White Paper

Networking Performance on RHEL with Cisco UCS 1240 & 1280 Virtual Interface Card (VIC)

Networking Performance on Red Hat Enterprise Linux (RHEL) Version 6.2

Ven Immani (TME, SAVTG)

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

Contents

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

Key Findings

This paper presents the networking performance characteristics of the Cisco Unified Computing System™ (Cisco UCS®) Virtual Interface Card (VIC) 1240 combined with an I/O Port Expander Card in a Cisco UCS blade server. The following observations are presented:

- The Cisco UCS VIC 1240 on Cisco UCS B200 Blade Servers can fully saturate a 10-Gbps Ethernet port for both Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) traffic in non-port-channel mode with a single IP flow.
- The Cisco UCS VIC 1240 can fully saturate 20-Gbps (2 x 10 Gbps) Ethernet ports across both fabrics with two separate IP flows.
- The Cisco UCS VIC 1240 can achieve close to 40 Gbps across both fabric connect paths with multiple IP flows from the user space using TCP/IP.
- The Cisco UCS VIC 1240 with I/O Expander can achieve close to 54 Gbps across both fabric connected paths with multiple IP flows from a kernel space using a packet generator (PKTGEN) kernel module.

Introduction

The Cisco UCS VIC is a virtualization-optimized Converged Network Adapter (CNA) mezzanine card designed for use with Cisco UCS B-Series Blade Servers. The Cisco® VIC cards support up to 256 Peripheral Component Interconnect Express (PCIe) standard-compliant virtual interfaces¬. These PCIe interfaces can be dynamically configured so that both their interface types (whether a network interface card [NIC] or host bus adapter [HBA]) and identity (MAC address and worldwide name [WWN]) are established using just-in-time provisioning. Complete network separation is guaranteed between the PCIe devices using network interface virtualization (NIV) technology.

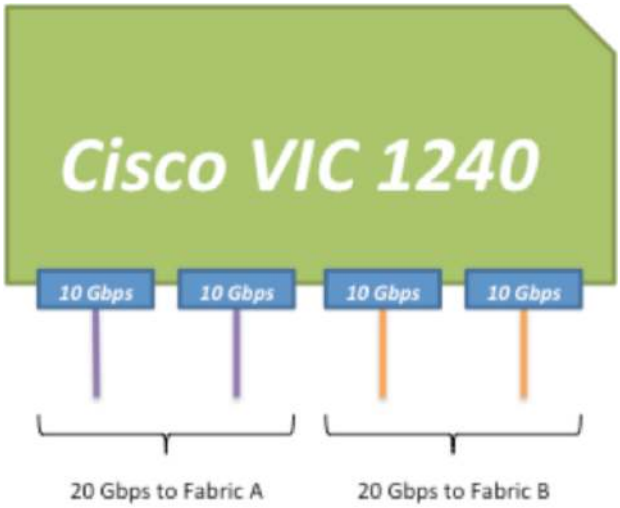
Cisco VIC 1240

Based on second-generation Cisco VIC technology, the VIC 1240 is a modular LAN on motherboard (LOM) that is designed specifically for the M3 generation of Cisco UCS B-Series Servers. The Cisco UCS VIC 1240 offers industry leading performance, flexibility, and manageability. The VIC 1240 is capable of aggregate 8 x 10 Gbps speed to the half-width blade slot when used with the Port Expander Card. Without the Port Expander Card, the VIC 1240 enables four ports of 10 Gbps network I/O to each half-width blade server.

The Cisco VIC 1240 is available in either a 4 x 10 Gbps hardware configuration (Figure 1) or 8 x 10 Gbps hardware configuration (Figure 2). In either case, half the ports are connected to Fabric A and the other half to Fabric B. A vNIC instantiated on the VIC can be explicitly pinned to either Fabric A or B. The vNIC can also be configured for failover. In this case, one fabric-connected path can be designated as a primary path and the other fabric-connected path as a secondary or failover path.

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

Figure 1. Cisco VIC 1240



In Figure 2, the boxes outlined in white are additional ports enabled with the use of an I/O Expander.

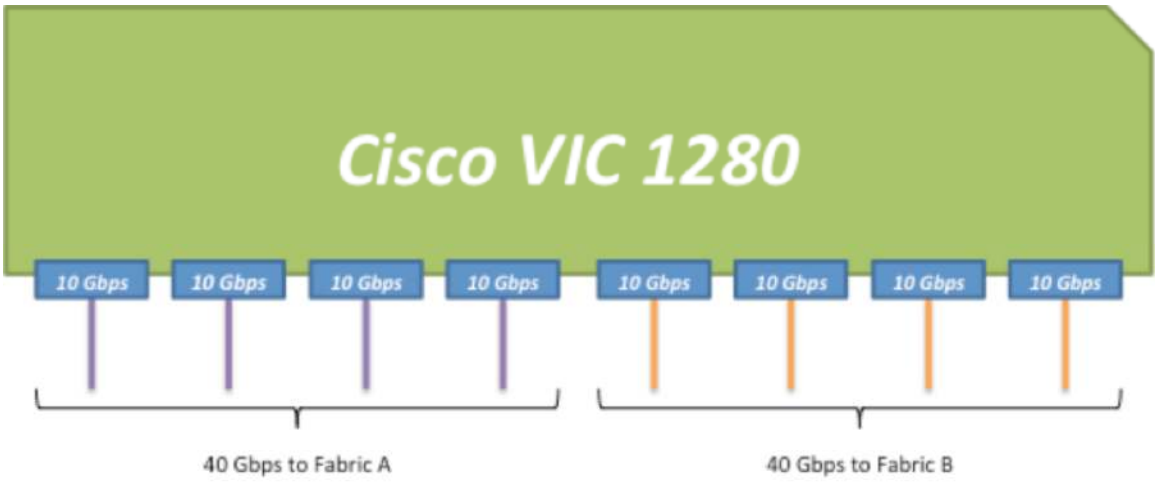
Figure 2. Cisco VIC 1240 with I/O Expander

