White Paper

Cisco Data Center Virtual Machine Fabric Extender (VM-FEX) Versus VMware vSwitch

Performance Comparison

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Contents

Key Findings ... 3

Introduction ... 3

TCP Stream Performance Results ... 5

Uni-directional TCP Performance ... 5
Bi-directional TCP Performance ... 5

TCP Transactional Performance ... 7

TCP Latency Performance ... 7
TCP Connect/Close Performance ... 8
TCP Connect/Request/Response/Close Performance ... 9

System Configuration ... 10

Fabric Topology ... 10
Hardware Configuration ... 11
System Settings ... 11

Performance Evaluation Tools ... 13

Uni-directional TCP Traffic ... 13
Bi-directional TCP Traffic ... 13
TCP Transactions ... 13

Conclusion ... 13

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Key Findings

This paper presents a networking performance comparison between Cisco Data Center Virtual Machine Fabric Extender (VM-FEX) and VMware vSwitch network connectivity technologies using the Cisco UCS® Virtual Interface Card (VIC) 1240 on Cisco UCS B200 M3 Blade Server. The following observations are presented:

- Cisco VM-FEX technology can transmit or receive 9.8 Gbps of uni-directional TCP network throughput while utilizing 44.80 percent system CPU for transmit and 65.60 percent system CPU for receive.
- Cisco VM-FEX uses 16 percent lower system CPU for transmit and 30.4 percent lower system CPU for receive compared to VMware vSwitch for the same amount of bandwidth.
- Cisco VM-FEX uses 65.60 percent of system CPU for transmit and receive while driving 10.89 Gbps of bi-directional TCP network throughput compared to VMware vSwitch, which uses 81.60 percent of system CPU while driving only 7.97 Gbps.
- Cisco VM-FEX takes 36 percent less time for an average round trip compared to VMware vSwitch.
- Cisco VM-FEX offers over 40 percent reduction in latency, compared to VMware vSwitch

Introduction

Cisco VM-FEX technology is a Cisco innovation that allows VMs to bypass the hypervisor networking stack and access the network directly.

Cisco VM-FEX utilizes the capability to create multiple vNICs in combination with VMware VMDirectPath and Intel® VT-d technologies. This, in turn, allows the VMs to bypass the hypervisor for their networking connectivity by allowing direct access to the underlying adapter hardware (see Figure 1). This approach avoids the overhead of the hypervisor software networking stack, resulting in lower system CPU utilization and higher networking throughput.

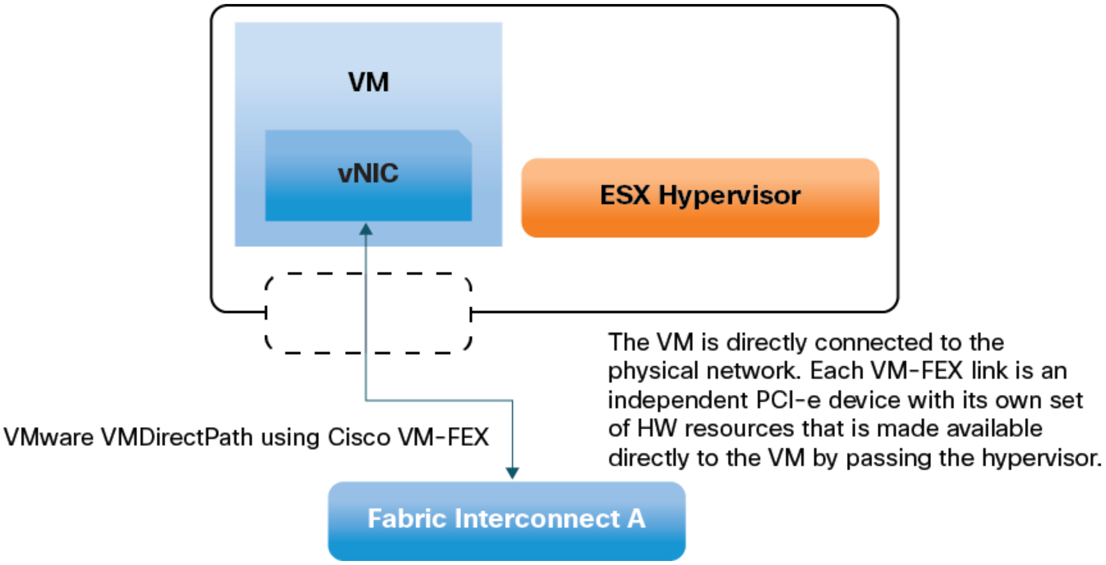
Cisco VM-FEX therefore enables the VMs to support higher networking traffic capacity and be more responsive, especially where TCP networking is used and/or the application is CPU-bound.

