What’s New in VMware vSphere® 5.1 – Networking

VMware vSphere 5.1

TECHNICAL MARKETING DOCUMENTATION
V 1.01 / UPDATED JUNE 2012
# Table of Contents

**Introduction** ................................................................. 3  
**VDS Operational Improvements** ........................................ 3  
  - Network Health Check ................................................ 3  
  - VDS Configuration Backup and Restore ........................... 4  
  - Management Network Rollback and Recovery .................... 4  
  - Distributed Port – Auto Expand .................................... 5  
  - MAC Address Management ......................................... 6  
  - Link Aggregation Control Protocol Support ..................... 6  
  - Bridge Protocol Data Unit Filter .................................. 6  
**Network Monitoring and Troubleshooting Enhancements** ............ 7  
  - Port Mirroring (RSPAN and ERSPAN) ......................... 7  
    - RSPAN ........................................................ 7  
    - ERSPAN ..................................................... 8  
  - IPFIX (NetFlow Version 10) .................................... 9  
  - Enhanced SNMP Support ......................................... 9  
**Other Enhancements** .................................................... 10  
  - Netdump ......................................................... 10  
  - Single Root I/O Virtualization .................................. 10  
  - Scalability Improvements ....................................... 11  
**Conclusion** ........................................................................ 11  
**About the Author** .......................................................... 11
Introduction

With the release of VMware vSphere® 5.1, VMware brings a number of powerful new features and enhancements to the networking capabilities in the VMware vSphere® (vSphere) platform. These new features enable users to manage their virtual switch infrastructure with greater efficiency and confidence. The new capabilities can be categorized into three main areas: operational improvements, monitoring and troubleshooting enhancements, and improved scalability and extensibility of the VMware vSphere® Distributed Switch (VDS) platform.

• Operational improvements made to the VDS platform increase the vSphere administrator’s ability to manage the virtual networking aspect effectively and coordinate with network administrators in case of configuration issues.
• Enhancements to troubleshooting features help network administrators centrally monitor virtual infrastructure traffic through the familiar set of network monitoring and troubleshooting tools.
• Improved scalability and extensibility of the VDS platform help users build larger cloud-scale deployments with the flexibility of hooking up with third party vendors’ network and security services solutions.

VDS Operational Improvements

The VDS is a centrally managed, datacenter-wide switch that provides advanced networking features on the vSphere platform. Having one virtual switch across the entire vSphere environment greatly simplifies management. However, users had faced some operational challenges with VDS with regard to recovering the virtual networking configuration in certain failure situations. These situations included configuration errors on the management network, causing loss of connectivity to the VMware vCenter Server™ (vCenter Server) system, and recovery after a datacenter-wide failure. vSphere 5.1 addresses these operational challenges and also provides new tools and features to help troubleshoot networking issues.

The following are the key features that address the operational aspects of the virtual and physical network infrastructure:

1) Network health check
2) VDS configuration backup and restore
3) Management network rollback and recovery
4) Distributed port – auto expand
5) MAC address management
6) Link Aggregation Control Protocol (LACP) support
7) Bridge Protocol Data unit filter

Network Health Check

The configuration of the network for the vSphere infrastructure is a two-step process whereby users must configure the physical network switches to which the vSphere host is connected and then configure the VDS. There are key network parameters such as VLAN, network adapter teaming and MTU that should be configured properly on physical and virtual switches. However, there are occasions when there are network connectivity issues in the user environment because these parameters are not configured properly. Network connectivity issues caused by configuration errors are often difficult to identify. This particularly is the case in an organization where the network administrators and vSphere administrators take management ownership of, respectively, physical network switches and vSphere hosts. In prior versions of vSphere, there were no tools available that could help resolve such misconfigurations across the physical and virtual switches.
The network health check in vSphere 5.1 monitors the following three network parameters at regular intervals:

1) VLAN
2) MTU
3) Network adapter teaming

The default interval for performing the configuration check is 1 minute. At this regular interval, layer 2 Ethernet probing packets are sent and received across the physical uplink interfaces of the VDS. Depending on the configuration of the network device, which is connected directly to the VDS through the physical uplink interfaces, REQ and ACK packets are received or dropped. When packets are dropped, it indicates that there is a configuration issue, and that warning is displayed in the VMware vSphere® Client™ (vSphere Client) view.