Cisco VIC 1280

Cisco UCS VIC 1280 (Figure 3) is identical in capability to Cisco VIC 1240 with an I/O Expander. The difference is that VIC 1280 comes with a default 8 x 10 Gbps hardware configuration. Therefore, the VIC 1280 has performance characteristics that are identical to the VIC 1240 with an I/O Expander.

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

Figure 3. Cisco VIC 1280



The Cisco VIC is dependent on the chassis I/O module (IOM) for external connectivity and bandwidth capabilities. The IOM may be configured to operate in either port-channel mode or non-port-channel mode. Configuring the IOM in port-channel mode enables up to 80 Gbps of fabric bandwidth per IOM or up to 160 Gbps of fabric bandwidth across both IOM.

However, the connectivity between the Cisco VIC and the IOM is always available in port-channel mode. When the IOM is configured in non-port-channel mode, the bandwidth per half- width slot per IOM is limited to 10 Gbps. When configured in port-channel mode, the bandwidth per half-width slot per IOM is limited to 20 Gbps on the Cisco VIC 1240 and to 40 Gbps on the Cisco VIC 1280 (or the Cisco VIC 1240 that has an I/O Expander Card).

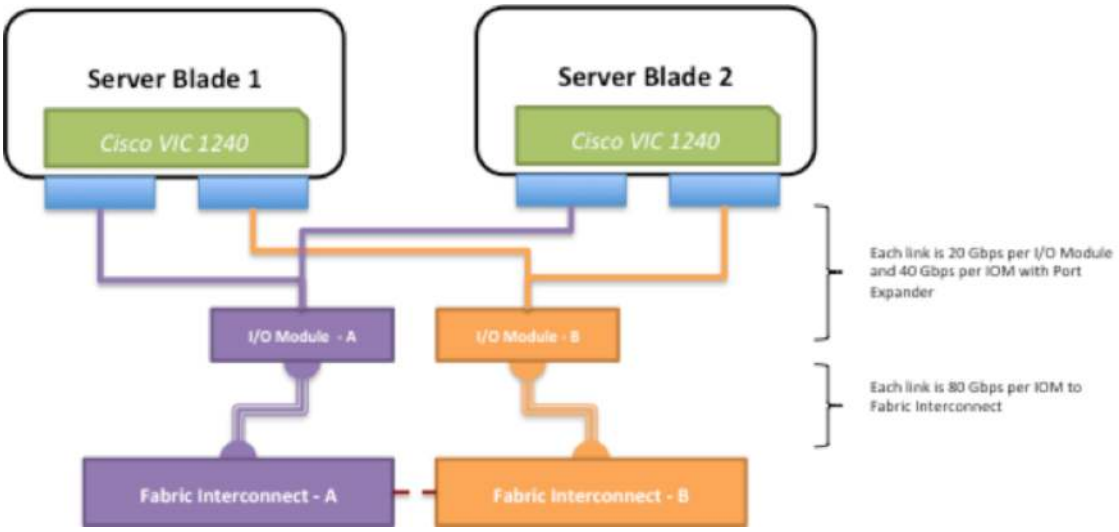
This paper presents networking performance characteristics of the Cisco VIC 1240 on a Cisco UCS B200 M3 Blade Server running the Red Hat Enterprise Linux (RHEL) Version 6.2 operating system.

Network Topology of the System Tested

Figure 4 shows the network topology used for this benchmark.

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

Figure 4. Network Topology



Two Cisco UCS B200 M3 Blade Servers were used for this performance evaluation. In addition, the same blades were used for both the VIC 1240 alone and VIC 1240 with an I/O Expander Card.

