White Paper

Unify Virtual and Physical Networking with Cisco Virtual Interface Card

Simplicity of Cisco VM-FEX technology and Power of VMware VMDirectPath

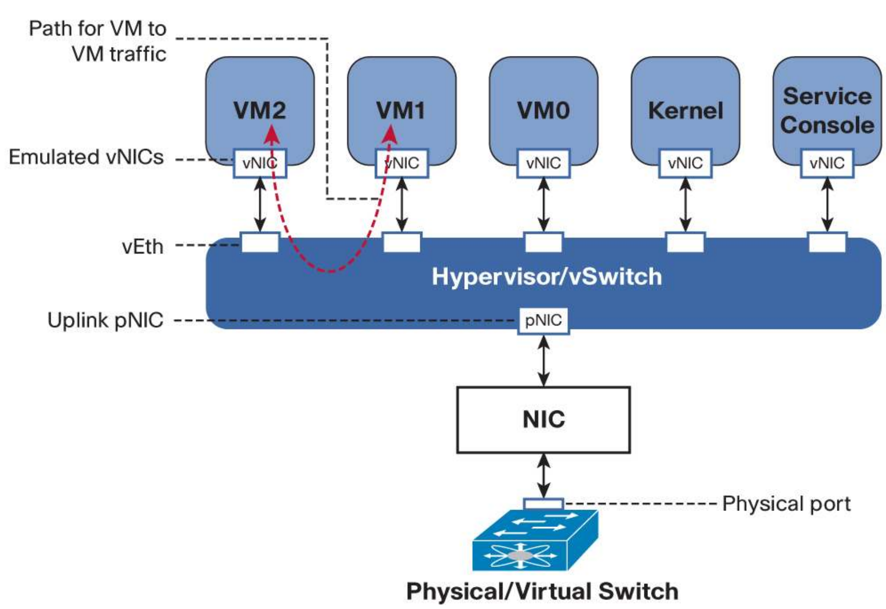
What You Will Learn

Server virtualization has added a new level of network management complexity. Virtual switches within each physical host have increased the number of network management points by an order of magnitude and have challenged the traditional administrator roles and workflows. Virtual switching within the hypervisor also taxes host CPU resources and has associated overheads that affect virtual machine I/O performance.

Cisco® VM-FEX technology addresses the manageability and performance concerns by consolidating the virtual switch and physical switch into a single management point. The number of network management points is dramatically reduced, and physical and virtual network traffic both are treated in a consistent policy driven manner. Cisco Virtual Interface Card also implement VMDirectPath technology from VMware to effectively pass through the hypervisor and significantly reduce associated overhead to improve virtual machine I/O performance.

Typical Virtual Switching Infrastructure

Figure 1. Hypervisor Virtual Switch and Physical Switch Managed Separately and Features and Performance Vary Between Physical and Virtual Switching



Server virtualization introduces the need to support local switching between different virtual machines within the same server. Each virtual machine instantiates a virtual network interface card (vNIC) logically and connects to the virtual switch to send and receive network traffic. If two virtual machines attached to the same virtual switch need to communicate with each other, the virtual switch performs the Layer 2 switching function directly, without sending the traffic to the physical switch. Traffic between virtual machines on a different physical host still needs to go through a physical switch.

This deployment model introduces the following complications:

- **Separate virtual and physical management points:** Virtualization increases the number of network management points. Each virtual switch within each host needs to be managed. With vSphere 4.0 VMware introduced the concept of the distributed virtual network (vDS), which allows a group of virtual switches to be managed together. Distributed virtual switches, such as Nexus 1000v, relieve some of the management pains.
- **Large broadcast domains:** The hypervisor layer inserts opacity between the physical network switch and the virtual machines. The physical switch is unaware of the virtual machine's presence behind the virtual switch and needs to be run in a wide open configuration to accommodate all the possible VLANs that may be associated with the virtual machines behind the virtual switch. All VLANs have to be trunked to every interface on every VMware ESX host. This approach leads to large broadcast domains.
- **Differential feature availability:** Traffic from one virtual machine to another virtual machine on the same host or to a virtual machine on a different host may not be able to use the advanced hardware features such as shaping and load balancing that may be available at the port level on the physical switch.
- **Increased latency and decreased throughput:** The hypervisor adds a layer of abstraction for emulating the virtual network interface card (vNIC), which results in additional latency and loss of throughput. Host CPU cycles are also used to process packets in the emulation layer and the packet forwarding path.