Cisco VM-FEX uses Cisco UCS Virtual Interface Card 1240 for hardware connectivity. Cisco UCS 1240 VIC is a 4 x 10 Gbps-capable networking adapter.

The Cisco VIC can also be used with vSwitch. The VIC is capable of supporting multiple, independent Peripheral Component Interconnect Express (PCIe) devices. They user can create multiple virtual network interface cards (vNICs) (in this case, PCIe devices) and associate them with one or more vSwitches to distribute interrupt load across multiple CPU cores if so desired. Cisco VM-FEX, however, allows for directly attaching these independent PCIe devices (vNICs) into the VM by bypassing the hypervisor networking stack.

The Cisco VIC is therefore a uniquely flexible networking adapter that offers both scalability and performance without compromise.

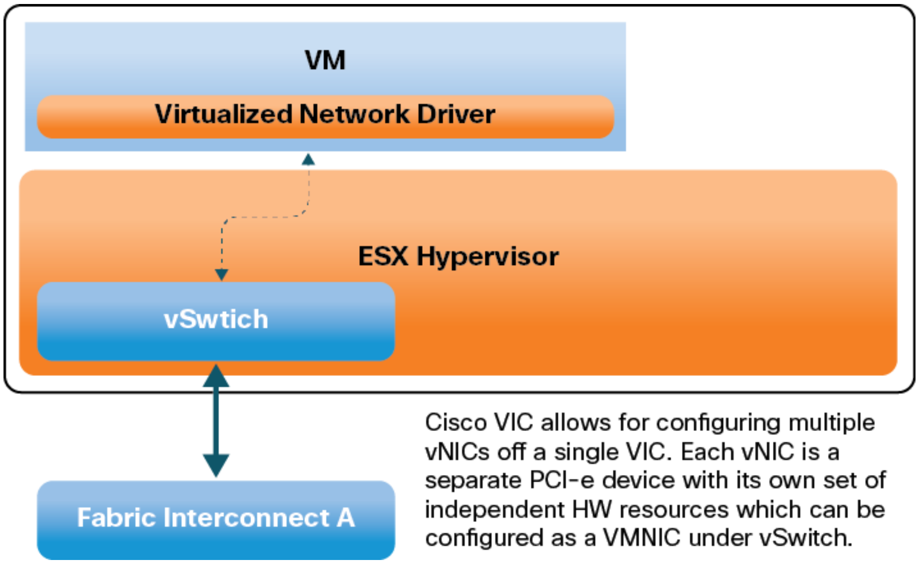
Figure 1. VM Network Connectivity with Cisco VM-FEX



The VM is directly connected to the physical network. Each VM-FEX link is an independent PCI-e device with its own set of HW resources that is made available directly to the VM by passing the hypervisor.

For the purpose of the benchmarking effort presented in this paper, only one vNIC was used for the vSwitch under consideration. Our goal was a simple, straightforward yet fair performance comparison between a VM using a Cisco VM-FEX derived vNIC and a VM using a VMware vSwitch derived vNIC (see Figure 2).

Figure 2. VM Network Connectivity with VMware vSwitch



Cisco VIC allows for configuring multiple vNICs off a single VIC. Each vNIC is a separate PCI-e device with its own set of independent HW resources which can be configured as a VMNIC under vSwitch.

