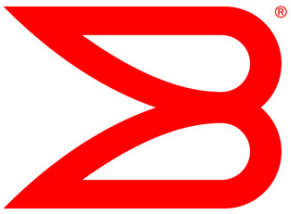

DATA CENTER

Brocade VCS Fabric Technical Architecture

This document is intended for data center networking and server architects. It describes the technical architecture of Brocade VCS Fabric technology. It explains how the control, data, and management planes function, how a Brocade VCS fabric operates, and how VCS Fabric technology supports Fibre Channel over Ethernet. Key concepts and important functions are explained. This document does not address specific feature support by release. Please refer to the Release Notes for these details.

BROCADE

CONTENTS

LIST OF FIGURES

AUDIENCE

This document is intended for data center networking and server architects. It describes the technical architecture of Brocade® VCS® Fabric technology, which includes virtual cluster switching, and explains how the control, data, and management planes function and how a Brocade VCS fabric operates. It also provides a comparison of the VCS fabric with the IEEE Transparent Interconnect of Lots of Links (TRILL) standard.

GOALS

The purpose of this document is to describe the technical architecture of the Brocade VCS Fabric technology platform. This document explains key concepts and provides descriptions of important functions without going into a detailed description of the implementation.

Because the scope of this document is architectural, capabilities that are described may not be available in a particular release of Brocade Network OS firmware. Feature support that is provided by a firmware release is described in the Release Notes and should be consulted to determine what specific features are available in any given Brocade Network OS release.

NOTES

VCS Fabric technology leverages the emerging TRILL standard as well as other standards from IEEE and ANSI, such as Data Center Bridging (DCB) and Fibre Channel over Ethernet (FCoE). Throughout this document, references are made to TRILL, RBridge, TRILL frames, and other terms that are found in the TRILL standard. However, the current implementation of VCS Fabric technology does not include all of the features that are found in the TRILL standard. See the Appendix A: Brocade VCS Fabric and TRILL section for more information about similarities and differences.

The Brocade VDX® switch family can operate in one of two modes: classic mode and VCS mode. This document is devoted to a discussion of the technical architecture supporting VCS mode.

INTRODUCTION

In a 2012 Gartner survey of more than 1,600 CIOs across the globe, participants were asked about their top current business priorities as well as looking three to four years into the future. The current focus was on improving business processes and cost savings. Looking out into the future, improving productivity, driving innovation, gaining competitive advantage, and attaining new customers were the priorities. These business concerns require data centers to deploy new applications quickly and efficiently, provide fast and reliable "around-the-clock" access to information, meet or exceed stringent service levels with no downtime, and do all this while driving down costs by maximizing investments. In short, IT must move at the speed of business to capitalize on new opportunities and respond to increasing global competition.

Addressing these business needs is a set of technology enablers, including high-density, multicore servers, server and storage virtualization, and a move towards service orchestration and cloud computing. Data centers are able to leverage these technologies to lower capital and operational expenditures, while at the same time creating an infrastructure that rapidly scales and responds to business needs. When data center operators leverage these technologies, there are added networking challenges that did not have to be dealt with when applications were tied to physical servers. Therefore, the network must evolve. It must move from management of physical ports to flows (virtual server to virtual server, or virtual server to virtual storage communication). It must be simpler to operate, more flexible, highly resilient, and much more scalable. These requirements are addressed with scale-out Ethernet fabrics, while classic Ethernet networks require complex architectures and protocols, which add higher levels of complexity and operational costs.

Brocade VCS Fabric technology is explicitly designed to meet these challenges, allowing users to greatly decrease the operational costs of networking by providing a highly reliable, simple, scalable networking infrastructure. The Brocade VDX switches deliver Brocade VCS Fabric technology and are revolutionizing the way data center networks are architected.

BROCADE VCS FABRIC TECHNOLOGY

This document describes the technical architecture of Brocade VCS Fabric technology. Brocade provides advanced Ethernet fabric technology, which replaces many of the limitations of classic Ethernet networks in the data center. In addition to standard Ethernet fabric benefits, such as logical flat (Layer 2) networks without the need for Spanning Tree Protocol (STP), Brocade VCS Fabric technology also includes unique services such as automatic Virtual Machine (VM) alignment and –highly efficient multipathing at Layers 1-3, with multiple Layer 3 gateways. The VCS architecture conforms to the Brocade strategy of "revolution through evolution"; therefore, all products with VCS Fabric technology can connect to existing data center Ethernet products, whether from Brocade or other vendors.

Ethernet Fabrics

Compared to classic hierarchical Ethernet architectures, Ethernet fabrics provide higher levels of performance, utilization, availability, and simplicity. They have the following characteristics, at a minimum:

Flatter. Ethernet fabrics eliminate the need for STP, while still being completely interoperable with existing Ethernet networks.

Flexible. Ethernet fabrics can be architected in any topology to best meet the needs of any variety of workloads.

Resilient. Multiple "least cost" paths are used for high performance and high reliability.

Elastic. Ethernet fabrics easily scale up and down at need.

More advanced Ethernet fabrics borrow further from Fibre Channel fabric constructs:

- They are self-forming and function as a single logical entity, in which all switches automatically know about each other and all connected physical and logical devices.
- Management can then be domain-based rather than device-based and defined by policy rather than repetitive procedures.
- These features, along with virtualization-specific enhancements, make it easier to explicitly address the challenges of VM automation within the network, thereby facilitating better IT automation.

Protocol convergence (for example, FCoE) may also be a feature, intended as a means of better bridging LAN and SAN traffic.

Ethernet fabrics have improved the way organizations deploy, use, and maintain their network access layers. However, the elastic scale of Ethernet fabrics and the applications that use them has outpaced traditional routing capabilities, constraining Layer 2 fabric growth and performance as a result.

Brocade VCS Fabric technology brings intelligent, efficient routing to VCS fabrics. It enables highly elastic Layer 2 and Layer 3 domains with extremely efficient load balancing and multiple active Layer 3 gateways, on top of Layer 2 Equal-Cost Multipath (ECMP) and Brocade Inter-Switch Link (ISL) Trunking. The result is more effective link utilization, which reduces overall cost, more resilience, which results in greater application uptime, and a more flexible and agile network, which helps organizations rapidly adapt to changing business conditions. In summary, Brocade VCS Fabric technology provides multiple load-balanced paths at Layers 1–3, with multiple active gateways.

Brocade VCS Fabric technology is available in the Brocade VDX family of switches. The Brocade VDX family is the first to deliver an Ethernet fabric and is designed with the following capabilities:

- Changes the way that data center networks are built, to support cloud computing initiatives
- Simplifies network architectures, dramatically reducing operating expenses
- Allows the virtual data center to scale, while reducing complexity and enabling seamless application mobility
- Increases network performance, utilization, and resiliency by building data center Ethernet fabrics

The Brocade VDX family of products includes four platforms:

- The Brocade VDX 6710 Switch is available in a 1U rack-mounted model with 48 GbE (Gigabit Ethernet) and six 10 GbE ports.
- The Brocade VDX 6720 Switch is available in either a 1U rack-mounted model with 24 10 GbE ports or a 2U rack-mounted model with 60 10 GbE ports.
- The Brocade VDX 6730 Switch is available in either a 1U rack-mounted model with 24 10 GbE ports and eight 8 Gbps Fibre Channel ports, or a 2U rack-mounted model with 60 10 GbE and 16 8 Gbps Fibre Channel ports.

The Brocade VDX 8770 Switch is available in two form factors: a 4-I/O slot system and an 8 I/O slot system with linecard support for 1-GbE, 10-GbE and 40GbE ports. The 4-slot chassis model supports up to 192 1/10 GbE ports, or 48 40 GbE ports, or a combination. The 8-slot chassis model supports up to 384 1/10 GbE ports, or 96 40 GbE ports, or a combination.

The Brocade VDX switches are scalable, flexible network building blocks that network architects can apply in these three important use cases:

1. 10 GbE Ethernet Access and Aggregation Within Traditional Network Architectures
    - Preserves existing hierarchical network design while eliminating STP, resulting in an active-active network and reduced management overhead.
    - Provides a two-switch VCS configuration at the top of each server rack.
2. Scale-out fabrics for Virtual Data Centers
    - Deploys scale-out fabrics instead of a hierarchical network to flatten the network design, provides seamless application mobility, and manages the entire fabric as a single logical chassis.
3. LAN/SAN Convergence
    - The fabric provides end-to-end DCB capabilities, which allows traditional IP and storage traffic (FCoE and/or iSCSI traffic), to exist on the same network. The Brocade VDX 6730 allows FCoE traffic in a VCS fabric to connect with Fibre Channel ports in a Fibre Channel fabric via Fibre Channel routing. This allows traffic flow between devices using FCoE and Fibre Channel, while avoiding the need to merge a VCS fabric with a Fibre Channel fabric.

BROCADE VCS TECHNICAL ARCHITECTURE

Overview

Brocade VCS Fabric technology is a revolutionary Ethernet fabric technology, which includes unique Layer 2 and Layer 3 services that improve network utilization, increase application availability, enhance system scalability, and drastically simplify the data center network architecture. VCS Fabric technology implements a new type of networking, the VCS fabric, using an extensible architecture for adding new services and capabilities. A VCS fabric provides a flexible interconnecting network between individual switches, which is called a "fabric." Switches that form a fabric create a virtual cluster of physical switches as seen by external classic Ethernet switches or devices, as shown in Figure 1.

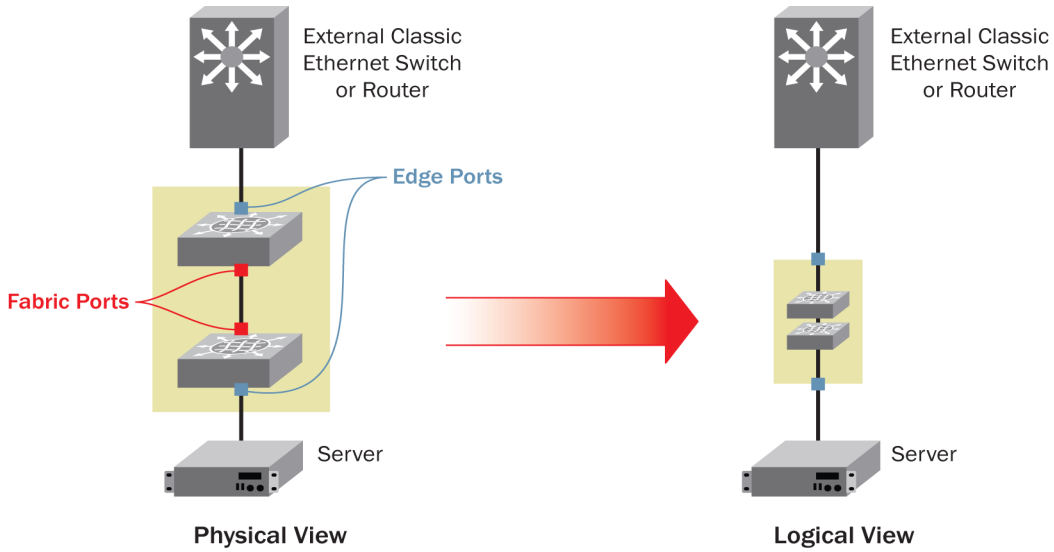

Figure 1. Brocade VCS technology, in which virtual switch clusters form a fabric.

The VCS architecture allows operators to define ports as either edge ports or as fabric ports. In Figure 1, the ports that are connecting the switches together are fabric ports and are transparent, as far as the external devices and classic Ethernet switches that are connected to the edge ports are concerned. Therefore, the fabric and its fabric ports behave like a single logical switch to the external network.

Because VCS Fabric technology removes the need for STP, yet interoperates with classic Ethernet switches supporting spanning tree, the entire fabric is transparent to any Bridge Protocol Data Unit (BPDU) frames and behaves like a transparent LAN service from the perspective of spanning tree.[1] Loop detection and active path formation are managed by spanning tree in the classic Ethernet switches. The VCS fabric is not involved, because it is transparent to the BPDU packets. The VCS architecture can be configured so that all BPDUs transit the VCS fabric, or to prevent them from transiting the VCS fabric.

Edge ports support industry-standard Link Aggregation Groups (LAGs) via Link Aggregation Control Protocol (LACP). Traditional Ethernet switches can use LAGs to eliminate STP on inter-switch links when connecting to a VCS fabric.

[1] On the Brocade VDX 6710, 6720. and 6730. Check current release notes for the Brocade VDX 8770.

VCS Fabric Layer 2

Operating Modes

The Brocade VDX switches support the following industry-standard features:

- Layer 2 data forwarding
- MAC learning and aging
- BPDU drop
- Ping and traceroute
- Per VLAN Spanning Tree Plus (PVST+), Per VLAN Rapid Spanning Tree Plus (PVRST+) [Cisco interoperability]
- LACP, Brocade ISL Trunking
- Link-Level Discovery Protocol (LLDP) and Data Center Bridging Exchange (DCBX)
- IEEE 802.1x
- sFlow
- Switched Port Analyzer (SPAN)
- Layer 2 Access Control Lists (ACLs)
- Management port IP ACL (standard and extended)
- Simple Network Management Protocol (SNMP)
- NETCONF support (RFC 4741)
- LDAP v3 (RFC 4510)
- Enhanced Transmission Selection (802.1Qaz)
- Priority-based flow control (802.1Qbb)
- Internet Group Management Protocol (IGMP) snooping
- Automatic Migration of Port Profiles (AMPP)
- In-band management
- TACACS+
- Internet Small Computer Systems Interface (iSCSI) DCBX support
- STP, Rapid Spanning Tree Protocol (RSTP), Multiple Spanning Tree Protocol (MSTP)[2]

The following additional features are available within the VCS fabric itself:

- Transparent LAN service
- Virtual Link Aggregation Groups (vLAGs)
- Distributed configuration management
- End-to-end FCoE
- Fibre Channel fabric connectivity
- VM-aware networking

[2] On the Brocade VDX 6710, 6720. and 6730. Check current release notes for the Brocade VDX 8770.

Control Plane

In a VCS fabric, the control plane is shared across all switches in the fabric. Specific protocols are used to discover directly adjacent switches in VCS mode and to form a VCS fabric with minimal user configuration. This describes how a VCS fabric forms:

- The "VCS Enable" setting determines if a switch will operate in traditional (STP) mode or VCS mode. "VCS Enable=OFF" means that the switch will operate in traditional mode. "VCS Enable=ON" means that the switch will operate in VCS mode.
- Upon power reset, if "VCS Enable=OFF," a switch goes into traditional mode and operates like a regular IEEE 802.1Q switch.
- By default, all interfaces are in a "shutdown" state when a switch is configured to run in traditional mode.
- Upon power reset, if "VCS Enable=ON," the switch will start the process of forming a VCS fabric.
- Each VCS fabric is identified by a VCS Fabric ID (VFID), and all switches joining a fabric must use the same VFID. The default VFID is set to "1" and can be manually changed.
- When "VCS Enable=ON," the switch executes the following sequence of steps:
    - All interfaces in the switch transition from the "shutdown" state and start operating as edge ports.
    - Brocade Link Discovery Protocol discovers if another switch with "VCS Enable=ON" is connected to any port. See the Neighbor Discovery section for details.
    - If another switch operating in VCS mode is found on a port, a merge operation attempts to form a VCS fabric at the link level.
    - A series of fabric formation protocols are initiated once the link level relationship has been established between two neighbor switches. See the Fabric Formation section for details.
    - A merge and join protocol starts merging switch configurations between the discovered switches once the VCS fabric has successfully formed. See the Fabric Formation section for details.
    - The VCS fabric formation is complete once the Join/Merge Protocol (JMP) is declared a success.

Neighbor Discovery

A port can be logically defined as either edge or fabric, but not both at the same time. An Inter-Switch Link (ISL) exists only between fabric ports. VCS-capable neighbor discovery involves the following steps. If a VCS-capable neighbor is discovered, the port comes up as a fabric port; otherwise, it comes up as an edge port.

1. Discover whether the neighbor is a Brocade VDX switch. If not, come up as an edge port.
2. Discover whether the Brocade neighbor switch is in VCS mode. If not, come up as an edge port.
3. Additionally, the discovery protocol ensures that all switches in a fabric have the same VFID. Only switches with the same VFID can form a fabric. If not, they come up as edge ports.
4. If discovery of an adjacent switch in VCS mode with the same VFID occurs, then that port transitions to become a fabric port, and an ISL is established.

Fabric Formation

As previously mentioned, the VCS fabric leverages proven Fibre Channel fabric protocols to build a TRILL-based fabric. These are the main functions of the fabric formation protocols:

1. Assign a fabric-wide unique RBridge ID (See the Switch ID Allocation section.)
2. Create the network topology database via a standard link-state routing protocol adapted for use in Layer 2.
3. Compute the broadcast tree to distribute broadcast traffic across the switches in the VCS fabric.

Brocade ISL Trunking

A Brocade ISL Trunk is a hardware-based Link Aggregation Group (LAG). These LAGs are dynamically formed between two adjacent switches using existing ISL connections. Brocade ISL Trunk formation does not use LACP. Instead, it

uses a special trunk protocol. Formation of a Brocade ISL Trunk does not require any user intervention or configuration. The command **fabric trunking <ON|OFF>**, which has a default value of ON, can be used to prevent formation of an ISL from joining a Brocade ISL Trunk.

When compared with a LAG implemented in software with IEEE 802.1ad LACP, a Brocade ISL Trunk with hardware-based load balancing distributes traffic evenly across all member links on a frame-by-frame basis without the use of a hashing algorithm. This provides very high-link utilization across all links in the trunk, with near-linear scaling of bandwidth as links are added.

Brocade ISL Trunks support VLAN tagged and untagged "native VLAN" traffic. Untagged traffic is assigned to the default VALN (ID = 1).

Requirements of a Brocade ISL Trunk include the following:

- There is a maximum of eight ISLs per trunk group.
- All ports in a switch must reside on the same ASIC hardware boundary.
- Multiple trunks can form between switches.
- A standard LAG or vLAG can be formed across Brocade ISL Trunks.

Switch ID Allocation

A unique switch ID is assigned to each switch within the fabric by the fabric formation protocol. The switch ID is equal to an RBridge ID. The RBridge ID that appears in the TRILL header is the same as a VFID.

Fabric Layer 2 Link-State Routing Protocol

After an RBridge ID is assigned to a switch, the link-state routing protocol starts forming adjacencies and collecting topology and interconnectivity information from its neighbors.

Fabric Interface Configuration

As previously mentioned, a physical interface can either be an edge port or a fabric port, but not both. Similar to the specific switch port configuration being captured on its physical interface, the fabric port configuration is also captured on a physical interface.

Additional interface configuration options are available for fabric ports.

Fabric ISL Enable

This setting controls whether an ISL (that is, a connection between two fabric ports) should form. The default setting allows automatic creation of an ISL between two switches.

Performing a **fabric isl enable** command on an interface of an operational ISL has no effect on traffic. Performing a **no fabric isl enable** command on an interface causes the following actions to be taken:

- The link goes offline.
- The ISL formation is disabled.
- The switch to inform its neighbor that the interface is an ISL is disabled.
- The link from a trunk group, if it was a member, is removed.
- The neighbor switch stops ISL formation activity, regardless of its current interface state.

Interface Shutdown

Performing a **shutdown** command on an operating ISL interface not only brings down the physical link but also removes its fabric adjacency information.

The main difference between **shutdown** and **no fabric isl enable** is that the link stays up after **no fabric isl enable**. The link stays down after **shutdown**. Use of the **no fabric isl enable** command is preferred to expedite ISL state transition, because its link state stays up.

Fabric Trunk Enable

This setting is enabled by default on an interface and allows the interface to automatically be added to a Brocade ISL Trunk group. A fabric trunk automatically forms from multiple ISL connections in the same Port Group (PG). A trunk is a single logical link between switches. If this attribute is disabled, each ISL acts as a single link between two VCS Fabric switches.

Reserved VLANs for Control Plane Traffic

The VCS Fabric architecture reserves two VLANs (VLAN: 4093 and VLAN: 4095) to carry VCS control plane traffic in the fabric. These VLANs use the VLAN in the TRILL outer header. The VLAN in the TRILL inner header is not changed when frames transit the VCS fabric.

Data Plane

The transparent switching service at Layer 2 is defined as Transparent LAN Service (TLS). The data plane forwarding is fully compliant with the TRILL standard.

TRILL Frames

There are two classes of TRILL frames that are exchanged over a fabric port: data frames (unicast; broadcast, unknown unicast, and multicast [BUM]; and FCoE), and VCS fabric control plane frames.

A data frame that is received on an edge port is encapsulated in a TRILL frame, as shown in Figure 2. Refer to the Internet Engineering Task Force (IETF) TRILL working group for details on the TRILL frame format.

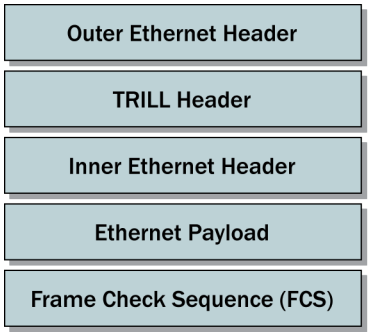

Figure 2. TRILL frame format.

Outer Ethernet Header

In the Outer Ethernet Header, the destination MAC address is the MAC address of the ISL port or Brocade ISL Trunk of the next-hop Brocade VDX switch. The source MAC address is the MAC address of the sending Brocade VDX switch ISL port or Brocade ISL Trunk. These are updated hop-by-hop. The tag includes the TRILL EtherType. A VCS fabric reserves two of the outer VLANs for VCS control plane traffic. The originating station VLAN tag in the Inner Ethernet Header is preserved.