The benefits of server virtualization have helped data center managers overlook the above limitations. The benefits include greater asset utilization, the capability to deploy a service quickly, high availability of applications, and the capability to dynamically rebalance workloads.

The Cisco Unified Computing System™ provides architecture based on Cisco VM-FEX technology helps data center managers address the limitations described earlier while preserving all the main benefits of virtualization.

Introducing the Cisco Virtual Interface Cards

Cisco Virtual Interface Card (VIC) (Figure 2) was developed to provide acceleration for the various new operational modes introduced by server virtualization. The VIC is a highly configurable and self-virtualized adaptor that can create up 128 PCIe endpoints per adapter. These PCIe endpoints are created in the adapter firmware and present fully compliant standard PCIe topology to the host OS or hypervisor.

Figure 2. Cisco Virtual Interface Card

Each of these PCIe endpoints that the VIC creates can be configured individually for the following attributes:

- **Interface type:** Fibre Channel over Ethernet (FCoE), Ethernet, or Dynamic Ethernet interface device
- **Resource maps that are presented to the host:** PCIe base address registers (BARs), and interrupt arrays
- **Network presence and attributes:** Maximum transmission unit (MTU) and VLAN membership
- **QoS parameters:** IEEE 802.1p class, ETS attributes, rate limiting, and shaping

Note: The VIC is capable of Single-Root I/O Virtualization (SR-IOV) at the hardware level, and Cisco will provide a smooth transition to a SR-IOV solution when operating systems and hypervisors support it.

Cisco VM-FEX technology

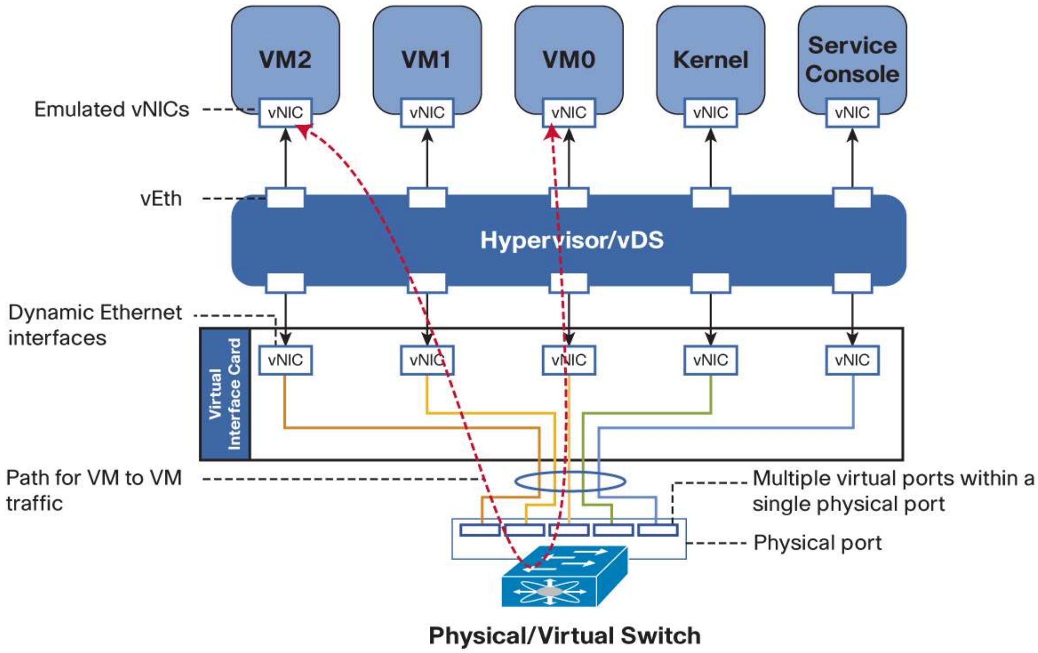
The Virtual Interface Card provides the first implementation of the Cisco VM-FEX technology technology. VM-FEX technology eliminates the virtual switch within the hypervisor by providing individual virtual machine virtual ports on the physical network switch. Virtual machine I/O is sent directly to the upstream physical network switch, in this case, the Cisco UCS 6100 Series Fabric Interconnect, which takes full responsibility for virtual machine switching and policy enforcement.