The following requirements must be met for the network health check feature to work:

1) VLAN and MTU check: There must be at least two physical uplinks connected to the VDS.
2) Teaming policy check: There must be at least two active uplinks in the teaming and at least two hosts in the VDS.

VDS Configuration Backup and Restore

As mentioned earlier, VDS configuration is managed through vCenter Server, and all the virtual network configuration details are stored in the VMware vCenter™ (vCenter) database. Previously, in the case of database corruption or loss events, users were not able to recover their network configurations and had to rebuild the virtual networking configuration from scratch. There also was no easy way to replicate the virtual network configuration in another environment or go back to the last working configuration after any accidental changes to virtual networking settings.

The VDS configuration backup and restore feature addresses all of these concerns. Users now can take snapshots of the VDS configuration as well as port group–level configuration. With these backup configurations, they can build a revision control system that tracks and provides controls over changes to virtual network configurations. This enables the user to restore any prior network configuration, including after a vCenter Server database failure situation. These stored VDS configurations also can be used as template to create similar VDS configurations in other environments.

The following operations are supported by means of the configuration backup and restore features:

1) Back up VDS or port group configuration asynchronously on disk. Users can choose to save data locally or on SAN via VMware vSphere® Web Client (vSphere Web Client).
2) Restore VDS or port group configuration from a backup.
3) Create a new entity (VDS or port group) from a backup.
4) Users also can immediately roll back the port group configuration to the previous settings. This is particularly useful when the configuration changes had impact on the operation, and users want to go back to the last configuration before the changes were made.

Management Network Rollback and Recovery

The management network is configured on every host and is used to communicate with vCenter Server and to interact with other hosts during VMware vSphere® High Availability (vSphere HA) configuration and operation. It is critical with regard to centrally managing hosts through vCenter Server. If the management network on the host goes down or there is a misconfiguration, vCenter Server can’t connect to the host and therefore can’t centrally manage the vSphere infrastructure.

In a vSphere standard switch (VSS) environment, users can recover from management network failure on the host by reconfiguring the host management network through a Direct Console User Interface (DCUI).
However, in the VDS environment, where multiple hosts are connected to a distributed switch, any network failure or misconfiguration of the management port group can potentially disconnect all hosts from the vCenter Server system. In this situation, vCenter Server can't centrally make any changes to the VDS port group configuration and push those changes to hosts. In vSphere 5.0, the only way for the user to recover from this situation is by going to individual hosts and building a VSS with a proper management network configuration. After all the hosts' management networks have been reconfigured with a VSS and are able to communicate on the management network, vCenter Server can again manage the hosts and reconfigure the VDS.

To avoid such operational issues of going back to a VSS, some users who don't have physical network interface card limitations on the hosts make use of a VSS for the management network and a VDS for all other virtual infrastructure and virtual machine traffic. In such deployments, users must have at least four network adapters on the hosts, two connected to the VSS and two to the VDS.

The automatic rollback and recovery feature introduced in vSphere 5.1 addresses all user concerns regarding use of the management network on a VDS. First, the automatic rollback feature automatically detects any configuration changes on the management network. If the host can't reach the vCenter Server system, it automatically reverts to the prior, working configuration. Second, users also have an option to reconfigure the management network of the VDS per host through the DCUI.

**Distributed Port – Auto Expand**

A distributed port is a port on a VDS that connects to the VMkernel interface or to a virtual machine's network adapter. Its configuration is determined by the distributed port group settings. As part of these settings, users can choose the number of distributed ports per port group. The port-binding options setting in the port group enables the user to dictate the way distributed ports get created. The following are the three port-binding options:

1) **Static binding:** With this option, vCenter Server immediately creates the number of distributed ports as defined in the distributed port group settings. When a virtual machine is connected to the distributed port group, one of the ports from the created pool gets assigned to the virtual machine. It remains assigned, even when the virtual machine is powered off.

2) **Dynamic binding:** The dynamic option is similar to the static in terms of creation of the distributed ports by vCenter Server. However, the distributed port is allocated only when a powered-on virtual machine is connected to the distributed port group.

