

# Networking's Path to Supporting Server Virtualization

Server virtualization is a growing reality in data centers. The economics are firmly behind the trend. Server virtualization reduces the total cost of ownership by reducing the number of physical servers, requiring less cooling and less power while increasing flexibility. This is all good for the business and the server group but what affect does it have on the management of the network? The truth is that it complicates network management.

There are two big network problems in supporting the dynamic nature of server virtualization. The first problem is configuring VLANs. The most immediate problem for network operations is making sure the VLAN used by the virtual machine (VM) is assigned to the same switch port as the physical server running the VM. One solution is for the server virtualization group to tell network management every possible server the VM can be started on and pre-configure the switch ports. This is not a perfect solution because it can cause the VLAN to be defined on a very large percentage of the switch ports. It can get even more complicated because the server group may not be aware of all the servers that images can be started on, especially during a recovery situation when they are taking emergency measures.

The second problem introduced by server virtualization is assigning QoS and enforcing network policies such as access control lists (ACLs). Traditionally this is done in the network switch connected to the server running the application. With server virtualization this is a software switch running under the hypervisor in the physical server - not the traditional physical network switch that connects to the physical server. It is still important that policy be enforced in the first switch - the software switch. A simple example shows why. Two VMs running on the server are not allowed to communicate with each other. If a hacker got control of VM1 they could open connections to VM2 and draining it of its data. If ACLs are applied by the soft switch in the server then this would be blocked. Before virtualization this is prevented because the applications in VM1 and VM2 would run in different servers and the ACLs defined in the network switch would prevent the communication. Having the policies applied in the software switch maintains the security. The issue is how to get the software to apply the policies or come up with a work around.

Overcoming these two challenges is critical to making server virtualization work smoothly. It would have been nice if the vendor community had worked together to create a uniform standard that works with all the different virtualization vendors. As is normally the case with rapidly growing new technology this did not happen. The industry has implemented four ways to make the port VLAN and applying policies problem manageable but not perfect.

## Virtualization Vendor's Solution

The market leading virtualization vendor is VMware but many other virtualization solutions exist including Citrix's Zen, Microsoft's Hyper-V, KVM and offerings from many other smaller vendors. The most widely available solutions to the networking problems are for VMware and it will be used as the example. The techniques outlined need to be applied to every one of the virtualization solutions but it should be noted in many cases the solutions exist for VMware only.

Figure 1 shows the VMware environment and demonstrates how they solved the problems. vCenter controls the virtualization process and directs where the VMs are started. The hypervisor controls the server and the VMs running on the physical server. vSwitch is a software layer 2 switch provided by VMware. Each VM has a virtual NIC, labeled vNIC. The vNIC uses a

MAC address from either the virtualization vendors pool of MAC addresses or one created and assigned by the enterprise. The diagram is not meant to show all the variations possible in a real environment such as servers with multiple NIC cards, just enough to show how the process works.