TRILL Header

The Egress RBridge ID is the RBridge ID of the destination switch. The ingress RBridge ID is the RBridge ID of the sending switch. See the Switch ID Allocation section for how the RBridge ID is set. The RBridge IDs are updated hop-by-hop, as is the hop count.

Ethernet Payload

This is the Ethernet data frame that is received on the edge port.

Inner Ethernet Header

This includes the destination and source MAC addresses that are assigned by the originating station, along with any tagging and VLAN assignments.

Frame Check Sum
This is computed across the entire TRILL frame and is updated hop-by-hop.

Non-TRILL Frames

VCS Fabric Formation Frames
At the link level, adjacent switches exchange VCS fabric-specific control frames with each other to form the VCS fabric via fabric ports. These control frames are sent using a well-known address.

FCoE Control Frames
All Fibre Channel (FC) protocols utilize FCoE as the data plane transport. This means that all FC control frames are first encapsulated with an FCoE header and then with an Ethernet header. These frames are not TRILL-encapsulated, because they are exchanged between switches before any RBridge ID is assigned. Note that FCoE data frames are, however, TRILL-encapsulated.

Equal-Cost Multipath Forwarding at Layer 2
A standard link-state routing protocol that runs at Layer 2 determines if there are Equal-Cost Multipaths (ECMPs) between RBridges in a VCS fabric and load-balances the traffic to make use of all available ECMPs.

Note: At launch in 2010, Brocade chose to initially use a well-known, stable, existing standard for the Layer 2 link-state routing protocol: Fabric Shortest Path First (FSPF). As defined by the ANSI standard, FSPF is a link-state routing protocol that is successfully used by all FC Storage Area Network (SAN) fabrics as well as FCoE traffic. In July of 2011, IETF finalized extensions to the Intermediate System-to-Intermediate System (IS-IS) standards for use at Layer 2, as required by the TRILL standard. As of this writing, Brocade VDX switches implement a VCS fabric in which the data plane is fully TRILL-compliant, while FSPF is used in the control plane. Alternative link-state routing protocols such as IS-IS may be supported in VCS in the future.

ECMP in the VCS fabric behaves slightly differently from traditional IP ECMP implementation. Although configurable via the CLI, the default link cost does not change to reflect the bandwidth of the interface. Any interface with a bandwidth equal to or greater than 10 Gbps has a predetermined link cost of 500. Thus, a 10 GbE interface has the same link cost as an 80 Gbps interface. As explained later, the VCS implementation of ECMP load-balances traffic and avoids overloading lower bandwidth interfaces.

If a neighbor switch is reachable via several interfaces with different bandwidths, all of them are treated as equal-cost paths. While it is possible to set the link cost based on the link speed, such an algorithm complicates the operation of the fabric. Simplicity is a key value of Brocade VCS Fabric technology, so an implementation was chosen that does not consider the bandwidth of the interface when selecting equal-cost paths.

The distributed control plane is aware of the bandwidth of each interface (ISL or Brocade ISL Trunk). Given an ECMP route to a destination RBridge, it can load-balance the traffic across the next-hop ECMP interfaces, according to the individual interface bandwidth. As a result, load balancing is based on the aggregate link speed that is available to an adjacent switch.

The other motivation of this implementation is to maximize the utilization of available links in the network. The VCS implementation of ECMP does not follow the traditional model for Layer 3 ECMP. In the traditional approach, an 80 Gbps interface, which has the least cost among all of the ECMP paths, is used as the only route to reach the destination. The lower-speed interfaces are not utilized, resulting in lower overall bandwidth. With VCS Fabric technology, lower bandwidth interfaces can be used to improve network utilization and efficiency.

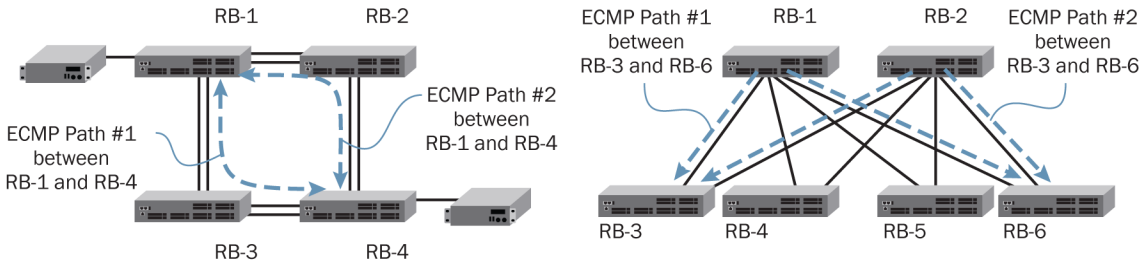

Figure 3. Equal-Cost Multipathing examples.

Figure 3 shows two examples of ECMP between two RBridges. In both examples, there are two non-overlapping paths between the source and the destination RBridges. The key to ECMP is that these paths are non-overlapping. The VCS fabric is capable of splitting the traffic between the two end nodes across these non-overlapping paths while maintaining in-order frame delivery. This approach ensures that all paths in the network are active and effectively and efficiently used. This ECMP method of full-link utilization not only uses all of the available network bandwidth but also increases the level of resiliency. A failure of a link or a switch on one of the paths does not affect connectivity between the end stations communicating across the fabric.

Primary to the efficient usage of ECMP is how a source RBridge can effectively split the incoming traffic to evenly distribute it on all available paths to reach the intended destination. While Brocade ISL Trunks do not use hashing when forwarding frames on multiple ISL connections, ECMP does use a hashing algorithm based on specific fields in frame headers of the incoming frame. Based on the hash, the frame is assigned to one of the available ECMP paths (such as vLAG).

These are the fields that are used for the hashing algorithm:

ECMP for LAN Traffic: <MAC-SA, MAC-DA, VID, IP-Proto, S-IP, D-IP, L4-SRC-Port, L4-DST-Port>

Configurable Load-Balancing Option

Load balancing allows traffic distribution on static and dynamic LAGs and vLAG. Although not common, some traffic patterns fail to distribute well, leading to only one ECMP path for all traffic. This causes underutilization of ECMP paths, resulting in congestion, even though ECMP paths are available to offload the traffic. A command is available to configure ECMP load balancing. This allows the user to select the parameters used to create the load-balancing hashing scheme. Refer to the Brocade *Network OS Administrator's Guide* for more information on the ECMP hashing scheme and how to customize the load-balancing hashing scheme.

MAC Learning

The Brocade VCS distributed control plane learns MAC addresses from data forwarding on edge ports, similar to any standard IEEE 802.1Q bridge. An edge switch learns about a MAC, its VLAN, and the interface on which the MAC was seen. It associates the learned information with the RBridge ID that is assigned to the switch containing the edge port. The frame is forwarded into the fabric on a fabric port with TRILL encapsulation, based on whether the destination address in the frame is known or unknown. As RBridges forward the frame, they also use data path MAC learning to populate their frame forwarding tables.

After the TRILL approach, the VCS distributed control plane helps synchronize aging and learning states across all fabric switches via the Ethernet Name Service (eNS), which is a MAC distribution service.

The following is an explanation of how MAC learning occurs for switches in a VCS fabric.

Source MAC Learning on an Edge Port

Assume a fabric has just been formed. None of the RBridges has seen any frames from server A or B. Each RBridge has been assigned a unique RBridge ID designated by "RB-#," shown in Figure 4. RB-1 is the root of the distribution tree that is used to flood frames for unknown destination MAC addresses. We assume that VLAN ID 1 has been configured on the edge ports of RB-1 and RB-3. The edge port on RB-1 learns the MAC address from station A (MAC-A) when the device transmits a frame.

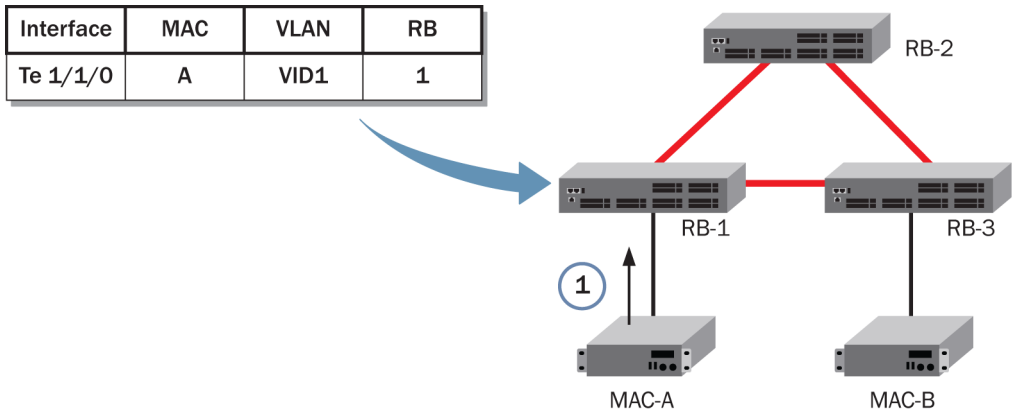

Figure 4. Learning at the edge.

Destination MAC Learning in the Fabric

The RB-1 edge port has not seen the destination MAC address before, so it sends a TRILL frame on the distribution tree to all other switches so that they can learn it, as shown in Figure 5. This is the way any IEEE 802.1Q switch learns MAC addresses.

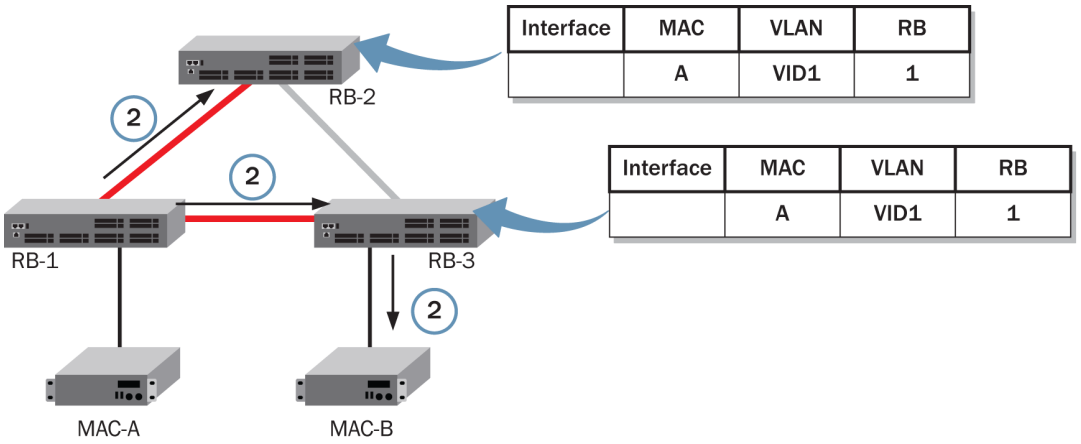

Figure 5. Learning on fabric ports from TRILL frames.

Forwarding Table Synchronization in the Fabric

Not all switches have complete forwarding table information for MAC-A. The eNS synchronizes the forwarding table information for all switches in the fabric, as shown in Figure 6.

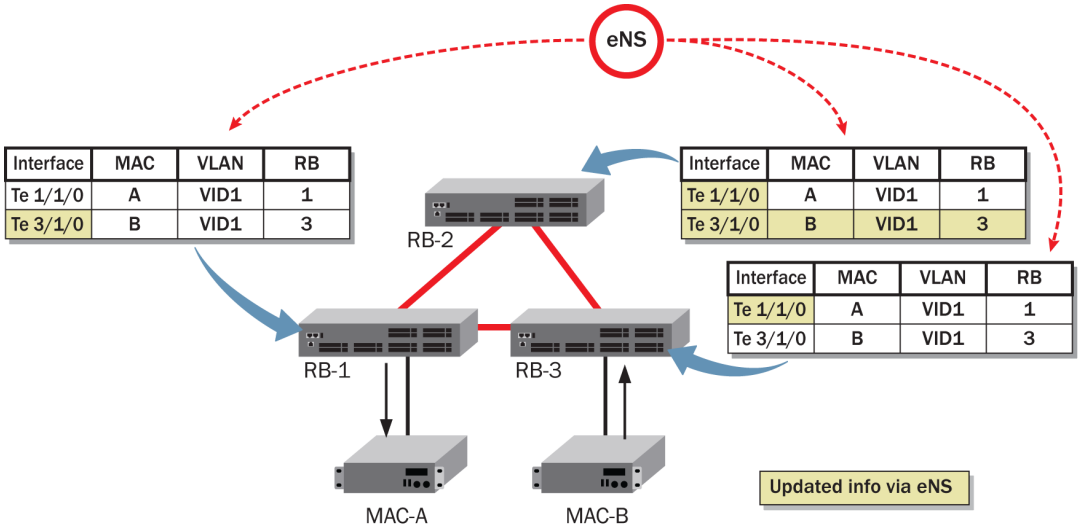

Figure 6. eNS synchronizing forwarding tables across fabric switches.

MAC Aging

MAC address aging works the same way as a standard IEEE 802.1Q switch (that is, addresses that are learned on an edge port are aged by the switch of the edge port, called the local RBridge). The local RBridge informs the eNS that it has aged out a MAC address from its forwarding table. The eNS informs all the other switches to remove the MAC address from their local forwarding tables.

eNS Support for MAC Mobility

Figure 7 shows MAC-A moving from RB-1 to RB-3, for example, when a virtual machine migrates from one server to another in a server cluster. The eNS ensures synchronization of all switch forwarding tables.

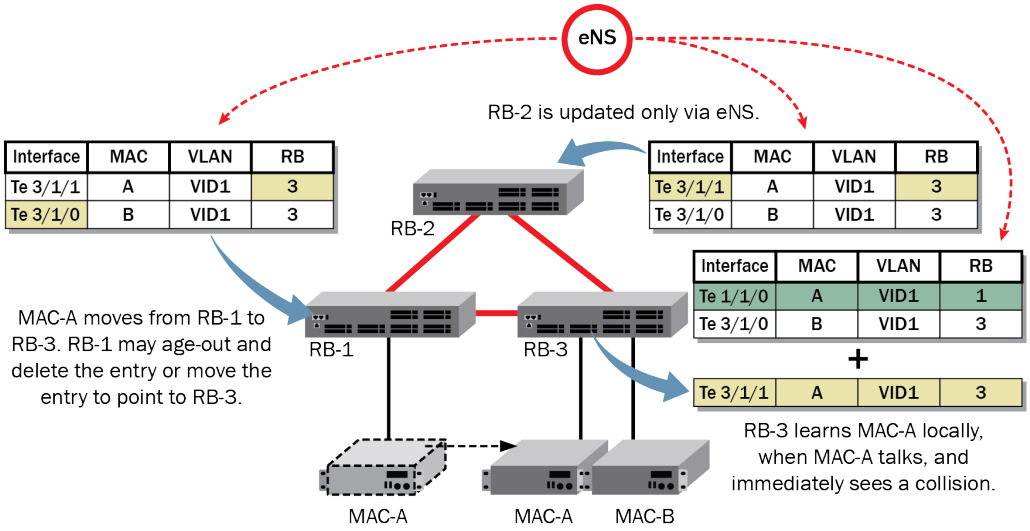

Figure 7. Distributed aging via eNS.

Root Bridge Selection for Multicast Tree

The VCS fabric defines the root bridge of the multicast tree, as follows:

- The root bridge of the distribution tree is the switch with the lowest RBridge ID. The selection process is automatic and does not require any user configuration.
- Each switch in the fabric also has a multicast root priority setting. This overrides the automatic selection of the multicast root based on the lowest RBridge ID. If the multicast root has to be a specific RBridge, the multicast root priority setting of that switch overrides root selection using the lowest RBridge ID. If two switches have the same multicast root priority value, then the RBridge with the lower RBridge ID is selected.
- An alternate multicast root is preselected as the switch with the next lowest RBridge ID. The alternate root bridge is automatically used by all RBridges, if the primary multicast root fails.

Broadcast, Unknown Unicast, Multicast Forwarding

All switches in a VCS fabric share a single multicast tree with the root bridge selected as described in Root Bridge Selection for Multicast Tree. All broadcast, unknown unicast, and multicast traffic that is received on any RBridge edge port is forwarded on the multicast tree, irrespective of the VLAN ID of the frame. The multicast tree includes a port for all RBridges in the fabric.

Following TRILL, the well-known multicast group MAC address is used in the Outer Header, and the multidestination bit is set. The frame is then forwarded along the multicast tree to all switches in the fabric.

For multicast and broadcast frames, if an RBridge has any edge ports in the same VLAN (shown in the Inner Header), which includes the multicast address learned from IGMP snooping, then the frame is forwarded on all appropriate edge ports with matching VLAN and multicast group membership.

IGMP Snooping

IGMP snooping supports dynamic multicast routers (m-routers) on a vLAG. M-router ports are learned dynamically via queries received on the edge ports. The edge port is automatically marked as an m-router port. Dynamic m-router ports are automatically removed if query messages are not received within the timeout period. IGMP snooping messages are sent based on the state of STP at the edge port, to optimize IGMP update processing.

VCS Fabric Services

The VCS fabric provides several services, such as the following:

- Interoperates with classic Ethernet switches running STP via a TLS
- Automatically reacts to the movement of MAC addresses across edge ports
- Provides link aggregation on edge ports to classic Ethernet switches
- Supports FCoE traffic across an arbitrary fabric topology

Transparent LAN Service

Similar to the Virtual Private LAN Service (VPLS), the TLS provides transparent Layer 2 connectivity services within the VCS fabric. Compared with VPLS, it offers improved provisioning and operational simplicity.

In Figure 8, a fabric of six switches is seen as a single Layer 2 switch by any switch or device that is attached to an edge port. Each VLAN at an edge port forms a single broadcast domain across the fabric. The fabric transparently forwards frames with a VLAN tag to edge ports that are configured for that VLAN. The VCS fabric of RBridges is transparent to any interconnected network of IEEE 802.1Q bridges running STP. The main difference between a network of 802.1Q switches and a VCS fabric of RBridges is the set of control path protocols controlling frame forwarding. In the case of 802.1Q, it is STP. In a VCS fabric, it is a link-state routing protocol for Layer 2 with specific enhancements to eliminate manual configuration of ISL and Brocade ISL Trunks.

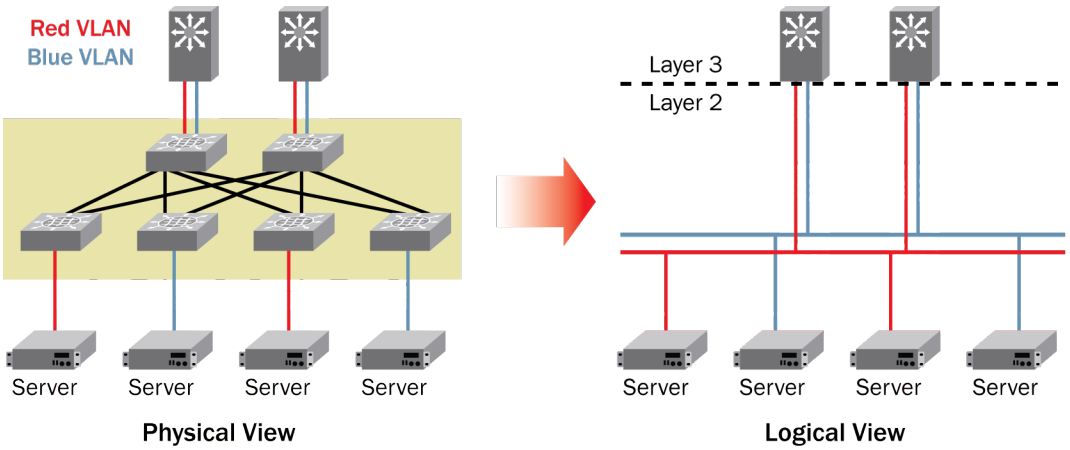

Figure 8. Transparent LAN services for multiple VLANs in a six-switch fabric.

Layer 2 Protocol Tunneling

The VCS fabric provides transparent forwarding of both unicast and multicast traffic. This presents a challenge for broadcast, unknown unicast, and multicast (BUM) traffic that is crossing the fabric when multiple classic Ethernet switches with STP are connected to edge ports. The VCS fabric provides TLS and is required to forward BUM traffic to all edge ports participating in a VLAN, to ensure that no loops form in this topology. Figure 10 shows a topology (physical and logical) with a VCS fabric (denoted by the dark grey block), with VCS Transparent LAN Service connected to classic Ethernet switches.

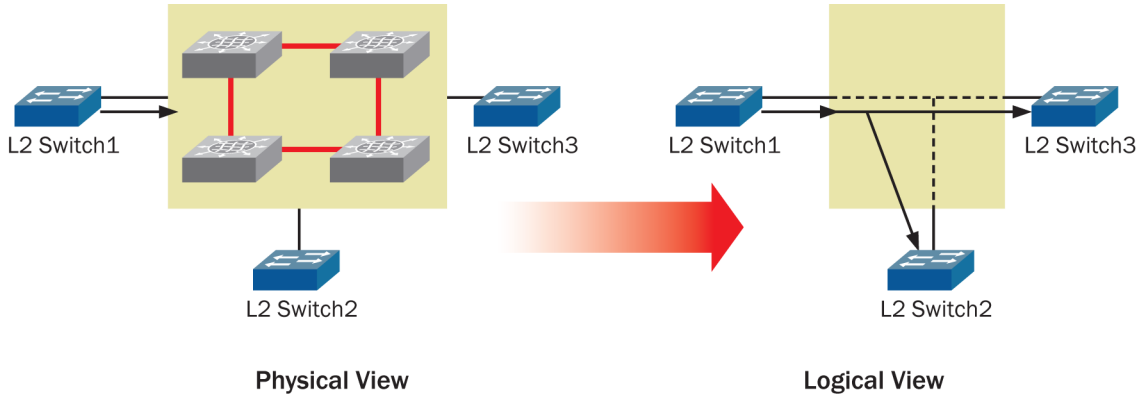

Figure 9. VCS fabric with classic Ethernet switches.

