) and VLAN 1002 (SAN B), but this configuration is not mandatory.

- If Spanning Tree is enabled on the CN4093 switch, disable Spanning Tree for the FCoE VLAN. The FCoE discovery VLAN 1 can have Spanning Tree enabled if it is needed.
- vLAG ISL ports must *not* include either of the two configured FCoE VLANs (such as 1001 and 1002) to prevent Fibre Channel fabric merging from occurring.
- Use loop guard across the vLAG ISL for loop detection and prevention.
- If Spanning Tree is enabled, use BPDU guard on the server facing ports.

Example 1 shows an example configuration script for a CN4093 switch with vLAG and FCoE in NPV mode.

Example 1 Example configuration script for a CN4093 switch with vLAG and FCoE in NPV mode

```
hostname "CN4093-CH1-SW1”
system port EXT15-EXT18 type fc
!
cee enable
fcoe fips enable
!
interface port INTA1-INTA14,EXT15-EXT18
   switchport mode trunk
   switchport trunk allowed vlan 1,10,20,30,1001
   exit
!
vlan 1001
   name "FCoE VLAN"
   npv enable
   npv traffic-map external-interface EXT15-EXT18
   exit
!
interface port EXT1,EXT2
   description vLAG-ISL
   switchport mode trunk
   switchport trunk allowed vlan 10,20,30,4090
   switchport trunk native vlan 4090
   spanning-tree loopguard
   lacp key 4344
   lacp mode active
   exit
!
interface port EXT11,EXT12
   description Uplink-To-ToR1
   switchport mode trunk
   switchport trunk allowed vlan 10,20,30,999
   switchport trunk native vlan 999
   lacp key 5354
   lacp mode active
   exit
!
interface port INTA1-INTA14
   bpdu-guard
!
no spanning-tree stp 26 enable
spanning-tree stp 26 vlan 4090
!
no spanning-tree stp 112 enable
```

```
spanning-tree stp 112 vlan 1001
!
vlan 10
   name Data-Network-10
vlan 20
   name Data-Network-20
vlan 30
   name Data-Network-30
!
interface ip 127
   ip address 1.1.1.1
   enable
!
vlag enable
vlag tier-id 10
vlag hlthchk peer-ip 1.1.1.2
vlag isl adminkey 4344
vlag adminkey 5354 enable
```

Important: The **`no spanning-tree stp`** ***`xxx`*** commands can vary regarding the instance number. The **`show run | section vlan`** command displays which STP instance is associated to which VLAN to correctly identify the STP instance ID. In Example 1 on page 7, spanning-tree for the ISL VLAN (4090) and the FCoE VLAN (1001) is disabled.

Flex Switch EN4093 to RackSwitch G8264CS

The EN4093, and top-of-rack switch equivalents such as the G8272 and G8296, support Ethernet traffic and can function as an FCoE transit switch. The G8264CS switch can support Ethernet and Fibre Channel as an FCoE Gateway device (FCF).

As in the previous example, the EN4093 uses INTxx and EXTxx port naming while the top-of-rack switches identify ports by number.

Figure 3 shows SAN A and a SAN B for Fibre Channel Multipath isolation. SAN A carries FCoE and Fibre Channel Traffic on the path vHBA-1 → EN4093-1 → G8264CS-1 → SAN A and SAN B carries FCoE and Fibre Channel Traffic on the path vHBA-2 → EN4093-2 → G8264CS-2 → SAN B. The vLAG Inter Switch Link (ISL) on the EN4093 and the G8264CS pairs of switches carries normal Ethernet traffic only and is required to prune FCoE VLANs.

