Global Knowledge®

Expert Reference Series of White Papers

# VMware vSphere Distributed Switches

# VMware vSphere Distributed Switches

Rebecca Fitzhugh, VCAP-DCA, VCAP-DCD, VCAP-CIA, VCP-DCV, VCP-DT, VCP-Cloud, Author, Global Knowledge Instructor

## Introduction

There are two types of virtual switches available using vSphere, the vSphere Standard Switch and the vSphere Distributed Switch. The vSphere Standard Switch (vSwitch or vSS) resides in and is manually configured and administered on each ESXi host. The vSphere Distributed Switch (dvSwitch or vDS) provides similar functionality but is centralized to vCenter Server and is more featured. Additionally, the vSphere Distributed Switch requires the use of vSphere Enterprise Plus licensing.

Both virtual switch types support the following features:

- Forwarding of L2 frames
- VLAN segmentation
- 802.1q encapsulation support
- NIC Teaming (support for more than one uplink)
- Outbound (Tx) traffic shaping
- Cisco Discovery Protocol (CDP) support

In addition, the vSphere Distributed Switch supports the following features:

- Datacenter level management
- Network I/O Control
- Traffic Filtering and Marking
- Inbound (Rx) traffic shaping
- Configuration backup and restore
- Private VLANs
- Link aggregation control support
- Port state monitoring
- NetFlow
- Port mirroring

This white paper will cover the vDS architecture as well as an overview of many of the different features that are exclusive to the vSphere Distributed switch.

## Architecture

A vSphere Distributed Switch is an aggregation of per-host virtual switches that are presented and controlled as a single distributed switch at the datacenter level through vCenter Server.

The vSphere Distributed Switch provides centralized management and monitoring of the ESXi host networking components that are associated with the dvSwitch. The purpose of this design is to establish a consistent switch configuration across the ESXi hosts in a virtual datacenter due to the dvSwitch being created and configured at the vCenter Server level and propagated to the ESXi hosts.

A vSphere Distributed Switch is made up of two architectural components—the control plane and the I/O plane:

The control plane exists at the vCenter Server level and is ultimately responsible for configuring and managing the dvSwitch, distributed port groups, uplinks, NIC teaming, PVLANs, and so on.

The I/O plane is a hidden virtual switch that exists on each ESXi host that manages the I/O hardware on the ESXi host and is responsible for forwarding frames to the correct uplink(s). Therefore, in the event that vCenter Server is unavailable, communications will persist.

When a virtual machine is connected to a port on a distributed switch, a folder named .dvsData is created on the datastore on which the virtual machine resides.  However, the .dvsData folder will not be created if no virtual machines on that datastore are attached to a distributed switch. This folder would not exist if the virtual machines are connected only to standard switches. There is at least one subfolder that matches the universally unique identifier (UUID) of a distributed switch. In that subfolder there may be one or more files that correspond to a port ID in which a VM is connected to. This file contains the port state and policy information.

Each distributed switch may have one or more distributed port groups assigned to it. A distributed port group associates multiple ports under a common configuration, defining how a connection is made to the network. A port can connect any networking entity, such as a virtual machine or a VMkernel interface.

## Features

This section will provide an overview to several features that are unique to the vSphere Distributed Switch.

## Network I/O Control

Network I/O Control (NIOC) is a traffic management capability that uses network resource pools to determine bandwidth allocation based I/O shares and limits. Network I/O Control was a feature that was released with vSphere 4.1 and is important in environments where 10 GigE cards are prevalent. Using Network I/O Control assists in facilitating sharing bandwidth by different traffic types across the same physical NIC(s).

When the NIOC feature is enabled, the dvSwitch traffic is divided into the following (system defined) network resource pools: management traffic, iSCSI traffic, NFS traffic, vMotion traffic, VSAN traffic, Fault Tolerance traffic, vSphere Replication traffic, and virtual machine traffic. You also have the ability to create custom (user defined) network resource pools should the system defined network resource pools not fit your exact needs. vSphere 6 introduces the ability to guarantee bandwidth not only at the distributed port group level but also to a vNIC at a virtual machine level.

The configurable options of a network resource pool include shares, limits, and QoS priority tags. Physical adapter shares are assigned to a network resource pool to determine the share value associated to the traffic type affiliated with that pool. Shares only apply when the physical adapter is saturated. A host limit may also be assigned; this is the upper limit of bandwidth that the traffic type with a related network resource pool can use. A QoS priority tag (802.1p) may also be applied to all outgoing traffic for a network resource pool. QoS tags will ensure that the traffic is prioritized properly as it reaches the physical switch.