As shown in Figure 10, the VCS fabric logically becomes a “wire” from the perspective of the classic Ethernet switches. The RBridges are not affected by the BPDU frames. This allows classic Ethernet switches to control their port states based on the tunneled BPDU that they receive from any other classic Ethernet switch. Since the classic Ethernet switches see the spanning tree BPDU, they can break any loops between them, as shown on the right side of Figure 10.

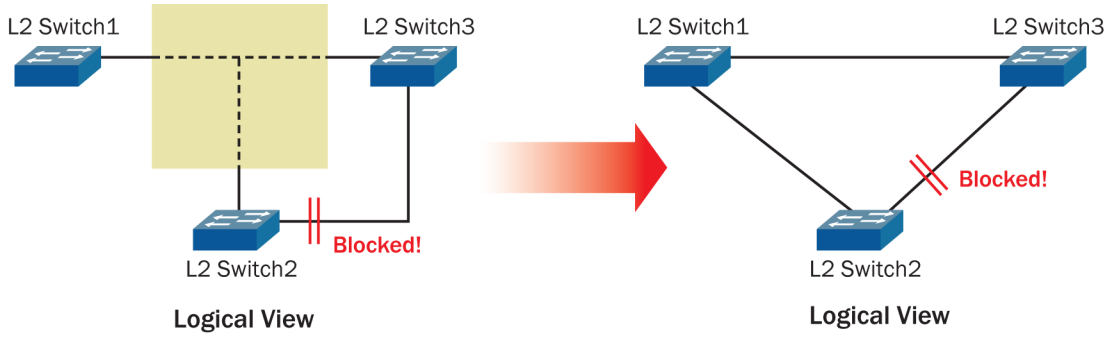

Figure 10. VCS Transparent LAN Service passes BPDU between classic switches to block loops.

Note that edge ports in a Brocade VDX switch do not process STP, because doing so would violate the nature of a TLS.

It is a best practice to connect a classic Ethernet switch that supports a multichassis LAG capability to the VCS fabric, in order to avoid any loops that are external to the VCS fabric.

Option to Turn Off BPDU Forwarding

As a user-configurable option, the edge ports on a Brocade VDX switch can be configured to block all BPDU traffic.

Standard Link-Level Protocol Support for Edge Ports

The following standard link-level protocols are supported on edge ports:

- LLDP (IEEE 802.1ab)
- LACP (IEEE 802.1ax)
- DCBX (IEEE 802.1Qaz)
- Port-based authentication (IEEE 802.1x)

Virtual Link Aggregation Group

Multi-Chassis Trunking (MCT) is an industry-accepted solution to avoid spanning tree on multiple ISL connections. LAG-based MCT is a special case of LAG, covered in IEEE 802.3ad, in which one end of a LAG can terminate on two separate switches. Virtual LAG (vLAG), an innovation that is included in Brocade VCS Fabric technology, extends the concept of LAG to include edge ports on multiple VCS Fabric switches. Edge ports in a vLAG support both classic Ethernet and DCB extensions. Therefore, any edge port can forward IP and FCoE traffic over a vLAG, as shown in Figure 11.

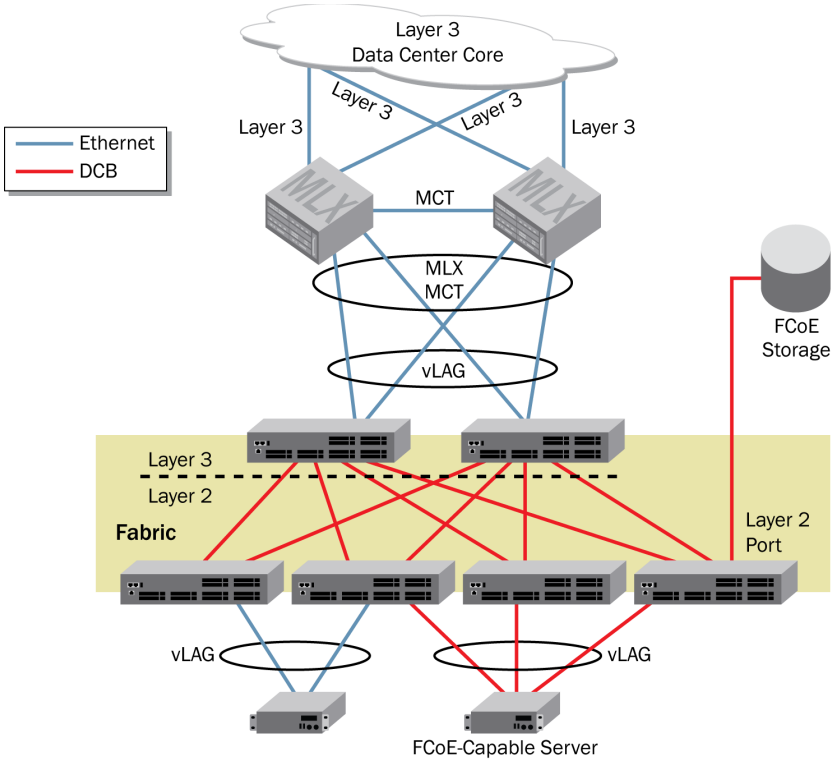

Figure 11. Using vLAG with VCS Fabric technology.

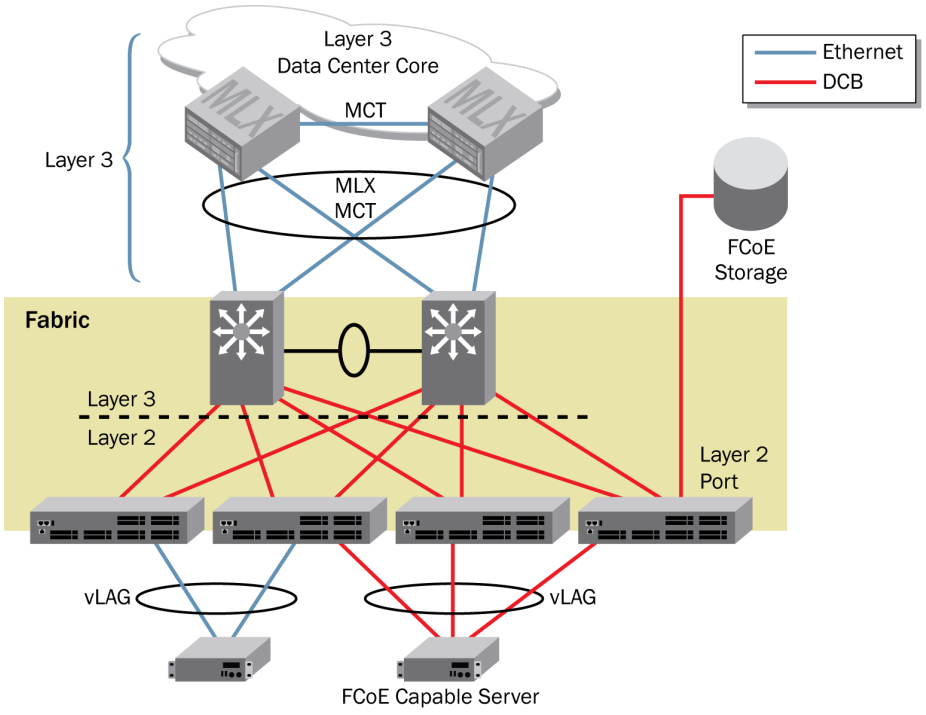

Figure 12. Using vLAG with Layer 3 services in a VCS fabric.

Note: In Figure 12 that both the Brocade VDX 6720 and VDX 6730 can also be used at the spine of the fabric, depending on the requirements.

The following are several important use cases for deploying vLAG.

vLAG to a Server

Servers with dual Network Interface Cards (NICs) that connect to two independent top-of-rack switches is an important use case, as shown in Figure 11. NIC teaming, commonly deployed in such scenarios, is subject to a significant limitation: lack of load balancing for traffic flowing from Top-of-Rack (ToR) to the server with dual active-active links.

Using vLAG between a server and switches in a VCS fabric provides these additional benefits over NIC teaming:

- Redundancy and load balancing based on IEEE 802.3ad LAG standards
- Deployment flexibility with the capability of including more than two ToRs within a LAG
- All MAC addresses behind the vLAG treated as multihomed to the VCS control plane, resulting in traffic being load-balanced across the TRILL network
- Adaptive load balancing in the TRILL network due to any vLAG port and switch membership change
- Auto-detection of improper configuration using LACP
- Adaptive load balancing from the server due to member port state change or MAC move.
- The server sends out a single IGMP join via vLAG. The VCS control plane adapts to the vLAG member change and delivers the multicast traffic through an available link in the vLAG, using IGMP snooping.

vLAG to a Server with Virtual Machines

An important use case is that of virtualized servers connecting to a VCS fabric via a hypervisor, since some hypervisors do not support LACP, so static LAG configuration at the server may be required. Virtual machine (VM) migration between servers that are connected to the same fabric is treated as a MAC address move from one vLAG logical interface to another vLAG logical interface. See the eNS Support for MAC Mobility section for further information.

vLAG to an External Switch

An IEEE 802.3ad standard-based LAG is formed when an external switch connects to any Brocade VDX switch in the VCS fabric via vLAG. All MAC addresses behind the vLAG are multihomed with active-active path protection.

vLAG to Switches Enabled with Multi-Chassis Trunking

An IEEE 802.3ad standard-based LAG is formed when an MCT-enabled external switch pair connects to any Brocade VDX switch in the VCS fabric via vLAG. All MAC addresses behind the vLAG are multihomed with active-active path protection.

vLAG Provisioning

Provisioning a vLAG is identical to provisioning LAG on a classic Ethernet switch. A channel group number is assigned to an interface to identify which vLAG the interface belongs to. Interfaces in a vLAG can include any edge port on any switch in the fabric. Consult the Brocade Network OS Release Notes for the maximum number of supported links in a vLAG and the number of vLAGs per VCS fabric.

vLAG Formation

The VCS distributed control plane provides LACP between participating switches when forming a vLAG. A vLAG membership set is represented in the form of {RB x.port y,...}. LACP uses the following information to form a vLAG:

LACP SYSTEM ID = Bridge MAC, Same Partner System MAC, Same Partner Operational Key, Same Partner Administrative Key
Any change to the partner attributes triggers removal of a member port from the vLAG.

vLAG Partner SID Validation

LACP Protocol Data Unit (PDU) frames are exchanged from a VCS-enabled switch and an end device (switch or host). Included in the exchange is a unique System ID (SID) for the vLAG. Any VCS-enabled switch participating in the vLAG

uses the same SID on its ports that participate in the vLAG. The remote device sends its own SID, which must be the same on all switches with ports participating in the vLAG. If the remote device SID received on a vLAG port is not the same, that port is not added to the vLAG.

vLAG Minimum Required Links

A further option is to require a minimum number of active links in the vLAG for it to form. If the minimum number is not available at any time, the vLAG does not form.

BUM Traffic on a vLAG

BUM traffic is carried on the primary interface of the vLAG. The primary interface can be a trunk with up to eight ports per trunk. The first operational interface of the vLAG becomes the primary link, and the RBridge with that interface becomes the primary RBridge. If the primary link goes down, and if there is another interface in the vLAG on the primary RBridge, then that interface becomes the primary link. If there are no other links on the primary RBridge in the vLAG, then a new RBridge and interface are selected as the primary RBridge and primary link.

Each switch with a link in a vLAG monitors the health of links in its vLAGs and notifies other switches in the fabric if any problem occurs with any of its vLAG links.

vLAG Interface

A vLAG is a port-channel interface and is identified with a port-channel number equal to the channel group number that is configured for its ports. A MAC is associated with each vLAG port-channel number. All protocols running over a vLAG use this MAC as the source address.

vLAG Learning and Aging

Fabric switches are aware of the vLAG RBridge ID, along with its member RBridge and port pair set, which are associated with a vLAG. A MAC address that is learned over a vLAG is treated as a multihomed MAC address and is reachable via all participating switches that have an interface in that vLAG. A MAC that is learned on a vLAG is only aged out if there is no data plane traffic on any interface in the vLAG.

MAC Movement with vLAG

A MAC address that is learned on a vLAG can move to another vLAG or to any another edge port in the fabric. The MAC move is detected by the new edge port and causes a fabric-wide update of the new MAC location for all switches in the fabric. See the eNS Support for MAC Mobility section for details.

vLAG and FCoE Traffic

The vLAG feature does not support FCoE traffic. However, FCoE can be enabled on the member links.

IGMP Snooping and Static Multicast MAC Addresses on a vLAG

An IGMP join on any of the vLAG interfaces results in the entire vLAG being one of the receivers of the multicast group. The multicast group membership is then distributed to all participating interfaces in the vLAG. The same holds true for static multicast MAC addresses that have a vLAG as one of the expansion ports.

vLAG Statistics

Statistics for vLAG are reported like regular LAG statistics.

Spanning Tree and vLAG

All spanning tree BPDUs are treated as data packets. A BPDU that is received on any vLAG interface is transparently forwarded as data through the fabric. When a BPDU arrives on a multicast tree in the fabric, the BPDU is transmitted on the vLAG primary link. As discussed earlier, an edge port can be configured to block BPDU traffic.

AMPP over vLAG

A vLAG can include a port profile port so that AMPP features are supported on a vLAG. See the Automated Migration of Port Profiles section for more about creating port profiles and assigning them to an edge port.

VM-Aware Network Automation

Server virtualization is used extensively in current data center environments. The server hosts, such as VMware ESXs, are connected directly to the physical switches through switch ports (or edge ports in the case of Brocade VCS Fabric technology). Many of these server hosts implement an internal switch called a vSwitch, which is created to

provide internal connectivity to the VMs. A new layer called the Virtual Access Layer (VAL) virtualizes connectivity between the physical switch and VM via a vSwitch. Virtual assets are not visible to the physical switch; thus, these VMs and other virtual assets remain hidden to the network administrator. Brocade VM-aware network automation provides the ability to discover these virtual assets.

With VM-aware network automation, a Brocade VDX switch can dynamically discover virtual assets and thus offer unprecedented visibility of dynamically created virtual assets. VM-aware network automation also allows network administrators to view these virtual assets using the Brocade Network OS CLI and the NETCONF interface. VM-aware network automation is supported on all Brocade VDX switch platforms.

Example: Data Center Network and vCenter

In a VMware-based environment, vCenter is primarily used to manage VMware ESX hosts. VMs are instantiated using the vSphere user interface. In addition to creating these VMs, the server administrator also associates these with a distributed Virtual Switch (dVS) and corresponding port group (PG). VMs and their network properties are primarily configured and managed on the vCenter/vSphere. Many of the VM properties, such as MAC addresses, are automatically configured by the vCenter, while some properties, such as VLAN and bandwidth, are assigned by vCenter through the VAL.

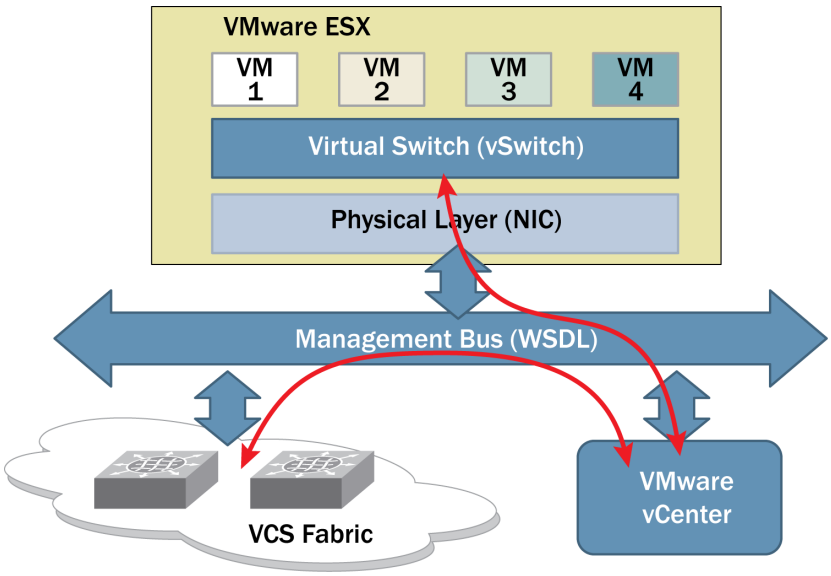

Figure 13: VM-aware network automation.

Brocade Network OS Virtual Asset Discovery Process

The Brocade switch connected to hosts/VMs needs to be aware of network policies in order to allow and disallow traffic. The discovery process starts upon boot-up. The switch is preconfigured with the relevant vCenter that exists in its environment. The discovery process entails making appropriate queries to the vCenter. The queries are in the form of Simple Object Access Protocol (SOAP) requests and responses sent to the vCenter.

The SOAP requests and responses conform to the vCenter SOAP-API. The procedure for moving Brocade VDX ports/vLAGs in profile mode is as follows:

- Each switch in the VCS fabric listens for CDP packets from ESX hosts on switch ports.
- Ports/LAGs/vLAGs are automatically put in port profile mode when the connected ESX host transmits the CDP.
- The port is removed from port profile mode in case the CDP is timed out on that port.

Authentication

Before any discovery transactions are initiated, the first order of transactions involves authentication with the vCenter. In order to authenticate with a specific vCenter, the following properties are configured at the switch: URL, login, and password. A new CLI is added to support this configuration.

Port Profile Management

After discovery, the switch/Brocade Network OS enters the port profile creation phase, where it creates port profiles on the switch, based on discovered distributed Virtual Switch port groups. The operation creates port profiles in the running configuration of the switch. Additionally, Brocade Network OS automatically creates the interface and VLAN that are configured in the port profiles, which end up in the running configuration.

The AMPP mechanism built into Brocade switches might provide a faster way to correlate the MAC address of a VM to the port with which it is associated. Brocade Network OS continues to allow this mechanism to learn the MAC address and associate the port profile with the port. The discovery process simply enhances this mechanism by providing automatic creation of port profiles or preassociating the MAC addresses before the VM is powered up. The VM-aware network automation feature manages the following manual configuration automatically (without VM-aware network automation, these steps need to be manually performed to create a port profile and associate a VM MAC address with it):

- Port profiles are automatically created for each distributed port group and standard port group.
- Port profiles are distributed across the VCS fabric.
- Necessary VLANs are created on all VCS Fabric members.
- vNIC MAC addresses are associated to the port profile.

Automatic Migration of Port Profiles

A hypervisor associates a Virtual Ethernet Bridge (VEB) port to each Ethernet MAC that is assigned to a VM. Some of the port attributes of the VEB or port profile include the following:

- The types of frames that are allowed on a port (whether all frames, only VLAN-tagged frames, or untagged frames)
- The VLAN identifiers that are allowed to be used on egress
- Rate-limiting attributes (such as port or access control-based rate limiting)

When a VM migrates from one physical server to another, the hypervisor ensures that the associated VEB port profile moves with it.

There is a gap between the network policy features of a classic Ethernet switch and a hypervisor VEB. The physical switch has advanced policy controls, compared to VEB implementations. These advanced policies are required in many environments. Furthermore, where policies are similar in the physical switch and the hypervisor VEB, it is desirable to have a single point of administration for uniform policies, rather than trying to coordinate between server and network administrators.

Classic Ethernet networks do not provide a mechanism for automatically migrating switch access and traffic control policies for an end device when that device migrates from one port to another. End-device migration can be physical, such as when an operating system image (application, middleware, operating system, and associated state) that is currently running on one system is moved to another system. Alternatively, the migration can be virtual, such as when an operating system image (OS image) that is currently running in a VM on one server moves to another.

The network needs a mechanism where a port profile that is resident in a switch and associated with an OS image moves between switch ports when the OS image moves from one server to another. AMPP provides that mechanism both in VCS and Classic modes.

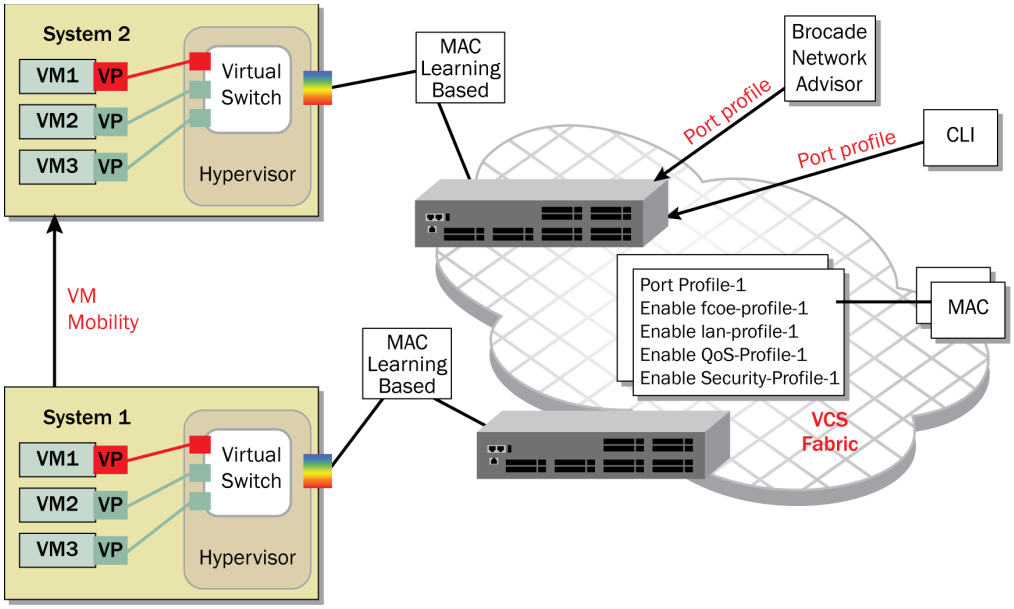

Figure 14. AMPP use with server virtualization.