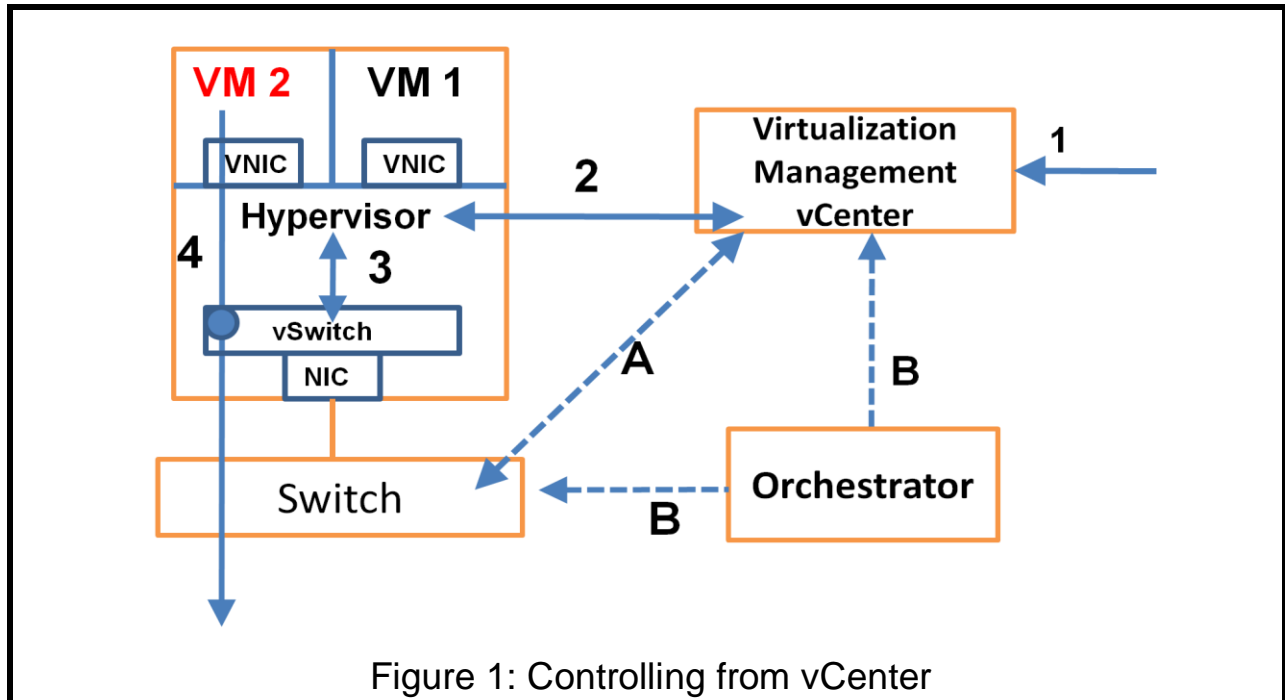


Figure 1: Controlling from vCenter

Lets walk through the process. The first step, labeled 1, is for the server group to define all network characteristics and policies for the VM machines. The operator tells vCenter to start VM2 in step number 2. This process includes multiple messages between vCenter and the hypervisor on the server, one of which pushes the network policy information to the hypervisor. In step 3 the hypervisor then configures the vSwitch with the correct VLAN, QoS and policy information. When the application on VM2 starts to send packets, the policy is applied in the vSwitch, represented by the blue dot.

This solves the problem of applying policies at the first switch but it does not solve the VLAN configuration problem in the network switch. The virtualization groups needs to tell network management to configure the VLAN on the switch port before the VM starts sending traffic which requires quick coordination or the switch has to be preconfigured. The coordination can get more complicated when the virtualization group moves the VM on the fly. Then the virtualization group needs to coordinate with the networking group as it moves the server and the network group needs to clean up the configuration on the old switch after a successful move.

One of the biggest concerns of this approach is the amount of coordination required between the virtualization and network groups. The complication begins because the virtualization group must configure parameters in vCenter that are controlled by the networking group; such as VLAN numbers, QoS and ACLs. This means that good ongoing coordination is needed between the server virtualization group and the networking group. Any change in VLANs or policies must

be immediately reflected in the virtual server configuration which introduces another possible failure point. Another concern is the lack of visibility to what is going on within a networking component, the vSwitch, by the networking group. The vSwitch is under vCenter's control, not traditional network management software. Additionally, the network management has little visibility into the VM. This visibility problem has been reducing by several networking vendors by having vCenter notify it of changes or polling for changes and then displaying this information along with the traditional network data which greatly helps with problem determination.

### **First Answer**

Blade Networks currently has an application that runs on their switch and Force10's next release of their OS addresses the VLAN problem. Their switches poll vCenter looking for any changes, shown as A in Figure 1, or alternatively listens for vCenter to send out a message announcing a change. If the switch finds any changes, such as a new VM2 using VLAN 5 when the network switch port is not configured for VLAN 5, the switch will automatically perform the configuration. The virtualization operator doesn't have to coordinate the change with network operation, allowing start up of the VM to go smoothly. The polling interval does need to be smaller than the time it takes to start a VM to make sure the switch sees the change fast enough. They will also clean up the VLAN definition after a successful move of a VM. In the first release the only parameter Force10 monitors for is VLAN. Blade Network goes further by also applying the full range of policies at the network switch based on the vNIC or VM's UUID. This solution still requires that policies be implemented in the vSwitch.