   **NOTE:** This option is deprecated and won't be available in future vSphere releases.

3) **Ephemeral binding:** There is no port binding with this choice. When a user chooses this option, the performance is similar to that of a VSS. The number of distributed ports on the distributed port group with the ephemeral binding setting is automatically set to 0. The distributed port group allocates one port for each powered-on virtual machine that gets connected. The maximum number of virtual machines that users can connect to a port group depends on the virtual switch port maximums. According to the vSphere Configuration Maximums guide, [https://www.vmware.com/pdf/vsphere5/r50/vsphere-50-configuration-maximums.pdf](https://www.vmware.com/pdf/vsphere5/r50/vsphere-50-configuration-maximums.pdf), a maximum of 4,096 virtual switch ports per host is possible, and only 1,016 of them can be active at one time on that host.

The static port-binding configuration on a distributed port group helps users to do stateful monitoring of distributed ports. This is not possible with ephemeral and dynamic port-binding configurations, where users lose the visibility and troubleshooting capability.
With releases before the vSphere 5.0 version, users who choose static port binding in their deployment must manually manage how many distributed ports are required per distributed port group. To reduce the management effort involved with the static port-binding option, VMware introduced the Auto Expand feature as part of the vSphere 5.0 release. This capability automatically increases the distributed port limits of the distributed port group when it runs out of distributed ports. To enable this feature, users must employ the vSphere API command or vSphere MOB, as described in VMware knowledge base article 1022312: http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1022312.

In vSphere 5.1, when users choose static port binding, the Auto Expand feature is enabled by default. Users now don’t have to worry about running out of distributed ports on distributed port groups or running a script on distributed port groups to manually enable the Auto Expand feature.

**MAC Address Management**

In the vSphere environment, vCenter Server allocates media access control (MAC) addresses for virtual machine virtual Ethernet adapters. MAC addresses allocated by vCenter Server should be unique in a layer 2 broadcast domain. vSphere 5.0 MAC address management has the following limitations:

1) Each vCenter Server can allocate up to 64K MAC addresses. This limit is too small in the cloud environment.
2) Duplicate MAC addresses are possible with multiple vCenter Server instances that are administered by different IT groups.

To make MAC address management simpler and scalable, vSphere 5.1 provides support for locally administered MAC addresses. This enables users to specify MAC prefix and/or MAC ranges and removes the limitation of using only VMware organizationally unique identifier (OUI)-based allocation. Users now have control over all 48 bits of a MAC address. They also still can use VMware OUI-based allocation.

The limit of 64K MAC addresses per vCenter Server instance is removed, and the duplicate MAC address issue will not arise, if prefix- or range-based allocation is used. These enhancements provide the required simplicity and scalability for the cloud environment.

**Link Aggregation Control Protocol Support**

Link Aggregation Control Protocol (LACP) is a standards-based method to control the bundling of several physical network links together to form a logical channel for increased bandwidth and redundancy purposes. LACP enables a network device to negotiate an automatic bundling of links by sending LACP packets to the peer. As part of the vSphere 5.1 release, VMware now supports this standards-based link aggregation protocol.

This dynamic protocol provides the following advantages over the static link aggregation method supported by previous versions of vSphere:

1) Plug and Play – Automatically configures and negotiates between host and access layer physical switch
2) Dynamic – Detects link failures and cabling mistakes and automatically reconfigures the links

**Bridge Protocol Data Unit Filter**

Bridge Protocol Data Units (BPDUs) are data messages or packets that are exchanged across switches to detect loops in a network. These packets are part of the Spanning Tree Protocol (STP) and are used to discover the network topology. VMware virtual switches (VDS and VSS) do not support STP and therefore do not participate in BPDUs exchange across external physical access switches over the uplinks.
In the vSphere environment, where VMware® ESXi™ (ESXi) host uplinks are connected directly to the physical switch ports, the suggested best practice is to enable PortFast and BPDU guard on the physical switch ports. With the PortFast setting, network convergence on the physical switch ports will take place quickly after power on and any failure because the port will enter the STP forwarding state immediately, bypassing the listening and learning states. This setting minimizes the time it takes for the ESXi hosts connected to the switch ports to come online, thus preventing problems with applications such as DHCP, DNS, PXE etc. The BPDU guard setting, on the other hand, enforces the STP boundary and doesn’t allow any invalid device (bridges/physical switches) connection on the physical switch ports. A device that generates BPDU packets, for example, a physical switch or a virtual machine, is considered an invalid device in this configuration.