In a VMware environment, the VIC presents itself as three distinct device types: a Fibre Channel interface, a standard Ethernet interface, and a special Dynamic Ethernet interface. The Fibre Channel and Ethernet interfaces are consumed by standard VMware vmkernel components and provide standard capabilities. The Dynamic Ethernet interfaces are not visible to vmkernel layers and are preserved as raw PCIe devices.

Using the Cisco vDS VMware plug-in and Cisco VM-FEX technology, the VIC provides a solution that is capable of discovering the Dynamic Ethernet interfaces and registering all of them as uplink interfaces for internal consumption of the vDS. As shown in 3, the vDS component on each host discovers the number of uplink interfaces that it has and presents a switch to the virtual machines running on that host. All traffic from an interface on a virtual machine is sent to the corresponding port of the vDS switch. The traffic is mapped immediately to a unique Dynamic Ethernet interface presented by the VIC. This vDS implementation guarantees the 1:1 relationship with a virtual machine interface and an uplink port. The Dynamic Ethernet interface selected is a precise proxy for the virtual machine’s interface.

The Dynamic Ethernet interface presented by the VIC has a corresponding virtual port on the upstream network switch, the Cisco UCS fabric interconnect (Figure 3).

Figure 3. Each Virtual Machine Interface Has Its Own Virtual Port on the Physical Switch



Cisco UCS Manager running on the Cisco UCS fabric interconnect works in conjunction with VMware vCenter software to coordinate the creation and movement of virtual machines. Port profiles are used to describe the virtual machine interface attributes such as VLAN, port security, rate limiting, and QoS marking. Port profiles are managed and configured by network administrators using Cisco UCS Manager. To facilitate integration with the VMware vCenter, Cisco UCS Manager pushes the catalog of port profiles into VMware vCenter, where they are represented as distinct port groups. This integration allows virtual machine administrators to simply select from a menu of port profiles as they create virtual machines. When a virtual machine is created or moved to a different host, it communicates its port group to the Virtual Interface Card. The VIC asks Cisco UCS Manager for the port profile corresponding to the requested profile, and the virtual port on the fabric interconnect switch is configured according to the attributes defined in the port profile.

Cisco VM-FEX technology addresses the concerns raised by server virtualization and virtual networking and provides the following benefits:

- **Unified virtual and physical networking:** Cisco VM-FEX technology consolidates the virtual network and physical network into a single switching point that has a single management point. Using Cisco VM-FEX technology, the number of network management points can be reduced by an order of magnitude.
- **Consistent performance and feature availability:** All traffic in now switched at the physical switch, leading to consistent treatment for all network traffic, virtual or physical. Each virtual machine interface is coupled with a unique interface on the physical switch, which allows precise decisions to be made related to the scheduling of and operations on flows to and from a virtual machine.

- **Reduced broadcast domains:** The virtual machine's identity and positioning information is now known to the physical switch, so the network configuration can be precise and specific to the port in question.

The Cisco VM-FEX technology implementation inserts itself as a completely transparent vDS implementation. No additional training cycles or best-practices guides are needed beyond those for the vDS framework and implementation that was introduced in VMware vSphere 4.0.

