WHITEPAPER

Optimizing Virtualized Servers with SR-IOV

- Advanced Virtualization with Emulex OneConnect Adapters

EMULEX®

Optimizing Virtualized Servers with Single Root I/O Virtualization (SR-IOV)

Server virtualization is being deployed on an almost universal scale to reduce costs and optimize data center resources. To underscore this point, a recent survey[1] of large midmarket and enterprise-class data centers in North America reported several key findings:

- **Server virtualization is widely used**—74% of respondents were currently using server virtualization and another 19% were planning deployments.
- **The number of virtualized servers is relatively small today**—58% of respondents have virtualized 30% or less of their servers
- **The number of virtualized servers is growing**—the percentage of virtual (VMs) run in production will increase from 39% to 58% within two years

This trend is further enhanced with the introduction of powerful multi-core servers that allow for "heavy weight" application workloads, either processing or I/O, that may use multiple cores. These servers also enable much higher virtualization ratios (VMs per physical server). The trend began in 2009 with quad-core servers from Intel and AMD, and will continue in the second half of 2011 with next generation servers that have up to 8 cores and run up to 16 simultaneous threads. These servers also support much higher memory capacities, another critical resource for virtualized servers.

In addition to cores and memory, the third component to be scaled is I/O resources. That involves high-performance 10Gb/s Ethernet (10GbE) ports, effective protocol offloading to save CPU cycles and the addition of new technologies like SR-IOV.

SR-IOV What?

SR-IOV is a PCI Special Interest Group (PCI-SIG) standard that was developed for virtualized servers. The SR-IOV specification allows a PCI Express (PCIe) I/O device to appear as multiple physical and virtual devices, using the concept of physical and virtual functions:

- **Physical function (PF)**—There is at least one PF for each physical port on an adapter. In some cases, adapters can partitioned into as many as four ports per physical port. In this example, there could be four PFs per port, or a total of 8 PFs for an adapter with two physical ports. The key differentiation is PFs have full configuration capabilities. They are associated with the hypervisor and can be managed like physical devices.
- **Virtual function (VF)**—VFs are associated with VMs and are limited to processing I/O streams, basically moving data. They don't support management of the physical device. The number of supported VFs will vary and will likely center around 64 VFs per physical adapter.

Although the SR-IOV standard applies to networking and storage I/O, the current and expected implementations are for networking only. Performance concerns for virtual servers are focused on Ethernet traffic that typically creates the largest I/O demand and uses the largest amount of server resources. Storage I/O uses far less overhead and usually achieves full line speed.

SR-IOV Why?

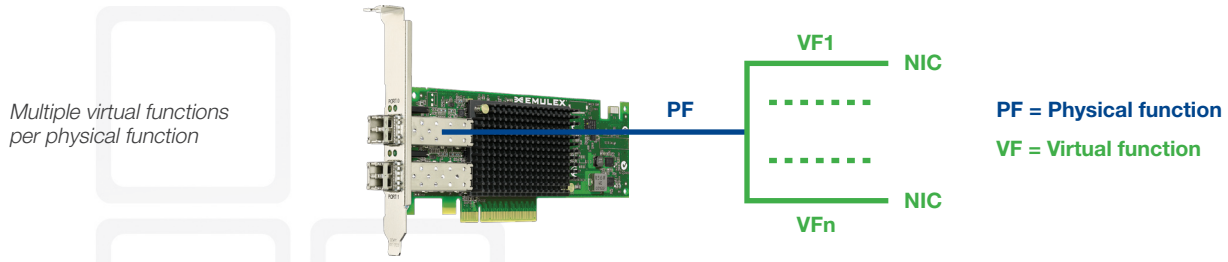
Virtual server hypervisors provide VMs with a set of resources that emulate the functionality of a physical server, allowing each VM to run independently on the physical server. For I/O, that means each VM uses a virtual I/O device that is presented by the hypervisor. Although this works well for virtualization, it's usually not efficient. This is especially true for received I/O.

The hypervisor performs the following steps to process received I/O with multi-core servers:

- One of the CPU cores is interrupted to inspect the packet and determine which VM should receive it
- The core that's servicing the VM is interrupted and processes the I/O
- The core that initially received the packet returns to its normal workload

Each of these steps slows down I/O and uses significant CPU resources.

[1] ESG Research Report, The Evolution of Server Virtualization (November 2010)



One of the solutions for improved performance is direct I/O (or pass-through) based on assignment of a unique physical port to each VM, bypassing the hypervisor. This method improves performance, but is limited by the number of physical ports that can be attached to the server. It also results in a complex and costly collection of adapter ports, switch ports and cables. This direct I/O option is not compatible with migration of a running VM. Migration requires a manual process to shut down, move and restart the VM. Clearly this is less than optimal.

With SR-IOV, VFs are assigned to VMs. This allows one adapter port, switch port and cable to support direct I/O for many VMs. If appropriate, multiple VFs can be assigned to one VM. For example, a VM could be assigned VFs from each physical port of a two-port adapter for high availability.