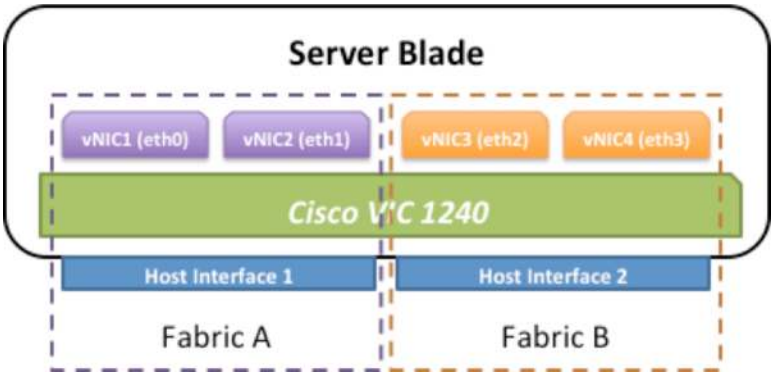
The Cisco VIC 1240, as explained earlier, comes with a default hardware configuration of 4 x 10 Gbps ports. Two of those ports connect to Fabric A and the other two to Fabric B (as just shown in Figure 4).

In Figure 5, the blue box (Host Interface 1) represents 2 x 10 Gbps ports.

When an I/O Expander is added to the VIC 1240, the blue box represents 4 x 10 Gbps ports. Again, each host interface is hard pinned to a specific fabric. Host Interface 1 and Host Interface 2 connect to Fabric A and Fabric B, respectively, in Figure 5.

Both the server blades were installed with standard-distribution RHEL Version 6.2. Each server blade was configured with four vNICs, with two vNICs pinned to Fabric A and the other two vNICs to Fabric B (Figure 5). All four vNICs were configured with a maximum transmission unit (MTU) of 9000 (see Table 1).

Figure 5. Host vNIC Configuration

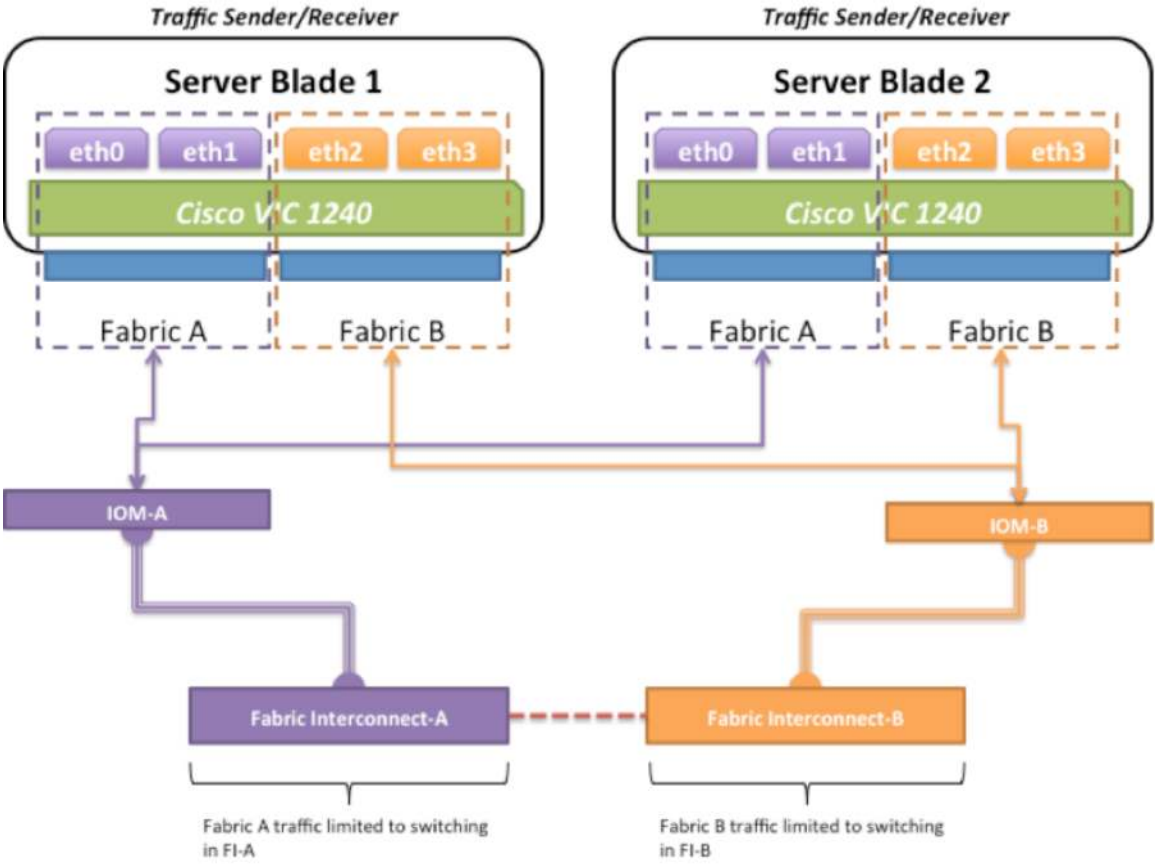


© 2013 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Table 1. vNICs and MTUs

vNIC	MTU (bytes)	Pinned to
vNIC 1 (eth0)	9000	Fabric Interconnect A
vNIC 2 (eth1)	9000	Fabric Interconnect A
vNIC 3 (eth2)	9000	Fabric Interconnect B
vNIC 4 (eth3)	9000	Fabric Interconnect B

Figure 6. Traffic Flow Topology



The hosts were set up in a manner that prevents switching traffic between the fabrics. Traffic flows from Host Interface 1 is restricted to Fabric A and traffic from Host Interface 2 is restricted to Fabric B.