Modes of Operations for VM-FEX technology

Cisco VM-FEX technology supports virtual machine interfaces run in two different modes:

- Emulated mode
- PCIe Pass-Through or VMDirectPath mode

Emulated Mode

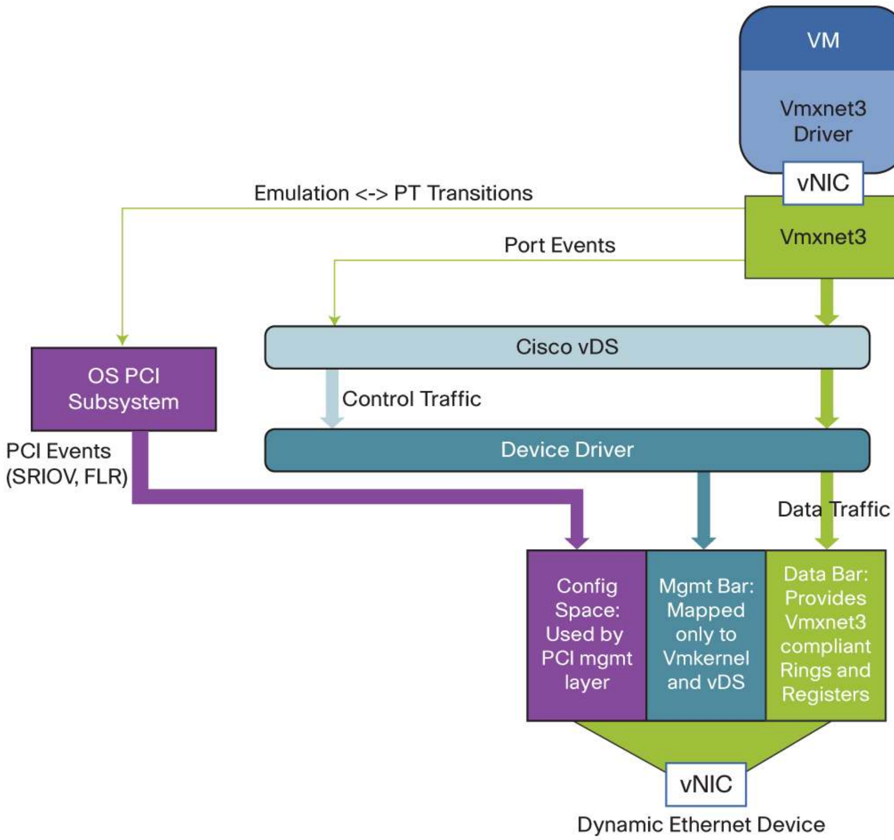
The hypervisor emulates a NIC (sometimes called a back-end emulated device) to replicate the hardware it virtualizes for the guest virtual machine. The emulated device presents descriptors, for read and write, and interrupts to the guest virtual machine just as a real hardware NIC device would. One such NIC device that VMware ESX emulates is the vmxnet3 device. The guest OS in turn instantiates a device driver for the emulated NIC. All the resources of the emulated devices host interface are mapped to the guest OS's address space.

The emulated NIC implementation within the hypervisor has two primary tasks:

- **Control path:** The implementation maintains and conveys the control information about the emulated interface to the virtual switching layer: information such as the MAC address of the vNIC as provisioned by VMware vSphere, the MAC address filters that will be implemented, and the connected or disconnected state of the vNIC to the virtual switch.
- **Data path:** The implementation within the hypervisor is also responsible for taking the data chains that are presented to it by the guest and packaging them so that they can be conveyed through the virtual switching layer and be placed on descriptors for transmission by the physical NIC itself. The receive path is a close approximation of the transmission path.

In emulated mode of VM-FEX technology, port events are used to create a 1:1 association between the emulated NIC implementation within the hypervisor and the Dynamic Ethernet interface that will be used as its proxy. The Cisco vDS component listens to all port events and maintains the coherency of this 1:1 association. Additionally, after association the control path is set up to track the virtual machine interface and can affect all change and control using the corresponding ports on the physical switch as shown in Figure 3 Each Virtual Machine Interface Has Its Own Virtual Port on the Physical Switch.