VMware virtual switches do not generate BPDU packets. But if a virtual machine sends them, they will be forwarded to the physical switch port over the uplink. When the physical switch port, configured with the BPDU guard setting, detects the packet, that port will be put in err-disabled state. In this err-disabled state, the switch port is completely shut down, which prevents effecting STP. However, the vSphere environment will detect this port failure and will move the virtual machine traffic over another uplink that is connected to another physical switch port. The BPDU packets will be seen on this new physical switch port, and the switch will block that port as well. This ultimately will cause a denial-of-service (DoS) attack situation across the virtual infrastructure cluster.

The BPDU filter feature available in this release enables users to filter the BPDU packets that are generated by virtual machines, preventing any DoS attack situation. This feature is available on VSS and VDS and can be enabled by changing an advanced kernel parameter “net” settings on the ESXi host.

**Network Monitoring and Troubleshooting Enhancements**

To address the network administrator’s need for visibility into virtual infrastructure traffic, VMware introduced the port mirroring and NetFlow features as part of the vSphere 5.0 release. In vSphere 5.1, the port mirroring feature is enhanced through the additional support for RSPAN and ERSPAN capability. The NetFlow feature now supports NetFlow version 10—also referred to as Internet Protocol Flow Information eXport (IPFIX), an IETF standard—rather than the old NetFlow version 5. This release also provides enhancements to SNMP protocol by supporting all three versions (v1, v2c, v3) with enhanced networking MIBs.

**Port Mirroring (RSPAN and ERSPAN)**

Users can employ the RSPAN and ERSPAN features when they want to centrally monitor network traffic and have a sniffer or network analyzer device connected multiple hops away from monitored traffic. This section will briefly describe the traffic flow and configuration requirements for the RSPAN and ERSPAN features.

**RSPAN**

To configure RSPAN, users must first create a dedicated VLAN for the RSPAN session. This VLAN is then configured on the physical switch infrastructure to enable the RSPAN session to be transported across multiple switches. As shown in Figure 1, VLAN 100 is configured as RSPAN VLAN, and a monitor or analyzer device is connected across the two physical switches. The orange broken arrows depict the port mirror traffic flow from the VDS with RSPAN source to a monitor device connected multiple hops away to the physical switch port.
ERSPAN

The encapsulated remote (ER) SPAN feature enables users to monitor the virtual infrastructure traffic across an IP network. Traffic to the remote monitoring device is sent through an encapsulation envelope, also referred to as a GRE tunnel. As shown in Figure 2, the ERSSPAN session is configured from the VDS to a remote monitoring device across an IP network. The configuration steps involve choosing the destination IP of the GRE tunnel. The orange broken arrows show the GRE tunnel established between the two endpoints and the mirror traffic flow.
IPFIX (NetFlow Version 10)

IPFIX, or NetFlow version 10, is the advanced and flexible protocol that enables users to define the NetFlow records that can be collected at the VDS and sent across to a collector tool. The following are some key attributes of the protocol:

1) Users can employ templates to define the records.
2) Template descriptions are communicated by the VDS to the collector engine.
3) IPv6, MPLS, VXLAN flows can be reported.

The vSphere 5.1 release now provides support for the latest version of the NetFlow protocol.

NOTE: *NetFlow version 5 is not supported in vSphere 5.1.*

Enhanced SNMP Support

ESXi hosts provide support for running an SNMP agent. SNMP is a standard protocol that enables a monitoring system to poll agents running on network devices for specific information. The information that devices report depends on the individual agents running on those devices as well as on the management information base (MIB) supported. The SNMP agent available on VMware ESXi 5.0 and earlier releases of VMware® ESX® (ESX) provides support for SNMPv1 and SNMPv2 protocols with VMware MIBs. Examples of VMware MIBs are “VMware-System-MIB” and “VMware-Env-MIB.” In vSphere 5.1, SNMP support is enhanced through the following key capabilities:

1) Better security through support for SNMPv3
2) Support for IEEE/IETF networking MIB modules that provides additional visibility into virtual networking infrastructure
A MIB is a collection of hierarchically organized information about a particular device. With this release, VMware supports MIBs that provide IEEE standard layer 2 monitoring information regarding virtual switches as well as layer 3 and higher protocol (TCP/UDP) MIB support. The following are some key parameters available regarding virtual switches through standard MIBs:

1) Per virtual switch (VSS, VDS) port state (up/down)
2) Traffic counters per port per VLAN
3) LAG data for network adapter teaming details
4) LLDP information for the layer 2 topology discovery

The following are some of the networking MIB modules that are supported in vSphere 5.1:

1) RFC 2863 IF-MIB (Interface Group MIB)
2) IEEE 802.1 BRIDGE-MIB and Q-BRIDGE-MIB
3) IEEE 802.3 LAG-MIB and LLDP-V2-MIB
4) RFC 4293 IP-MIB
5) RFC 4292 IP-FORWARD-MIB
6) RFC 4022 TCP-MIB
7) RFC 4133 UDP-MIB and ENTITY-MIB

Through this enhanced MIB support, network administrators can collect important virtual network data that can help in the monitoring and troubleshooting of network issues. This networking MIB support, along with the NetFlow and port mirroring capabilities of VDS, provides a comprehensive network monitoring and troubleshooting solution for the vSphere platform.

Other Enhancements

The following section provides an overview of the additional features and scalability improvements in vSphere 5.1.

Netdump

Netdump is an ESXi platform capability that redirects VMkernel core dumps to a designated server on the network. It consists of client and server components and utilizes UDP protocol with the server listening on port 6500. On the client side, netdump utilizes the vmknic to send the core dump over the network and also provides capability to VLAN-tag the packets.

In vSphere 5.0, enabling netdump on an ESXi host with the management network configured on a VDS was not allowed. In vSphere 5.1, this limitation has been removed. Users now can configure netdump on ESXi hosts using management network on VDS.

Single Root I/O Virtualization

Single Root I/O Virtualization (SR-IOV) is a standard that enables one PCI Express (PCIe) adapter to be presented as multiple separate logical devices to virtual machines. The hypervisor manages the physical function (PF) while the virtual functions (VFs) are exposed to the virtual machines. In the hypervisor, SR-IOV-capable network devices offer the benefits of direct I/O, which include reduced latency and reduced host CPU utilization. The ESXi platform’s VMDirectPath passthrough functionality provides similar benefits to the user, but it requires a physical adapter per virtual machine. In SR-IOV, this functionality can be provided from a single adapter to multiple virtual machines through virtual functions.
Users who are concerned about network latency can employ this feature and off-load the processing onto the hardware of the PCIe adapter. However, they can’t make use of the VMware vSphere® vMotion® (vMotion), VMware vSphere® Fault Tolerance (VMware FT), and vSphere HA features while using this functionality.

**Scalability Improvements**

The following table lists the new VDS scale numbers for vCenter Server:

<table>
<thead>
<tr>
<th>VDS PROPERTIES</th>
<th>5.0 LIMIT</th>
<th>5.1 LIMIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of VDS per vCenter Server</td>
<td>32</td>
<td>128</td>
</tr>
<tr>
<td>Number of Static Port Groups per vCenter Server</td>
<td>5,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Number of Distributed Ports per vCenter Server</td>
<td>30,000</td>
<td>60,000</td>
</tr>
<tr>
<td>Number of Hosts per VDS</td>
<td>350</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 1. VDS Scale Numbers for vCenter Server

**Conclusion**

The VMware vSphere 5.1 release substantially improves the operational aspect of the virtual network by providing tools such as network health check, which detects misconfiguration across the virtual and physical switch environment. And with the ability to store virtual switch network configuration, users now can recreate their virtual network quickly after failures or a virtual center database corruption scenario. New monitoring and troubleshooting enhancements and SNMP MIB support help network administrators centrally collect all virtual network-related information for monitoring and debugging purposes. And finally, scalability improvements in VDS provide users with additional flexibility to expand their virtual network infrastructure.

**About the Author**

Venky Deshpande is a senior technical marketing manager in the Cloud Infrastructure Product Marketing group at VMware, where his technology focus area is networking. He has been working in the networking industry for more than 15 years and currently is responsible for technical marketing activities related to networking features in the vSphere platform. He is enthusiastic about networking technology, has a strong belief in the software-defined networking (SDN) vision and is excited to be involved in the changes that are occurring in that field.


Follow Venky on Twitter: @VMWNetworking.