

PCI-SIG Single Root I/O Virtualization (SR-IOV) Support in Intel® Virtualization Technology for Connectivity

Efficient Native Sharing of I/O Devices with Virtual Machines for enhancing I/O Performance

White Paper
PCI-SIG Single Root I/O
Virtualization

Introduction

As virtualized server deployment increases, virtualization technologies continue to evolve especially in the area of I/O performance. Within the industry, significant effort has been expended to increase the effectiveness of hardware resource utilization (i.e., application execution) through the use of virtualization technologies. The Single Root I/O Virtualization and Sharing Specification (SR-IOV) defines extensions to the PCI Express* (PCIe*) specification suite to enable multiple System Images (SI) or Virtual Machines (VMs/Guests) in the virtualized environment to share PCI hardware resources.

I/O Virtualization Goals

There are numerous trends in virtualization driving the need for more effective I/O virtualization solutions:

- Due to reducing Virtual Machine Monitor (VMM) overheads through Intel® Virtualization Technology (Intel® VT) and increasing power efficient performance through Intel® microarchitecture and multi-core processors, the number of virtual machines per server is increasing.
- Enhancing processing capability and server utilization require faster, scalable I/O.
- Isolation of device direct memory access (DMA) offers enhanced security and robustness.
- As hardware assists in processors, including Intel® VT-x, reduce software overhead on the processor side and bridge the gap to native performance, software-based sharing of high performance I/O devices among VMs will not be sufficient.

The goal of a holistic I/O virtualization solution hence would be to provide the:

- Same isolation that was found when the environment was running on a separate physical server.
- Scalability to support the number of virtual machines (VMs) necessary to take advantage of physical resources on I/O devices. They should also provide near native performance for I/O operations.

Intel® Virtualization Technology for Connectivity

Intel's latest addition to its suite of virtualization technologies is Intel® Virtualization Technology for Connectivity (Intel® VT for Connectivity). This new collection of I/O virtualization technologies improves overall system performance by improving communication between host CPU and I/O devices within the virtual server. This enables a lowering of CPU utilization, a reduction of system latency, and improved networking and I/O throughput.

Intel® VT for Connectivity is supported by Intel Ethernet adapters and it includes:

- PCI-SIG SR-IOV implementation
- Virtual Machine Device Queues (VMDq)
- Intel® I/O Acceleration Technology (Intel® IOAT)

For more information on Intel® VT for Connectivity, please refer to:
www.intel.com/go/vtc.

Today's Scenario

So far all the technologies available in the industry for a virtualized server are sharing and virtualizing a single physical port of the network adapter via software emulation in order to satisfy the I/O needs of the virtual machines. The multiple layers of the emulated software have been making all the I/O decisions for the virtual machines, thereby causing a bottleneck in this environment and impacting the I/O performance. It has also impacted the number of virtual machines a physical server can run in order to balance the system's I/O performance.

Solving the Problem

Current I/O virtualization techniques have many challenges. These challenges include:

- High I/O performance impact on a virtualized server
- The need for the software emulation layer to work on all communication and processing information, thus increasing CPU utilization
- The distribution of the interrupts and data via a single CPU core creates an I/O bottleneck on the system.

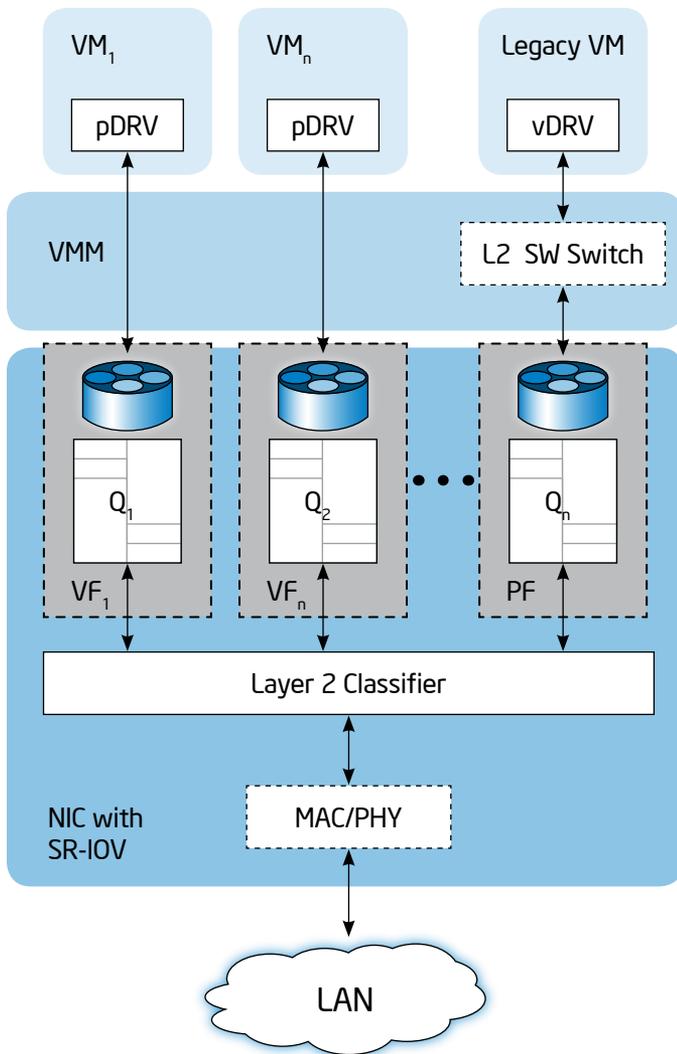


Figure 1: A typical network adapter supporting SR-IOV functionality

NOTE:

- pDRV: Physical Driver
- vDRV: Virtual Driver

To address these challenges, two solutions were created: VMDq and SR-IOV.