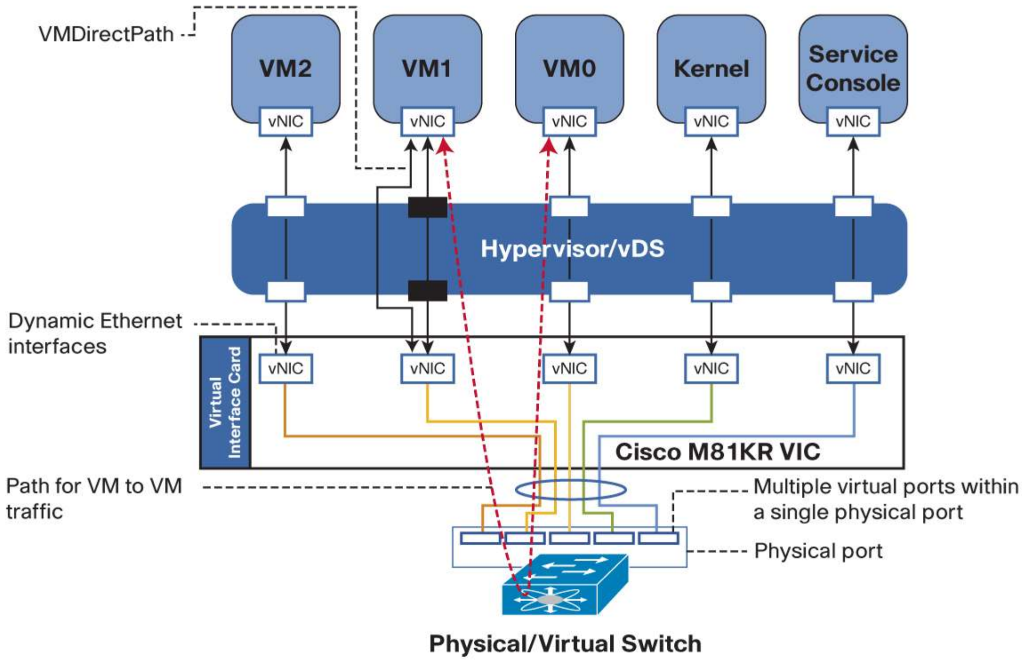
Figure 4. Major System Components for Cisco VM-FEX technology Using Emulated Mode



