

SR-IOV Improves Server Virtualization Performance

Cavium FastLinQ 3400/8400 Series Adapters Enhance I/O Throughput and Reduce Processor Utilization



SR-IOV alleviates bottlenecks in virtual operating systems and enables “bare-metal” performance of virtualized resources.

INDUSTRY CHALLENGE

While virtualization improves server processor utilization tremendously, it introduces new challenges to network I/O performance. Software-based sharing of hardware resources adds overhead to each I/O operation due to the emulation layer between the virtual machines (VMs) and the underlying I/O hardware. This virtualization introduces latency, consumes compute cycles, and reduces overall network and system performance.

In a virtualized server, the hypervisor abstracts and shares physical NICs among multiple virtual machines, creating virtual NICs for each virtual machine. Another piece of software in the hypervisor—the Virtual Switch or vSwitch—manages these virtual NICs. Using a software-based vSwitch has a number of disadvantages, including a tax on valuable CPU and memory bandwidth. The higher the traffic load, the greater the number of CPU and memory cycles required to move traffic through the vSwitch, reducing the ability to support larger numbers of VMs in a physical server.

In server virtualization, the ultimate goal is for virtual machines to perform as if they were running on bare-metal physical machines. For virtual networking that translates into achieving near-native I/O performance. This means reducing CPU utilization, lowering latency, and boosting I/O throughput as if the VM were talking directly to the physical network adapter.

WHAT IS SR-IOV?

PCI-SIG, the special interest group that owns and manages PCI specifications as open industry standards, introduced a suite of specifications for Single Root I/O Virtualization (SR-IOV) to allow multiple operating systems to share a physical interconnect. The SR-IOV standard allows multiple VMs to share an I/O device, while allowing for close to bare-metal runtime performance.

The SR-IOV specification details how a single PCI Express (PCIe) device can be shared between various guest operating systems—the VMs. Devices capable of SR-IOV functionality support multiple virtual functions on top of the physical function. Virtual functions are enabled in hardware as a lightweight PCIe function that can be directly assigned to a VM without hypervisor mediation. These VFs operate in the context of a VM, and must be associated with a physical function (PF), a full-featured PCIe function that operates in the context of the hypervisor or parent partition.

SR-IOV provides direct VM connectivity and isolation across VMs. It allows the data to bypass the software virtual switch (vSwitch) and provides near-native performance. The benefits to deploying hardware-based SR-IOV-enabled NICs include reduction of CPU and memory usage compared to vSwitches. Moving the network virtualization into hardware (SR-IOV adapters) relieves the performance problems associated with vSwitches. By directing VM I/O directly to VFs and bypassing the hypervisor, the vSwitches are no longer part of the data path. In addition, it significantly increases the number of virtual networking functions for a single physical server.

With Cavium™ FastLinQ® 3400/8400 Series Adapters, SR-IOV can establish up to 64 VFs per physical port. These SR-IOV VFs provide the mechanism to bypass the virtual server’s hypervisor. Setting up virtual functions on the network adapter using SR-IOV allows a VM to be free of the underlying hardware. This flexibility means the VM is also not tied to a physical server.

SR-IOV SUPPORT IN MICROSOFT® WINDOWS SERVER 2012/2012 R2 HYPER-V

Hyper-V in Windows Server 2012/R2 enables support for SR-IOV-capable network devices and lets an SR-IOV VF be assigned directly to a virtual machine, bypassing the root partition. This increases network throughput and reduces network latency while also reducing the host CPU overhead that is required for processing network traffic.