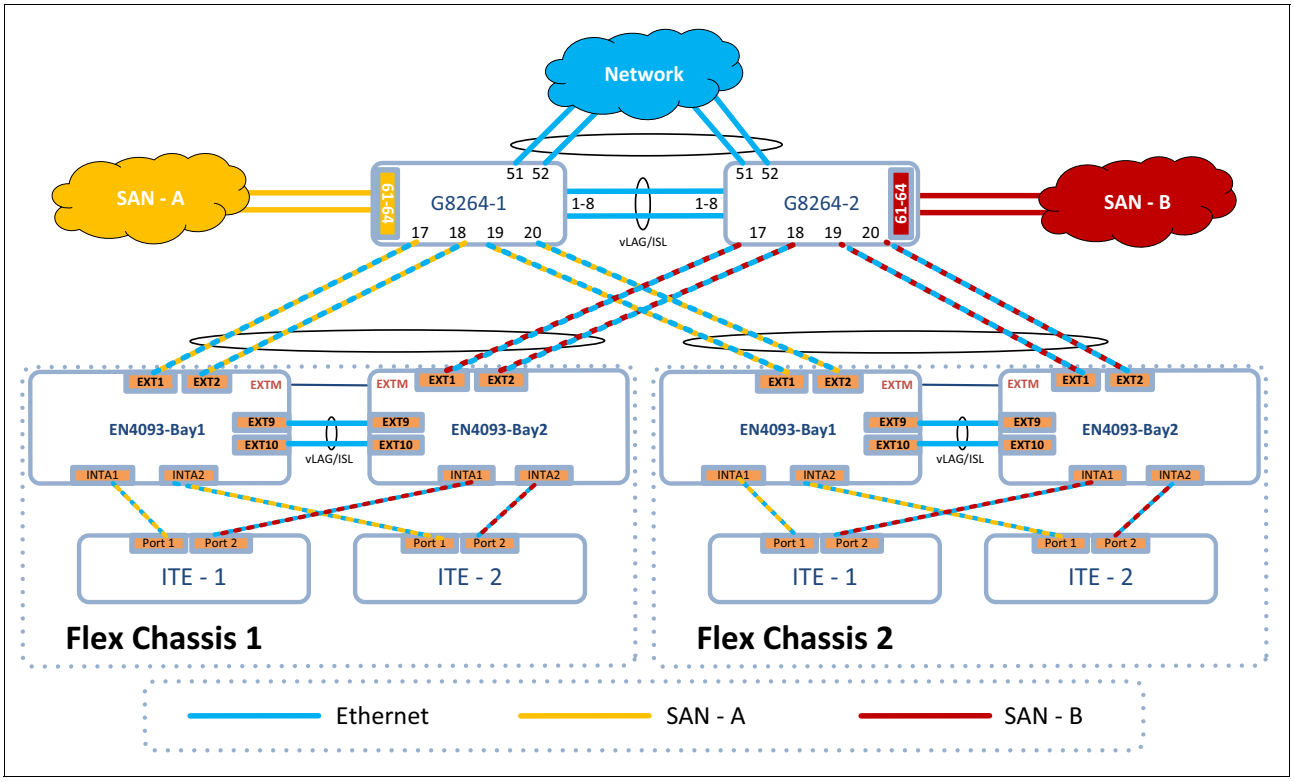

Figure 3 FCoE with vLAG that uses EN4093 and G8264 switches

Consider the following preferred practice guidelines:

- Ensure that vLAG ISL ports contain a non-production VLAN as the native VLAN and PVID; for example, VLAN 4090 or VLAN 4094. Disable Spanning Tree for this VLAN to prevent it from entering an STP block state or participating in Spanning Tree. Do not add this VLAN to any other ports other than the vLAG ISL ports.
- Internal-node-facing ports (for example, EN4093's INTA ports) on each of the access switches must have VLAN 1 as the Native VLAN and PVID. It is used as the FCoE discovery VLAN.
- Uplink ports, that face the network, must contain VLAN 1, but can be trunked and tagged to carry the FCoE discovery information to the G8264CS switch for processing.
- Internal Node facing ports (for example, INTA ports) on each of the EN4093 switches must contain the FCoE VLAN. It is usually VLAN 1001 (SAN A) and VLAN 1002 (SAN B).
- If Spanning Tree is enabled on the EN4093 switch, G8264CS switch, or both, disable Spanning Tree for the FCoE VLAN. Optionally, if it is required to be enabled on the vLAG ISL ports, VLAN 1 can have Spanning Tree enabled if it is needed.
- vLAG ISL ports must not include either of the two configured FCoE VLANs to prevent Fibre Channel merging from occurring.
- Use loop guard across the vLAG ISL for loop detection and prevention.
- If Spanning Tree is enabled, apply **`bpdu-guard`** on the compute node (server) facing ports.

Example 2 uses the Easy Connect feature to configure an EN4093 switch. A similar configuration can be used on G8272 and G8296 switches running ENOS firmware (versions up to 8.x). CNOS firmware (version 10.x and higher) do not currently support this feature.

Easy Connect is an implementation of 802.1q double tagging, also referred to as Q-in-Q. It allows the switch to ignore the VLANs defined and used on the servers which are also used on the switches upstream from the EN4093's. This enables the configuration of the EN4093's in this example to be simpler than the standard configurations which would otherwise be used. Easy Connect does this by adding an additional VLAN tag - in the example, this is for VLAN 4091 - to incoming traffic and then removing it when the packets are forwarded. VLANs 10, 20, and 30 are not defined in the configuration in Example 2, but they would be defined on the servers, and are defined in the configuration for the G8264CS, shown in Example 5 on page 15.

Example 2 Example configuration of an EN4093 in EC Mode with vLAG and FCoE

```
hostname EN4093-Sw1
spanning-tree mode disable
cee enable
!
interface port ext9,ext10
   description vLAG-ISL
   switchport mode trunk
   switchport trunk allowed vlan 4090,4091
   switchport trunk native vlan 4090
   lacp key 5152
   lacp mode active
   exit
!
vlan 4090
   name Peer-Link
   exit
!
vlan 4091
   name EasyConnect
   exit
!
interface port inta1-inta14,ext1-ext2
   switchport access vlan 4091
   tagpvid-ingress
   exit
!
interface port ext1-ext2
   description Uplink-To-G8264CS-1
   lacp key 4344
   lacp mode active
   exit
!
interface ip 127
   ip address 1.1.1.1
   enable
   exit
!
vlag ena
vlag tier-id 10
vlag hlthchk peer-ip 1.1.1.2
```

```
vlag isl adminkey 5152
vlag adminkey 4344
enable
```

Example 3 shows an EN4093 configuration similar to the above but not using EasyConnect. Note that the production VLANs are included but VLAN 4091 is not.