Therefore, traffic from Server Blade 1 on Host Interface 1 is destined to Host Interface 1 on Server Blade 2. Traffic from Server Blade 1 on Host Interface 2 is destined to Host Interface 2 on Server Blade 2.

Once again, the bandwidth capacity is limited to 20 Gbps (2 x 10 Gbps) per host interface per blade when using the Cisco UCS VIC 1240. The bandwidth capacity for the VIC 1240 using the I/O Expander or VIC 1280 is limited to 40 Gbps (4 x 10 Gbps).

As Figure 6 shows, traffic flow between the traffic Receiver and traffic Sender was constrained to a single fabric interconnect. The Cisco VIC 1240 allows for the flow of traffic from vNICs on both the fabric interconnects in an

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

active/standby mode. It is possible to configure vNICs so that host interface ports connected to both fabrics are used for a cumulative bandwidth of 40 Gbps with VIC 1240 and 80 Gbps with a VIC 1240 that has an I/O Expander.

Hardware Configuration

Table 2 shows the hardware configuration used for benchmark testing.

Table 2. Hardware Configuration for Benchmark Testing

Hardware
Cisco UCS B200 M3 Blade Server • 2 8-Core Intel Xeon E5-2690 processors at 2.90 GHz • 128 GB DDR3 1600 MHz RAM • Cisco UCS VIC 1240 with I/O Expander
Cisco UCS • Cisco UCS 5108 Chassis • Cisco UCS 2208XP I/O Modules • Cisco UCS 6248XP Fabric Interconnects

Table 3 shows the firmware versions used for benchmark testing.

Table 3. Firmware Used in Testing

Firmware
UCS Software Release 2.0(4a) was used • Cisco UCS VIC 1240 • Cisco UCS B-200 M3 Server Blade BIOS • Cisco UCS 6248 Fabric Interconnects • Cisco UCS 2208 I/O Modules

System BIOS Settings

Table 4 shows the system BIOS settings used in testing.

Table 4. BIOS Settings Used in Testing

CPU Configuration	
Hyper Threading	Disabled
Turbo Mode	Enabled
Intel SpeedStep	Enabled
Direct Cache Access	Enabled
Processor C State	Disabled
Processor C1E	Disabled
Processor C3 Report	Disabled
Processor C6 Report	Disabled
Processor C7 Report	Disabled
CPU Performance	High-performance computing (HPC) mode
Package C State Limit	No limit
Memory Configuration	
Memory RAS Config	Maximum performance
NUMA	Enabled

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

LV DDR Mode	Performance-mode

Adapter Settings

Table 5 shows adapter settings.

Table 5. Adapter Settings

Adapter Settings

Resources
• Transmit Queues: 1
  ◦ Ring Size: 256
• Receive Queues: 1
  ◦ Ring Size: 512
• Completion Queues: 2
• Interrupts 4

Options • Transmit Checksum Offload: Enabled • Receive Check Sum Offload: Enabled • TCP Segmentation Offload: Enabled • TCP Large Receive Offload: Enabled • Receive Side Scaling: Disabled • Failback Timeout (Seconds): 5 • Interrupt Mode: MSI-X • Interrupt Coalescing Type: min • Interrupt Timer: 125 usecs [microseconds]