Figure 14 illustrates how AMPP can be used to ensure that network policies that are set in the VCS fabric continue to be applied to VMs when they move. An AMPP profile is defined and bound to the MAC address of one or more VMs to which the policies in the profile apply. The AMPP profile is made available to all switches in the fabric using VCS Distributed Intelligence. When a VM moves, its MAC address moves from one edge port to another. See the eNS Support for MAC Mobility section for details on how MAC mobility is detected and switch forwarding tables are synchronized. The AMPP database can be distributed across switches. When a switch learns a MAC address, it can look up the associated AMPP entry and apply the network policies to the ingress VM traffic.

Port Profile Contents

A port profile is a container that can hold one or more classes of network policy, as shown in Figure 15. If a policy class does not apply, its profile is not defined in the port profile container.

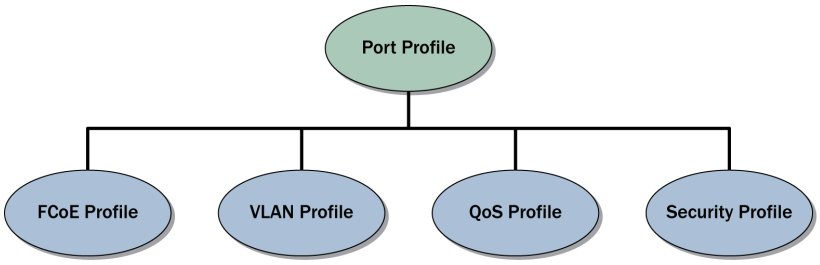

Figure 15. Port profile container showing policy classes.

A port profile container can include the following combinations of policy classes:

- Port profile for VLAN and quality of service (QoS) policy profiles
- Port profile for FCoE policy profile
- Port profile with FCoE, VLAN, and QoS policy profiles
- In addition, any of the above combinations can be mixed with a security policy profile.

A port profile does not contain some of the interface configuration attributes, including LLDP, SPAN, or LAG. These are associated only with the physical interface.

A port profile is a policy configuration container. When applied to a switch, it is sufficient to configure the global and local interface policies and allow the switch to forward traffic without further intervention. For details about configuring a port profile and configuring a port to be a port profile port with the CLI, refer to "Configuring AMPP" in the Brocade *Network OS Administrator's Guide*.

VLAN Policy Profile

The VLAN policy profile defines the following:

- VLAN membership, which includes the tagged VLANs and an untagged VLAN
- Protocol-based VLAN classification

QoS Policy Profile

A QoS policy profile defines the following:

- The incoming IEEE 802.1p priority to internal queue priority mapping. If a port is in QoS untrusted mode, all incoming priorities are mapped to the default best effort priority.
- The incoming priority to outgoing priority
- The mapping of incoming priorities to strict or Weighted Round Robin (WRR) traffic classes
- The enabling of flow control on a strict or WRR traffic class

The QoS profile can be one of two types: a Converged Enhanced Ethernet (CEE) and Data Center Bridging (DCB) or a standard Ethernet QoS policy.

FCoE Policy Profile

An FCoE policy profile defines all the attributes that are needed for a port to support FCoE traffic and defines the following:

- Fabric Map
- Fabric Map Priority
- Fabric Map Advertisement Interval
- Fabric Provided MAC Address (FPMA)
- FCoE Map Address Prefix (FCMAP)

Security Policy Profile

A security policy profile defines all the security rules that are needed for a switch port. However, the security rules can be different at different ports, so some of the locally configured ACLs are allowed to override any conflicting rules from a port profile. A typical security profile contains MAC-based standard and extended ACLs.

Editing a Port Profile

A port profile can be locked to prevent editing it. When a port profile is unlocked, any changes are immediately applied to the data plane.

Please note that the activation of a port profile is compulsory for the port profile to be applied to any port, making it a port profile port.

MAC to Port Profile Association Methods

Figure 16 shows MAC addresses that are associated with different port profiles. The port profile database is distributed to all switches in the VCS fabric. Any edge port can use a MAC address to locate the associated port profile and apply the profile policies within it to the ingress traffic.

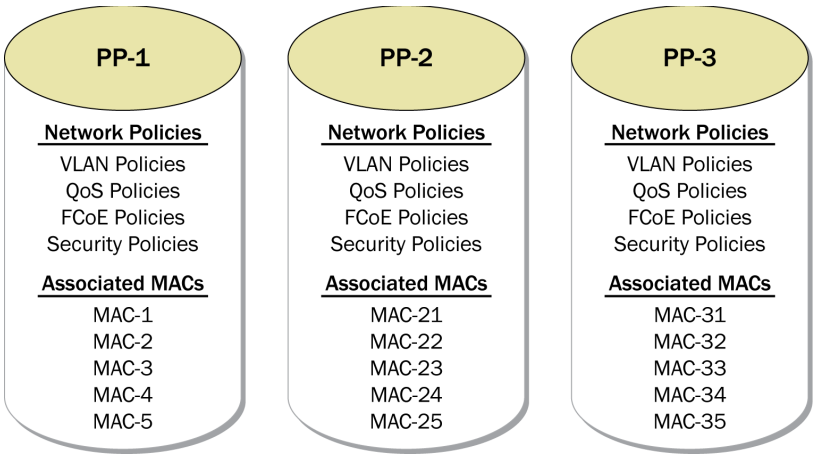

Figure 16. Associating port profiles with MAC addresses.

Association via MAC Learning

The CLI can be used to associate a MAC to a particular port profile. The application of the profile in hardware occurs when the MAC address is learned. If the MAC moves to another edge port, the port profile is consistently applied to traffic from that MAC.

Association via Management Interface

The CLI can be used to associate a MAC to a particular profile on a specific edge port. In this case, the profile is fixed to that edge port. If the MAC is seen on a different edge port, the port profile is not applied to that traffic, and the switch default forwarding behavior applies.

Automatic Association Using VM-Aware Network Automation

See the VM-Aware Network Automation section for details about this method of automatic creation and association of port profiles for VMware environments.

MAC to Port Profile Unbinding

There are several methods that are available to unbind a port profile from a MAC address. They are as follows:

Unbinding via Association Timeout

When a MAC to port profile association has been created based on MAC learning at an edge port, the association is removed via a mechanism similar to MAC aging. However, the timeout value is different than that used for MAC aging. Since a VM move could take a period of hours, the association timeout should be in the same order of magnitude. Optionally, an administrator can set the association timeout to zero, which makes the MAC to port profile association permanent.

Unbinding via Management Interface

If the MAC association to a port profile is created manually via the CLI, then the association is permanent until it is manually removed.

FCoE MAC Addresses and Port Profiles

FCoE uses two types of MAC addresses: the ENode MAC address and the fabric-assigned FPMA MAC address. Unlike the virtual NIC (vNIC) MAC addresses that are assigned per VM, the current FCoE implementations exist inside the hypervisor.

- The ENode MAC address is registered via DCBX, which is added to the allowed MAC addresses on a port.
- All FPMA addresses are assigned by the fabric and therefore allowed by default.
- By default, all FCoE traffic passes through the FCoE ACLs.
- All control plane protocols using the well-known destination MAC address are not affected by a port profile.

FCoE Port Profiles with LAG or vLAG

A port profile can contain both FCoE and LAN (VLAN and QoS) profiles. AMPP is designed to support FCoE on LAG or vLAG connections, but this also requires the server Converged Network Adapter (CNA) to provide support. Verify the Release Notes for a list of CNAs that support port profiles on LAG and vLAG connections.

Port Profile Port and SPAN

A port that is configured as a port profile port can be used with SPAN for traffic monitoring and diagnostics.

Directly Connecting VCS Fabrics Together

Connectivity between separate VCS fabrics is supported. However, due to the lack of loop detection, it is highly recommended that the following best practices are carefully followed.

- The topology must not have any loops. Any local loop within a single VCS fabric will cause broadcast storms and bring down the network. A local loop in one VCS fabric will also impact other VCS fabrics connected to it.
- Multiple VCS fabrics can be directly connected to each other, but the topology is restricted to leaf-spine. For example, a VCS fabric at the aggregation layer can connect to multiple VCS fabrics at the access layer. However, multiple aggregation layer VCS fabrics should not be connected to each other. Refer to the Brocade *Network OS Administrator's Guide* for scalability limits when connecting multiple VCS fabrics together.
- All links connecting two VCS fabrics *must* be a part of the same vLAG. This provides link resiliency and high availability without STP. Even though multiple 10 GbE links can be configured in the vLAG, all multicast control traffic and data traffic will be limited to a single 10 GbE link within the vLAG. This is due to the fact that multicast traffic is always sent out on the primary link of the vLAG. See the BUM Traffic on a vLAG section for more details.

Fibre Channel over Ethernet

Brocade VDX switches are convergence-ready. This means that they are capable of providing the lossless service that is expected of a network supporting Data Center Bridging (DCB). A VCS fabric can be used for a multiswitch, multihop fabric for FCoE traffic. That is, any FCoE initiator can communicate with any FCoE target across multiple switches (hops), as shown in Figure 17. And, with the Brocade VDX 6730, an FCoE device can communicate with a Fibre Channel device in a separate Fibre Channel fabric. See the Fiber Channel Device Connectivity section for more information about connecting FCoE devices in a VCS fabric to Fibre Channel devices in a Fibre Channel fabric.

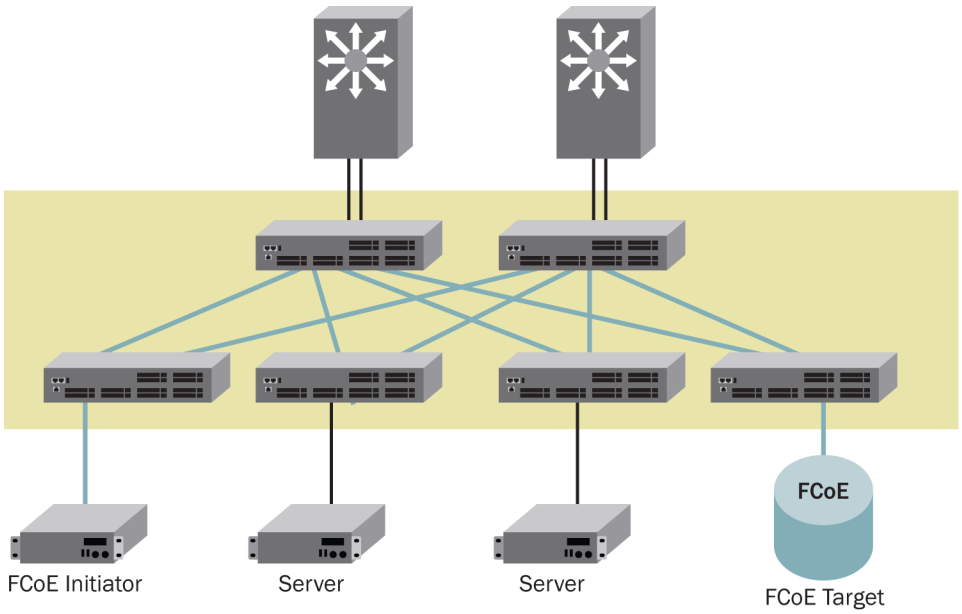

Figure 17. Brocade VCS Fabric multihop FCoE support.

FCoE Feature Support

ISL connections between switches form automatically in VCS mode. All information that is required to form lossless links, such as the PFC information and DCB-map, are sent over the ISL links. The following features are supported for FCoE traffic:

- FCoE Initialization Protocol (FIP) version 1 (FC-BB-5 rev 2.00). FIP version 0 is also supported.
- FC fabric services for FCoE VN_Port devices
- The FCoE VF_Port services provide access to FCoE VN_Port devices similar to those provided by FC F_Ports to FC N_Port devices.
- ISL port services
- ISL ports are formed between two switches in VCS Fabric mode. These carry data and control traffic between the switches.
- Multipath support for FCoE traffic. There can be multiple paths (ISL ports and/or Brocade ISL Trunks) between two switches, and ECMP will distribute the traffic on the available paths.
- Connectivity for FIP Snooping Bridge (FSB) Devices. This feature allows FSB-enabled devices such as blade switches or ToR, to connect to Brocade VDX switches. This feature supports FCoE traffic through non-VDX DCB switches to be connected between the Brocade VDX and CNA. This feature essentially allows multiple FCoE logins per physical interface VF_Port on Brocade VDX, support for dynamic binding between the Ethernet port and VF_Port, and LAG support for FCoE. In this scenario, the intermediate DCB switch aggregates flows from multiple CNAs. Refer to the Brocade *Network OS Administrator's Guide* and *Network OS Command Reference* and the Brocade Network OS Release Notes for additional details.

FCoE Operations

Each switch in the fabric can act as a full-function Fibre Channel Forwarder (FCF). All FC services to support a VN_Port run in every switch, and each switch in the fabric acts as if it were a separate domain in an FC SAN, as shown in Figure 18. A VCS fabric provides a network similar to a FC fabric from the perspective of the FCoE initiators and targets attached to the VCS fabric. Each switch is assigned a domain ID and, once the VCS fabric forms, all the FC services (such as Name Server, Login Controller, and Domain Controller) are available to each switch.

FCoE traffic forwarding across the fabric follows the same ECMP rules as described earlier for LAN traffic. See Equal-Cost Multipath Forwarding at Layer 2 or details.

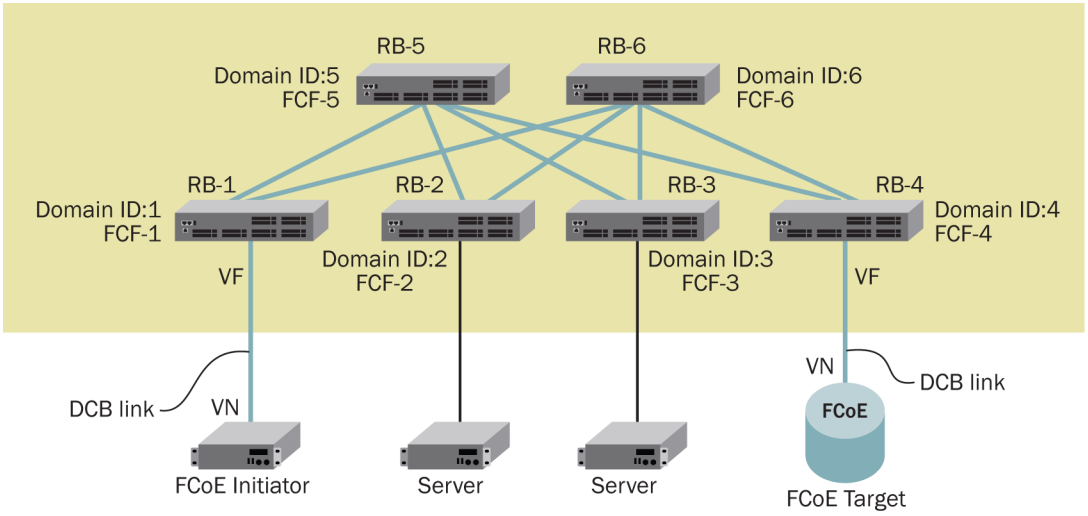

Figure 18. VCS fabric and FCoE traffic.

FCoE Frame Forwarding

FCoE data frame forwarding between two FCoE devices attached to a VCS fabric uses link-state routing and encapsulates the FCoE data frame inside a TRILL frame. See TRILL Frames for more details. The TRILL header specifies the source and destination RBridge, while the Outer Header is modified hop-by-hop until the frame reaches its final destination.

Fibre Channel R_A_TOV and E_D_TOV

The resource allocation timeout value (R_A_TOV) and error detect timeout value (E_D_TOV) implementations for the VCS fabric are the same as those used in an FC fabric. For example, the VCS fabric R_A_TOV is implemented by enforcing a maximum transit delay in each switch (similar to an FC switch). Based on the ANSI FC-BB-5 standard, the maximum transit delay per switch is no more than 500 ms. The default value of R_A_TOV is 10 seconds.

Maximum Switch Hops for FCoE Traffic

Refer to the Release Notes to determine the tested and supported limit for a specific Brocade Network OS firmware release.

Lossless Forwarding for Class F Traffic and FCoE Data Traffic

The Brocade VDX 6720, 6730, and 8770 switches provide lossless links for Fibre Channel Class F traffic (control traffic) and the FCoE data traffic. (The VDX 6730 also supports native Fibre Channel interfaces.) All the control traffic between switches is assigned to Priority 7. PFC is enabled for Priority 7 and also enabled for the priority that is assigned to FCoE traffic.

DCB Layer 2 Configurations Required for FCoE

FCF MAC addresses

A single FCF MAC address is used per switch.

FCoE VLANs

The FCoE VLAN is configurable through the CLI. The fabric-map context is edited to set the desired VLAN value. Refer to the Release Notes for the number of VLANs that are supported for FCoE traffic.

DCBX

The DCB parameters that are required for FCoE traffic must be negotiated before the FCoE traffic can start flowing on an edge port. This is managed by the DCBX protocol. An FCoE application—Type, Length, Value (TLV)—is exchanged using the DCBX protocol. Switches can discover and negotiate parameters such as FCoE priority, priority group information, and so forth. A successful exchange of the TLV allows the logical link to initialize.

Default Values

When a switch boots up, several FCoE settings are preconfigured to default values. They are as follows:

- FCoE VLAN is set to 1002.
- PFC-enabled priority for FCoE is set to 3.
- Default VFID is set to 1.
- FCMAP is set to 0x0EFC00,
- The FIP Keep Alive (FKA) is disabled. This means that the devices will not be logged out, irrespective of whether FKA frames are received from those devices.

Spanning Tree Restrictions

Spanning tree must not be enabled on any direct server connections to the front-end to an edge port that can pass FCoE traffic. If spanning tree is enabled, this may result in logins being lost or dropped.

FCoE Control and Data Traffic Flows

FCoE F_Port Services and Traffic Flows

FCoE fabric services are switch-resident. With the exception of fabric discovery, FCoE VN_Port device access to fabric services is similar to FC N_Port access to the same services in an FC fabric. The FCoE VN_Port device sends all the traffic that is destined to the FC Well-Known Addresses (WKAs) to the FCoE edge port.

Unlike FC N_Port devices that automatically know how to reach their F_Port interface (because they are directly attached to it), FCoE ENode (VN_Port) devices need to learn the Ethernet MAC address of their FCoE edge port interface or interfaces. Multiple intermediate switch hops may exist between the FCoE VN_Port device and the edge ports in an FCoE-aware switch. FIP discovery protocol is used to discover the edge port.

FCoE VN_Port traffic that is destined to another FCoE VN_Port device is sent via the FCoE edge port. The hardware looks at the FC frame header in the FCoE frame and makes forwarding decisions based on its tables. When the decision is made, it sends the frame to the appropriate egress edge port.

FCoE Discovery

FCoE ENodes learn the Ethernet MAC addresses of reachable FCF-capable FCoE devices in the fabric through the FIP.

The FIP discovery mechanism consists of solicitations and advertisements. The ENodes send solicitation messages to a well-known Ethernet multicast address. All the reachable FCFs receive those solicitations. If the service provided by the FCF matches the service requested by the ENode, the FCF responds by sending a unicast advertisement message to the ENode.

Fabric Login

All FCoE VN_Port devices must log in to the fabric. FCoE login is accomplished by the VN_Port devices by encapsulating the standard FC login frames (FLOGI, FDISC) in a FIP frame with Ethernet headers and sending them to an FCF MAC address that was learned during the FIP discovery phase. The FIP encapsulated login frame takes the same path in the FCoE subsystem as the FIP discovery frame.

FCoE VN_Port to FCoE VN_Port Traffic

FCoE VN_Port devices communicate with each other via the FCoE edge port. This ensures that all traffic between the FCoE VN_Port devices is intercepted, examined, and authenticated by the fabric.

FC Cyclic Redundancy Check

The FC Cyclic Redundancy Check (CRC) is considered to be a part of the FC frame during encapsulation. It is maintained though the FCoE and FCoE/FC data paths and modified only if the FC frame is modified.

FCoE and VCS Fabric Ports

A port may be used as an edge port (FCoE VF_Port) or a fabric port (FCoE VE_Port).

To configure a port as an FCoE VF_Port, the following steps are needed:

- Use the **interface fcoe x/y/z** command, where x, y, and z refer to the switch, slot, and port of the Logical Chassis.
- Then use the **no shut** command to activate the port.

If the device that is connected to this edge port has a CNA, it can log in as long as the other properties such as fabric map, DCB map, and so forth have also been defined.

Issuing the **shut** command on the edge port will force the CNA to log out.

For switch ports that are operating as fabric ports, issuing the command **fabric isl enable/disable** has no effect on the CNA.

Fibre Channel Device Connectivity

This section describes how connectivity can be established between FCoE initiators (servers) connected to a Brocade VDX 6720, VDX 6730, or VDX 8770 switch and storage devices connected to a Fibre Channel fabric. The Brocade VDX 6730 is required to create the gateway to the Fibre Channel fabric.