This paper focuses on networking performance with a single VM configured with a single vCPU and a single virtual NIC (vNIC). In the case of Cisco VM-FEX the vNIC is an actual hardware PCIe device plumbed directly into the VM. In the case of VMware vSwitch the vNIC is a virtualized network driver (VMXnet3).

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

TCP Stream Performance Results

The VMware esxtop tool presents CPU utilization for each individual CPU core, as well as all the cores in the system. Core utilization percent is the percentage of an individual CPU core that is used. A core utilization percent value of less than or equal to 100 percent denotes a single core and 1600 percent denotes all 16 cores.

Uni-directional TCP Performance

In this test, a single TCP stream is sent from the source VM to the destination VM.

Transmit CPU utilization (TX CPU%) was captured on the hypervisor hosting the source VM and receive CPU utilization (RX CPU%) was captured on the hypervisor hosting the destination VM (see Table 1).

For TCP transmit, Cisco VM-FEX consumes 16 percent less CPU when compared to VMware vSwitch. For the same stream on the receive side, Cisco VM-FEX consumes 30.4 percent less CPU when compared to vSwitch (see Figure 3).

Table 1. Uni-directional TCP Performance

Uni-directional TCP Performance (Single vCPU)—8192B Payload / 9000B MTU					
	TX CPU%	**Core Utilization%**	**RX CPU%**	**Core Utilization%**	**Gbps**
VMware vSwitch	3.80%	60.80%	6.00%	96.00%	9.80
Cisco VM-FEX	2.80%	44.80%	4.10%	65.60%	9.80

Figure 3. CPU Utilization Difference between VM-FEX and vSwitch for Uni-directional TCP

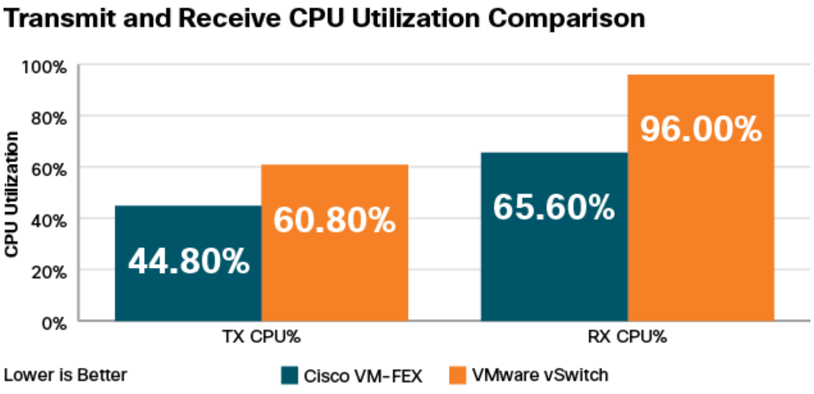


Figure 3 shows the CPU utilization difference between VM-FEX and vSwitch. For the same amount of network bandwidth, VM-FEX clearly consumes less CPU. This is simply because with VM-FEX, the traffic stream does not have to traverse the hypervisor networking stack on the sender or on the receiver. By avoiding the software overhead while performing direct memory access (DMA) from the VM to the hardware vNIC, VM-FEX can save valuable CPU cycles.

Bi-directional TCP Performance

In this test, TCP streams were sent to and from the source VM simultaneously. Bi-directional TCP traffic requires more CPU than uni-directional TCP traffic.

In this test, both transmit CPU utilization (TX CPU%) and receive CPU utilization (RX CPU%) were captured on the hypervisor hosting the source VM (see Table 2).

