Realtime
publishers

The Essentials Series: Best Practices in Virtual Network Management

Correctly Managing Virtual Switches

by Eric Beehler

Copyright Statement

Correctly Managing Virtual Switches

Virtualization is changing the data center. The traditional bounds of network management and server management are becoming blurred as virtual machines are deployed across enterprises to save money and increase efficiencies. The simplicity presented by virtualization—that a virtual machine is no different than a physical machine—only gets you so far. This simplistic view may apply to the operating system (OS) and applications running in a virtual machine, but the mechanics of the system require a deeper look.

Virtual servers are introducing new devices into the typical inventory of managed components. Thus, although virtualization reduces the complexity from one vantage point, it adds new layers for others. What this means for network managers is the need for an understanding of the familiar technology of switches from a new perspective. The switch is the backbone of our networks; virtualization adds a new wrinkle by putting a virtual switch inside the virtual machine. Now the server is not just an endpoint but a complex network device that must be integrated into the entire network solution. In addition, it's not exactly the same as other physical network devices on your network. Rather than ignore this addition to the network environment, a much better option is to manage the virtual switch as a part of your infrastructure.

The Virtual Switch

The key to understanding the VMware ESX virtual switch, also known as a vSwitch, is to understand that the switch is pure software. To communicate with the outside world, the virtual switch uses physical ports on a server, but even that is not necessary. The switch could just allow virtual machines to communicate with each other within the virtual host. The vSwitch itself is necessary because all networking is virtualized to the server. Thus, when looking at a virtual machine, the server's network interface card (NIC) is actually a virtualized software network card whose traffic is either sent through the virtual switch to interface with the internal network of the virtual machine host or with the outside IP infrastructure.

These virtual network components are all part of the kernel of VMware's hypervisor. In fact, any modern hypervisor from other vendors such as Sun, Microsoft, or Citrix have similar setups. So, the normal parameters of the physical switch are now variables in the virtual environment. For example, when managing a 48-port switch, you know there will only ever be 48 ports. To make it any different would require a hardware change of some sort, if it's even possible with that particular switch. In a virtual host, the number of virtual ports on a vSwitch can change. A vSwitch can have a minimum of 8 ports and a maximum of 1016 ports, with a default port count of 52. In fact, the number of vSwitches can change dynamically as well.

However, a virtual switch is a switch, so there are some consistent comparisons to a physical switch. A vSwitch maintains a MAC address table, checks each frame's MAC address for destination, and forwards those frames to avoid unnecessary traffic. There are still ports that network cards connect to; they are just virtual ports and virtual network cards.

VMware Port Groups

The virtual switch handles all traffic inside the virtual host. There are several port groups set up by default in different port groups on the vSwitch. The administration network port group is designed to connect admin-only functions such as the virtual console, remote access (typically known as iLO or DRAC cards on a physical server), iSCSI storage, and the virtual server host itself. Some sensitive data on this network must travel unencrypted, and this network provides the segmentation to accomplish that goal.

There is also the VMotion network port group. VMotion allows machines to move between different virtual hosts. This process moves the storage and memory contents of a virtual machine across the network to another computer, running hot and moving without an interruption of service. The need for security in this instance is obvious. Allowing traffic to traverse a normal network path not only is a possible drain on bandwidth but also introduces the very real potential for a security leak. Best practice is to keep this switch on a private physical network or VLAN.

The latest versions of VMware ESX include the VMKernal network, which carries NFS storage traffic. Finally, the VM network carries all the traffic for the virtual machines to the outside world by connecting to physical ports.

The ports of a virtual host server will not physically correspond to the number of virtual ports on the vSwitch. Configuring a separate physical port for each port group is an option; however, the consideration of having enough NIC ports comes into play. As you've seen, there are at least four distinct and separate networks to support. Some servers may have sufficient capacity to support dual NIC ports per port group, but many will be limited to two or four ports, especially smaller 1U or blade servers. The ports of a NIC do not need IP addresses. Instead, the NICs are placed into bridge mode and the IP addresses are assigned to the service console and the virtual machines attached to each vSwitch. You do not assign IP addresses to vSwitches.