A Brocade VDX 6730 can connect to an EX_Port on an 8 Gbps or a 16 Gbps Fibre Channel platform running Brocade Fabric OS (FOS) v7.0.1 or later. Typically, a Brocade VDX 6730 switch connects a VCS fabric as an edge fabric to a Fibre Channel routing backbone fabric. In turn, the Fibre Channel backbone fabric connects to one or more Fibre Channel edge fabrics. This topology is commonly referred to as an edge-to-edge topology, in which Fibre Channel devices are not directly connected to the backbone fabric. Figures 19 and 20 show options for an edge-edge topology with the Brocade VDX 6730 in a VCS fabric to existing Fibre Channel fabrics.

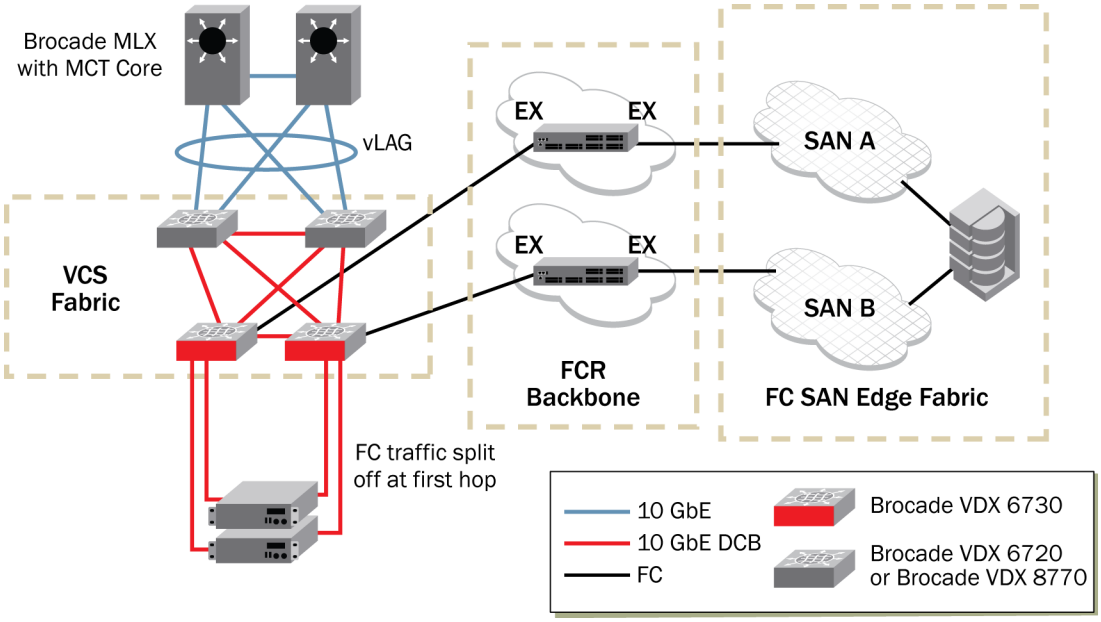

Figure 19. Fibre Channel traffic split at Top of Rack.

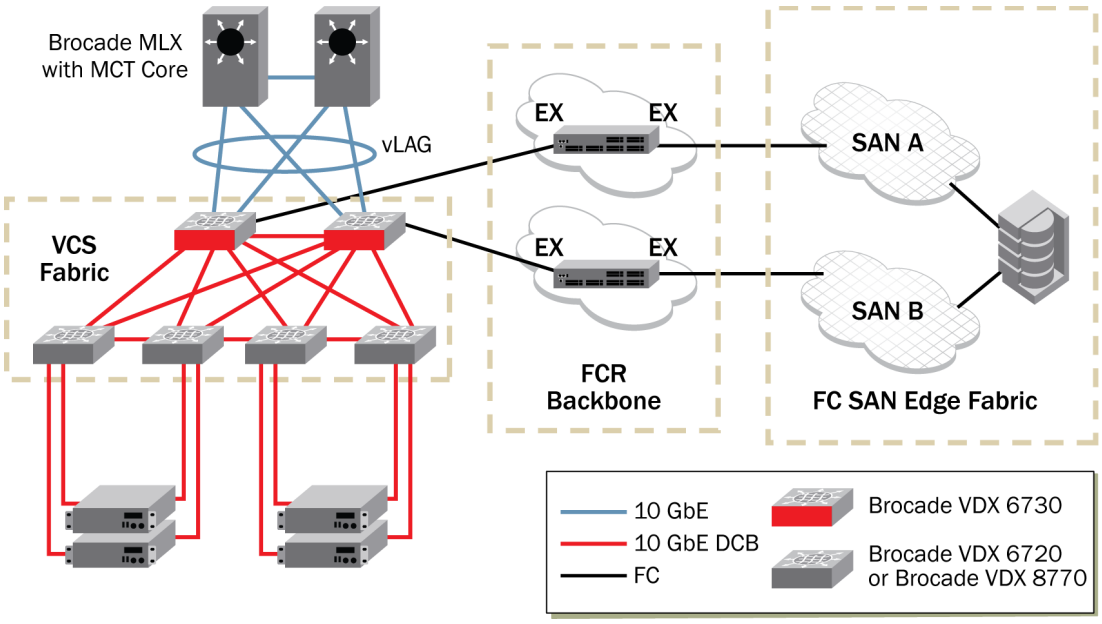

Figure 20. Fibre Channel split at the edge of the Brocade VCS fabric.

It is also possible to use a Brocade VDX 6730 to connect a VCS fabric directly to backbone fabric that contains Fibre Channel devices. This topology is commonly referred to as an edge-to-backbone topology. There are constraints on this configuration. Refer to the Brocade FOS scalability section in the Brocade FOS Release Notes for further guidance. Figure 21 shows a typical edge-backbone configuration, where Fibre Channel devices are directly connected to the backbone fabric. Although the figure shows only Brocade VDX 6730 switches, the Brocade VDX 6730 could be connected to other Brocade VDX switches in the VCS fabric, as shown in Figures 19 and 20.

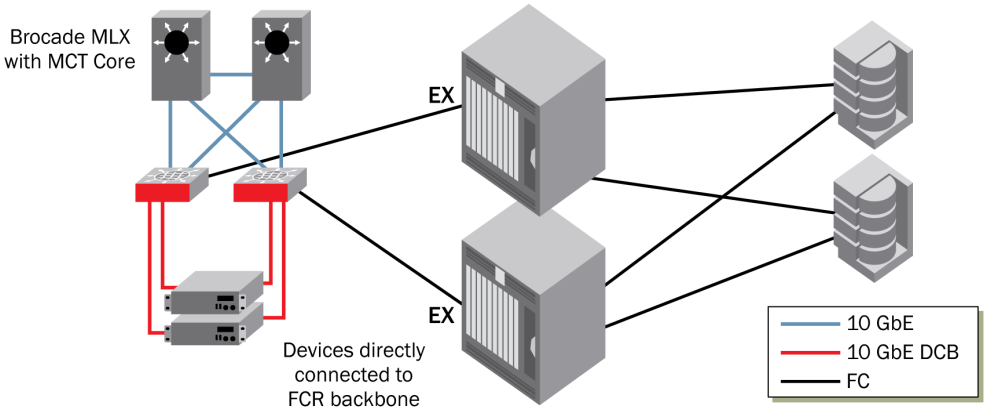

Figure 21. Connectivity to FCR backbone with devices.

Please refer to the Brocade *Fabric OS Administrator's Guide* for details about Fibre Channel routing and configuring Logical SANs.

Fibre Channel WWN Zoning

Zoning support can be used to create Fibre Channel zones within a Brocade VCS fabric and to create LSAN zones when configuring Fibre Channel routing. WWN zoning also provides Registered State Change Notification (RSCN) storm suppression, seen in Brocade Network OS 2.0.0 for native FCoE devices within VCS fabrics. CLIs support zoning

database queries, zone transaction display, zone entry management (creation, add, delete, remove, save, disable and enable), default zone access mode, and ability to abort transactions. Both distributed service mode and management cluster mode are supported. Zoning is not supported in standalone mode. Any residual Zone database is cleared and disabled when going from supported to unsupported mode. Zoning is supported in both FC with distributed service enabled mode and management cluster mode.

RBridge Reboot

When a principal RBridge reboots, a new principal RBridge is selected, and an active transaction is aborted. As a part of the transaction abort, "show running-config zoning" reflects the previously committed Zone database. When a non-principal RBridge reboots (or goes down), any changes initiated by the RBridge are still reflected as part of an active transaction maintained by the principal RBridge, thus maintained by the fabric.

Global Transaction Support

Zoning is implemented as a global transaction, where execution of the CLI is performed and serialized at the principal RBridge regardless of where the CLI is initiated. When a first edit zone CLI—such as create, add, or remove—is received at the principal RBridge, a transaction is opened, and any subsequent edit is also received at the principal RBridge, regardless of where it originated, and is subsequently added to the transaction. Once the user issues a commit (such as save, enable, or disable), the transaction is committed, regardless of where the commit originated. Abort also follows the same transaction model.

As edit zone CLIs are executed, the change is distributed and reflected in the database of all RBridges. Therefore, changes triggered by zone CLIs being executed in parallel on different RBridges are reflected on all RBridges within the fabric, as they happen. However, a commit action is still required to commit the change to the Zone database, which keeps a single master copy.

VCS Fabric Layer 3

Brocade VCS Fabric technology provides a new level of bandwidth scale with the introduction of the Brocade VDX 8770 modular chassis. Brocade Network OS 3.0 provides greater design flexibility with new, enhanced Layer 3 capabilities which further improve performance by clustering bandwidth and using it efficiently through multipathing.

Routing can be enabled on all individual RBridges in the VCS fabric. Each RBridge that has Layer 3 enabled on it within the VCS fabric is an independent router. From a routing perspective, all routers enabled on a given VLAN appear as if they are on the same LAN. This is because TRILL provides a flat Layer 2 topology.

Layer 3 in VCS fabrics is delivered in both the fixed configuration and modular Brocade VDX platforms. Layer 3 support in the VCS fabric provides the following:

- Inter-VLAN routing in the VCS fabric
- Router port to VE routing capabilities in the VCS fabric
- Router port to Router port routing in the VCS fabric

Note: The fixed-configuration platform (Brocade VDX 6700 Series) and modular chassis (Brocade VDX 8770) have different Layer 3 scalability numbers. Please be aware that if there is a hybrid VCS fabric with Layer 3 enabled on both platforms, then the scalability will be limited to the fixed configuration Brocade VDX 67xx platform Layer 3 scalable numbers. If scale is a primary design concern, you should consider a multifabric implementation with separate Brocade 6700 and 8770 fabrics.

A significant benefit of Layer 3 routing in VCS is the simplification of the topology from the traditional three-tier structure (access-aggregation-core) to a two-tier (VCS fabric-core). Additionally, Brocade VCS Fabric technology with Layer 3 routing takes advantages of all the existing benefits of the Layer 2 Brocade VCS Fabric solution and provides simplified configuration and management.

A traditional data center uses the three-tier architecture deploying the typical access-aggregation-core topology. With the introduction of VCS Fabric technology, Brocade has transformed the access layer into a single logical Layer 2 switch. The addition of routing to the VCS Fabric technology is the next logical step in the evolution of VCS fabrics. Adding routing capabilities has effectively flattened the traditional access and aggregation layers into a single layer. In other words, the traditional three-tier architecture is now flattened into two layers.

Figure 22 shows the topology of a traditional three-tier data center that can be built with Brocade products.

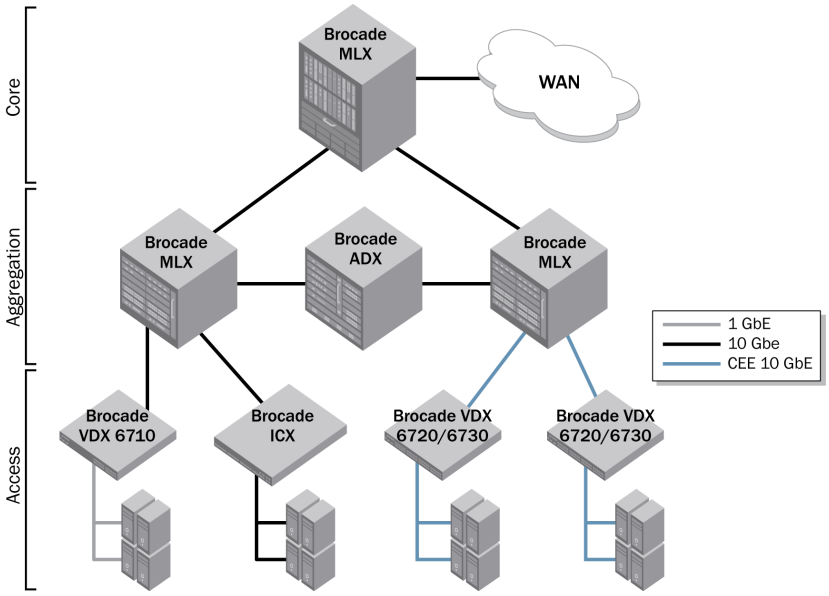

Figure 22. Brocade traditional data center topology.

Figure 23 shows how the introduction of Layer 3 services in Brocade Network OS 3.0 has effectively flattened the data center and simplified the topology from three tiers to two tiers.

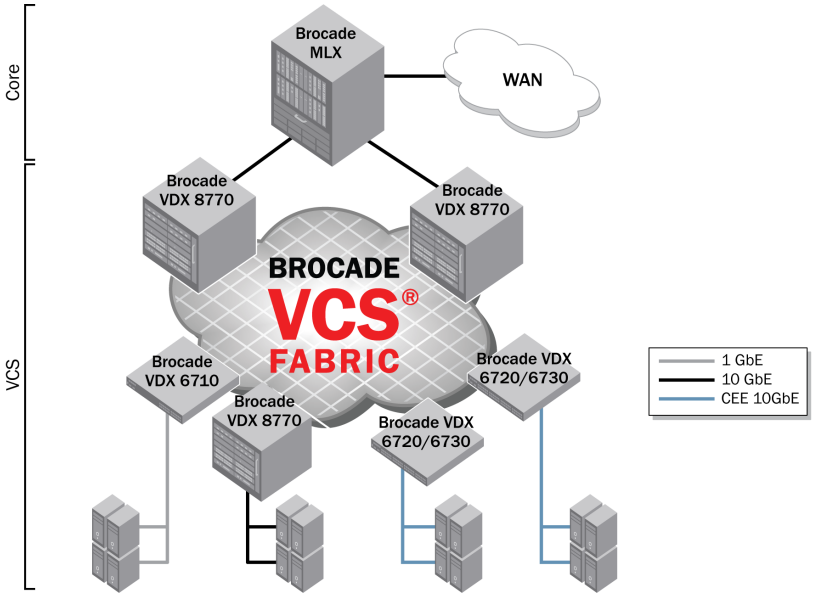

Figure 23. VCS fabric with Layer 3.

Brocade VCS Fabric Technology with Layer 3 services uses TRILL for its inter-switch connectivity and delivers a transparent LAN service. In keeping with the original goal of a highly available and highly redundant control plane for VCS fabrics, routing is enabled on an individual RBridge. In other words, each RBridge that has Layer 3 enabled on it within the VCS fabric is an independent router. From a routing perspective, all RBridges—irrespective of the physical connectivity and the number of Layer 2 hops between them—appear as if they are all on the same LAN. This ability to translate any physical topology into a single LAN is the key enabler of the many unique differentiating features for routing in the VCS fabric.

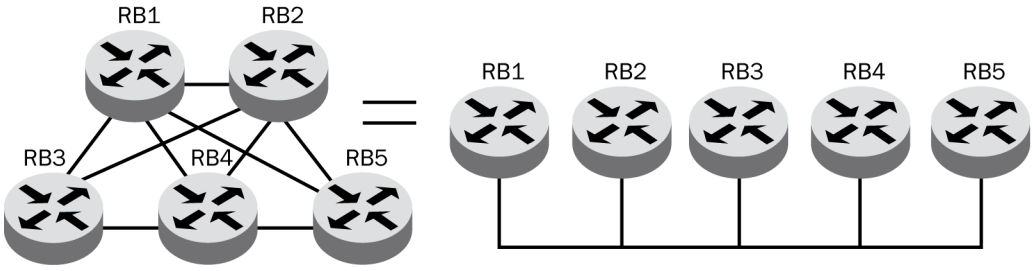

Figure 24. (TLS) Layer 2 topology: All routers participating in the same VLAN.

Layer 3 Features in Brocade VCS Fabrics

The following Layer 3 features are supported in Brocade Network OS 3.0:

- IPv4 unicast routing
- IP static routes
- Open Shortest Path First (OSPF)
- Virtual Router Redundancy Protocol (VRRP)
- Brocade Virtual Router Redundancy Protocol-Extended (VRRP-E)
- Route-maps
- Prefix lists
- Address Resolution Protocol (ARP)
- IGMP snooping
- IPv4 services/inband management
- Layer 3 Access Control Lists (ACLs) (ingress and egress)
- Routing on VE, router ports (no support for vLAGs)

Configuration Model

Since the routing control plane is fully distributed, and each router is a unique entity for a given RBridge within the VCS fabric, routing can potentially be enabled on every single RBridge within a VCS fabric.

This architecture gives the user the ability to deploy routing in the VCS fabric in different ways. We envision the following two models as the most commonly deployed routing scenarios.

Centralized Routing

In this mode, it is assumed that the majority of the traffic for the end user is the traditional north-south traffic, that is, extending from the access layer towards the core.

- To get optimal results, it is expected that the user will typically enable routing on a single RBridge within the VCS fabric—preferably towards the core.
- The key requirement here is that users need to ensure that they have Layer 2 connectivity for all routed traffic all the way to this RBridge. In other words, the user has to ensure that any VLAN on which the user wants to enable

routing is configured on every single RBridge within the VCS fabric. Once users have ensured seamless Layer 2 connectivity, they can then enable routing for that VLAN.

- On all the other switches connectivity is achieved through TRILL. At the ingress of the VCS fabric, the native Ethernet packet is encapsulated in a TRILL frame and is Layer 2 forwarded to the routing RBridge, where it is decapsulated, routed, and re-encapsulated to get forwarded to its Layer 3 next-hop. See Forwarding Architecture for a detailed explanation of Layer 3 forwarding.

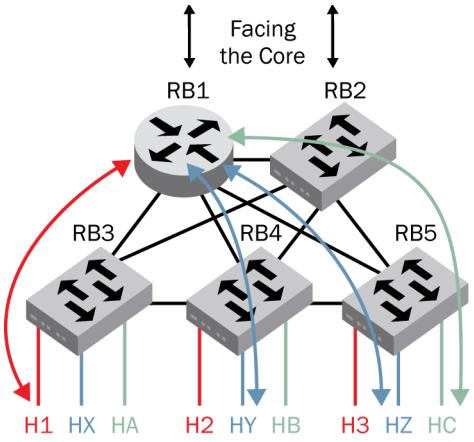

Figure 25. Centralized routing.

Figure 25 shows a five-node VCS fabric:

- Routing is turned on RB1.
- If Host H1 on the red VLAN needs to communicate with the core, traffic gets switched all the way to RB1 and then routed from there towards the core.
- Similarly, if Host HC on the green VLAN wants to communicate with Host HZ on the blue VLAN, even though they both are hanging off the same RBridge RB5, their traffic gets switched all the way to RB1 and then routed back to RB5 with the necessary next-hop changes.

Distributed Routing

In this model, routing can be turned on any and every RBridge within the VCS fabric. Please note this model simply illustrates that the architecture provides the flexibility to configure routing on any and every RBridge, depending on the specific requirements. This deployment model looks very similar to a traditional aggregation layer in legacy data centers and has the following key benefits:

- This approach scales very well with respect to the control plane as it is truly distributed.
- ARP is maintained only for locally attached VLANs.
- Simply enabling routing in the core, as in the traditional three-tier model, results in suboptimal Layer 2 forwarding, since traffic must reach the routed RBridge even if you have paths to locally route the traffic. The model described here allows the user to envision a more optimal routing topology.
- Similar to the traditional three-tier approach, the key requirement here is that users ensure that they have Layer 2 connectivity for all routed traffic all the way to this RBridge.

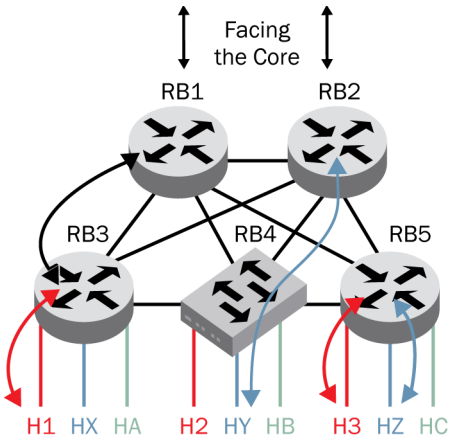

Figure 26. Distributed routing on multiple RBridges.

Figure 26 shows a five-node VCS fabric:

- Routing is turned on RB1, RB2, RB3 and RB5. (This is done to simply illustrate the flexibility of the VCS Fabric architecture. As described previously, routing can be turned on or off depending on the requirements.)
- If Host H1 on the red VLAN needs to communicate with the core, traffic will get routed on RB3, sent to RB1, and routed again towards the core. Essentially the packet gets routed as many times as there are VLAN boundary changes between the host and the core.
- If Host HZ on blue VLAN wants to communicate with Host H3 on red VLAN, there is a more optimal solution. Since RB5 is a router, it can locally route the traffic between the two hosts.
- Another benefit is that RB5 can limit the ARP broadcast domain to itself when using the BUM suppression feature. In this case, there is no need for the other routers RB1 and RB3 to learn about the ARPs for H3 and HZ. With proxy ARP turned on, they simply need to learn the ARP for RB5.
- If the same VLAN spans more than one RBridge, we recommend that the user enable VRRP on that VLAN to elect a master for that subnet. This requirement is needed because it provides a deterministic and redundant gateway for the hosts to reach the outside world.