The second way to solve the configuration problem is with orchestration software from vendors such as HP's DCM and Juniper's Junos Space solutions for their own switches or from management vendors such as Scalent and CA. This is also shown in Figure 1 as the B process. The VLAN and other policies are defined in the orchestration software. The orchestration software talks to both the network switch and vCenter and coordinates changes in configuration between the two environments. This approach has the potential benefit of being able to work with a wide range of switch and virtualization vendors.

### **Cisco's Answer**

Cisco has developed a third way to solve the problem. Cisco provides its own soft switch solution to replace vSwitch called the 1000V. There are two components to the 1000V. The VSM is the Virtual Switch Module and replaces the vSwitch software running inside the hypervisor. The Virtual Element Manager (VEM) is software where the network policies for the VSM are configured and stored.

Figure 2 shows how the process works. The VM's VLAN and policies are first configured based on the VM's UUID or vMAC addresses in the VEM. vCenter starts a new VM or moves a VM in step 2. The Hypervisor informs the 1000V VSM in step 3. The VSM then retrieves the policy information from the VEM in step 4. If the network switch is from the Nexus product line it also retrieves the necessary VLAN and policy information from the VEM. At this point both the switch in the hypervisor and the Nexus switch have all the correct information on how to handle VM2. When VM2 starts sending traffic, step 5, all the correct policies are applied in the 1000V switch in the hypervisor, the blue dot.