VLANs and Trunks from the Virtual Host

The most common method to address limited port capacity is to configure the ports to support VLANs and apply the best practice of segmentation of the port groups on the vSwitch. VLANs are nothing new to switches. In fact it's a way to provide a form of virtualization inside the switch; but VLAN trunks on servers are still not a common occurrence. Conceptually, you are connecting to a VLAN trunk port on a switch, not just another NIC on a server.

Of course, the trunking protocol used should be selected with intent. Leaving the default in place could bring incompatibilities between the network and server or may not otherwise serve the needs of the network or the virtual host.

The vSwitch supports VLAN tagging and has the option of using the 802.1q standards of external switch tagging (EST), virtual guest tagging (VGT), or virtual switch tagging (VST). External switch tagging sends each network in the trunk over a pair of network cards. This limits the number of networks the trunking can support. VGT requires that the guest OS on the virtual machines handle the trunking. As this requires special drivers and configuration and puts trunking into the guest OS, it's usually avoided.

The most common and recommended method used is VST. This acts as a typical trunk, tagging all the packets and running all port groups through a single set of network ports. This allows you to configure two ports and give the host 2GB of bandwidth on a typical gigabit switch, and simplifies the configuration on the physical network switch. The configuration complexity now falls to the virtual server administrator, requiring them to configure the vSwitch and port groups properly.

Redundancy and Load Balancing to the Physical Network

As with a regular server, the virtual host should be set up with redundant connections to the physical network. Two NICs can be set up in a team in order to provide for the following options:

- Redundant NICs—The virtual host has the option to be set to use redundant NICs, providing failover support.
- Load-balanced NICs—The virtual host can be set to be load balanced inside the NIC teaming settings, spreading the traffic between the NICs and providing failover.

The vSwitch gives the ability to fail over a network connection if more than one physical NIC is connected to it. As with a physical switch, best practice is to connect two virtual NICs to a vSwitch to provide for failover and load balancing. The physical connection is set up in the same way a connection on any production server should be set up—with a separate connection to two physical switches.

Best practice dictates that 2GB ports from two physical NICs be assigned to the VM network for full performance. If there is an issue with one of the switches, the other switch will take over. In fact, if additional pairs of network ports are available, setting up a separate set of physical NIC ports for the VMotion network or IP storage network such as iSCSI will help with bandwidth and security. This also translates into your physical switch, where you can have one physical trunk connected to one switch and the other to a separate physical switch. It's also common to have a separate NIC for the service console network.

The VMotion port group should be separated from other networks to avoid a virtual machine's memory contents. In fact, the VMotion network should be separate from the production network and put onto separate physical switches with its own physical NICs as a whole to avoid bandwidth and security concerns if possible.

Load balancing is often used so that you don't waste a port on just failover but instead provide additional bandwidth. There are several types of load balancing available to the virtual host. The option to route based on the originating virtual port ID is the default and uses the rarely changing port ID on the virtual switch. Routing based on the source MAC hash (out-mac) is another option that uses the MAC address of the VM but is limited to the bandwidth of the virtual NIC and cannot take advantage of physical NIC teaming bandwidth. Routing based on IP hash (out-IP) uses source and destination IP utilizing Etherchannel or 802.3AD bonding, making it able to take advantage of additional bandwidth but also making it more CPU intensive. The default of Port ID is preferred or out-mac in previous versions of ESX. These are preferred because the other option of out-ip is a sequential method based on a fixed volume of packets. If IP load balancing will be used, Etherchannel or 802.3AD must be configured on the physical switch.

vSwitch Limitations Compared with Physical Switches

Virtual switches have several limitations over a standard physical switch. For example, they are not able to operate above layer 2. This serves to isolate networks but also hinders the ability of vSwitches on the same virtual host to talk to one another. Virtual switches cannot be connected to one another. You might just drop an Ethernet cable from one port to another on a physical switch; this is not possible in the virtual switches. This isolates any possible loops that could have occurred, but this limitation brings the issue of routing and connection of different vSwitches to the outside network devices.