In either model, enabling routing in Brocade VCS fabrics maintains all the benefits of Layer 2 VCS fabrics, including these:

- **Zero Configuration:** Automatic detection and configuration of the fabric
- **Zero Assumption:** Support for any arbitrary topology
- **Multipath:** Layer 2 ECMP to utilize Layer 2 bandwidth more effectively
- **Efficient Paths:** Layer 3 ECMP to optimally utilize the bandwidth of the underlying fabric

Configuration Modes

Brocade VCS Fabric technology has two configuration modes within the VCS fabric.

Global Configuration Mode

This set of commands has a global scope and are fabric-wide. Today, in Brocade Network OS 3.0, global configuration commands are required to be entered on every single switch in the VCS fabric. A typical example of this command would be the creation of a Layer 2 VLAN. For the complete list of global configuration commands please see the Brocade *Network OS Command Reference*.

Physical Interface Configuration
This set of commands has an interface local scope. These commands are typically those that get executed on a particular RBridge for a physical interface, for example, setting an MTU for a physical interface.

RBridge Configuration
Enabling routing in VCS fabrics has introduced an intermediate configuration type—the RBridge Level Configuration.

Any set of commands that are scoped to a particular RBridge fall under this category. This configuration model is used to enable routing. As mentioned before, all Layer 3 commands fall into this category. A typical example of this command would be configuration of an IP address on a VLAN. The complete list of RBridge level commands is beyond the scope of this document. *The exception to this rule is the configuration of Layer 3 commands on a physical interface. They remain under the physical interface scope.*

Enabling Routing
By default, Brocade Network OS 3.0 comes with the Layer 3 capabilities described here. When an IP address is enabled on an interface, we consider that interface to have Layer 3 enabled. There is no specific command to enable routing in the software. The software also has the flexibility of allowing out of order configurations; in other words, the user is allowed to configure protocol configuration without the IP address being programmed on that particular interface. The protocol configuration, although accepted by the software, remains in an inactive state until the IP address is configured, and then it gets activated.

The following interfaces support routing.

VE Interface
Brocade is introducing the concept of a logical Layer 3 interface, called the VE interface. The VE interface is a Layer 3 interface on which an IP address or multiple IP addresses are configured. The "ID" of the VE interface for Brocade Network OS 3.0 can map from 1 through 4094. This ID only has a local significance to a given router. For Brocade Network OS 3.0, the ID of the VE is tightly coupled with the Layer 2VLAN ID. In other words, the VE interface inherits the Layer 2 port hierarchy of the Layer 2 VLAN.

The VE interface provides the user the flexibility to build a hierarchy of any grouping of Layer 2 interfaces underneath the Layer 3 interface.

The VE interface needs to be created like any other interface in the RBridge mode. For Brocade Network OS 3.0, since the VE interface relies on the Layer 2 VLAN for its hierarchy, the following behavior ties the two together: The VE interface cannot be created if the underlying Layer 2 VLAN is not present.

Physical Interface, or Router Port
Brocade VCS Fabric technology supports the configuration of an IP address on a physical interface. The only requirement from the end user is to make this interface reachable to the rest of the VCS fabric. To prevent the logical isolation of this physical interface, it is necessary for at least one VE to be enabled on the RBridge that has one or more router ports enabled. For a multinode VCS fabric, a VE interface is required between the router ports.

LAGs
Brocade Network OS 3.0 will not support the configuration of an IP address on a LAG interface.

Loopback Interfaces
Loopback interfaces are logical interfaces. The primary use case for a loopback interface is to be used as the "router-id" or termination address for a particular protocol like OSPF or BGP. Since this interface never goes "down," this gives additional robustness for the protocol, as it is always reachable, even if the physical interfaces on the routers cannot be reached.

Protocol commands typically have an RBridge level scope as well as an interface level scope. For a detailed discussion on Layer 3 routing in VCS fabrics, please see the Brocade *Network OS Administrator's Guide* and *Network OS Command Reference*.

Forwarding Architecture

Layer 3 forwarding in Brocade VCS fabrics relies on the capabilities of Brocade custom-built ASICs to decapsulate, route, and re-encapsulate Layer 3 traffic on a TRILL network.

Figure 27 depicts a typical operation on the *header of a Layer 3 packet* as it enters the VCS fabric, gets routed, and eventually exits the VCS fabric.

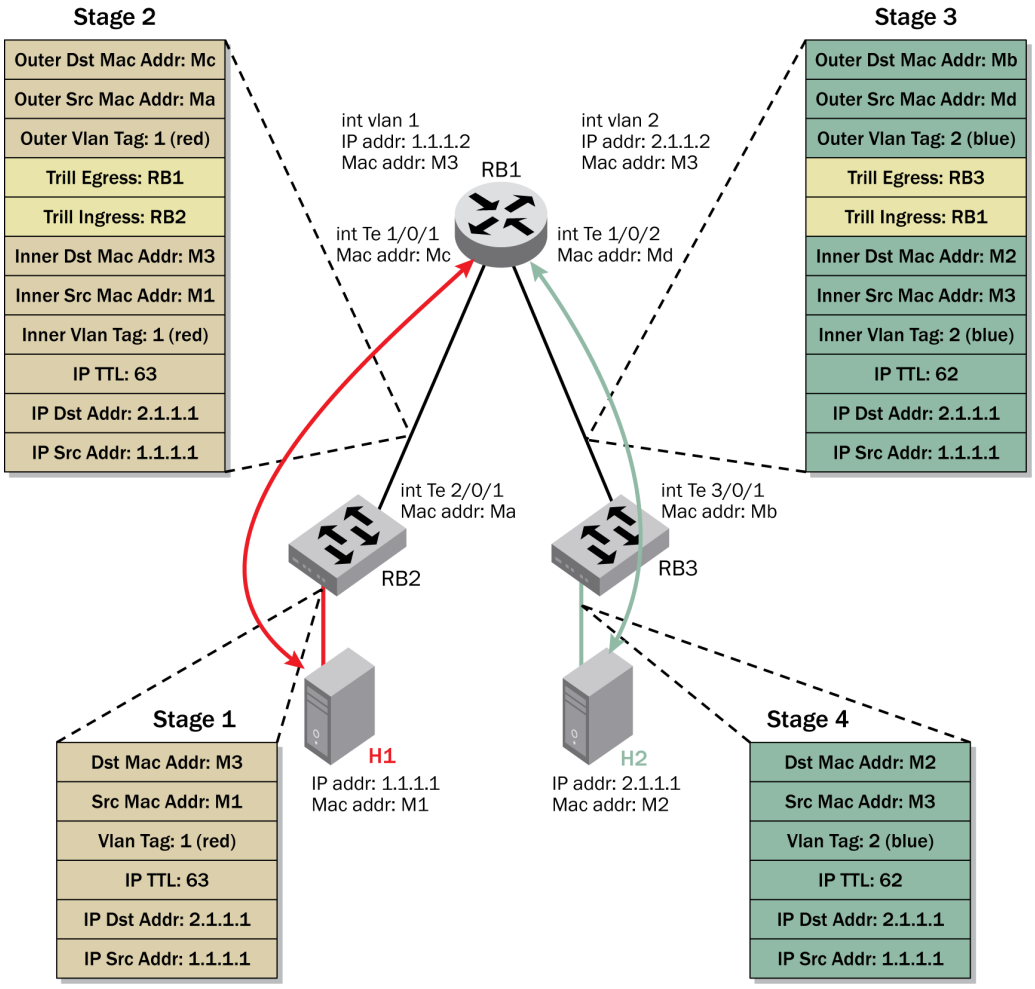

Figure 27. Life of a packet.

Figure 27 shows a three-node VCS fabric. Only RB1 has routing enabled. RB2 and RB3 are acting as pure Layer 2 switches. You see two VLANs active in the fabric: VLAN 1 (red) and VLAN 2 (blue). The router has VEs for VLAN 1 and VLAN 2 configured. You can assume that ARP has already been resolved and that H1 knows how to reach H2 via RB1. The router RB1 has proxy-ARP turned on so that it can respond to ARP requests for either VLAN 1 or 2 on any interface that has routing enabled.

Assume that H1 now wants to talk to H2.

- **Stage 1**
  - Before the packet enters the VCS fabric, you can see that H1 has the destination MAC address for the packet as M3—the RB1 VE MAC address.
  - As in a traditional Ethernet switch, all VEs within a single RBridge share the same MAC address. Note that both VLAN 1 and VLAN 2 on RB1 have the same MAC address.
  - The packet has an IP TTL of 63.
  - The packet is tagged on VLAN 1 (red).
- **Stage 2**
  - The packet first hits RB2, which is a pure Layer 2 switch.
  - RB2 then looks up its Layer 2 forwarding tables and determines where it needs to forward the packet destined for MAC address M3 on VLAN 1 (red).
  - Assume that RB2 has already learned of this particular destination.
  - RB2 will encapsulate the frame in a TRILL frame and forward it along to its next hop.
  - The outer header has the destination MAC address of Mc, which is the address for Te 1/01/ (the T_Port of RB1).
  - The outer header has a source MAC address of Ma, which is the MAC address for Te 2/0/1 (the T_Port of RB2).
  - The Outer Header has a destination RBridge of RB1 and a source RBridge of RB2.
  - Please note that the Inner Header is untouched at this point.
- **Stage 3**
  - At this point the packet has reached RB1.
  - RB1 determines that the packet has reached its initial destination by looking at the Egress RBridge address RB1 and the outer MAC address of Mc.
  - RB1 decapsulates the packet and looks at the inner VLAN tag of VLAN 1 and inner MAC address M3 and realizes that M3 is its own MAC address, as well as the fact that routing is turned on VLAN 1.
  - RB1 takes the inner packet and performs an IP lookup.
  - Based on the IP lookup, RB1 determines the next hop of the packet.
  - RB1 changes the inner destination MAC address to M2 (the address of H2).
  - RB1 changes the inner source MAC address to M3 (its own MAC address).
  - RB1 decrements the TTL by 1.
  - The router also realizes that the destination is sitting behind the fabric.
  - RB1 now encapsulates this packet into a TRILL header.
  - The TRILL encapsulation will include the egress RBridge of RB3 and ingress RBridge of RB1 (itself).
  - The outer header has the destination MAC address of Mb, which is the address of Te 3/0/1 (the T_Port of RB3).
  - The outer header has the source MAC address of Md, which is the address of Te 1/0/2 (the T_Port of RB1).
- **Stage 4**
  - At RB3, the switch looks at the outer MAC address Mb and egress RBridge of RB3 and realizes the packet has reached its intended destination.
  - RB3 decapsulates the packet and looks at the inner VLAN tag of 2 (blue) and inner destination MAC address M2 and realizes that it needs to do a Layer 2 lookup.
  - Assuming the Layer 2 table is already populated, RB3 will know that it needs to forward the packet to host H2.
  - RB3 will forward the decapsulated packet towards H2.

Finally H2 receives the packet, as shown in stage 4.

Layer 3 Interface Management

From a forwarding perspective, the Layer 3 interface within a VCS fabric is very similar to that on a regular router. Since routing is local to a given RBridge, all Layer 3 interface management is local.

- All the router MAC addresses are allocated from a common pool that is burned into every single box at the time of manufacturing. This MAC address is a globally unique MAC address.
- All VEs are assigned the same Layer 3 MAC address. The reason for assigning the same MAC address for all VEs is because the VLAN ID provides the distinguishing identifier. By definition, this means that every VE has a unique <MAC, VID> identifier across the entire VCS fabric.
- Every single router port has a unique Layer 3 MAC address. By definition, this means that every router port has a unique MAC identifier across the entire VCS fabric.
- As per the regular VCS fabric operations, all T_Ports are automatically assigned to VEs.
- Due to the fact that all T_Ports are automatically part of all VLANs that are enabled on a router, every router appears to be part of the same LAN within a contiguous VLAN domain.
- This unique ability allows you to form router adjacencies across T_Ports, achieve Layer 3 adjacencies, resolve ARPs, and enable two RBridges to talk to each other seamlessly.

Preventing Layer 2 flooding for Router MAC Address

Every single router MAC address that is enabled in the VCS fabric is synced throughout the VCS fabric. The reason for this syncing is to ensure that every single router MAC address is treated as a known MAC address within the fabric. This ensures that when any packet enters the VCS fabric destined towards a router, it is never flooded and is always unicast to its correct destination.

Similarly, when routing is disabled, the router sends a message to withdraw that particular router MAC address from the VCS fabric. This behavior prevents the periodic issue of Layer 2flooding caused by the router MAC address being aged out. We no longer require the administrator to ensure that the Layer 2 aging time is greater than the ARP aging time interval.

ARP in VCS Fabrics

ARP is used to resolve IPv4 addresses into Ethernet MAC addresses. ARP packets are not IP packets, and they have their own ether type. Figure 28 illustrates the various ARP scenarios within a VCS fabric.

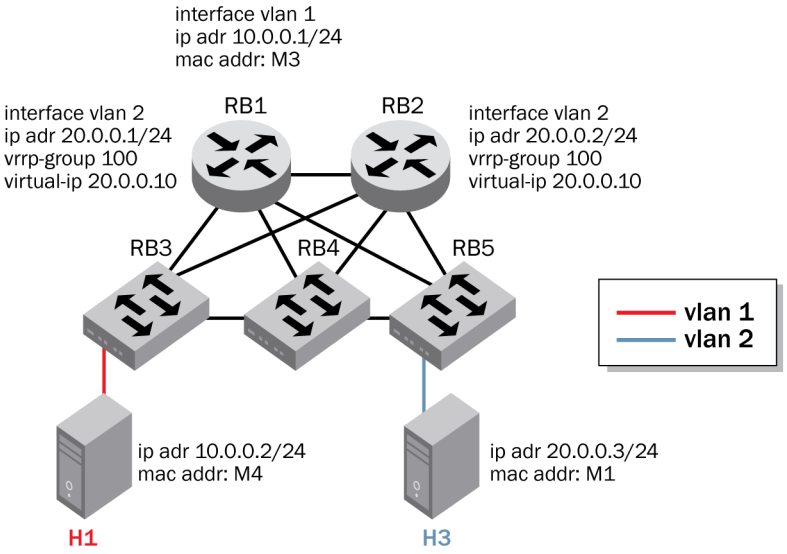

Figure 28. ARP in a VCS fabric.

The topology shown in Figure 28 has a five-node VCS fabric with routing turned on RB1 and RB2.

Host H1 is attached to RB3 on VLAN 1 and has an IP address of 10.0.0.2/24 with a host MAC address of M4.
Host H1 has 10.0.0.1 configured as its default gateway.
Host H3 is attached to RB5 on VLAN 2 and has an IP address of 20.0.0.3/24 with a host MAC address of M1.
Host H3 has 20.0.0.10 configured as its default gateway.
RBridge RB1 has VE 1 with an IP address of 10.0.0.1 and a MAC address of M3.
RBridge RB1 also has VE 2 with an IP address of 20.0.0.1 and a MAC address of M3.
RBridge RB2 has VE 2 with an IP address of 20.0.0.2 and a MAC address of M2.
Since VE 2 spans RB1 and RB2, you have to run VRRP across both of them to elect a master.
In this example, you should assume that RB1 has become the master.
H1 wants to communicate with H3.

1. ARP resolution for 20.0.0.3 on H1:
    - H1 sends out an ARP request for 20.0.0.10 because that is its default gateway.
    - The request enters the VCS fabric on RB3.
    - RB3 looks at this broadcast packet, encapsulates it in a TRILL frame, and floods it on the multicast tree in the entire VCS fabric.
    - This ARP request reaches RB1.
    - On RB1 for VLAN 1, a trap entry is programmed for all ARP requests to get trapped to the CPU.
    - RB1 responds to the ARP request with its own MAC address.
    - It is possible that by the time RB1 replies to the ARP request from H1, it has learned the MAC address M4 for H1. If Layer 2 learning has occurred, then RB1 will unicast its ARP reply back to H1. If not, then RB1 will multicast its response towards H1 on VLAN 1 on the multicast tree.
    - In either case, eventually RB3 receives the ARP response from RB1, decapsulates the TRILL packet, and forwards the ARP response out of the edge port towards H1.
    - H1 has now learned the ARP for 20.0.0.10, thus it thinks that it can now reach 20.0.0.3.
2. ARP resolution for 20.0.0.3 on RB1:
    - Now that H1 has learned the ARP for “20.0.0.3” (actually 20.0.0.10), it starts sending its data traffic on VLAN 1 with a destination MAC address of M3.
    - As described earlier, RB3 is already aware of the router MAC address M3 on VLAN 1. It unicasts this packet to RB1.
    - RB1 receives the packet, decapsulates it, performs the Layer 3 lookup, and realizes that it not aware of the directly attached host 20.0.0.3.
    - RB1 initiates an ARP request for 20.0.0.3.
    - RB2 does not respond to this ARP request, since it is not the master on that VE.
    - RB5 receives this ARP request, decapsulates it, and forwards it towards H3.
    - Host H3 responds to the ARP request with its own MAC address. It responds back to RB1 on VLAN 2 with a destination MAC address of the virtual MAC address.
    - RB5 once again encapsulates this ARP response towards RB1, since it is aware of RB1 and the virtual router MAC address.
    - RB1 installs the next-hop information.
3. Native forwarding:
    - The next set of data traffic is natively routed at RB1 towards H1 and follows the regular TRILL routed path as described in the Forwarding Architecture section.

Interactions with Layer 2 Forwarding

While ARP informs the Layer 3 routing infrastructure about the destination VLAN and the Ethernet encapsulation, the Layer 3 system still has to rely on the base Layer 2 infrastructure to figure out the eventual Layer 2 header information. This information includes the TRILL header details, as well as the Layer 2 egress port through which the packet will eventually exit.

There are quite a number of combinations that are possible. Figure 29 illustrates all of these scenarios.

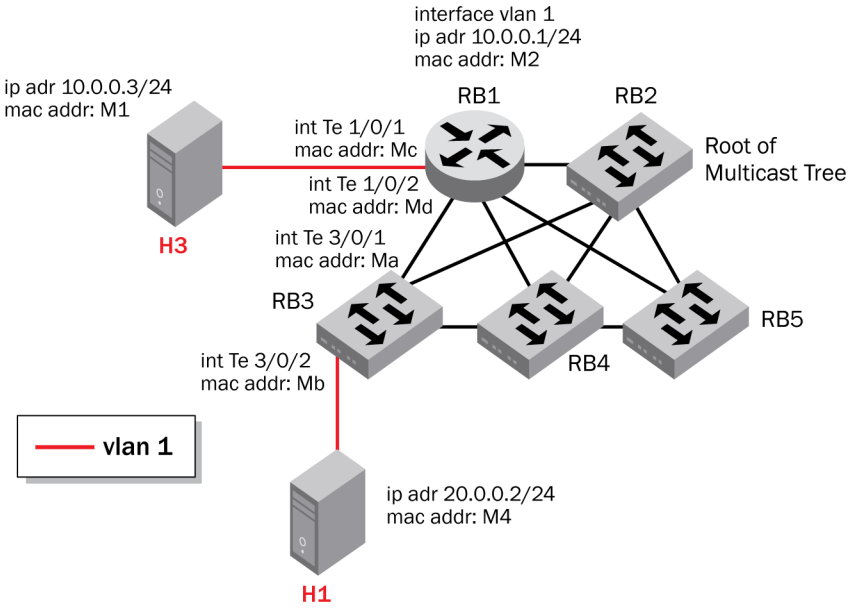

Figure 29. Layer 2 interactions.

For all the scenarios, assume that ARP has already resolved, and both the hosts H1 and H3 and the router RB1 are reachable from an Layer 3 connectivity point of view. RB1 is initiating the communication towards H1 and H3.

Scenario 1: H3 MAC address M1 has been learned and distributed in the VCS fabric.
This scenario is similar to a regular Layer 3 router. H3 is connected to an edge port of the RBridge RB1. Since the MAC address is already known to Layer 2, RB1 simply has to do a regular Ethernet encapsulation and forward the frame towards H3 on the edge port. Figure 30 shows the frame format of the packet as it egresses RB1 towards H3 on Te 1/0/1.

Dst Mac Addr: M1
Src Mac Addr: M2
Vlan Tag: 1 (red)
IP TTL: decrement by 1
IP Dst Addr: 10.1.1.3
IP Src Addr: 10.1.1.1

Figure 30. Known Layer 2 MAC address hanging off an E_Port.

Scenario 2: H3 MAC address M1 has not been learned in the VCS fabric.
This scenario is possible if either the H3 MAC address has been aged out or if the Layer 2 learning is slow. In this case, although RB1 knows how to reach H3 from an Layer 3 perspective, this is an unknown unicast destination case. RB1 will multicast its packet on VLAN 1 within the fabric, since it is not aware of the exact Layer 2 location of the H3 MAC M1 on VLAN 1.

The frame that exits on the E_Port Te 1/0/1 will look exactly like the frame shown in Figure 30.

The frame that exits on the T_Port Te 1/0/2 will look like Figure 31.

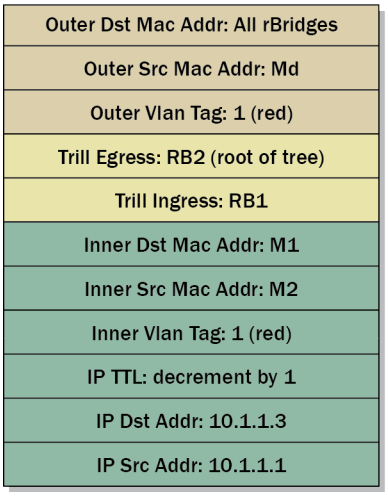

Figure 31. Frame existing on a T_Port for an unknown Layer 2 MAC address hanging off an E_Port.