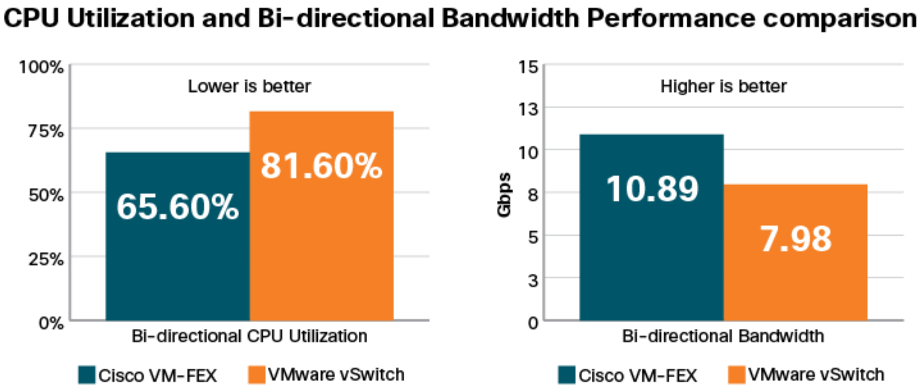
© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

The bandwidth metric was captured from within the source VM and verified against the esxtop network view of the hypervisor (Figure 4). The RTT (round trip time) was captured within the source VM (Figure 5).

Table 2. Bi-directional CPU Performance

Bi-directional TCP Performance (Single vCPU)—8192B Payload / 9000B MTU				
	TX/RX CPU%	**Core Utilization%**	**Gbps**	**Avg RTT (usecs)**
VMware vSwitch	5.10%	81.60%	7.97	542.81
Cisco VM-FEX	4.10%	65.60%	10.89	397.28

Figure 4. CPU Utilization Difference between VM-FEX and vSwitch for Bi-directional TCP



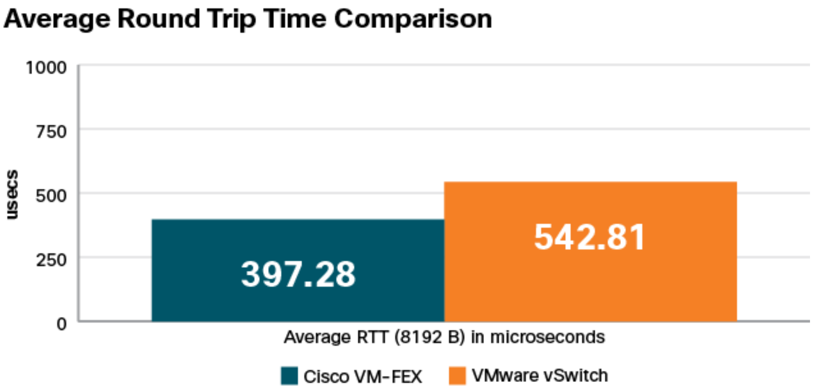
Cisco VM-FEX consumes less CPU while delivering higher bandwidth.

From a CPU utilization perspective, Cisco VM-FEX consumes 65.60 percent of a single CPU compared to 81.60 percent consumed by VMware vSwitch for the same workload.

And Cisco VM-FEX can deliver up to 10.89 Gbps of bi-directional bandwidth compared to VMware vSwitch which can deliver up to 7.97 Gbps while consuming more CPU.

In addition, as Figure 5 shows, the average RTT of a request/response of 8192B TCP packet with Cisco VM-FEX is significantly lower when compared to VMware vSwitch. This is the cumulative side effect of using fewer CPU cycles and bypassing the hypervisor networking stack for traffic flows.

Figure 5. RTT Difference between VM-FEX and vSwitch



© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

TCP Transactional Performance

TCP Latency Performance

Both the source and destination VMs were configured for low latency (see the section VM Settings).

The actual latency numbers were captured inside the source VM. The netperf TCP_RR test was used to derive the latency numbers.

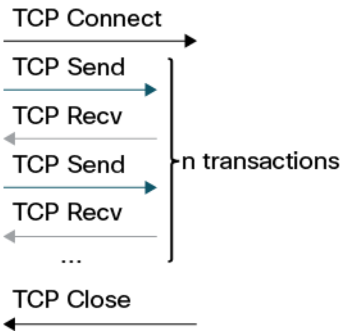
The netperf tool reports total number of such send/receive transactions per second. With the number of TCP_RR transactions:

```
    Latency = (1000000 / TCP_RR_transactions) / 2
```

TCP_RR is a request/response test where the source VM sends a packet to the destination VM and waits to receive the packet before re-sending the same packet. Each send/receive operation is a single TCP_RR transaction.