Host Side Settings

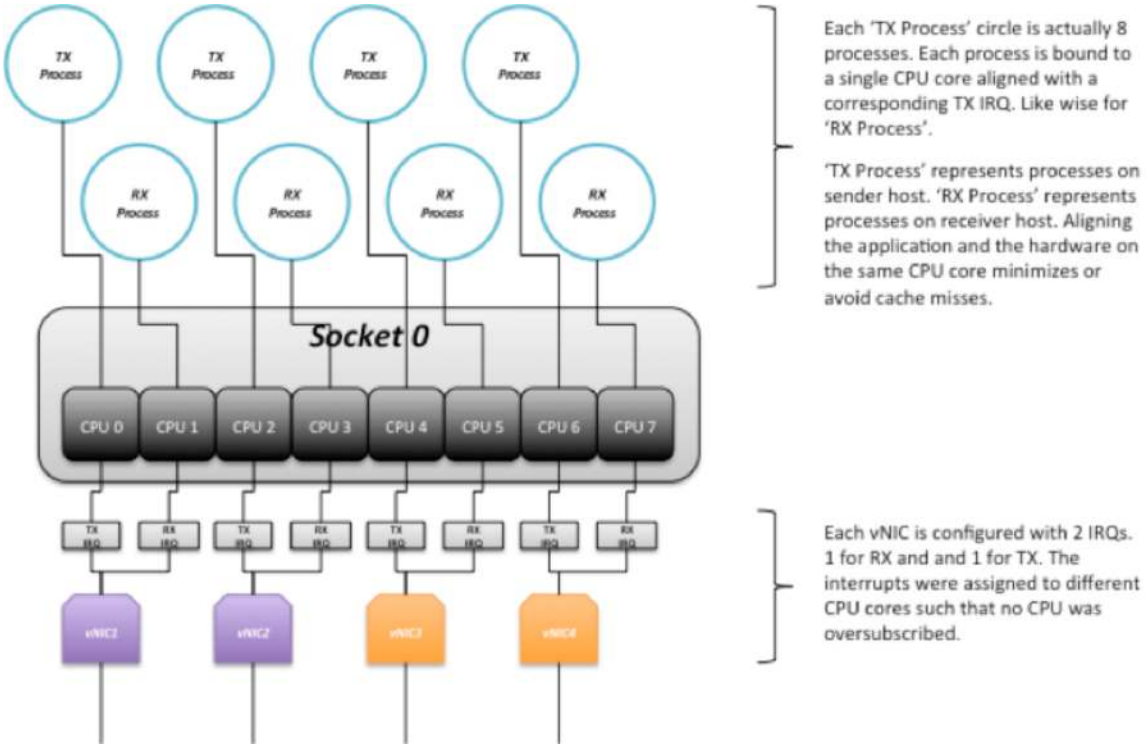
SYSCTL Settings

Default sysctl settings were used. It is possible to achieve better performance by tuning sysctl settings.

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

CPU Affinitization

Figure 7. CPU Affinitization



All the transmitter (Tx) and receiver (Rx) information requests (IRQs) of the four vNICs were affinitized to cores on CPU Socket 0. On the Cisco UCS B200 M3 Blade Server, the VIC 1240 hardware is connected to a PCIe I/O HUB on CPU Socket 0.

This has a bearing on performance, since the VIC hardware is local to Socket 0. If the operating system designated a core from CPU Socket 1 to service the IRQs, the user might experience lower throughput due to the QuickPath Interconnect (QPI) traverse between the sockets. When the IRQs are localized to CPU Socket 0, this traverse is avoided.

CPU Settings

cpuspeed service was explicitly configured for performance mode. The usual default is ondemand. In “high-performance computing (hpc)” mode, the CPU is always maintained in P0 state. P0 is the highest power/performance state of the CPU.

Performance Evaluation Tools

Table 6 lists performance evaluation tools.

Table 6. Performance Evaluation Tools

Benchmark Tool	Version	OS Platform
netperf	2.4.5	RHEL 6.2
PKTGEN kernel module	N/A	RHEL 6.2

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

The netperf tool provides tests for unidirectional throughput for TCP and UDP. netperf was used for benchmarking the performance on the RHEL 6.2 operating system for single-flow TCP and UDP performance and multi-flow TCP performance. Since it operates in the user space, it is a close representation of an application that one might use in a production environment. For more information, visit: http://www.netperf.org

PKTGEN is a traffic generator that is available as a Linux kernel module. PKTGEN has very low overhead for traffic generation and is therefore able to push large amounts of traffic. This feature is helpful in evaluating and observing the real capability of networking hardware under consideration. More information is available as part of the Linux kernel documentation.

Performance Results

The Cisco VIC 1240 has four 10 Gbps Data Center Ethernet (DCE) ports: two for Fabric A and two for Fabric B. With an I/O Expander, the number of ports is doubled to four for Fabric A and four for Fabric B.

The Cisco VIC 1240 supports port channels for egress and ingress traffic to the vNIC. A single vNIC is represented as a unique PCIe device and is backed with either two or four DCE ports connecting to the fabric. With port channels, both egress and ingress flows are distributed across the DCE ports for better bandwidth capacity on a per-vNIC basis.

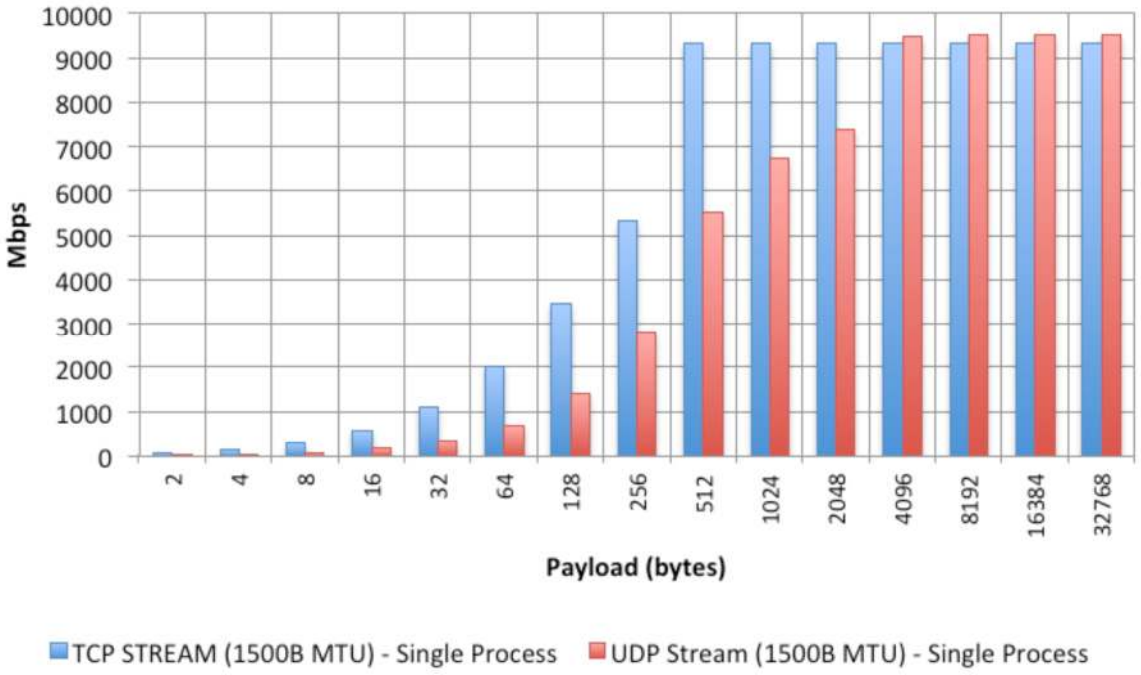
The Cisco VIC port-channel mode is based on source (Port/IP/MAC) and destination (Port/IP/MAC) flow hashing. With multiple flows, a unique hash for each flow is derived and used to direct the flow on one of the available adapter egress DCE ports. On Ingress flow, the hash is calculated by the IOM and steered into a specific VIC DCE port. A higher and disparate number of flows will result in unique hash values or better flow distribution across the available DCE ports.