Intel developed VMDq technology to offload the data sorting functionality from the Hypervisor to the network silicon. This improves I/O performance to deliver near line rate throughput and lower CPU utilization.

In addition, an industry standards body, PCI-SIG, created the PCI-SIG Single Root I/O Virtualization (SR-IOV) specification to further improve I/O performance on a virtualized system. Intel actively participates with other industry leaders within the PCI-SIG working group.

SR-IOV defines a method to share a physical function of the I/O port of the I/O device without software emulation. This process creates a number of virtual functions per physical port of the I/O device. Each virtual function is directly assigned to a virtual machine, thereby achieving near native performance.

In summary, SR-IOV allows for the partitioning of a PCI function into many virtual interfaces for the purpose of sharing a PCIe device's resources in a virtual environment. SR-IOV enables network traffic to bypass the software emulation layer and to be assigned to the VM directly. By doing so, the I/O overhead in the software emulation layer is diminished.

PCI-SIG SR-IOV Advantage

PCI-SIG SR-IOV provides a standard mechanism for devices to advertise their ability to be simultaneously shared among multiple VM's. The SR-IOV spec allows an Independent Hardware Vendor (IHV) to modify their PCI card to define itself as several devices of the same type to a VMM (Hypervisor). A benefit of SR-IOV is the creation of a lightweight interface that allows an IHV to efficiently implement interfaces that can be directly assigned to VM's.

SR-IOV Overview

The goal of the SR-IOV specification is to standardize on a way of bypassing the VMM's involvement in data movement by providing independent memory space, interrupts, and Direct Memory Access (DMA) streams for each virtual machine. SR-IOV architecture is designed to allow an I/O device to support multiple Virtual Functions, while minimizing the hardware cost of each additional function. SR-IOV introduces two new function types:

- **Physical Functions (PFs):** These are PCIe functions that support the SR-IOV Extended Capability. The capability is used to configure and manage the SR-IOV functionality.
- **Virtual Functions (VFs):** These are 'lightweight' PCIe functions that contain the resources necessary for data movement but have a carefully minimized set of configuration resources.

The direct assignment method of virtualization allows a VM to interface directly to an I/O device. Therefore, direct device assignment provides a native experience and very fast I/O because it can reuse existing drivers or other software to talk directly to the device. However, it restricts the sharing of I/O devices. SR-IOV provides a mechanism by which a function (for example: a single Ethernet port) can appear as multiple separate physical devices.

An SR-IOV-capable device can be configured (usually by the VMM) to appear in the PCI configuration space as multiple virtual functions (VFs), each with its own PCI configuration space. The VMM then can assign one or more VFs to a VM by emulating the configuration space.

Each virtual function can support a unique and separate data path for I/O-related functions within the PCI Express hierarchy. Use of SR-IOV with a networking device, for example, allows the bandwidth of a single port (function) to be partitioned into smaller slices that may be allocated to specific virtual machines, or guests, via a standard interface. A common methodology for configuration and management is also established to further enhance the interoperability of various devices in a PCIe hierarchy. Such sharing of resources increases the total utilization of any given resource presented on a SR-IOV-capable PCIe device, thus reducing the cost of a virtual system.

Intel's implementation of PCI-SIG SR-IOV functionality requires the VMM software to configure the direct assignment of the virtual function to the virtual machine using Intel® Virtualization Technology for Directed I/O (Intel® VT-d). Memory Translation technologies in Intel® VT-d provide hardware assisted techniques to allow direct DMA transfers. Intel VT-d helps with secure translation, and SR-IOV provides separate data spaces for the virtual machines. For more details on Intel® VT-d, please refer to: <http://www.intel.com/technology/magazine/45nm/vtd-0507.htm>.

Intel® Server Adapters supporting PCI-SIG SR-IOV functionality can work with any vendors' platform solution that has the capability to configure the direct assignment of the virtual functions from the network's silicon to the virtual machines.

Summary

In summary, the key benefits of PCI-SIG SR-IOV on a virtualized platform are:

- A standard way of sharing the capacity of any given I/O device thus allowing for the most efficient use of that resource in a virtual system
 - Near native I/O performance for each virtual machine on a physical server
 - Data protection between the virtual machines on the same physical server
 - Smoother transition of a virtual machine between the physical servers thus enabling a dynamically provisioned I/O environment
- Intel's latest and upcoming PCIe Gigabit Ethernet adapters will support PCI-SIG SR-IOV functionality. Support must also be integrated into operating systems by the different VMM vendors in order to achieve the benefits of the technology.

To find out more about Intel® Server Adapters go to www.intel.com/network and www.intel.com/go/vtc.

*Other names and brands may be claimed as the property of others.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web Site.

Intel, Intel. Leap ahead, Intel. Leap ahead. logo, Xeon, and Xeon Inside are trademarks of Intel Corporation in the U.S. and other countries.

Printed in USA

SLT/046NTL/CMM

 Please Recycle

© 2008 Rev. 06/08-001US