The number of microseconds in a second is 1000000. Dividing (1 million / TCP_RR transactions) will give the RTT of a single transaction in microseconds. Further dividing the number by 2 will give the one way latency of a single transaction (Figure 6).

Figure 6. TCP_RR Test



The TCP_RR test reports on multiple send/receive operations over single persistent TCP connection. Table 3 shows the performance results.

Table 3. Latency Results for VM-FEX and vSwitch

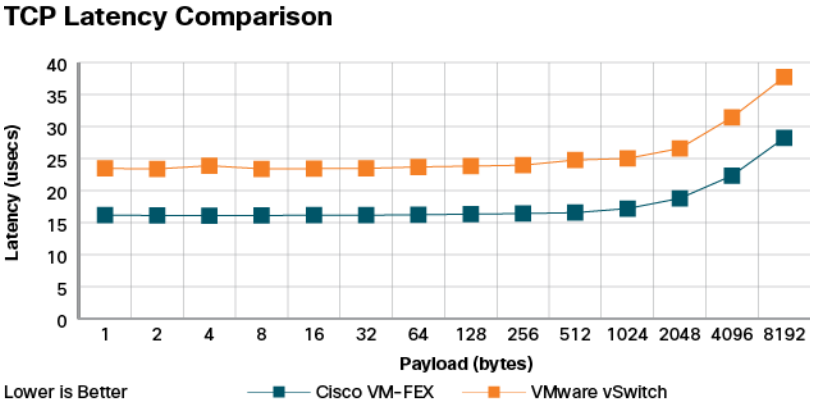
Payload (Bytes)	Cisco VM-FEX (usecs)	VMware vSwitch (usecs)
1	16.20	23.47
2	16.12	23.40
4	16.11	23.90
8	16.12	23.41
16	16.18	23.44
32	16.17	23.48
64	16.24	23.71
128	16.34	23.84
256	16.43	23.98

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Payload (Bytes)	Cisco VM-FEX (usecs)	VMware vSwitch (usecs)
512	16.57	24.79
1024	17.21	25.03
2048	18.80	26.61
4096	22.35	31.46
8192	28.23	37.72

Across the board, Cisco VM-FEX delivers lower latency compared to VMware vSwitch. With packet payloads of 1 byte through 512 bytes, Cisco VM-FEX offers up to 46 percent lower latency (Figure 7).

Figure 7. TCP Latency Comparison between VM-FEX and vSwitch

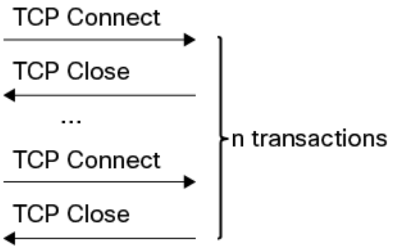


TCP Connect/Close Performance

TCP connect/close transaction results were derived using netperf TCP_CC test. This test reports the number of TCP connect/close transactions per second (Figure 8). There are no request/response operations within the connect/close operations.

```
TCP_CC Latency = (1000000 / Number of TCP_CC transactions) / 2.
```

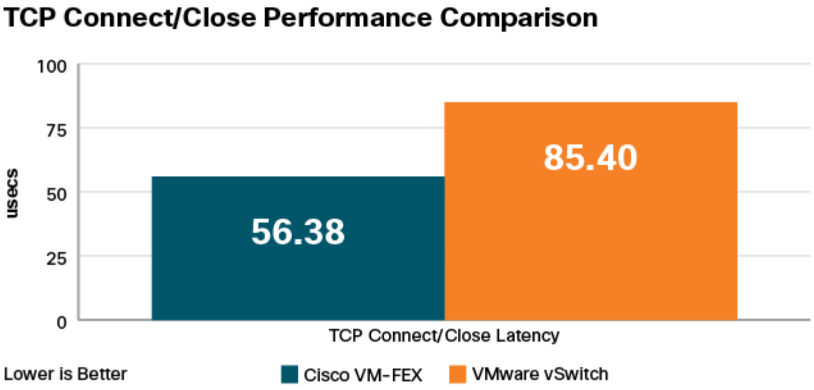
Figure 8. TCP_CC Test



TCP_CC test shows the performance results for TCP connection setup and close (Figure 9).