For received I/O, the server core that is assigned to the VM and its associated VFs executes all of the processing for a packet. There's no need to interrupt cores that are assigned to other VMs. To further enhance performance, I/O's between VFs on the same PF can be processed by the adapter using an internal Layer 2 switch, eliminating routing through a physical switch. Finally, SR-IOV is compatible with migration of running VMs.

SR-IOV When?

The ecosystem for SR-IOV is in the process of bootstrapping itself. It requires support by suppliers of adapters, switches and hypervisors. It also requires support by server vendors that will be adding management tools to fully enable robust solutions. SR-IOV is currently supported with Kernel Virtual Machine (KVM) in Red Hat Enterprise Linux 6 and SUSE Enterprise Linux 11 (and later).

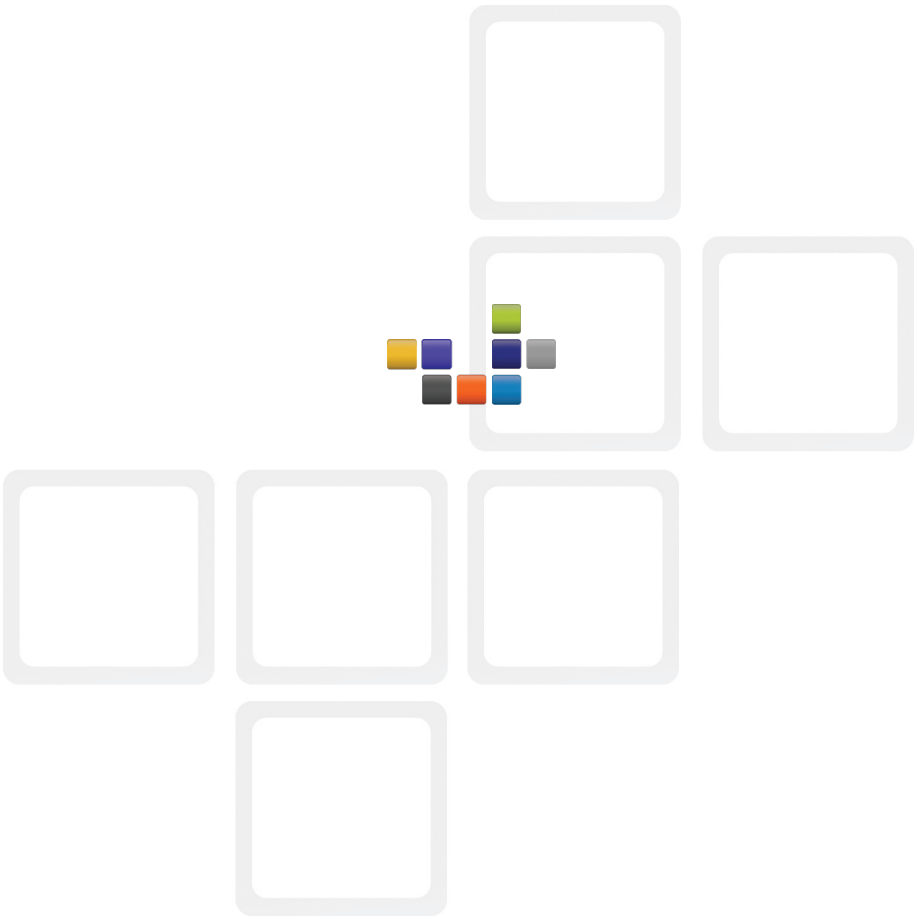
Other hypervisor providers are expected to announce SR-IOV support.

SR-IOV and Emulex OneConnect™ Network Adapters

Emulex is working closely with leading hypervisor and server suppliers as they develop plans for SR-IOV support. The OneConnect OCe11000 family is the third generation of Emulex 10GbE adapters and ships today with SR-IOV capabilities and full offload for optimized use of CPU resources. Data centers can deploy OCe11000 adapters knowing they can use SR-IOV when it's supported by their hypervisor of choice. Any changes based on final hypervisor solutions can be implemented with firmware and driver updates.

Conclusion

SR-IOV will be a key technology to optimize I/O for virtualized servers, enabling higher virtualization ratios for maximum cost savings. SR-IOV will provide a much more cost-effective solution than multiple physical ports and will be fully compatible with VM migration. New third-generation adapters like the Emulex OneConnect OCe11000 are a future-proofed investment that will allow data centers to use SR-IOV when it's supported by hypervisors.



EMULEX®

www.emulex.com

World Headquarters 3333 Susan Street, Costa Mesa, California 92626 +1 714 662 5600
Bangalore, India +91 80 40156789 | **Beijing, China** +86 10 68499547
Dublin, Ireland+35 3 (0)1 652 1700 | **Munich, Germany** +49 (0) 89 97007 177
Paris, France +33 (0) 158 580 022 | **Tokyo, Japan** +81 3 5325 3261
Wokingham, United Kingdom +44 (0) 118 977 2929

11-0919 · 6/11