VMDirectPath or Pass-Through Mode

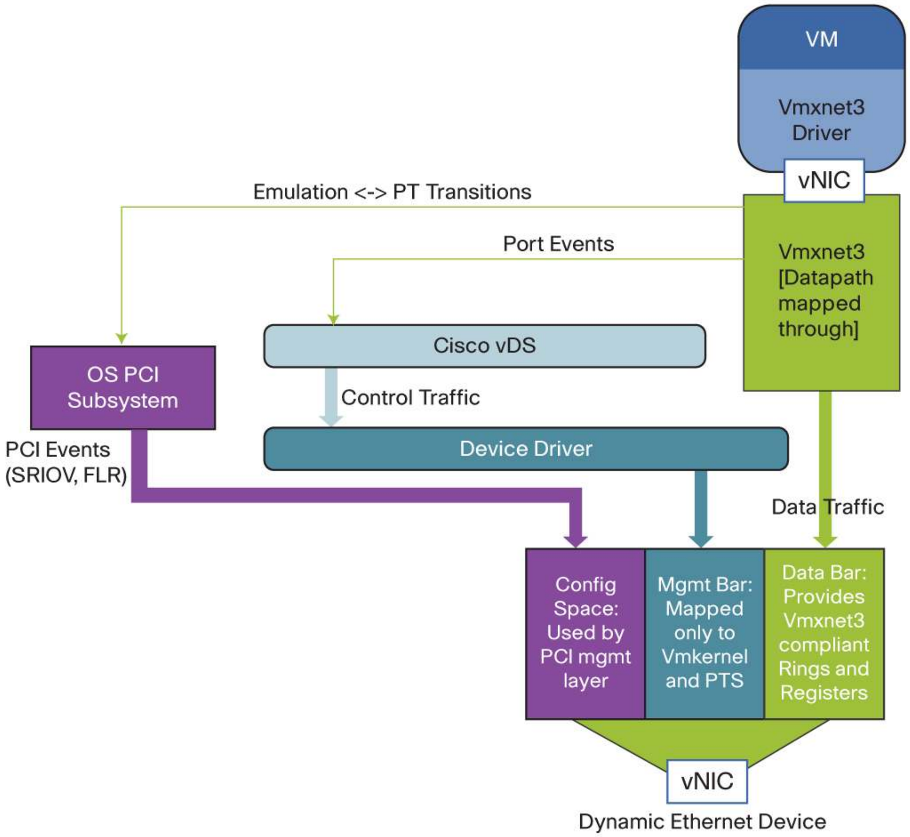
Virtual Interface Card uses PCIe standards-compliant IOMMU technology from Intel and VMware's VMDriectPath technology to implement PCIe Pass-Through across the hypervisor layer and eliminate the associated I/O overhead (Figure 5). Pass-Through mode can be requested in the port profile associated with the interface using the "high-performance" attribute.

Figure 5. Building VMDirectPath on Cisco VM-FEX technology



The 1:1 association between emulated NICs within the guest and the dynamic Ethernet interfaces is established using the emulated mode described earlier. The VIC can then migrate the data path from internally emulated rings and registers to a uniquely assigned Dynamic Ethernet interface's physical registers and rings on an instruction boundary (Figure 6). The VIC does not need to insert any shell or device-specific translation code into the guest virtual machine. This behavior is possible because the Dynamic Ethernet interfaces presented by the VIC can export the exact descriptor format and control registers of the vmxnet3 device, one of the devices the ESX emulates.

Figure 6. Major System Components Showing the VMDirectPath Model



If high performance is requested on an interface and the vmkernel deems the interface ready, the vmkernel triggers a transition from emulated mode to PCIe Pass-Through mode. As part of the transition, the Dynamic Ethernet interface selected by the Cisco VM-FEX technology driver will be paused, and a PCIe reset will be applied. Then the vmxnet3 emulated device implementation in the hypervisor is quiesced and checkpointed. As a result of the checkpoint, the complete state of the network interface is saved. The state of the emulated device is then transferred to the Dynamic Ethernet interface presented by the VIC, and the virtual machine I/O operations are continued. While in PCIe Pass-Through mode, if any events need handling, vmkernel will reverse the process and put the interface into emulated mode again and handle the event.

With ESX 5.0 VMware has introduced the support of vMotion with VMDirectPath. As VMware vMotion is triggered in the VMDirectPath mode, the vmkernel quiesces and checkpoints the Dynamic Ethernet interface presented by the VIC. The state is then transferred to the vmxnet3 emulated device in the hypervisor before VMware vMotion processing proceeds. If any other events need handling, vmkernel follows the same process to transfer the interface from VMDirectPath mode to emulated mode before handling the event.

Conclusion

Cisco® VM-FEX technology addresses the primary networking concerns that arise due to server virtualization. It simplifies networking infrastructure by unifying the virtual and physical network, increases the integrity of the networking infrastructure and provides network administrators full visibility of the virtual machine traffic without increasing the number of management or monitoring points. Virtual Machine I/O performance is also boosted with VM-FEX technology allowing data center customers to enjoy the full benefits of virtualization, including dynamic workload management, without any compromises.

For More Information

For more information about the Cisco Virtual Interface Card visit http://www.cisco.com/en/US/products/ps10331/index.html or contact your local account representative.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-618838-00 06/11