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Figure 9. Connect/Close Performance Comparison between VM-FEX and vSwitch

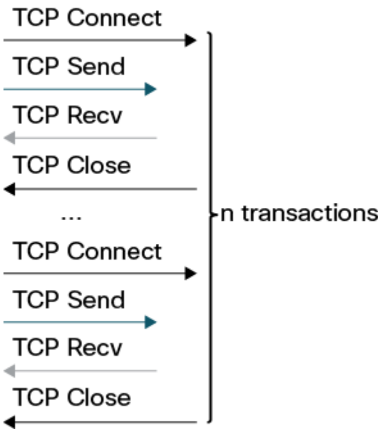


TCP Connect/Request/Response/Close Performance

TCP Connect/Request/Response/Close results were derived using netperf TCP_CRR test. This test reports the number of TCP connect/request/response/close transactions per second (Figure 10).

```
TCP_CRR Latency = (1000000 / Number of TCP_CC transactions) / 2.
```

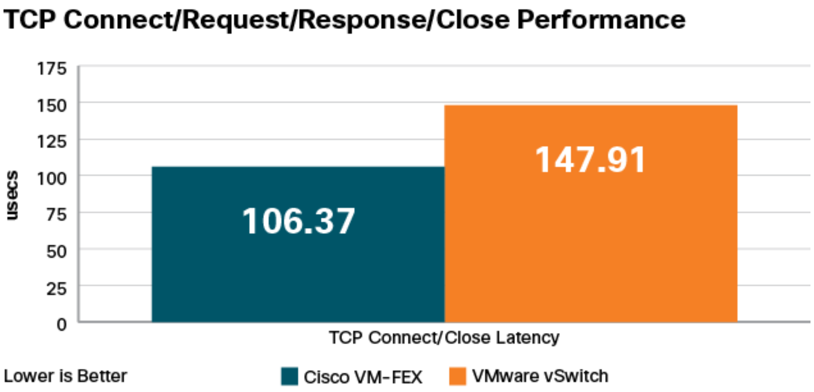
Figure 10. TCP_CRR Test



TCP_CRR test shows the times for TCP connection setup, request send, response received and connection close (Figure 11). The TCP_CRR test is similar to what happens with HTTP.

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Figure 11. TCP Connect/Request/Response/Close Performance Comparison between VM-FEX and vSwitch