Example 3 EN4093 configuration for transit upstream to G8264CS - without EasyConnect

```
hostname EN4093-Sw1
spanning-tree mode disable
cee enable
!
interface port ext9,ext10
   description vLAG-ISL
   switchport mode trunk
   switchport trunk allowed vlan 10,20,30,4090
   switchport trunk native vlan 4090
   lacp key 5152
   lacp mode active
   exit
!
vlan 4090
   name Peer-Link
   exit
!
vlan 10
   name Production-Vlan-10
   exit
!
vlan 20
   name Production-Vlan-20
   exit

vlan 30
   name Production-Vlan-30
   exit

interface port inta1-inta14,ext1-ext2
   switchport mode trunk
   switchport trunk allowed vlan 1,10,20,30,1001
   switchport trunk native vlan 1
   exit


interface port ext1-ext2
   description Uplink-to-G8264CS
   lacp key 4344
   lacp mode active
   exit
!
interface ip 127
   ip address 1.1.1.1
   enable
   exit
```

```
!
vlag ena
vlag tier-id 10
vlag hlthchk peer-ip 1.1.1.2
vlag isl adminkey 5152
vlag adminkey 4344 enable
```

Example 4 shows a configuration of a G8264CS switch with vLAG plus FCoE and FC enabled. Note that VLAN 4091 is not defined in this configuration but VLANs 10,20,30, and 1001 are defined.

Example 4 Example configuration of a G8264CS with vLAG plus FCoE and FC

```
hostname G8264CS-Sw1
!
interface port 17-20,61-64
   switchport mode trunk
   switchport trunk allowed vlan 1,1001
   exit
!
cee enable
fcoe fips enable
!
system port 61-64 type fc
!
interface port 17,18
   description To-Ch1-Sw1
   lacp key 1718
   lacp mode active
   exit
!
interface port 19,20
   description To-Ch2-Sw1
   lacp key 1920
   lacp mode active
   exit
!
vlan 1001
   name "FCoE VLAN"
   npv enable
   npv traffic-map external-interface 61-64
   exit
!
interface port 1-8
   description vLAG-ISL
   switchport mode trunk
   switchport trunk allowed vlan 10,20,30,4090
   switchport trunk native vlan 4090
   spanning-tree loopguard
   lacp key 1080
   lacp mode active
   exit
!
interface port 51,52
   description To-Core-Network
   switchport mode trunk
```

```
   switchport trunk allowed vlan 10,20,30,999
   switchport trunk native vlan 999
   lacp key 5152
   lacp mode active
   exit
!
interface port 17-20
   switchport mode trunk
   switchport trunk allowed vlan add 10,20,30
   bpdu-guard
   exit
!
no spanning-tree stp 26 enable
no spanning-tree stp 112 enable
!
vlan 10
   name Data-Network-10
vlan 20
   name Data-Network-20
vlan 30
   name Data-Network-30
!
interface ip 128
   ip address 1.1.1.1
   enable
!
vlag enable
vlag tier-id 1
vlag hlthchk peer-ip 1.1.1.2
vlag isl adminkey 1080
vlag adminkey 5152 enable
vlag adminkey 1718 enable
vlag adminkey 1920 enable
```

Important: The **`no spanning-tree stp 26 enable`** command correlates to VLAN 4090, which is used as the ISL VLAG Native VLAN ID.

The **`no spanning-tree stp 112 enable`** command correlates to VLAN 1001, which is used as the FCoE VLAN ID.

However, the **`show run | section vlan`** command displays the `stp` instance that is associated to which VLAN to correctly identify the `stp` instance ID.

Flex Switch EN4093 and G8272 to a Cisco Nexus switch

The EN4093 switch allows for Ethernet and for FCoE transit traffic. This access switch function can also be performed by the RackSwitch G8272 and G8296 switches.

Several Cisco Nexus family switches can support Ethernet traffic and also support FCoE gateway (FCF) functionality. Nexus switches which have this capability also have ports which can be configured to operate as Ethernet or Fibre Channel ports.

Figure 4 shows SAN A and SAN B for Fibre Channel Multipath isolation. SAN A carries FCoE and Fibre Channel Traffic on the path vHBA-1 → EN4093-1 → Nexus 5548-1 → SAN A. SAN B carries FCoE and Fibre Channel Traffic on the path vHBA-2 → EN4093-2 → Nexus 5548-2 → SAN B. The vLAG Inter Switch Link (ISL) on the EN4093 switch and the vPC ISL on the Nexus 5548 pairs of switches carry only normal Ethernet traffic and are required to prune any FCoE VLANs.