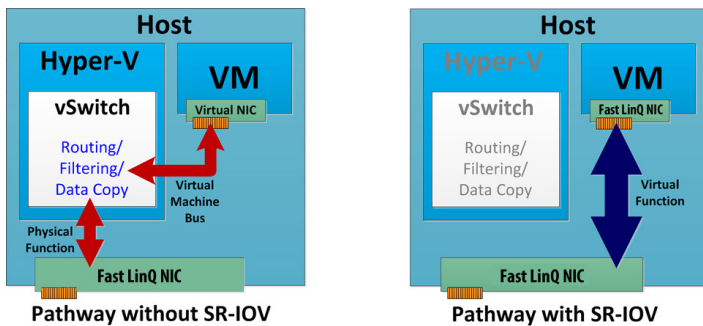


Figure 1. Hyper-V with and without SR-IOV

SR-IOV technology in Hyper-V provides flexibility with live migration and as a solution for workloads that need higher throughput, lower latency, and lower CPU utilization for network traffic.

SR-IOV can be deployed without losing flexibility. For example, live migration is fully supported for SR-IOV. You can live migrate a VM using SR-IOV to another host that either does or does not support SR-IOV, and back again. The VM will use SR-IOV if it is available on the target host, and if SR-IOV is unavailable, it will use the traditional software network path.

For Cavium FastLinQ 8400 Series Adapters with data center bridging (DCB) enabled ports, all SR-IOV traffic is automatically routed through the default traffic class and will follow that traffic class’s minimum bandwidth (DCB enhanced transmission selection or ETS) and losslessness (DCB priority flow control or PFC) settings.

SR-IOV from a specification standpoint doesn’t mention anything about an I/O class. Because the overhead of storage I/O is significantly less than that of networking I/O, a design choice in Windows Server 2012/R2 Hyper-V was to exclusively focus on SR-IOV for networking as the only supported device class.

SR-IOV SUPPORT ON CAVIUM FASTLINQ 3400/8400 SERIES ADAPTERS

To create an SR-IOV capable PCIe device, following the SR-IOV specifications alone is not enough. The NIC must support networking to and from multiple sources (PF and VFs), as well as on the wire. To enable Ethernet frames to be routed between local VFs, for example, parts of an Ethernet switch must be embedded onto the physical adapter, above and beyond the SR-IOV specification. Such an embedded switch (eSwitch) within the Cavium FastLinQ 3400/8400 Series Adapters enables Ethernet switching between the VMs, from the same port VF to VF/PF, and to or from the external port (see Figure 2).

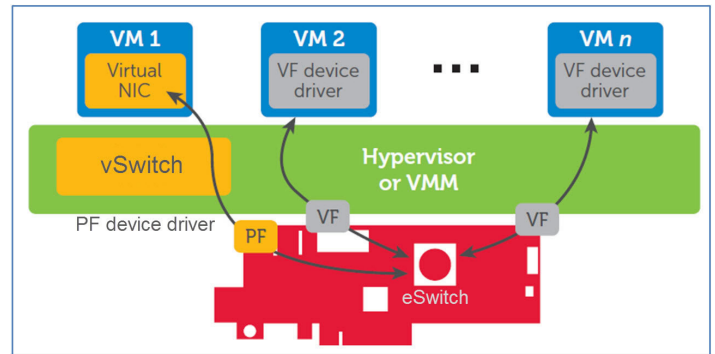


Figure 2. Cavium FastLinQ 3400/8400 PF/VFs eSwitching

The Cavium FastLinQ 3400/8400 Series Adapter’s SR-IOV VFs can also use Windows Server 2012/2012 R2 new receive segment coalescing (RSC) feature. This combined with the adapter’s hardware transparent packet aggregation (TPA) will greatly increase an SR-IOV VFs bandwidth performance (see Figure 3) and overall efficiency (see Figure 4).