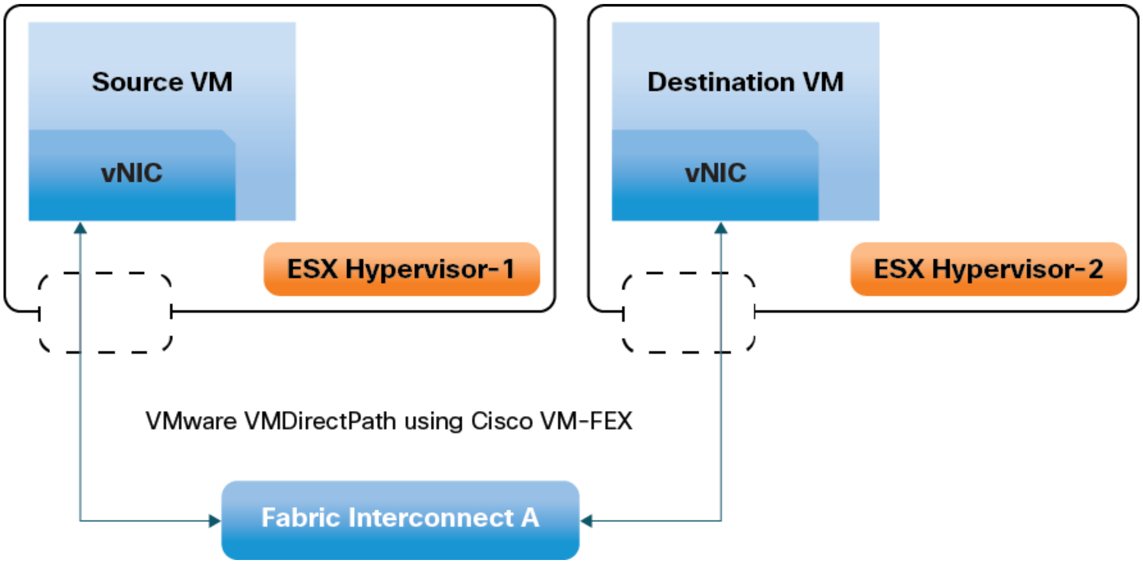


System Configuration

Fabric Topology

For both vSwitch and VM-FEX, two separate ESX hosts were each configured with a source and destination VM. Both the VMs were isolated on a common physical Layer 2 network, as shown in Figure 12 and 13.

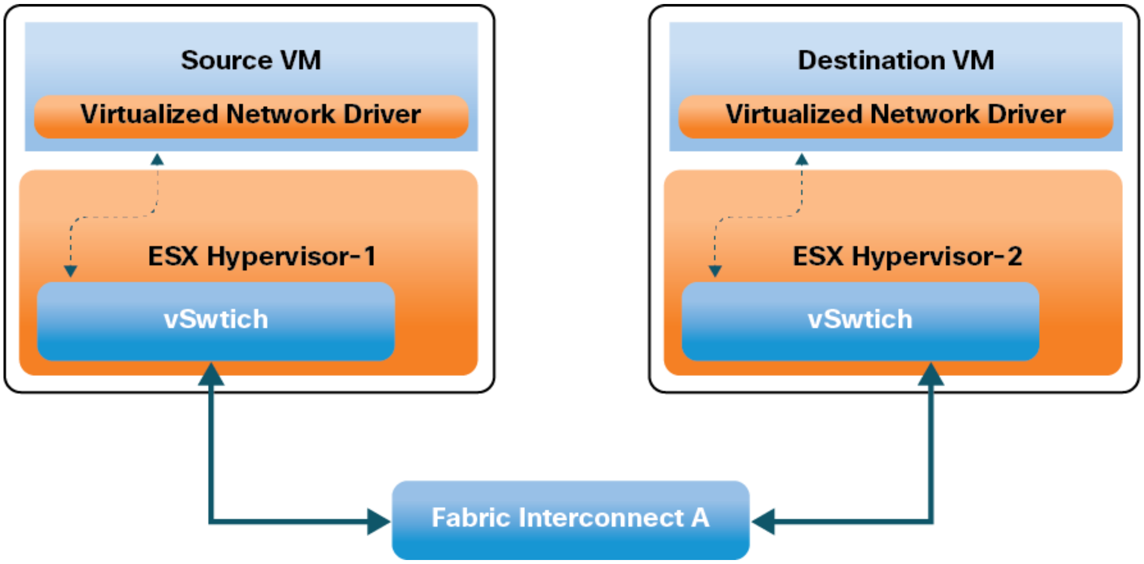
Figure 12. Cisco VM-FEX Fabric Topology



With Cisco VM-FEX, both VMs bypass the hypervisor networking stack. However, with VMware vSwitch, traffic between the source and destination VMs has to traverse two sets of networking stacks, which has implications for performance.

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Figure 13. VMware vSwitch Fabric Topology



Hardware Configuration

Identical compute hardware was used for both vSwitch and VM-FEX performance testing (see Table 4).