Another point to note is the vSwitch does not understand or negotiate the Spanning Tree Protocol. Thus, it is best practice to disable Spanning Tree Protocol on any port connecting virtualized switches on a virtual host and just use PortFast to reduce the downtime associated with failing over to another port if failover options are set up. The forwarding table of a virtual switch is unique to that switch.

Security Features

There are policies that can be applied to virtual switches that allow for some optimization. To monitor the traffic of a physical switch, you can put a port into promiscuous mode and attach a monitor. This can be accomplished on a vSwitch by setting promiscuous mode and passing the traffic to a virtual machine, where a network monitor can be installed. MAC addresses are unique on each virtual server, but like many server network drivers, the MAC can be changed. You can set the security policy of MAC address changes to deny forcing a match of the MAC address received to the MAC address specified in the virtual machine configuration file. Forged transmits is another security option that, when set to deny, will deny outgoing IP traffic if the MAC address doesn't match the configuration file.

Traffic Shaping

The vSwitch does have the ability to limit traffic but in a very limited capacity. These are basic settings that can be applied to a vSwitch on a port group or on a per virtual machine basis in ESX versions before 3.5. Average bandwidth, peak bandwidth, and burst size can all be specified. Average bandwidth can be set between 0Kbps and 102,400Kpbs, which is a 1Gb network connection. This setting allows the burst above the set speed, but overall if you want to limit a port group to half a gigabit, you can. Peak bandwidth will set the upper limit and will never be allowed to rise above that limit. Finally, burst size will limit the size of a burst of data, including not just the data but also the packet size.

So why would someone want to traffic share? Well, the answer depends on the situation, but several examples exist where traffic sharing is used to simulate the effect of WAN links on a server environment. This is one example of possibly many reasons to utilize the feature. For most applications though, this will likely remain disabled in production scenarios.

Detecting Failures

Integrated into the vSwitch are basic features to ensure internal monitoring and management of failure. The first method relied upon is link state, which basically tells the vSwitch if the associated link is up or down. This will initiate any failover that needs to happen based on the configurations of load-balancing or failover settings.

Another option that can help determine a failure under the NIC teaming settings is called *beacon probing*. A beacon is sent down all network paths and if the beacon does not return down one of those paths, the path is then considered dead. Be careful with this setting. If the beacon is trapped by the switch and not returned, the path will be considered dead. Testing with existing network gear should take place before deploying this setting. In fact, there is little reason to use beacon monitoring if network monitoring information is available from other sources. If beacon monitoring will be used, don't connect multiple connections to the same physical switch, which would could cause the same data to be sent to the same device. Consider using it with multiple physical switches that exist in the same broadcast network.

The vSwitch does have the ability to talk the language of the Cisco Discovery Protocol. This protocol is recognized by enabled network gear to identify directly-connected CDP devices. This information may be basic, but it is one of the keys to gaining information on these virtual network devices remotely and automatically. CDP might not be turned on by default, especially if the hosts have been upgraded to ESX 3.5 from a previous version. Once enabled, it will be able to either just listen or listen and advertise. This can greatly reduce the need to translate information from the ESX administrator to the network administrator. Details such as ports, IP addresses, and VLANs will all be visible to network management and monitoring. Previous to 3.5, CDP was not available, but it can now be used to help address the network devices as a whole.

If additional management needs are required, VMware makes an additional product called the vNetwork Distributed switch. This is the framework to manage virtual network components centrally. This requires the data center-centric vSphere and allows the addition of virtual appliances on the virtual host such as the Cisco Nexus 100V virtual switch, which allows for more seamless management between physical and virtual network devices. This is not an integrated feature, so it is a significant cost to an environment to have this integration. This product allows management across VMware servers centrally.

Conclusion

Although a virtual server is really just multiple computers running on top of the same hardware, no virtual server will run without the virtual network components. The vSwitch is a necessary component to configure before any virtual machines can get on the network. It adds a layer of management for the network administrator as well as the virtual server administrator. Knowing what configuration and management options are available is the first step to ensuring network administrators understand and provide proper information to properly configure and manage the network connected to the VM. The network will appear the same when looking at the cables hooked up to any server, but the required configurations between the physical and virtual network are certainly different.