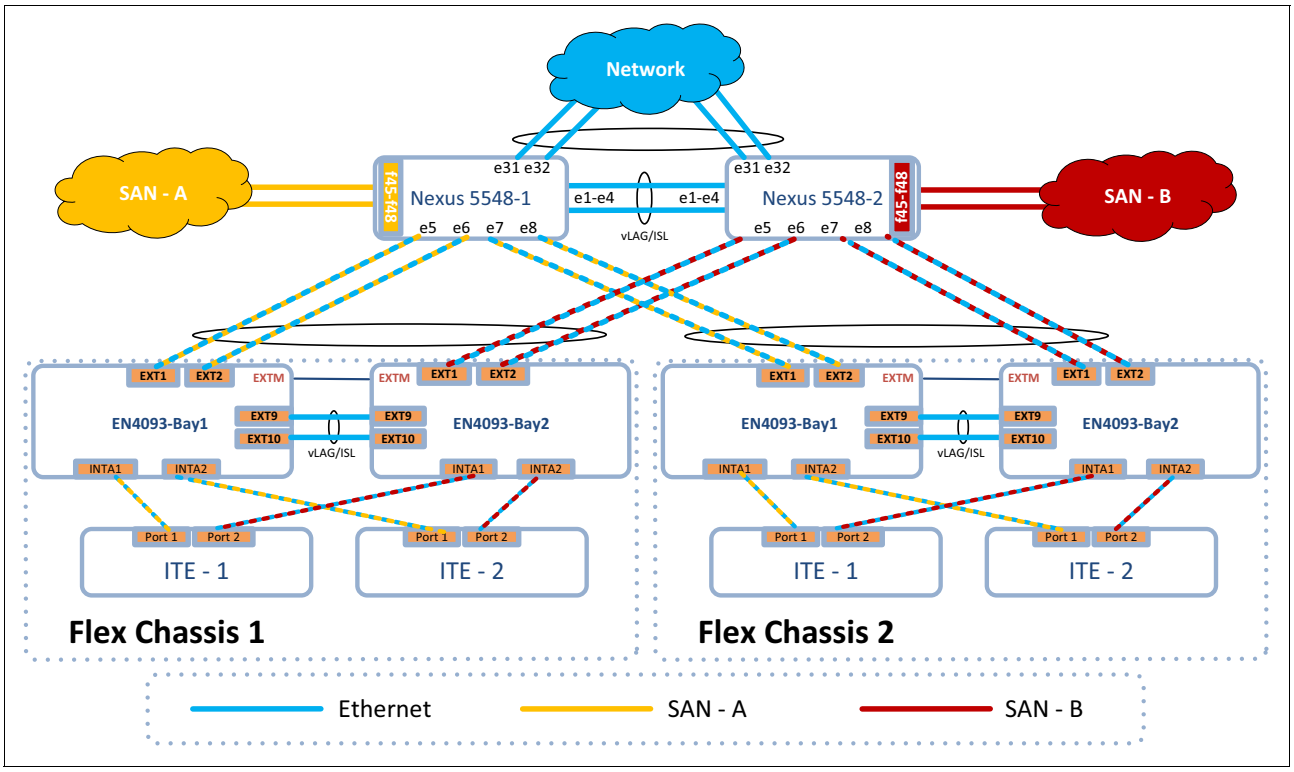

Figure 4 FCoE with vLAG that uses EN4093 and Nexus switches

Consider the following preferred practice guidelines:

- For more information about Nexus Peer-Link and vPC preferred practices, see the Cisco documentation. An example of a vPC and PortChannel is provided in Example 5 on page 15 and Example 7 on page 17.
- Ensure that vLAG ISL ports should contain a non-production VLAN as the native VLAN and PVID; for example, VLAN 4090 or VLAN 4094. This VLAN should also have Spanning Tree disabled to prevent it from ever entering an STP block state or participating in Spanning Tree. This VLAN should not be added to any other ports other than the vLAG ISL ports.
- Server-facing ports (for example, EN4093's INTA ports) on each of the access switches must have VLAN 1 as the Native VLAN / PVID, which is used as the FCoE discovery VLAN.
- EN4093 external ports that are facing the Network must contain VLAN 1 but can be trunked or tagged to carry the FCoE discovery information to the Nexus 5548 Switches for processing.
- Server-facing ports (for example, INTA ports) on each of the EN4093s must contain the FCoE VLAN; usually VLAN 1001 (SAN A) and VLAN 1002 (SAN B).
- If Spanning Tree is enabled on the access switches, the FCoE VLAN should have Spanning Tree disabled. Optionally, VLAN 1 (if required to be enabled on the vLAG ISL ports) can have Spanning Tree enabled if needed or is required.

- vLAG ISL ports do not include either of the two configured FCoE VLANs to prevent Fibre Channel merging from occurring.
- Loop Guard across the vLAG ISL for loop detection and prevention.
- If Spanning Tree is enabled, applying `bpdu-guard` on the ITE facing ports is recommended.
- The Nexus interfaces that are facing the Lenovo access switches must have the **`priority-flow-control mode on`** command used to interoperate with the Lenovo Switches when FCoE is enabled over those interfaces. (Omitting this command can result in PFC not being properly negotiated and can cause severe FCoE performance issues to occur).

Example 5 shows an EN4093 with vLAG and FCoE enabled. Note that while this example is meant to be seen in concert with the Cisco configuration in Example 7 on page 17, it is essentially the same as the configuration which would be used if the use of EasyConnect is not desired. It can therefore substitute for the configuration in Example 2 on page 10.

As in previous examples, a configuration for a top-of-rack switch would be essentially the same as this but the ports would be identified only by their numbers.

Example 5 Example configuration of an EN4093 with vLAG and FCoE without EasyConnect

```
hostname EN4093-Sw1
spanning-tree mode disable
cee enable
!
interface port ext9,ext10
   description vLAG-ISL
   switchport mode trunk
   switchport trunk allowed vlan 10,20,30,4090
   switchport trunk native vlan 4090
   lacp key 5152
   lacp mode active
   exit
!
vlan 4090
   name Peer-Link
   exit
!
vlan 10
   name Production-Vlan-10
   exit
!
vlan 20
   name Production-Vlan-20
   exit

vlan 30
   name Production-Vlan-30
   exit

interface port inta1-inta14,ext1-ext2
   switchport mode trunk
   switchport trunk allowed vlan 1,10,20,30,1001
   switchport trunk native vlan 1
   exit
!
```

```
interface port ext1-ext2
   description Uplink-to-Nexus-1
   lacp key 4344
   lacp mode active
   exit
!
interface ip 127
   ip address 1.1.1.1
   enable
   exit
!
vlag ena
vlag tier-id 10
vlag hlthchk peer-ip 1.1.1.2
vlag isl adminkey 5152
vlag adminkey 4344 enable
```

Example 6 shows a configuration of a G8272 providing FCoE transit and running CNOS. Note that the EN4093 does not run CNOS, and that CNOS does not support EasyConnect at the present time.