Table 4. Hardware Configuration Used in Testing

Item	Description
Server Model	Cisco UCS B200 M3
Networking Adapter	Cisco UCS 1240 Virtual Interface Card
CPU Model	Intel® Xeon E2690 @ 2.93 GHz/Core
Memory Configuration	16GB x 8 1600 MHz DDR3 RAM

System Settings

BIOS Configuration

Default BIOS configurations were used. By default, Cisco UCS B200 M3 has Intel VT-x and VT-d extensions enabled. These options are available under Advanced CPU Configuration section of the BIOS. Enabling these options is a mandatory requirement for VM-FEX functionality.

Adapter Configuration

The default adapter configuration was used.

VM Settings

Table 5 shows the VM configuration settings used in testing.

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Table 5. VM Configuration Settings Used in Testing

	VMware vSwitch	Cisco VM-FEX
Guest OS	RHEL 6.2	RHEL 6.2
Number of vCPUs	1	1
Number of vNICs	1	1
Networking Driver	VMXNet3	VMXNet3
MTU	9000B	9000B

Note that Cisco VM-FEX requires the VMXNet3 guest network device driver. Even though VM-FEX bypasses the hypervisor, it still relies on VMXNet3 to bring up and manage the device and also during VMware vMotion VM migration process. Once the VM is active, the device is relinquished from hypervisor and attached directly to the VM. When the user initiates vMotion, the device is reattached to the hypervisor using VMXNet3, migrated to the target host, and then unattached from the hypervisor and re-attached for the direct access to the VM.

Low-Latency Configuration for VMware vSwitch

Table 6 shows the low-latency configuration for vSwitch.

Table 6. vSwitch Low-Latency Configuration

VM Guest Settings	
ethernetX.coalescingScheme	Disabled
monitor_control.halt_desched	False
ESX Settings	
Net.CoalesceDefaultOn	0

Additionally, from within the ESX console, the following command was used to ensure the interrupt coalescing timer for vmnic0 was set to 0 (turned off).

```
ethtool –c vmnic0 rx-usecs 0
```

Low-Latency Configuration for Cisco VM-FEX

See Table 7.

Table 7. VM-FEX Low-Latency Configuration

VM Guest Settings	
monitor_control.halt_desched	False

Additionally, the interrupt coalescing timer was set to 0 (turned off) in the VIC Adapter Policy in UCSM. This is the Adapter Policy assigned to the dynamic vNICs created using UCSM.

The ‘monitor_control.halt_desched = False’ option configures the hypervisor to never de-schedule the VM process. This can result in higher CPU utilization.

© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Performance Evaluation Tools

netperf-2.6.0 (available at http://www.netperf.org) was used for performance evaluation.

Uni-directional TCP Traffic

For uni-directional TCP traffic, the following netperf command options were used on the source VM:

```
netperf –H $remote_vm –l 60 –t TCP_STREAM –– –m 8192 –s 262144 –S 262144
```

Bi-directional TCP Traffic

For bi-directional TCP traffic, the following netperf command options were used on the source VM:

```
netperf –H $remote_vm –l 60 –t TCP_RR –v2 –– –b 8 –r 8192,8192 –s 262144 –S
262144
```

TCP Transactions

TCP Latency

For TCP latency, the following netperf command options were used on the source VM:

```
netperf –H $remote_vm –l 60 –t TCP_RR -- -r $payload
```

Refer to “TCP Latency Performance” section for details on how latency was calculated.

TCP Connect/Close

For TCP Connect/Close transactional performance, the following netperf options were used on the source VM:

```
netperf –H $remote_vm –l 60 –t TCP_CC
```

TCP Connect/Request/Response/Close

For TCP Connect/Request/Response/Close transactional performance, the following netperf options were used on the source VM:

```
netperf –H $remote_vm –l 60 –t TCP_CRR -- -r $payload
```

Conclusion

Cisco VM-FEX with Cisco UCS VIC 1240 offers a significant savings in CPU utilization without compromising performance.

Cisco VM-FEX also offers significant latency performance benefits compared to VMware vSwitch. These benefits can translate to application transaction speed up and better response times.

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.**

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

Printed in USA C11-727581-00 04/13