Scenario 3: H1 MAC address M4 has been learned and distributed in the VCS fabric.
In this scenario, RB1 is already aware of the Layer 2 location for H1. Since it knows H1 sits behind RB3, it will send a unicast TRILL-encapsulated frame out of the T_Port Te 1/0/2 towards RB3. The frame will be TRILL-encapsulated and will look like Figure 32.

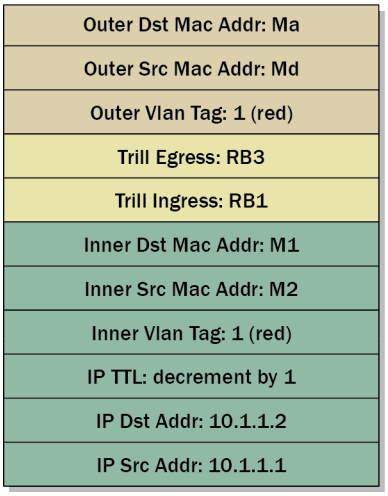

Figure 32. Frame existing on a T_Port for an unknown Layer 2 MAC address hanging off an E_Port.

Scenario 4: H1 MAC address M4 has not been learned and distributed in the VCS fabric.
In this scenario, since RB1 does not know the Layer 2 location for H1, it is similar to Scenario 2. The router will send two copies of the packet—one on the E_Port Te 1/0/1 and the other on the T_Port Te 1/0/2. The frame formats will be similar to Scenario 2 except for the inner destination MAC address, which will be that of M4.

This multicast frame will reach RB2. Since RB2 itself is not aware of this destination, it will decapsulate it and flood this packet on all members of VLAN 1, including Te 3/0/2, which is connected to the host H1.

VRRP and Brocade VRRP-E in VCS Fabrics
VRRP eliminates a single point of failure in the static route environment. It is an election protocol that dynamically assigns responsibilities of a virtual router to one of the VRRP enabled routers on the LAN. VRRP provides a higher availability default path without requiring configuration of dynamic routing or router discovery protocols on every end-host. VRRP-E (Extended) is the Brocade proprietary extension to the standard VRRP protocol. It does not interoperate with VRRP.

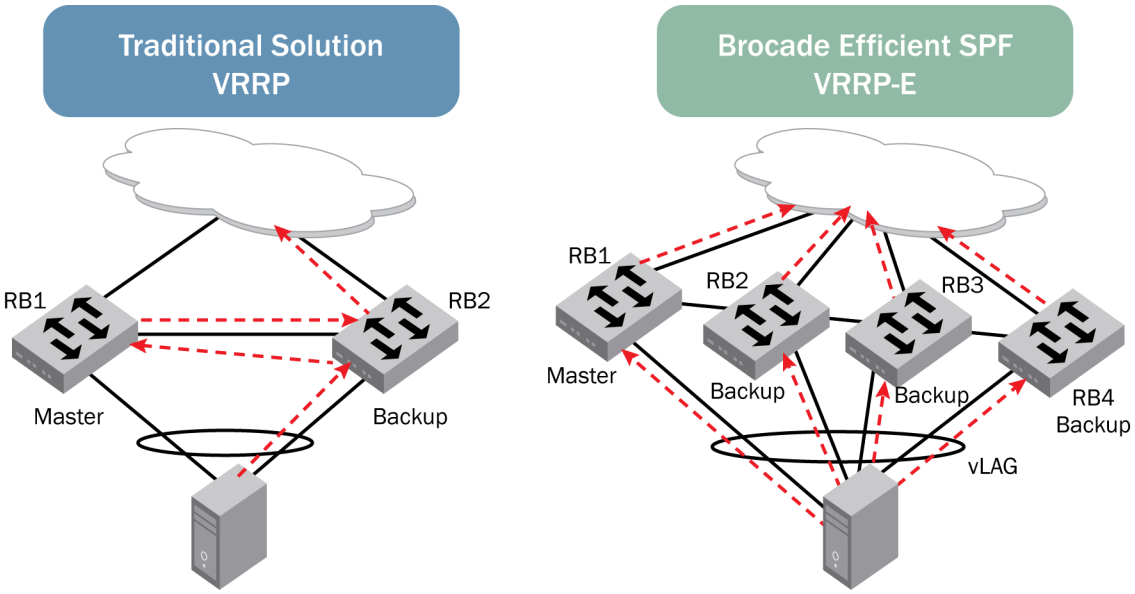

Figure 33. Comparison of VRRP and VRRP-E implementations.

In Figure 33, R1 and R2 form a VRRP group. R1 is the Master and R2 is the standby. VRRP multicast advertisement frames are broadcasted by R1. Since TRILL ports on all of the RBridges are all part of the fabric, R1 and R2 are able to see each other's advertisement frames.

VRRP is required to provide gateway redundancy for Layer 3 targets connected to the VCS fabric. It provides failover between active/standby routers. The standard VRRP protocol (compliant with RFC 3768) is supported in Brocade Network OS 3.0.

At the same time, organizations need to be able to extend their Layer 2 fabrics on demand, as the need for new servers grows, and provide inter-VLAN routing without recourse to external Layer 3 devices. VRRP limits the number of Layer 3 gateways, artificially compromising the scale of Layer 2 domains and limiting VM mobility. In contrast, the Brocade VRRP-E (VRRP-Extended) protocol provides support for four or more *active/active* Layer 3 gateways, yielding scalability, higher bandwidth, and easier migration, without the need for an extra aggregation layer. VRRP-E also adds short-path forwarding for reduced latency and a shared virtual IP gateway and virtual MAC address across the entire fabric to ease administration.

The administrator can select VRRP to be configured on both VE interfaces and router ports. However, VRRP–E is supported only on VE interfaces. There are some differences in how VRRP and VRRP-E handle ARP and VRRP control packets. For more information about these differences, please refer to the "VRRP/VRRP-E Packet Behavior" section of the Brocade *Network OS Administrator's Guide.* Since Brocade does not support the configuration of an IP Address on a vLAG, VRRP and VRRP-E on vLAG router ports are not supported.

As with any Layer 3 configuration, VRRP and VRRP-E are always local to a particular RBridge. All VRRP/VRRP-E configurations are provided under the RBridge mode. A typical VRRP configuration is shown in Figure 34.

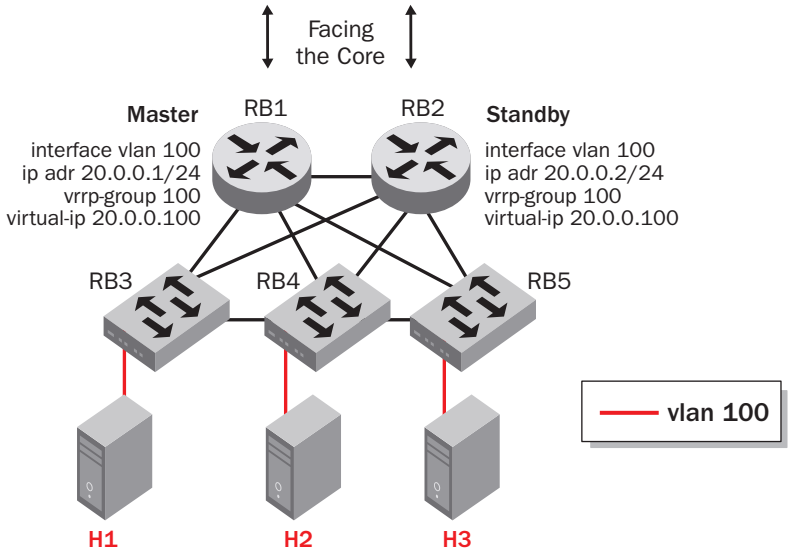

Figure 34. Typical VRRP deployment in a VCS fabric.

For the hosts H1, H2, and H3, VCS Fabric technology provides a Transparent LAN service, with routers being the only devices visible to them. Thus the topology is simplified and actually looks like the diagram in Figure 35.

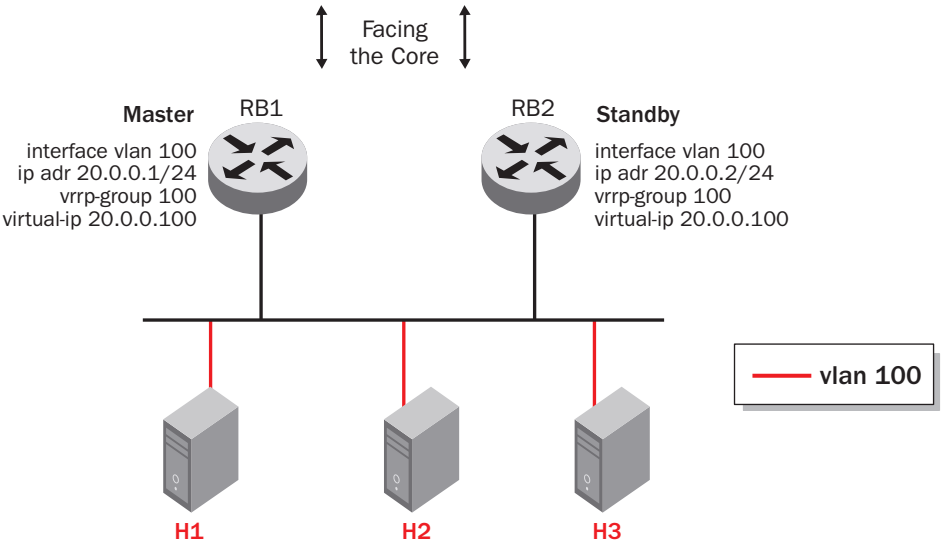

Figure 35. Equivalent VRRP topology for the hosts.

In Figure 35, RB1 and RB2 form a VRRP group. In this particular example. RB1 is the master and RB2 is the standby. VRRP multicast advertisement frames are broadcasted by R1 on VLAN 100. Since TRILL ports on all of the RBridges are part of all the VLANs, both RB1 and RB2 are able to see each other's advertisement frames.

Hosts resolve the ARP for the gateway virtual IP address by injecting ARP requests into the fabric. ARP is resolved by the master RBridge, and hosts send data traffic destined to the virtual MAC address for routing. When VRRP is configured on the VE interface, the VRRP advertisement frames are transported in the TRILL network, similar to any other multicast Ethernet frame. If VRRP is configured on the router port of an RBridge, the topology is similar to what you would observe on traditional Ethernet routers.

Distribution of vMAC in the VCS Fabric

As explained in the ARP in VCS Fabrics section, the VRRP virtual MAC address, or vMAC, has to be learned by all of the RBridges in the fabric in order to unicast the frame destined to the vMAC towards the VRRP master and avoid unnecessary Layer 2 flooding. This is done through the typical distribution service that is used to keep the Layer 2 MAC address database in sync in the VCS fabric. When an RBridge becomes the VRRP master, it triggers the distribution of the vMAC. All the RBridges add static Layer 2 forwarding entries pointing towards the master RBridge.

In the example above, RB1 triggers the distribution of the vMAC in the VCS fabric using the MAC distribution service. All of the RBridges, including RB2, install a static forwarding entry in hardware for the vMAC. When hosts H1, H2, or H3 send frames destined to the vMAC, the RB3 and RB4 RBridges will unicast the packet in the TRILL network towards the VRRP master RB1. When the standby becomes the master, the old vMAC is withdrawn from the VCS fabric, and a new vMAC forwarding entry overwrites the previous entry in all of the RBridges in the VCS fabric.

VRRP over vLAG

Brocade does not support the configuration of VRRP directly over a vLAG. However, a vLAG can be part of a VE, and VRRP could be running over that vLAG. The topology diagram in Figure 36 explains this particular scenario.

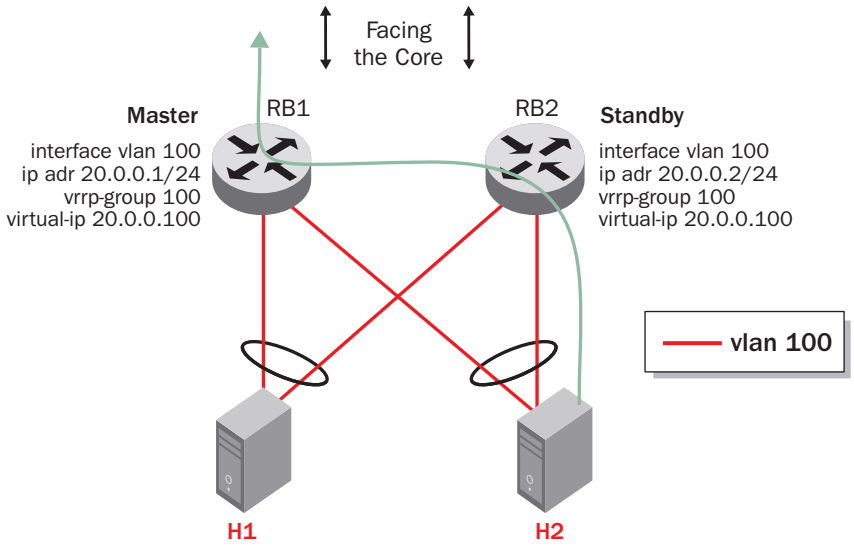

Figure 36. VRRP on a VE that has a vLAG as a member.

The example shown in Figure 36 considers a two-port vLAG. This scenario can be extended to an N_Port vLAG without any changes.

RBridge RB1 is the master and RB2 is the standby. Hosts H1 and H2 are connected to RB1 and RB2 through vLAGs terminating in Port Channel 1 on RB1 and Port Channel 2 on RB2, respectively. These vLAGs are tagged in VLAN 100. VRRP is running on VE 100. ARP will be resolved by the master. The vMAC is installed only on RB1 and,

as explained, RB2 will have a static entry programmed to direct any traffic destined to the vMAC on VE 100 towards RB1.

As a result of this, any traffic originating from H1 or H2 can potentially get hashed between RB1 and RB2. If the traffic hits RB1, it will get routed correctly towards the core. If the traffic gets hashed to RB2, it will result in RB2 forwarding the traffic towards RB1.

Brocade also supports a scenario in which the user wants to configure VRRP across a VCS Fabric router and an external router.

IP Static Route in VCS Fabrics

The user has the ability to configure static routes on the devices in the VCS fabric. Static routes can be used to avoid overhead of running dynamic protocols in simple networks. The user can overwrite the routes calculated by routing protocols, can inject networks that do not have routing protocols enabled, and can keep default backup paths when routing protocol instabilities affect the network.

Brocade Network OS 3.0 offers the following functions for static routes:

- Recursively resolves the next hop using static or dynamic routes.
- Egress interface status decides the activeness of the static route.
- Several options in selection criteria for installing static routes in RIB:
  - The route with the least metric value is preferred.
  - If the metric value is the same, the route with smallest distance value is preferred.
  - If distance and metric value are the same, the routes will be combined to the ECMP route if it is enabled. However, the non-recursive routes will not form ECMP with recursive routes—it is the optimization to make the recursive route take preference.
- The static route can be configured with a tag value. This value can be used to filter the static routes during redistribution to routing protocols.
- The static route pointing to the null interface is used to drop the traffic.
- Default route (0.0.0.0/0) is used to handle traffic to unknown destinations.
- /31 masks are supported.

Open Shortest Path First (OSPF) in VCS Fabrics

OSPF is a popular link-state dynamic routing protocol that uses an SPF algorithm to compute routes. This feature will be supported on both modular and fixed-port Brocade VDX platforms. Brocade Network OS 3.0 implements OSPFv2 protocol for IPv4. The following additions are added to the OSPF feature, along with the standard features as described in RFC 2328 and RFC 1587.

- Interoperates with other vendors
- Supports manual configuration of neighbors on broadcast links
- Performs partial route computations when only external routes are injected
- Performs incremental SPF runs to recompute the affected tree only

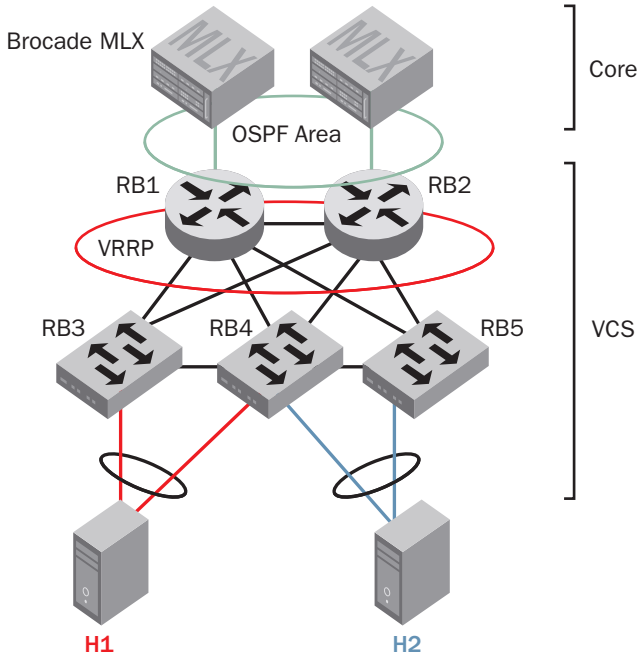

Figure 37. VCS fabric with OSPF to the core.

Layer 3 ACLs in VCS Fabrics

Layer 3 ACLs, as in every other Layer 3 policy, are local in scope to a particular RBridge. It is expected that the user will configure a unique Layer 3 ACL on every RBridge and apply them either on the router port or VE interface.

All of the Brocade custom ASICs support deep packet inspection and the application of Layer 3 filters on switched traffic. This is a particularly powerful tool that can be used to ensure the security of a VCS fabric.

General Rules for ACLs

- When an ACL is bound to an interface, the rules of the ACL will be instantiated in the hardware. If there are not enough available resources to instantiate all rules of the ACL, then as many rules as possible will be instantiated into the hardware. Also, a syslog message will be generated, indicating that the ACL is only partially instantiated and providing the interface index and/or the BP slot number on which the binding failed.
- Layer 2 ACLs may be applied only on Layer 2 interfaces and VLANs. Layer 3 ACLs may be applied to either Layer 3 or Layer 2 interfaces or VLANs, regardless of whether the port is being used in switch or routing mode.
- In Brocade Network OS 3.0, Layer 3 ACLs will apply to all traffic on the interface, regardless of the port's mode of operation (switch or routed). For example, an ACL applied to a router interface will apply to all traffic regardless of whether the traffic is addressed to the MAC address of the interface.
- Layer 3 ACLs may coexist with Layer 2 ACLs on the same Layer 2 interface or VLAN. Layer 3 ACLs may coexist with Layer 2 ACLs on a member Layer 2 interface.
- In case a traffic stream satisfies both Layer 2 ACL and Layer 3 ACL rules, the final action taken to this traffic stream is determined by the priority of action (in other words, DENY > any other action).
- Rules can be inserted, deleted, and appended anywhere in the ACL without disruption of traffic.
- The default action of "permit any" will be inserted at the end of a bounded Layer 2 ACL.
- The default action of "deny any" will be inserted at the end of a bounded Layer 3 ACL.
- A physical interface that has an ACL bound to it cannot be made a member of an existing LAG.

Figure 38 illustrates the capabilities of a Layer 3 ACL within a VCS fabric.

RB1 is a router and has an ingress Layer 3 ACL applied on it that prevents host H1 from sending IP traffic within the fabric. RB3, on the other hand, is functioning as an Layer 2 switch within the VCS fabric. However, you can apply an Layer 3 ACL on RB3 and filter on the basis of Layer 3 fields. In Figure 38, H4 cannot send IP traffic to the fabric through switch RB3. Also, note that both the ACLs have the same name and are applied on VLAN 1, underscoring the fact that Layer 3 ACLs have only local significance to an individual RBridge.

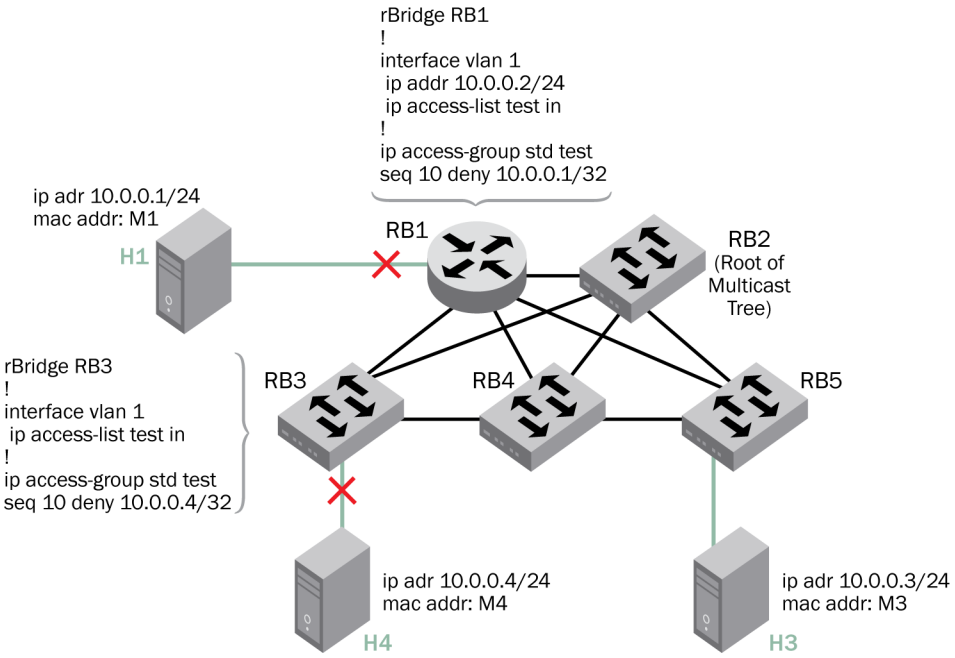

Figure 38. Layer 3 ACLs in a VCS fabric.