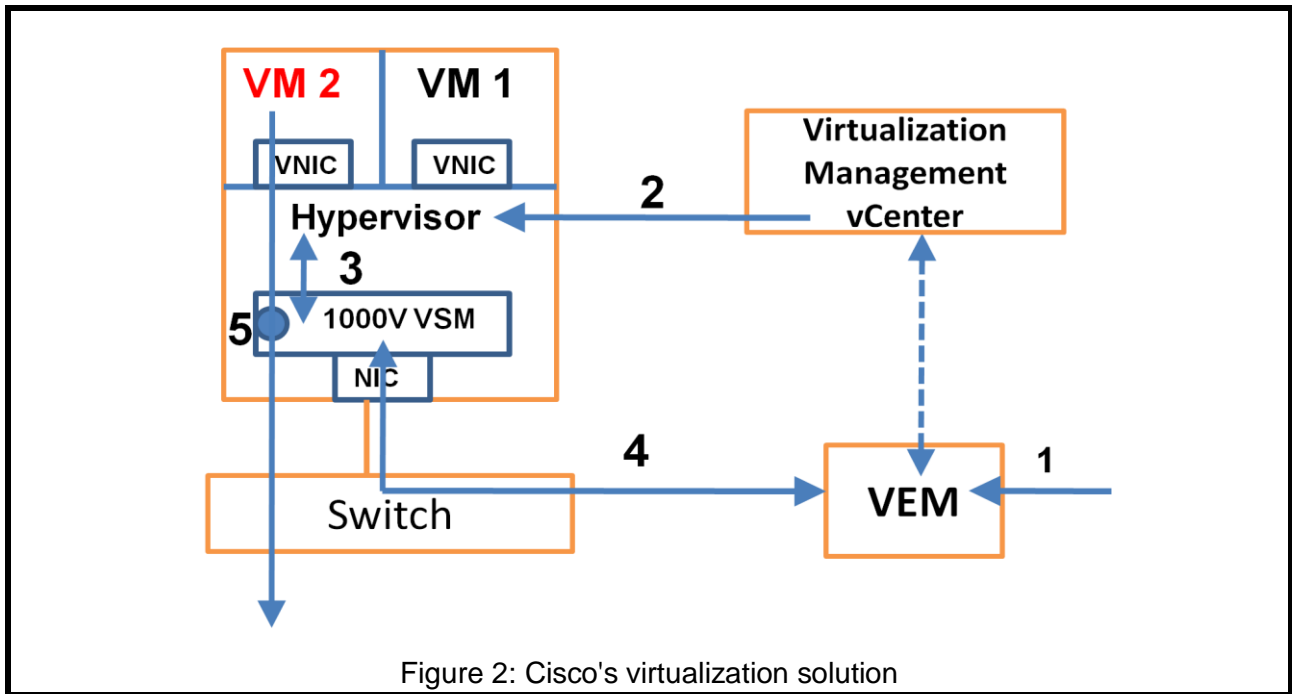


Figure 2: Cisco's virtualization solution

The benefits of Cisco's approach are the same as with the first approach. It blocks any communication between the two VM if it is not allowed and applies the appropriate policies the first time traffic hits a switch. If the 1000V is used with the Nexus switch that is virtualization ready it will solve the VLAN problem in the network switch otherwise the VLAN problem still has to be solved. It has the additional benefit of moving the switch in the hypervisor under the control of the network management software, returning the clear accountability to the network group. There are downsides. First a solution is needed for each virtualization solution. Currently Cisco only has a solution for VMware and not for other solutions such as Zen and HyperV.

### The Fourth Way

The fourth approach takes a network equipment centric view and is shown in Figure 3. In step 1 the VM is defined in the network management software by its virtual NIC. In step 2 vCenter directs the hypervisor to start up VM. The hypervisor sends out an advertisement packet announcing it is starting VM2 in step 3. The advertisement has VM2's vNIC and its UUID. In step 4 the switch sees the advertisement and sends a request for its VLAN and other policy information. The switch then applies the policies to any traffic entering the network, step 4. The key point is the switch only applies policy in the network switch, shown by the blue dot, and not at the vSwitch. The switch also monitors for messages from the hypervisor that indicate the VM has moved and then removes the VLAN and policy information associated with the vNIC. Vendors employing this solution include Arista Networks, Blade and Enterasys along with HP and Juniper through their orchestration approach. Other vendors such as Brocade planning to offer this solution. Extreme Networks uses this technique for QoS and policy but not for VLANs.