## Traffic Filtering and Marking

The traffic filtering and marking policy is available with dvSwitches that are version 5.5 or later. This will effectively allow for the creation of Access Control Lists (ACLs) as well as allow the tagging of traffic to pass Quality of Service (QoS) or Differentiated Services Code Point (DSCP) values for network prioritization.

An ACL allows for granular control of what traffic is allowed in or out of a specific VM, set of VMs, or even a port group. The rules are applied on the data path between the vNIC and the distributed port or between the uplink port and the physical NIC. The VMkernel processes these rules, thus eliminating the need for any kind of external application and allowing for faster rule processing.

These rules can be created using the following qualifiers:

- MAC Source Address and Destination Address qualifiers
- IP qualifiers, such as protocol type, IP Source Address, IP Destination Address, port number
- System Traffic qualifiers, such as vMotion traffic, management traffic, FT traffic, and so on

A rule generally consists of one of the aforementioned qualifiers and an action of whether to restrict or prioritize the matching network traffic.

## Private VLANs

A VLAN divides a broadcast domain into multiple logical broadcast domains; a Private VLAN (PVLAN) is an extension to the standard VLAN, further segmenting the logical broadcast domain into "private" groups. Ultimately this functionality allows for the extension of a single VLAN (primary PVLAN) into secondary PVLANs, with these secondary PVLANs residing only within the domain of the primary VLAN.

There is one type of primary PLAN—promiscuous. Any node attached to a promiscuous PVLAN may send and receive network traffic to any node in any secondary PVLAN associated the same primary. The promiscuous PVLANs have the same VLAN ID for both the primary and secondary VLAN.

The secondary PVLANs are broken up into two types, community and isolated. Any node attached to a port in a community secondary PVLAN can send to and receive network traffic from any other port in the same secondary community PLAN, as well as send to and receive network traffic from the promiscuous PVLAN. Any node attached to an isolated secondary PVLAN may only send to and receive network traffic from the promiscuous PVLAN, even if there are other nodes attached to the same isolated secondary PVLAN.

Network traffic between virtual machines that reside on different ESXi hosts but are on the same PVLAN must traverse a physical switch. The physical switch must be PVLAN aware and must be configured so that the secondary PVLANs reach the destination.

## NIC Teaming Policies

vSphere NIC teaming policies enable the distribution or load balance of network traffic across the physical NICs by providing a mechanism that logically binds multiple physical NICs. This will result in greater throughput and availability. There are a few NIC teaming policies that are available using the vSphere Distributed Switch but not the vSphere Standard Switch. These policies include load-based teaming and Link Aggregation Control Protocol support.

vSphere 4.1 introduced a NIC teaming policy called load-based teaming (LBT) or "route based on physical NIC load." This policy is traffic load aware and reshuffles port binding dynamically based on load and the uplink usage in order to make the most efficient use of the bandwidth. Earlier releases provided several load balancing choices, still available today, which base their routing on a source MAC hash, an IP hash, or the originating virtual port ID. While these are valid load-balancing options in most environments, they each have limitations. Each of these policies statically maps the virtual NIC to an affiliated physical NIC rather than based on the current networking traffic. Because of this, these policies may not effectively distribute the network traffic across the uplinks. Load-based teaming can address this shortcoming.

Link Aggregation Control Protocol (LACP) support began with vSphere 5.1 and has been enhanced since then. LACP support on a vSphere Distributed Switch 5.5 allows ESXi hosts to connect to physical switches using dynamic link aggregation. Multiple link aggregation groups (LAGs) can be created to aggregate the physical NIC bandwidth on ESXi hosts connected to LACP port channels. A LAG consists of two or more uplink ports and connects physical NICs to the ports. LAGs are used to increase network redundancy, bandwidth, and load balancing to the port groups. Up to sixty-four LAGs can be created on a vDS and an ESXi host can support up to thirty-two LAGs. LACP configuration for the vSphere 5.1 vDS only supported the IP hash load balancing policy whereas the vSphere 5.5 vDS supports all load-balancing algorithms.