Figure 38 gives a snapshot of an ACL config on the system.

Here you see two ACLs, a MAC standard access-list called “test” and an IP standard access-list called “test_ip_acl”.

The MAC ACL is applied on VLAN 100 and interface te 1/1.

The IP ACL is applied on VLAN 100.

Routing Scenarios

In the first scenario you see a distributed router configuration. (These are simply two scenarios being used for illustration purposes, out of many possible scenarios.)

Distributed Routing Deployment Towards the Aggregation Layer

In the scenario depicted in Figure 39, you see a five-node VCS fabric. RB1 and RB2 are the RBridges with routing enabled. They have OSPF configured towards the core to provide connectivity out of the data center. RB3, RB4, and RB5 are switches that provide the Layer 2 connectivity to servers hanging off them. H1 and H3 have vLAGs connecting them to RB3-RB4 and RB4-RB5, respectively.

To provide redundancy for servers H1 and H3, the gateway is a virtual IP address. RB1 and RB2 have VRRP running between them. Since H1 and H3 are in different subnets, there are different VRRP groups running between RB1 and RB2.

It is possible that both RB1 and RB2 become the master for each of the VRRP groups, and this can provide load balancing along with redundancy in the VCS fabric.

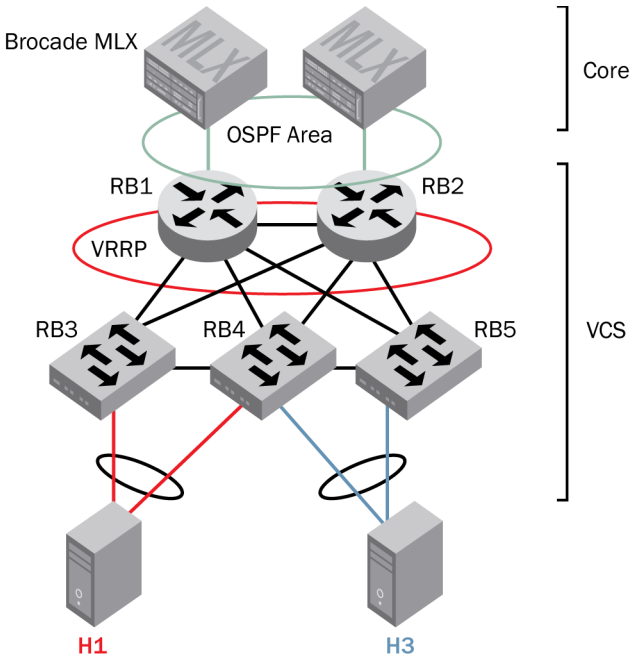

Figure 39. Distributed routing deployment towards the aggregation tier.

Distributed Routing Deployment on the Access Layer

In Figure 40 you can see the VCS fabric deployed at the access layer and enabling Layer 3 directly on that layer.

RB1 and RB2 are the two routers that are part of the VCS fabric. Hosts are directly connected to these two routers through vLAGs. Redundancy is provided on the VCS-enabled device by running VRRP between RB1 and RB2. The distribution layer has most probably a high-density router like the Brocade MLX® Router. RB1 and RB2 are running a routing protocol OSPF to provide connectivity towards the distribution layer.

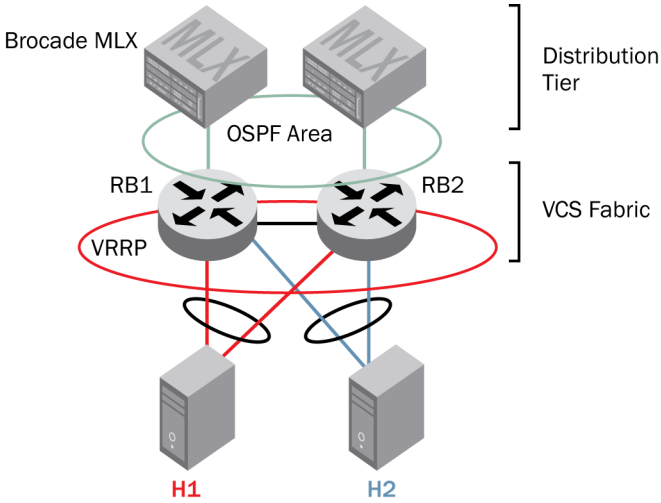

Figure 40. Distributed routing deployment on the access tier.

Management

Brocade Fabric Watch Monitoring

Brocade Fabric Watch monitors the health of the components and, based on the threshold, each component can be declared as marginal or down. Once a component gets declared as marginal or down, Brocade Fabric Watch generates user notification using one or more of the following mechanisms:

- RASlog Message
- Email

Factory Default Thresholds

Default thresholds are defined for fans, power supplies, compact flash, and temperature sensors.

RASlog

Brocade Fabric Watch generates entries in the RASlog for changes in the switch status. The entries include an ID, severity code, and a description/cause classification indicating if the message is information only or a warning.

FRU Monitoring

The field-replaceable unit (FRU) alert monitoring feature monitors the insertion and removal of the FRUs. Typical FRUs are fans and power supplies, and each Brocade VDX model has a defined set of FRUs. Brocade Fabric Watch monitors each FRU of the switch, and if an FRU is removed, inserted, or goes into a faulty state, then Brocade Fabric Watch sends an alert. An FRU can be in one of the following states: removed, inserted, on, or off.

The following sections cover management capabilities for VCS Fabric mode. These do not apply when a Brocade VDX switch operates in standalone mode.

Fabric Join/Merge

The Fabric Join/Merge Protocol (JMP) is responsible for adding in a new switch to the Logical Chassis when the switch is physically attached to another switch in a VCS fabric. JMP distributes configuration information to all switches that have joined the fabric. The following rules are enforced during those operations:

- Boot order. The JMP is executed after the principal switch has been defined. The principal switch is primarily responsible for centrally allocating the RBridge IDs for every RBridge in the fabric and detecting and resolving any RBridge ID collisions.
- State Change Notification (SCN) triggers. A join operation is initiated by a “Domain Capabilities Available” SCN from the Export Switch Services (ESS). Disconnects are initiated by a SCN “Domain Unreachable” from ESS.
- Every switch in a fabric must use a single VFID, and it must match the VFID of the other switches in the fabric. If a new switch is connected to a fabric, and the VFID is not the same as the other switches, it will not be allowed to join the fabric.
- If a principal switch goes offline, or if the fabric reassigns the principal switch role to a different switch, no configuration updates will be processed until completion of the principal switch election process.

VCS Fabric Node Merge

If multiple switches are added to the fabric, they are added serially to avoid potential deadlocks due to conflicting configurations.

Forming an Initial VCS Fabric (Two Switches)

When two switches connect to form a VCS fabric, a principal switch must be selected before the merge process can be initiated. The newly elected principal switch can then “invite” the other switch to join using the “Node Merge” procedure.

Non-Principal Switch Rejoins a Fabric

A switch that has left a fabric and then joins again is processed differently to maintain interface naming consistency. For instance, if a CLI output shows an interface as Te 3/2/2, you want to ensure that this naming does not change

when the same switch joins the same fabric again. The switch ID in the configuration table is used to determine the logical slot in the Logical Chassis that will be assigned to the switch. In a VCS fabric, interfaces are formatted as follows:

TE <SwitchID>/2/2

In this format, "SwitchID" is a unique value that is assigned to a switch when it is part of the fabric, which should be the same as the RBridge ID that is assigned to that switch.

Interface Numbering

Interface numbering is unique within a VCS fabric. Every interface is prefixed with the unique switch ID of the physical switch of which it is part. Initially, the switch ID is identical to the RBridge ID. If a switch leaves a fabric and then joins again, the RBridge ID can change, but the switch ID will be persistent and reapplied based on matching the unique worldwide name of the switch.

Centralized vs. Distributed Cluster Management Access

All switches are connected to an external management network. The management network is not involved in communicating any internal fabric configuration or maintenance functions. The management network is used by a switch for communication with the RADIUS, syslog, and Simple Network Management Protocol (SNMP) managers via the local management interface port in the switch.

Edge and Fabric Port Configuration Behavior

As previously mentioned, when a new switch joins a fabric, it discovers its neighbor, automatically forms ISL connections on the fabric ports that are connected to a neighboring switch to which it is physically attached, and joins the VCS fabric. The switch applies a set of fixed configuration parameters to every fabric port as part of fabric port bring-up (that is, it is self-configuring).

If a port transitions from being an edge port to a fabric port, the edge port configuration settings in the configuration file are retained.

For example, when a switch is removed from the VCS fabric by disconnecting its fabric ports, they become edge ports, and each port assumes any previous edge port settings. Or, if a port was configured as an edge port, then is used as a fabric port and once again becomes an edge port, the edge port configuration settings are reapplied.

Default Configuration of Fabric Ports

The switch applies the following set of configuration changes to fabric ports when an ISL dynamically forms:

- Configure the interface to be in Brocade ISL Trunking mode.
- Any VLAN that is configured on any of the edge ports is applied to the fabric port.
- Add the default VCS fabric control VLANs 4093, 4095.
- Apply DCB default settings (PFC-enable on Priority 3 [FCoE], default ETS parameters).
- Fabric control frames are classified as Priority 7 and are treated as strict priority on the fabric ports. If any existing traffic uses or requires Priority 7, then that traffic is required to configure priority classifiers to remap 7 to any other available priority class. Users can use standard priority remapping QoS CLIs to make this change.
- Because Priority 7 is used for internal VCS Fabric operations, it is recommended that users define priority remapping classifiers on the switch in environments where Priority 7 is absolutely required.

Brocade Network Advisor

Brocade Network Advisor is an easy-to-use network management platform for advanced management of Brocade VCS fabrics and Brocade VDX switches across the entire network life cycle. Organizations can use Brocade Network Advisor to manage a VCS fabric as a single entity or to drill down to individual Brocade VDX switches for fault, inventory, or performance management—and to manage multiple VCS fabrics in parallel. More important, Brocade Network Advisor manages both SAN and IP networks and can be used to do zoning across Brocade VDX switches and Fibre Channel SANs.

Brocade Network Advisor also provides centralized management of AMPP configurations in multifabric deployments, and integrity checks can be performed across the virtual configurations and physical Brocade VDX configurations. In addition, Brocade Network Advisor enables VM-level monitoring and can help identify top-talker applications leveraging sFlow across the fabric. Finally, Brocade Network Advisor provides VCS fabric diagnostics, including visualization of VCS fabric traffic paths and network latency monitoring that enables fault isolation via hop-by-hop inspection. For details, visit www.brocade.com/management.

APPENDIX A: BROCADE VCS FABRIC AND TRILL

TRILL, an IETF protocol that is designed to scale Ethernet networks, defines RBridges and the connections between them. A link-state routing protocol is used at Layer 2 to forward traffic between RBridges. One of the primary reasons that a fabric of RBridges scales better than classic Ethernet with STP is the use of a link-state routing protocol. A fabric of interconnected RBridges is as efficient as a link-state routed IP network. TRILL enables Layer 2 networks to behave like routed Layer 3 networks, while at the same time providing the LAN services that are typically expected in Layer 2 networks.

TRILL also defines native support for forwarding both unicast and multicast traffic and therefore unifies support for both classes of applications over a single transport. RBridges that are connected via a connectionless TRILL fabric natively operate at Layer 2 on the non-TRILL connected ports. This native Layer 2 capability that is inherent in RBridges eliminates the need to overlay another layer of protocols to achieve bridging functionality. For example, Multiprotocol Label Switching (MPLS)-based approaches, which predominantly rely on IP protocols, do not operate like bridges and, therefore, need another overlay set of protocols and services to achieve the transparent bridging functionality.

In summary, TRILL operates like a Layer 2 network from the perspective of the devices that are attached to bridges supporting TRILL and does not require additional protocols to emulate a Layer 2 network. Because it natively supports both unicast and multicast traffic forwarding at Layer 2, there is no need to overlay any point-to-point or point-to-multipoint tunnels over the TRILL network. This significantly reduces the number of protocols that are required to provide TLS.

The VCS fabric data plane forwarding is in full compliance with TRILL. The control plane protocols that are defined in TRILL require manual intervention to assign RBridge IDs and add a new set of protocols for RBridge discovery and TRILL network formation. The VCS fabric control plane improves upon the TRILL control plane in that it provides automatic assignment of RBridge IDs, automatic resolution of duplicate IDs, and automatic ISL formation and topology discovery.

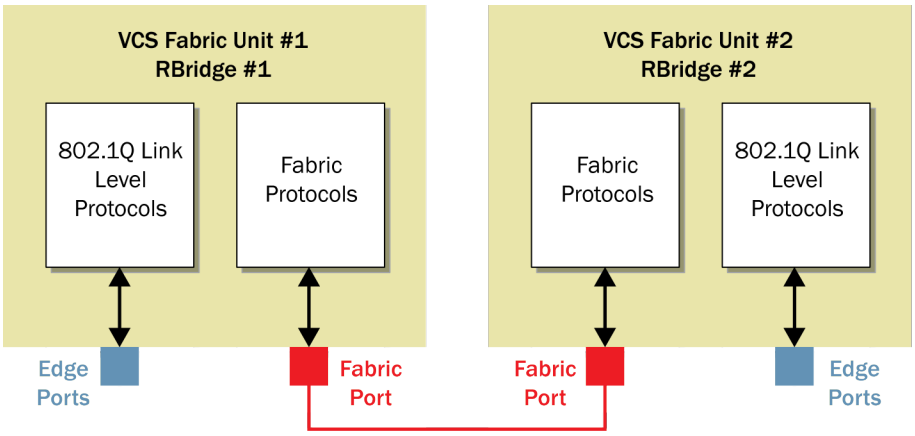

Figure 41. Two-switch VCS fabric.

Comparison of Brocade VCS Fabric and TRILL

Attributes	VCS	TRILL
No Spanning Tree Protocol	Yes	Yes
Link-State Routing Protocol	FSPF	IS-IS
ECMP	Yes	Yes
vLAG	Yes	No
LAG	Yes	Yes
Hardware-Based Frame-Level Trunking	Yes	No
Inter-Switch Links	Point-to-point	Point-to-point and Multiaccess

APPENDIX B: VCS FABRIC FRAME TYPES

This section summarizes various frame types used in a VCS fabric.

Generic TRILL Frame

A TRILL frame is used to encapsulate Ethernet data frames (unicast, multicast, and broadcast), FCoE data frames, and VCS fabric control frames. A TRILL frame is created to encapsulate ingress traffic received on an edge port prior to forwarding it on a fabric port to another RBridge in the VCS fabric. The generic TRILL frame is shown here. Refer to the TRILL standard for a description of the TRILL frame format.

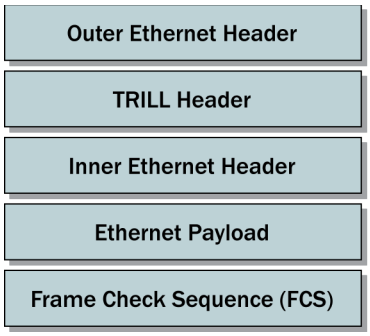

Figure 42. TRILL frame format.

FCoE Frame

FCoE data plane traffic is encapsulated in a TRILL frame before being forwarded on a fabric port to another RBridge. The destination RBridge decapsulates the FCoE data frame from the TRILL frame prior to forwarding it on an edge port. All switches in a VCS fabric have an FCF service.

Fibre Channel Protocol Frames

All Fibre Channel protocol frames are encapsulated in an FCoE frame and forwarded inside an Ethernet frame to another RBridge. Fibre Channel protocol frames do not use generic TRILL frames, because they are only exchanged between ports on adjacent switches.

An example frame format destined for a well-known address is shown here.

Destination Address (DA)	Source Address (SA)	Type/Length	Ethernet Payload
ALL-FCF-MAC	*MY-FCF-MAC*	*802.1Q Tag | EtherType*	*FCOE Header | FC Frame*

Figure 43. Fibre Channel frame format.

APPENDIX C: GLOSSARY

Term	Description
ACL	Access Control List
AMPP	Automatic Migration of Port Profiles
BLDP	Brocade Link Discovery Protocol
BPDU	Bridge Protocol Data Unit
Bridging	Layer 2 switching
BUM	Broadcast, unknown unicast, and multicast
CNA	Converged Network Adapter
CLI	Command-Line Interface
CoS	Class of Service, Layer 2 priority marking
DCB/CEE/DCE	Data Center Bridging Ethernet. This is an enhanced version of Ethernet with Priority-based Flow Control (PFC) and congestion notifications.
DCM	Distributed Configuration Management
EAP	Extensible Authentication Protocol
ECMP	Equal-Cost Multipath
Edge Port	Any non-fabric port. Typical service protocols operating on those edge ports include Standard IEEE 802.1Q link-level protocols, and so forth.
ENode	FCoE Node: An FCoE device that supports FCoE VN_Ports. Servers and target devices are ENode devices.
eNS	Ethernet Name Server
ET port	TRILL port, also known as ISL
Fabric	A network with interconnecting individual switching elements and their associated protocols
Fabric port	Ports connecting one switch in a VCS fabric to another switch. Forwarding on these ports is controlled by VCS Fabric protocols.
FC	Fibre Channel
FCF	FCoE Forwarder: An FCoE device that supports FCoE edge ports and/or FCoE VE_Ports. It is the equivalent of an FC switch.
FCoE	Fibre Channel over Ethernet
FCoE VE_Port	The FCoE equivalent of an FC E_Port
FCoE VF_Port	The FCoE equivalent of an FC F_Port. An FCoE VF_Port can occur on an edge port.
FCoE VN_Port	The FCoE equivalent of an FC N_Port
FIP	FCoE Initiation Protocol
FLOGI	Fibre Channel LOGIN
FSPF	Fabric Shortest Path First
IEEE	Institute of Electrical and Electronics Engineers standards organization

Term	Description
IGMP	Internet Group Multicast Protocol
IP	Internet Protocol (see Internet RFC791)
IS-IS	Intermediate System-to-Intermediate System
ISL	Inter-Switch Link, which is a link between switches.
JMP	Join/Merge Protocol
LACP	Link Aggregation Control Protocol
LAG	Link Aggregation Group (see IEEE 802.3ad)
LEP	Link End Point
LLDP	Link-Level Discovery Protocol (see IEEE 802.1AB)
LLS	Logical Link Status TLV
MCT	Multi-Chassis Trunking
MLX	Brocade/Foundry Core L2/L3/MPLS Router
MPIO	Multipath IO
MPLS	Multiprotocol Label Switching
MQC	Modular QoS Command-Line interface
NIC	Network Interface Card
PBF	Policy-Based Forwarding
PFC	Priority Flow Control
P-GID	Policy Group Identifier
PSS	Principal Switch Selection
PVST+/PVRST+	Per-VLAN Spanning Tree Protocol (Cisco proprietary protocol)
QoS	Quality of Service
RB	Routed Bridge as specified in IETF TRILL base protocol specification
RDI/DIA	Request for Domain ID/Domain ID Allocated
TCP	Transmission Control Protocol (see Internet RFC793)
TRILL	Transparent Interconnection of Lots of Links
TTL	Time To Live field in an Ethernet or IP packet
VCS	VCS Fabric technology is proprietary to Brocade.
VCS Node	A physical switch participating in the VCS fabric
VEB	Virtual Ethernet Bridge
VE Interface	Logical Layer 3 Interface
VE_Port	The FCoE equivalent of an FC E_Port
VF_Port	See FCoE VF_Port.
VFID	Brocade VCS Fabric ID
VID	VLAN ID

Term	Description
vLAG	Virtual Link Aggregation Group
VLAN	Virtual LAN
VLAN ID	VLAN Identifier
VM	Virtual Machine
VN_Port	The FCoE equivalent of an FC N_Port
VPLS	Virtual Private LAN Services
VRRP	Virtual Router Redundancy Protocol
VRRP-E	Virtual Router Redundancy Protocol–Extended
VSI	Virtual Station Interface
WKA	Well-Known address

APPENDIX D: RELATED DOCUMENTS

For more information about Brocade VCS Fabric technology, please see the following documents:

The following documents are located on www.brocade.com:

The Brocade *Network OS Administrator's Guide*

The Brocade *Network OS Command Reference*

The Brocade *Fabric OS Administrator's Guide*

The following document is located on my.brocade.com:

The Brocade Network OS Release Notes: my.brocade.com

For more information about the Brocade VDX Series of switches, please visit:

- Brocade VDX 6710 Switch
  www.brocade.com/vdx6710
- Brocade VDX 6720 Switch
  www.brocade.com/vdx6720
- Brocade VDX 6730 Switch
  www.brocade.com/vdx6730
- Brocade VDX 8770 Switch
  www.brocade.com/vdx8770

ABOUT BROCADE

As information becomes increasingly mobile and distributed across the enterprise, today's organizations are transitioning to highly virtualized infrastructure, which often increases overall IT complexity. To simplify this process, organizations must have reliable, flexible network solutions that utilize IT resources whenever and wherever needed—enabling the full advantages of virtualization and cloud computing.

As a global provider of comprehensive networking solutions, Brocade has more than 15 years of experience in delivering Ethernet, storage, and converged networking technologies that are used in the world's most mission-critical environments. Based on the Brocade One® strategy, this unique approach reduces complexity and disruption by removing network layers, simplifying management, and protecting existing technology investments. As a result, organizations can utilize cloud-optimized networks to achieve their goals of non-stop operations in highly virtualized infrastructures where information and applications are available anywhere.

For more information, visit www.brocade.com.