Single-Flow IP Performance

With single flow, the egress bandwidth is limited to a single DCE port or 10 Gbps. Figure 7 shows bandwidth performance for TCP and UDP STREAM with a single flow. The X-axis denotes the packet payload size, and the Y-axis denotes the bandwidth achieved.

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

Figure 8. Single-Flow IP Performance



Raw Packet Performance

Higher-bandwidth gains are realized when multiple flows are initiated over the vNIC. With multiple-flows, the hash is calculated in the VIC hardware and the flow is steered out of a DCE port. The graph in Figure 8 represents bandwidth performance with the PKTGEN kernel module. The X-axis represents packet size. The Y-axis represents bandwidth achieved. There are two data sets represented in the graph. The data set in cyan represents the Send performance (in Mbps) of VIC 1240 with an I/O Expander. The data set in purple represents Send performance of a VIC 1240 without the I/O Expander.

The horizontal lines across the graph represent the real physical limits or the maximum performance capability. The cyan line represents the maximum bandwidth capability for a VIC 1240 with an I/O Expander. The maximum bandwidth capability in this case is 64 Gbps. The purple line represents the maximum bandwidth capability for VIC 1240 without the I/O Expander. The maximum bandwidth capability in this case is 40 Gbps.

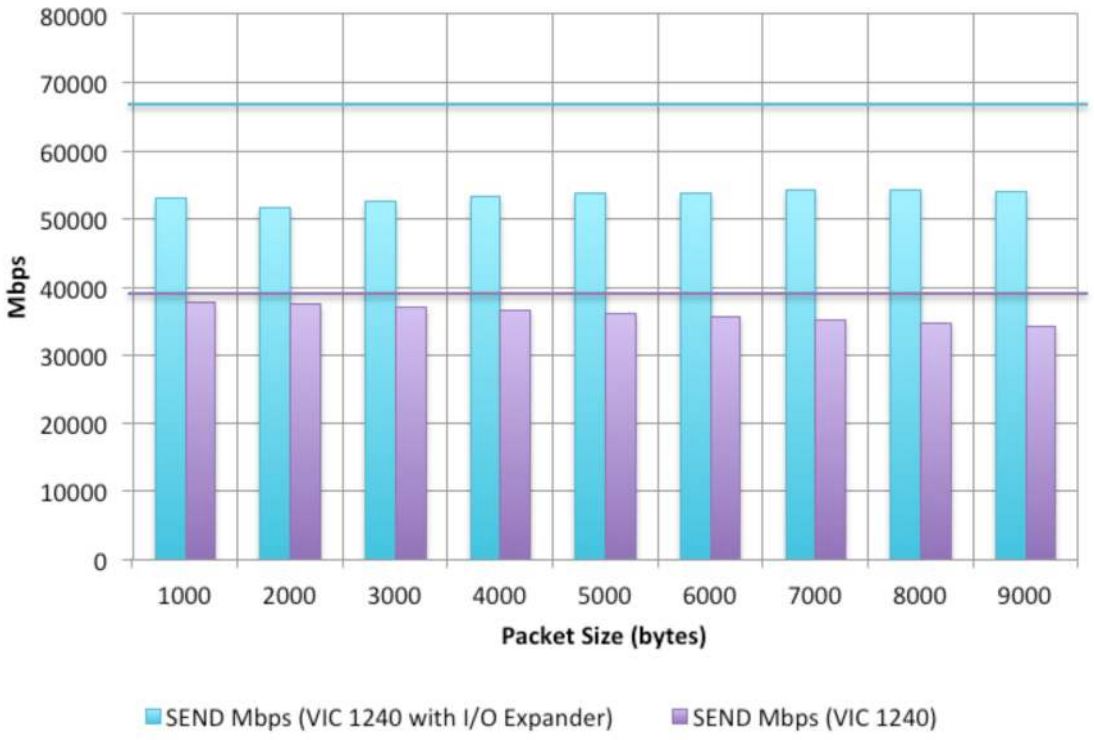
Even though the VIC 1240 with the I/O Expander has a physical outward connectivity of 80 Gbps (8 x 10 Gbps), it is limited to 64 Gbps due to the limitations of the PCIe bus. The Cisco UCS B200 M3 Blade Servers have PCIe 2.0 16x connectivity to the VIC. PCIe 2.0 16x is limited to 64 Gbps of maximum bandwidth in one direction.