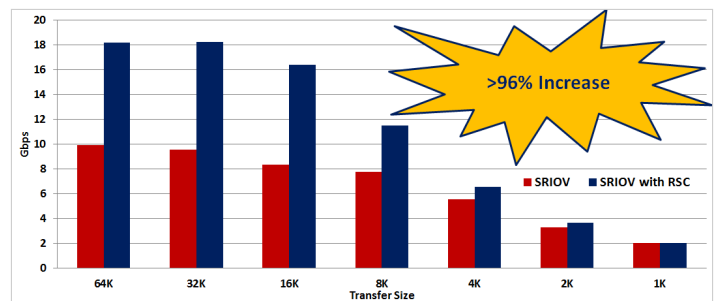


Figure 3. Cavium FastLinQ 3400/8400 Single Port Bidirectional Bandwidth (Gbps)

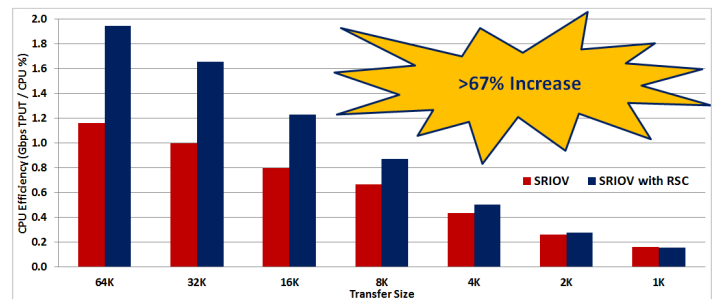


Figure 4. Cavium FastLinQ 3400/8400 Single Port Bidirectional Efficiency (Gbps/Total CPU Utilization %)

TEST CONFIGURATION

- **Operating System:** Windows Server 2012 R2 Hyper-V
- **Server:** Standard x86 Server – Quad Socket 2.0 GHz CPUs (four cores each)
- **Virtual Machines:** One Windows Server 2012 R2 – 2v CPUs, 4GB RAM using 20 threads/direction
- **10GbE Adapter:** Cavium FastLinQ 3400 Series

The Cavium FastLinQ 3400/8400 Series Adapter’s RSC aggregates multiple received packets from the same IPv4/IPv6 TCP connection into a single indication to reduce the amount of headers in the stack (similar to the way large receive offload works) and also reduces the number of interrupts the CPU must handle. TPA also moves the aggregation work from the driver to the NIC’s hardware, further enhancing RSC’s effectiveness.

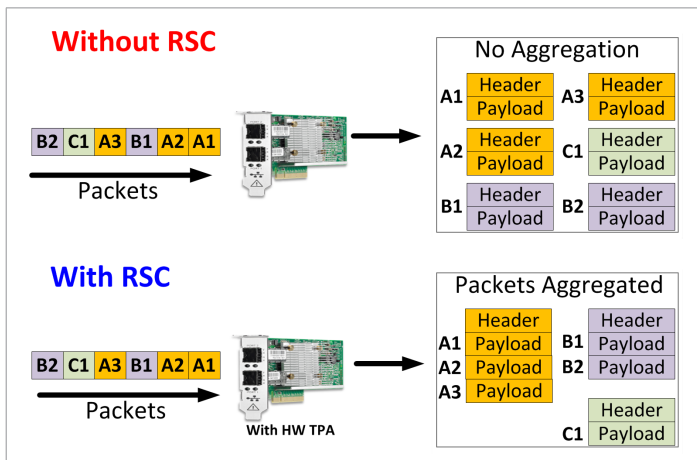


Figure 5. Receive Segment Coalescing with Hardware Transparent Packet Aggregation

Multi-core VMs can take advantage of receive side scaling (RSS)/transmit side scaling (TSS) when using Cavium FastLinQ 3400/8400 Adapter’s SR-IOV VFs to further enhance their performance.

SUMMARY

Virtual networking technologies are becoming increasingly important and complex because of the growing use of virtual machines, distributed applications, and cloud infrastructures.

The network is a key component of virtualized environments, and network adapters that are enhanced to meet the demands of virtualization can help maximize performance. Cavium capabilities in SR-IOV and I/O pass-through functionality for Ethernet network adapters can provide near-native performance while reducing network processing overhead and latency. The Cavium FastLinQ 3400/8400 also provides RSC capabilities along with SR-IOV, which further improves network throughput and CPU efficiency. These results translate into greater VM density per host and increased network bandwidth on the host and guest VMs on servers running hypervisors—such as Microsoft Windows Server 2012/2012 R2 Hyper-V, RHEL 6.5 KVM, SLES 11.3 XEN, Citrix® XenServer 6.2, and VMware® ESXi 5.5 (see Table 1).