## Monitoring

It's not uncommon for different teams to be managing the virtual switch and physical switch configurations. This can make it very difficult to troubleshoot unless each configuration parameter has been gone through manually. There have been enhancements to the vSphere Distributed Switch over the past few years to address these operational challenges.

vSphere 5.0 introduced support for Link Layer Discovery Protocol (LLDP); earlier version of vSphere had support for Cisco Discovery Protocol (CDP). CDP is supported for both Standard and Distributed vSwitches, whereas LLDP is supported only for Distributed vSwitches. Both discovery protocols provide information about neighbor network devices, such as the device ID, software version, timeout, and so on. CDP is Cisco proprietary, so there are obvious incompatibility issues when using network devices from other vendors. LLDP is a vendor neutral discovery protocol.

The Network Health Check feature was introduced with vSphere 5.1. This feature detects any misconfiguration of VLAN, MTU, and NIC Teaming parameters across the virtual switch and the connected physical switch (access layer switch). When enabled, layer 2 Ethernet frames are exchanged across the ESXi host uplinks each minute to detect misconfiguration. In order for this feature to operate correctly, there should be at least two uplinks configured on the vDS and at least two ESXi hosts using the vDS. This feature is not enabled by default and can only be enabled using the vSphere Web Client.

Port mirroring is the capability of a network switch to send a copy of network traffic seen on one switch port to another switch port that may have a network-monitoring device connected. Port mirroring is sometimes referred to as Switch Port Analyzer (SPAN) on Cisco switches. The vSphere Distributed Switch provides a similar port mirroring function. A port mirroring session is configured with a destination and once configured, the vDS will copy the network traffic to the destination. Port mirroring sessions may be created between virtual machines on the same ESXi host, virtual machines on different ESXi hosts, from a source VLAN to a destination port, from a source port to a destination IP address, or from a source port to a destination uplink. This feature can assist in troubleshooting or debugging network issues in the virtual infrastructure.

NetFlow is a networking protocol that collects IP traffic information as records and sends them to a collector for traffic analysis. It gives visibility into traffic between virtual machines on the same ESXi host, virtual machines on different ESXi hosts, and virtual machine to physical infrastructure network traffic. NetFlow support gives an administrator the ability to monitor network traffic while assisting with network forensics, to include intrusion detection, compliance monitoring, and more. This feature can help to give real insight to the virtual networking piece of the virtual infrastructure.

## Backup and Restore

vSphere Distributed Switch and distributed port group configurations can be exported to a file. This file will preserve all valid network configurations, enabling the ability to restore in case of issue, loss of vCenter Server, or even use this file to distribute the configurations to other deployments.

A new vDS can be created with the configuration settings from the exported file. If distributed port groups were included in the configuration file then those will also be created. Restoring a distributed switch will overwrite the current settings of the distributed switch and its port groups that were included in the exported file. Any port group not part of the configuration file will remain.

This functionality is available using the vSphere Web Client.

## Conclusion

The vSphere Distributed Switch extends the capabilities and features of virtual networks while simplifying configuration, management, and monitoring by centralizing the dvSwitch. Virtual switches can be divided up into two logical sections, the data plane (I/O plane) and the management plane (control plane). Each vSphere Standard Switch contains both the management and data planes, which are configured and managed individually. The vSphere Distributed Switch eases the management burden by treating the network as an aggregated resource, abstracting the virtual switches into a dvSwitch spanning multiple ESXi hosts at a datacenter level. The data plane remains local to each ESXi host but the management is centralized. The vDS also provides enhanced network monitoring and troubleshooting capabilities, like port mirroring, NetFlow, and network health check capabilities. PVLANs, Network I/O Control, and several other features are also made possible by use of the vSphere Distributed Switch.

## Learn More

Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge through training.

VMware vSphere: Install, Configure, Manage [V6.0]

VMware vSphere: Optimize and Scale [V6.0]

Visit **www.globalknowledge.com** or call **1-800-COURSES (1-800-268-7737)** to speak with a Global Knowledge training advisor.

## About the Author

Rebecca Fitzhugh is a VMware Certified Instructor and consultant whose primary focus is on VMware virtual infrastructure products as well as the vCloud and Horizon suites. Prior to becoming an instructor and consultant, she served five years in the United States Marine Corps where she assisted in the build-out and administration of multiple enterprise networks residing on virtual infrastructure. Packt Publishing recently published her book, *vSphere Virtual Machine Management*.