In Figure 8, the Cisco VIC 1240 with an I/O Expander achieves Send bandwidth ranging up to 54.15 Gbps. This is about 84.60 percent of PCIe 2.0 16x line-rate efficiency.

The Cisco VIC 1240 without the I/O Expander achieves Send bandwidth ranging up to 37.8 Gbps. This is about 94 percent of line-rate efficiency.

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

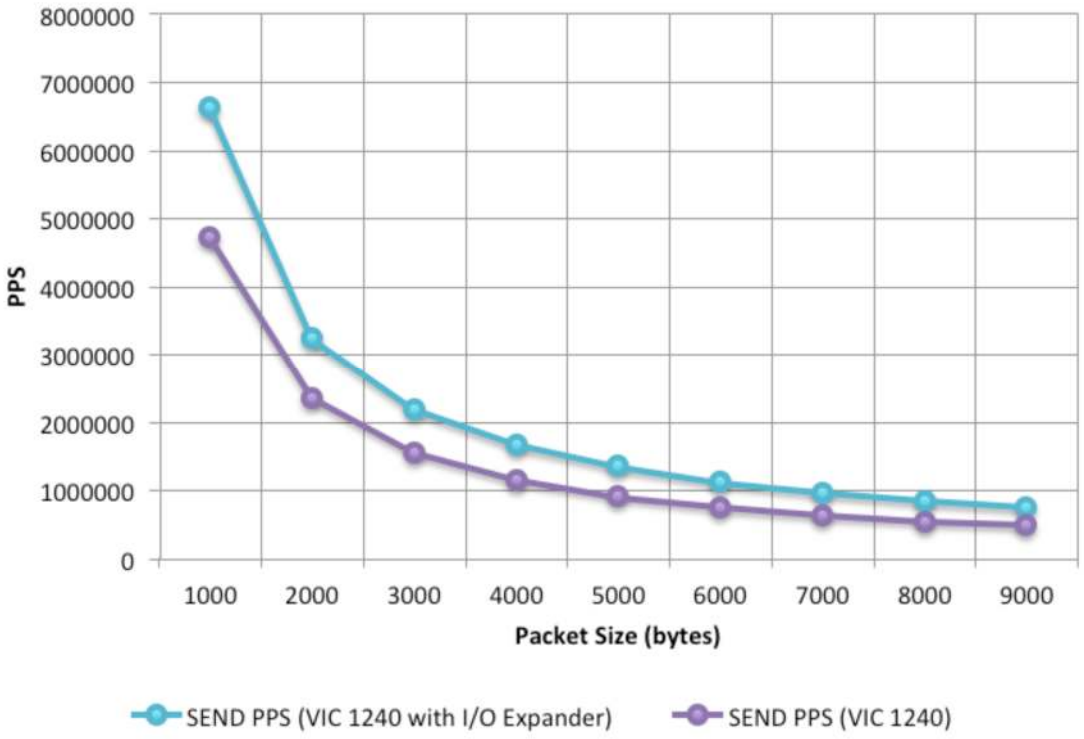
Figure 9. Send Bandwidth for Cisco VIC 1240 with I/O Expander



In Figure 9, the Cisco VIC 1240 with an I/O Expander achieved a Send packets per second (pps) rate ranging from 6.6 million packets to 752K packets for 1000-byte and 9000-byte packet sizes, respectively. The VIC 1240 achieved a Send pps rate ranging from 4.7 million packets to 477K packets for 1000-bytes and 9000-byte packet sizes, respectively.

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

Figure 10. Send PPS for Cisco VIC 1240 with I/O Expander



The VIC 1240 with the I/O Expander performs about 43 percent better than the VIC 1240 without the expander.