Example 6 G8272 CNOS configuration for FCoE transit to FCF on a pair of Nexus switches

```
hostname G8272-Sw1
spanning-tree mode disable
cee enable
!
interface Ethernet 1/19-20
   description vLAG-ISL
   bridgeport mode trunk
   bridgeport trunk allowed vlan 10,20,30,4090
   bridgeport trunk native vlan 4090
   aggregation-group 910 mode active
   exit
!
vlan 4090
   name Peer-Link
   exit
!
vlan 10
   name Production-Vlan-10
   exit
!
vlan 20
   name Production-Vlan-20
   exit

vlan 30
   name Production-Vlan-30
   exit

interface Ethernet 1/1-14,1/17-18
   bridheport mode trunk
   bridgeport trunk allowed vlan 1,10,20,30,1001
   bridgeport trunk native vlan 1
   exit
```

```
!
interface Ethernet 1/17-18
   description Uplink-to-Nexus-1
   aggregation-group 99 mode active
   exit
!
interface mgmt 0
   ip address 1.1.1.1/24
   vrf context management
   ip route 0.0.0.0/0 1.1.1.254/24
   exit
!
vlag ena
vlag tier-id 10
vlag hlthchk peer-ip 1.1.1.2 vrf management
vlag isl port-aggregation 99
vlag instance 1 port-aggregation 910
vlag instance 1 enable
```

Example 7 shows a Nexus 5548 interface/PortChannel with vPC and FCoE enabled.

Example 7 shows a configuration of a Nexus 5548 interface/PortChannel with vPC and FCoE

```
interface Ethernet1/5
   description PureFlex-Ch1-Sw1-Port-EXT1
   switchport mode trunk
   switchport trunk allowed vlan 1,10,20,30,1001
   channel-group 5 mode active
   priority-flow-control mode on
!
interface Ethernet1/6
   description PureFlex-Ch-1-Sw1-Port-EXT2
   switchport mode trunk
   switchport trunk allowed vlan 1,10,20,30,1001
   channel-group 5 mode active
   priority-flow-control mode on
!
interface port-channel5
   description PF-CH-1-Sw1-Ports-EXT1&EXT2
   switchport mode trunk
   switchport trunk allowed vlan 1,10,20,30,1001
   spanning-tree port type edge trunk
   priority-flow-control mode on
   speed 10000
   vpc 5
```

FCoE configuration notes

In order for a server to support FCoE, a Converged Network Adapter (CNA) is required. These adapters present both NIC and HBA functions to the server's operating system during PCI discovery, and typically use standard Fiber Channel drivers as well as Ethernet drivers.

Note that some Lenovo CNA adapters for System x and Flex System servers require a Features on Demand license upgrade to enable FCoE or iSCSI storage functions.

Adapter configuration

The Emulex CNAs have several configuration options which must be set in UEFI for FCoE to be enabled and for the appropriate functions and drivers to be loaded by the server operating system.

The two key parameters are the adapter Personality and Multichannel support.

- Personality

  The adapter personality option must be set to **FCoE** in order for FCoE functionality to be available. This option will not be available if the appropriate license has not been installed. The default option for this parameter is **NIC** which enables Ethernet functionality only; **iSCSI** is the other option available.

- Multichannel

  There are several options for this parameter, and FCoE can be used with all of them. However, if any of the multichannel options are enabled, turning on the virtual NIC capabilities of the adapter, then one virtual instance will be seen by the operating system as an HBA rather than a NIC. For vNIC and UFP multichannel modes, this will be instance number 2 and will need to be configured appropriately on the switch as shown immediately below.

Switch configuration

There are different switch configuration requirements for original Virtual Fabric (vNIC) mode and for UFP mode as specified on the multichannel option. In addition, the commands in Example 8 are always required for FCoE to function properly.

Example 8 FCoE configuration commands

```
cee enable
fcoe fips enable
```

vNIC configuration

Note that vNIC has less functionality than UFP, and the use of UFP is recommended when possible.

The FCoE instance will always be vNIC instance number 2 when FCoE functionality is enabled. The bandwidth allocation for this instance, which is seen by the operating system as an HBA, is configured as shown in Example 9, ranging from 10% to 100% of the 10Gb physical port.

Example 9 vNIC Configuration excerpt for FCoE

```
vnic enable
vnic port <port number> index 2
   bandwidth <10-100>
   enable
   exit
```

UFP configuration

FCoE will always be carried on instance (vport) number 2 when FCoE functionality is enabled. Despite this, it is necessary to configure instance 2 on the switch as an FCoE instance. It is also possible to set minimum guaranteed and maximum allowed bandwidth, as shown in Example 10 on page 19.