Table 1. SR-IOV Features Summary

SR-IOV Features	Metrics	Benefits
Maximum VFs per Device	128	Allows bypassing hypervisors for better performance.
Maximum PFs per NPAR Mode Enabled Device	8	Allows multiple PF instances each with their own QoS (Minimum and Maximum bandwidth) controls.
SR-IOV over NPAR	Yes	Allows the coexistence of NIC Teaming, storage, and SR-IOV on the same adapter.
Physical Port eSwitch	Yes	Same port VF/PF to VF/PF switching, offloads hypervisor (CPU and memory resources) leading to greater VM density.
RSC with TPA	Yes	Reduces CPU utilization by offloading received packet aggregation in hardware.
Broad Hypervisor support	Yes	Supports all major SR-IOV OS vendors: <ul style="list-style-type: none"> • Windows 2012/2012 R2 Hyper-V • Linux <ul style="list-style-type: none"> – RHEL 6.4+ with KVM – SLES 11.3+ with XEN – Citrix XenServer 6.2+ • VMware ESXi 5.5/6.0

The portfolio of SR-IOV capable Cavium FastLinQ 3400/8400 Series 10GbE Adapters include:

- QLE3440-CU Single 10Gb SFP+ Port PCIe Gen3 x8 Networking Adapter
- QLE3442-CU Dual 10Gb SFP+ Ports PCIe Gen3 x8 Networking Adapter
- QLE3442-RJ Dual 10GBASE-T Ports PCIe Gen3 x8 Networking Adapter
- QLE3440-SR Single 10Gb SR Optical Port PCIe Gen3 x8 Networking Adapter
- QLE3442-SR Dual 10Gb SR Optical Ports PCIe Gen3 x8 Networking Adapter
- QLE8440-CU Single 10Gb SFP+ Port PCIe Gen3 x8 Converged Networking Adapter
- QLE8442-CU Dual 10Gb SFP+ Ports PCIe Gen3 x8 Converged Networking Adapter
- QLE8440-SR Single 10Gb SR Optical Port PCIe Gen3 x8 Networking Adapter
- QLE8442-SR Dual 10Gb SR Optical Ports PCIe Gen3 x8 Networking Adapter

ABOUT CAVIUM

Cavium, Inc. (NASDAQ: CAVM), offers a broad portfolio of infrastructure solutions for compute, security, storage, switching, connectivity and baseband processing. Cavium's highly integrated multi-core SoC products deliver software compatible solutions across low to high performance points enabling secure and intelligent functionality in Enterprise, Data Center and Service Provider Equipment. Cavium processors and solutions are supported by an extensive ecosystem of operating systems, tools, application stacks, hardware reference designs and other products. Cavium is headquartered in San Jose, CA with design centers in California, Massachusetts, India, Israel, China and Taiwan.



Follow us:      

Corporate Headquarters Cavium, Inc. 2315 N. First Street San Jose, CA 95131 408-943-7100

International Offices UK | Ireland | Germany | France | India | Japan | China | Hong Kong | Singapore | Taiwan | Israel

Copyright © 2015 - 2017 Cavium, Inc. All rights reserved worldwide. Cavium and FastLinQ are registered trademarks or trademarks of Cavium Inc., registered in the United States and other countries. All other brand and product names are registered trademarks or trademarks of their respective owners.

This document is provided for informational purposes only and may contain errors. Cavium reserves the right, without notice, to make changes to this document or in product design or specifications. Cavium disclaims any warranty of any kind, expressed or implied, and does not guarantee that any results or performance described in the document will be achieved by you. All statements regarding Cavium's future direction and intent are subject to change or withdrawal without notice and represent goals and objectives only.