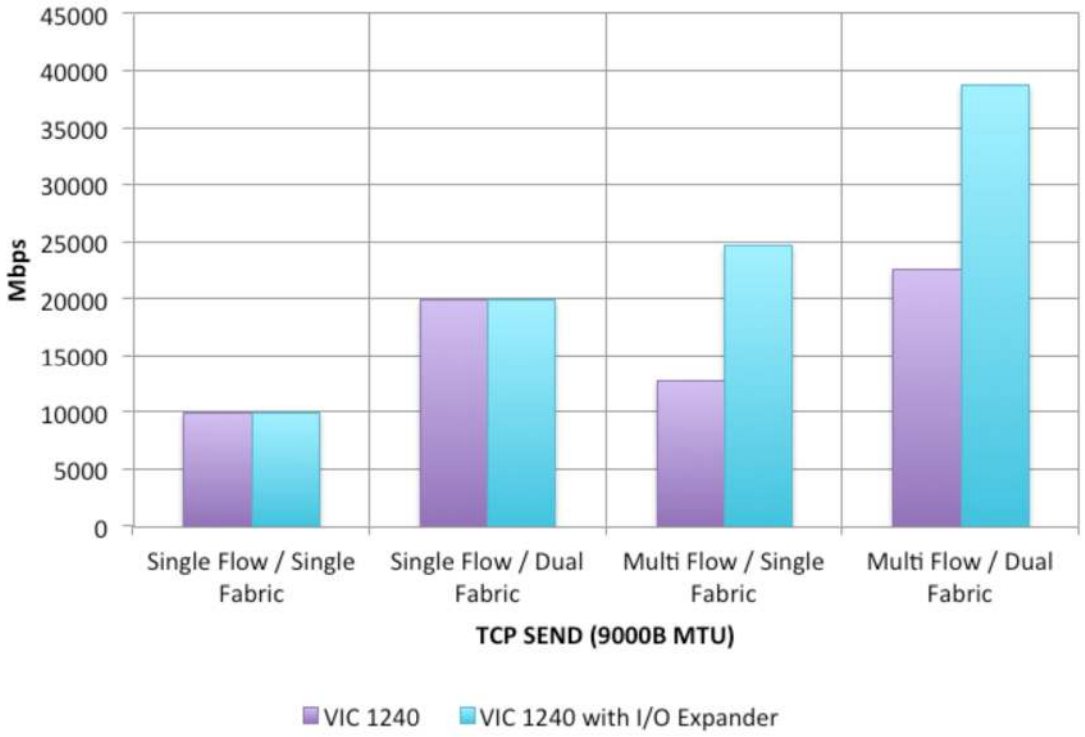
Multi-flow IP Performance

As explained earlier, the Cisco VIC supports hardware port channels based on source (Port/IP/MAC) and destination (Port/IP/MAC) flow hashing.

These IP flows may be limited to one fabric or spread across both fabrics (see Figure 10).

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

Figure 11. Comparison of Multiflow IP Performance between VIC 1240 with I/O Expander and without I/O Expander



As Table 7 shows, with a single IP flow, both the VIC 1240 and the VIC 1240 with the I/O Expander behave similarly. This is because a single IP flow cannot take advantage of the hardware port channel and therefore cannot utilize all the available egress bandwidth capacity. For a single IP flow for both the VIC 1240 and the VIC 1240 with the I/O Expander, the numbers are identical (as they should be) on single-fabric and dual-fabric configurations. However, with multiple IP flows, the VIC 1240 with the I/O Expander provides 94 percent better performance compared to the VIC 1240 by itself on a single fabric. And in a dual-fabric configuration, the VIC 1240 with the I/O Expander performs 70 percent better in comparison to the VIC 1240 by itself.

Table 7. Comparison of Single and Dual Fabrics and VIC 1240 with and without I/O Expander

Benchmark Tool	Single Fabric		Dual Fabric	
	Single Flow	Multi Flow	Single Flow	Multi Flow
VIC 1240	9966	12749	19931	22655
VIC 1240 with I/O Expander	9966	24747	19931	38698
In Mbps				

A few things to note: The multiflow TCP Send numbers are lower than the multiflow PKTGEN numbers. This is because TCP has a lot more overhead with the connection setup, packet framing, and packet acknowledgement. Additionally, TCP in the case just described is run from the user space. PKTGEN, on the other hand, has a lot lower overhead. It is unreliable packet Send, with almost no connection setup overhead and no packet acknowledgement. Since PKTGEN uses UDP packets, the packet-framing overhead is also quite low. And it is run in the kernel space.

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

The numbers from PKTGEN are, therefore, an indication of the real capability of the adapter without the typical-use overheads. However, the numbers from TCP Send are an indication of practical capabilities of the adapter when used in real-world scenarios.

At 38698 Mbps with TCP multiflow, the Cisco VIC 1240 performs at about 60 percent line-rate efficiency. This number by itself is unprecedented for TCP performance out of a single host. Still better performance is possible with specific tuning.

In both cases, the data presented here illustrates the capabilities of the Cisco VIC 1240 and the VIC 1240 with the I/O Expander. The VIC 1240 with the I/O Expander clearly outperforms the VIC 1240 alone due to the availability of higher physical port bandwidth.

Conclusion

The results presented in this paper illustrate that the Cisco UCS VIC 1240 with the I/O Port Expander Card running on RHEL Version 6.2 can effectively achieve close to 40 Gbps of TCP Send traffic and close to 54 Gbps of kernel mode UDP traffic.

The Cisco UCS VIC 1240 is a fully standards-compliant network adapter that delivers industry-leading network throughput performance. The Cisco UCS VIC 1240 with the I/O Port Expander Card is an excellent choice of network adapter for I/O-intensive server workloads.

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 C11-720526-01 01/13

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