Example 10 UFP Configuration excerpt for FCoE

```
ufp enable
ufp port <port number> enable
ufp port <port number> vport 2
   network mode fcoe
   network default vlan 1001 (or other FCoE vlan)
   qos bandwidth min <1-100>
   qos bandwidth max <1-100> (optional, defaults to 100)
```

Other networking topologies with FCoE

There are several other topologies which can be used with FCoE, which are discussed in detail in Chapter 2 of *Lenovo Networking Best Practices for Configuration and Installation*. One which is commonly used with topologies which do not have multi-device link aggregation is shown below.

Traditional STP design with blocking

This topology was commonly used when functions, such as those provided by vLAG and stacking, were not available. It uses a partial mesh between the embedded (or server adjacent) switches and two upstream switches that are cross-connected to each other. The loops, which are built in to this design, are blocked by STP, which puts some ports into a blocking status to prevent a broadcast storm. Operationally, this design resembles an inverted-U topology; however, the blocked links can take over if there are switch or link failures.

The major drawback of this design is that it does require the use of STP and results in wasted bandwidth owing to blocked links. There are several versions of Spanning Tree protocols, including proprietary ones from Cisco, which are supported on Lenovo switches. The choice of one of these protocols will have effects on all or most of the network, not just any portions of it which are carrying converged (storage/FCoE) traffic.

Because ports that are blocked by STP do not carry production traffic, sufficient bandwidth must be built into the topology to carry the expected loads with these ports idle. This topology uses the available links inefficiently.

By design, FCoE traffic from a dual-homed server will flow on two parallel paths until it reaches switches which have the FCF function. This traffic should stay on those parallel paths - traffic on the path on the left should not cross over to the right hand side, and vice-versa. This means that mesh topologies using failover - via Spanning-Tree or other techniques - can experience reachability issues if the FCoE VLANs can cross over. In some cases the easiest way to deal with this is to provide a distinct physical Ethernet connection to the next switch upstream which is dedicated to carrying FCoE traffic. This approach is recommended if it is acceptable to the customer.

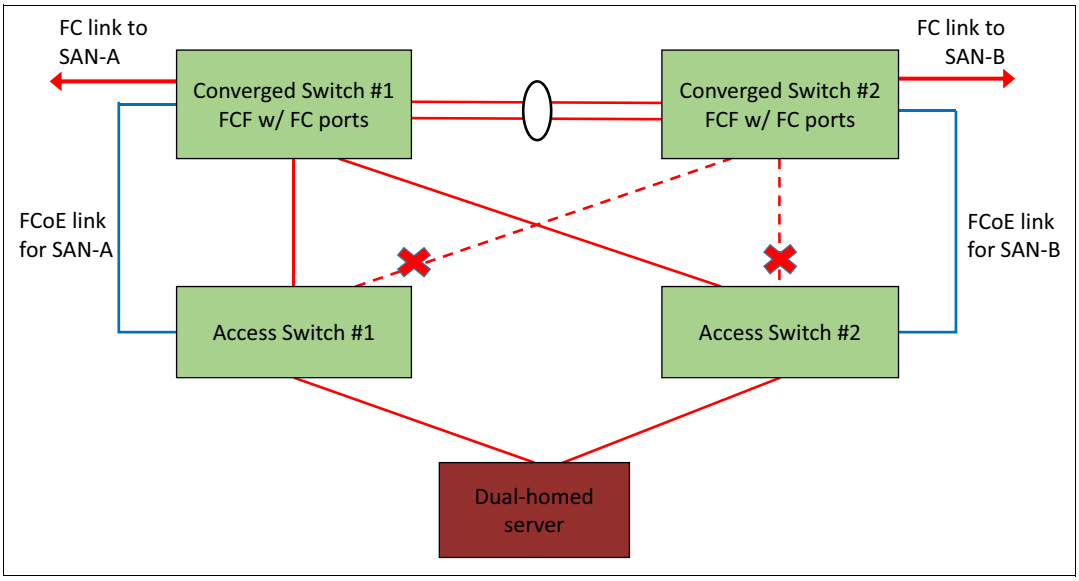

Figure 5 Network topology with spanning tree and blocked links

Limitations and scaling

Converging Ethernet and Fibre Channel networking requires implementing a Fibre Channel engine inside the Ethernet switch. As with Ethernet, too many nodes on the same FCoE VLAN can diminish performance. When a network solution is implemented in which an G8264CS or CN4093 is functioning as the FCF, care must be taken to avoid overloading a single VLAN with FCoE traffic. To maximize total system performance and avoid problems that are associated with a congested Fibre network (latency, fibre login failures, and so on), create FCoE VLANs as the number of physical FCoE hosts increases.

The guidelines for an G8264CS or CN4093 operating in NPV or full fabric mode are as follows:

- < 30 hosts: 1 VLAN with up to 12 uplinks to the Fabric
- 31-70 hosts: 2 VLANs with up to 6 uplinks to the Fabric
- 71-140 hosts: 4 VLANs with up to 3 uplinks to the Fabric
- 141-160 hosts: 6 VLANs with up to 2 uplinks to the Fabric

The number of hosts refers to the number of *physical* hosts. The number of *virtual* hosts that are running on each physical host does not affect the guidelines.

Figure 6 shows multiple FCoE VLANs.