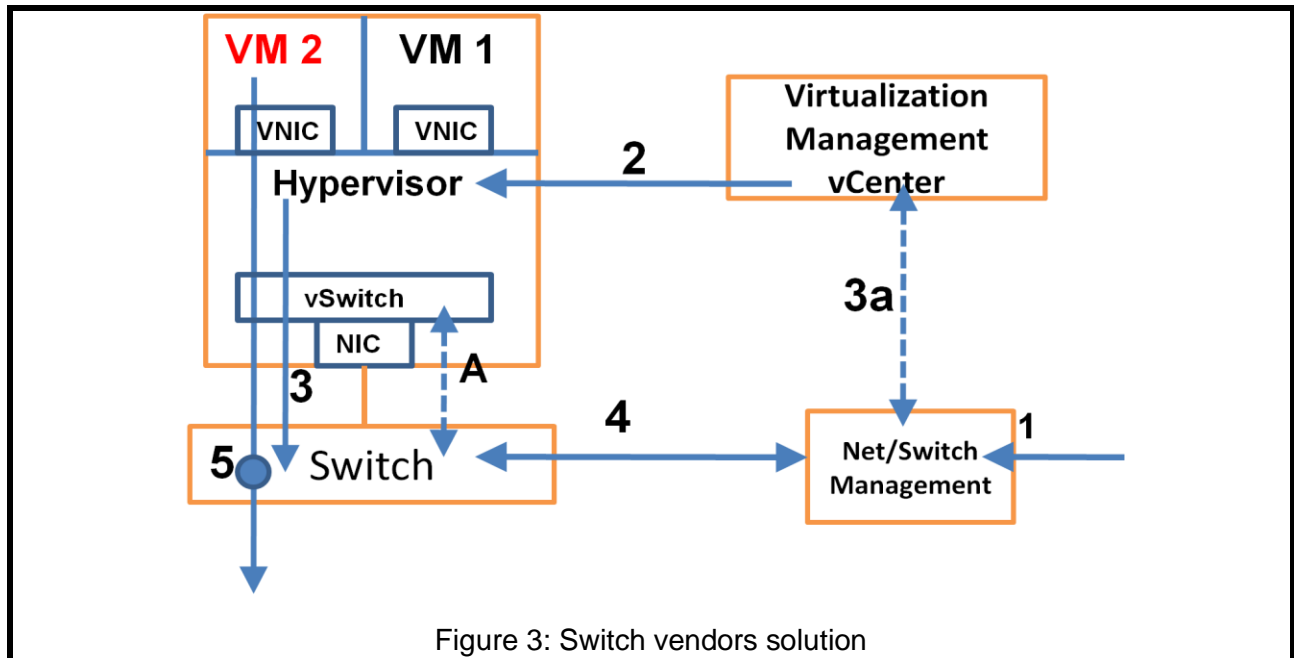


Figure 3: Switch vendors solution

This approach has the benefit of preserving the current way the network operates by removing the need to involve the virtualization group in enforcing network policies. There are two problems. The first is that the server virtualization and networking groups must still coordinate VLAN numbers. Currently Enterasys has the ability to automatically provide the vSwitch with the VLAN number with Arista planning to add it shortly.

The biggest problem with this approach is that it does not apply policies at the vSwitch thus allowing traffic between VMs on the server to bypass ACLs and other security policies. Enterasys and Arista plans to overcome this problem by adding the ability to push the policies down to the vSwitch, shown by A in the diagram.

A future way to overcome this problem is by allowing "hairpin turns" in the switch. With hairpin turns the vSwitch is configured to send all traffic, even VM1 to VM2 traffic, directly to the network switch. The network switch then applies the policies and assigns QoS. The VM1 to VM2 traffic would be sent back to the vSwitch which then delivers it to VM2. This would make vSwitch a "dumb" switch with only a forwarding role. The problem is that 802.1D, the bridging standard that all layer 2 switches are based on, does not allow traffic that came from a port to be sent back down the same port the traffic came from. Thus under the current rules the network switch could not return the packet from VM1 addressed to VM2 since it would break this rule. This was added to prevent loops from being formed. The IEEE is working a revision to 802.1D that allows the switch to perform hairpin turns and additional work is underway to standardize the dumb switch in the hypervisor. When this becomes widespread it addresses the problems with this approach and has the benefit of removing most of the coordination between the network and server group.

Enterasys currently has a work around that directs the vSwitch to put each VM in a separate VLAN. They select VLAN numbers that are not currently being used to prevent any potential problems. Since the VMs are in different VLANs they cannot communicate with each other. When the packet arrives at the network switch the switch replaces the VLAN numbers with the

real one assigned to the VM making it appear to the network and their destination that they have always been in the correct VLAN.

## **Conclusion**

There is another VLAN configuration problem that the techniques outlined above do not address. When a new VLAN number is assigned to a port it needs to be connected to all the other ports with that VLAN number. This requires that all the aggregation switches in the path between all them need to have the VLAN defined on them. A simple example demonstrates the problem. VLAN 5 supports an accounts payable application. All the VMs that support the application are located on one top-of-rack switch which has been configured for VLAN 5. For work load reasons one of the VM's running the accounts payable application is moved to another rack with its own switches in another part of the data center that is reached by way of several aggregation switch. Preserving the VLAN means that all the intermediate switches need to be configured with VLAN5; if they are not then the VLAN is broken. Currently none of the solutions outlined deal with how to automatically assign the VLAN in the aggregation switch. This means that if a VM can move across a data center the VLAN number must be pre-configured in all the aggregation and core switches. Additionally it is not likely that this problem will be solved in the immediate future.

The industry has worked out a set of adequate solutions for the port VLAN and policy problems. There is no magic, most require the networking and virtualization groups to coordinate their activities in the short term to make it work smoothly. The hairpin turn the best long term solution and the industry is moving toward it. It is important that network managers understand how the various solutions work with the different virtualization vendors they support as the full range of solutions are not available for all the virtualization solutions currently in the market.