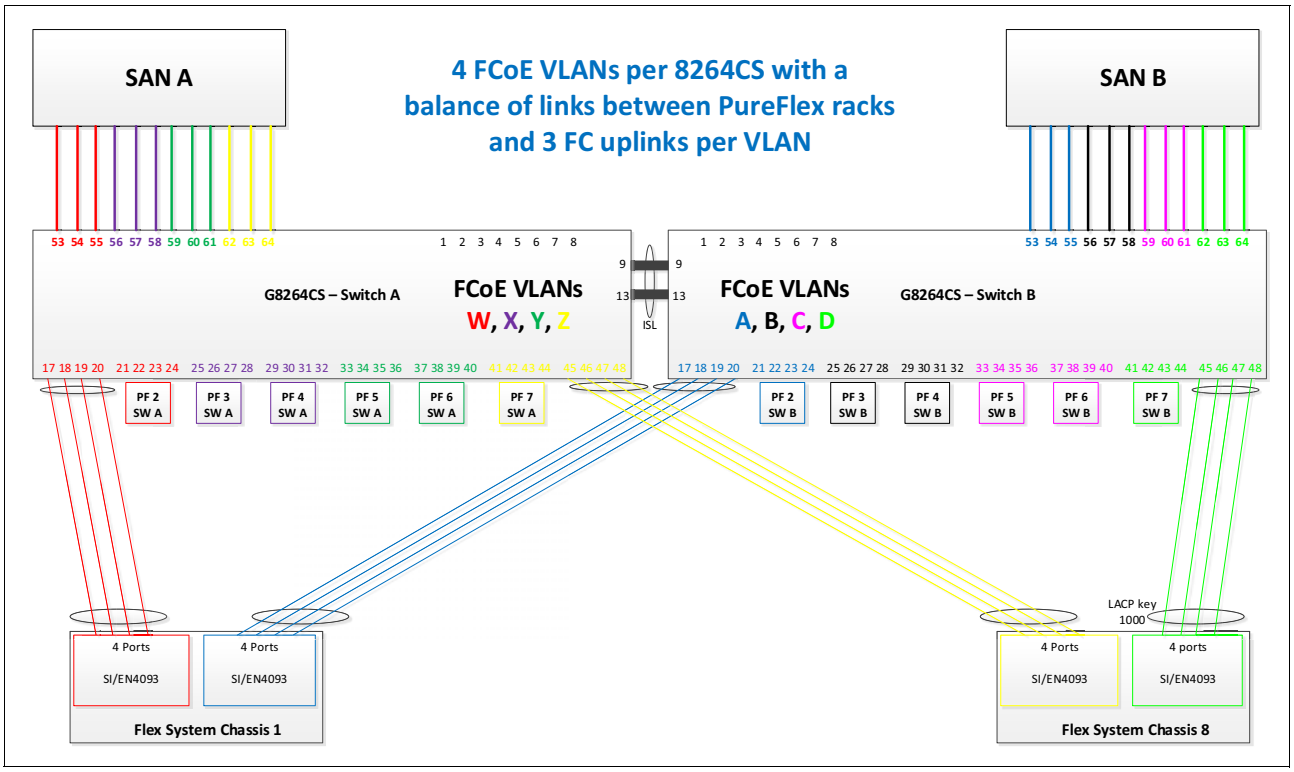

Figure 6 Multiple FCoE VLANs

None of the FCoE VLANs should flow across the vLAG ISL if vLAG is in use, but all of the data-bearing VLANs should cross the ISL. Allowing the FCoE VLANs to cross the ISL can result in ports on the Fiber Channel Forwarder being discovered twice and can result in reachability and stability issues.

Ethernet SAN considerations

Considers the following points regarding any Ethernet-based SAN:

- Reliability: Data paths must be redundant and handle failures efficiently.
- Performance: A high-speed and uncongested network must support Jumbo Frames.
- Latency: Cut through performance and high-speed links is important.

The Lenovo 10 Gbps and 40 Gbps Ethernet switches satisfy all of these requirements by using the redundancy, link aggregation, and jumbo Ethernet packet features. Virtual Link Aggregation (VLAG) is another feature that improves reliability and performance.

Another performance-enhancing feature to consider is Converged Enhanced Ethernet (CEE) if every device, end-to-end, supports CEE, including the NICs and Switches. CEE improves the Ethernet traffic with early congestion notification and traffic metering features to even out the flow of any traffic that is using high bandwidth. This feature can be useful in eliminating slow TCP traffic patterns that result from traffic loss because of congestion and retransmissions.

A TCP retransmission results in a slow start congestion avoidance that is defined in the TCP protocol. Repeated retransmissions because of congestion results in what is known as a *saw*

tooth traffic pattern, as shown in Figure 7. The use of CEE can reduce the maximum relative throughput, but eliminates the saw tooth pattern. Enabling CEE on Lenovo Networking switches is accomplished by using the **`cee enable`** command.

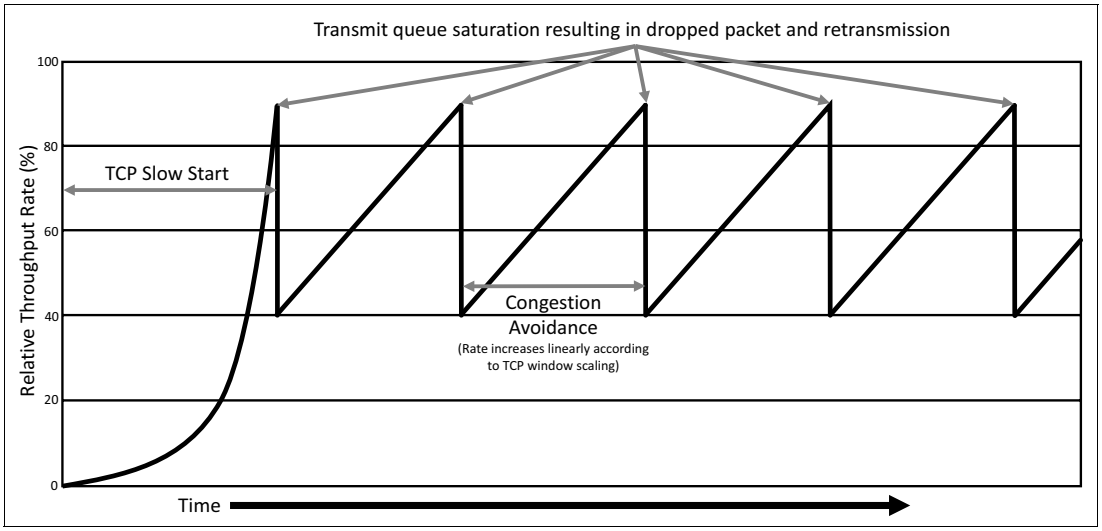

Figure 7 TCP saw tooth traffic pattern because of congestion

IBM Spectrum Scale

IBM Spectrum Scale, previously known as IBM General Parallel File System (GPFS), optimizes file access by splitting the file access across multiple nodes. When implemented over Ethernet, the connections are accomplished by using multiple TCP connections. The use of more than one NIC in an active-active configuration rather than active-standby or transmit load balancing (TLB) is preferred for redundancy and for increasing the bandwidth. Active-active NICs are accomplished by using link aggregation.

The use of VLAG enhances the reliability because the LAG is split across two switches. VLAG can also increase the performance because more connections between multiple switch hops can be used. Because there are multiple TCP connections, it is preferred to add Layer 4 ports into the trunk hash calculations by using the **`portchannel thash l4port`** configuration command.

Because GPFS is implemented by using TCP, enabling CEE on all nodes can improve the overall performance by reducing the network congestion.

It is important to also verify that the links are not over-used, which results in dropped traffic. These links can be monitored by monitoring the interface counters and looking for IBP/CBP and HOL-blocking discards. It is also good to review the unicast packets that are being transmitted and received across LAGs to verify that a good distribution is achieved. Egress distribution is controlled on the local device, whereas ingress distribution is controlled on the remotely attached device. It is useful to clear the counters by using clear interfaces and monitor them at intervals.

iSCSI

iSCSI is another Ethernet SAN that is implemented by using TCP similar to GPFS, except that the file access is accomplished by using a single-server interface. Because a single file server is used, the use of LAGs is more critical for the server connections. Implementing the network by using VLAG is important for active use of all connections and enabling Layer 4

port into the trunk hash calculations might increase even distribution across the links (**portchannel thash l4port**).

The use of CEE improves the performance if all of the switches support CEE. There also are CEE enhancements available for iSCSI by using Data Center Bridge Exchange (DCBX) protocol. This configuration is similar to the DCBX protocol enhancements that are available for FCoE in which optimal connections are negotiated through the network. For an iSCSI network, CEE and iSCSI DCBX support must be enabled by using the following commands:

```
configure terminal
cee enable
cee iscsi enable
```

It is important to monitor the port counters to verify even distribution and congestion control. The use of 40 Gbps network connections can be helpful in optimizing an iSCSI network design because multiple TCP connections are not used; therefore, LAGs might not have effective traffic distribution. Also, 40 Gbps links reduce the transmission latency over the links.

RoCE

Remote Direct Memory Access Over Converged Ethernet (RoCE) performance optimization is similar to SAN optimization. Latency is critical to RoCE; therefore, it is important to limit the number of hops through which the traffic passes. As the protocol name indicates, CEE is used; therefore, it must be enabled by using the **cee enable** command.

About the author

Scott Lorditch is a Consulting System Engineer for Lenovo. He performs network architecture assessments and develops designs and proposals for solutions that involve Lenovo Networking products. He also developed several training and lab sessions for technical and sales personnel. Scott joined IBM as part of the acquisition of Blade Network Technologies® and joined Lenovo as part of the System x® acquisition from IBM. Scott spent almost 20 years working on networking in various industries, as a senior network architect, a product manager for managed hosting services, and manager of electronic securities transfer projects. Scott holds a BS degree in Operations Research with a specialization in computer science from Cornell University.

Thanks to the following people for their contributions to this project:

- David Watts, Lenovo Press

Portions of this paper we taken from *Lenovo Networking Best Practices for Configuration and Installation*. Thanks to the authors of that publication:

- Scott Irwin
- Scott Lorditch
- Ted McDaniel
- William Nelson
- Matt Slavin
- Megan Gilge

Notices

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

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

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

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

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

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document was created or updated on October 6, 2017.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp0624

Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at http://www.lenovo.com/legal/copytrade.html.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Blade Network Technologies®	Lenovo®	Lenovo(logo)®
Flex System™	RackSwitch™	System x®

The following terms are trademarks of other companies:

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