IDC
Analyze the Future

WHITE PAPER

Optimizing I/O Virtualization: Preparing the Datacenter for Next-Generation Applications

Sponsored by: Intel Corporation

Gary P. Chen Jean S. Bozman

September 2009

EXECUTIVE SUMMARY

Rapid adoption of virtualization technology in the x86 server market is bringing a wider array of applications into the virtualized IT infrastructure. Driving this adoption is the need to fully utilize computers, manage those resources more efficiently, and reduce the number of server "footprints" in the datacenter. All of this brings business benefits, such as reduced power/cooling requirements, reduced IT staff requirements, and greater IT flexibility. Computing tasks can be provisioned to available resources, as needed — and resources can be put to other uses, as business requirements change over time.

The rate at which data can flow from one device to another — long referred to as I/O, for input/output — is fast becoming a bottleneck in this growing virtualized infrastructure. If I/O is not sufficient, then it could limit all the gains brought about by the virtualization process.

With powerful new server hardware and software optimizations in the hypervisor, virtual machine (VM) density is increasing and is consolidating more I/O traffic onto the server bus. Also, many users are beginning to virtualize high-performance tier 1 applications that can be I/O intensive. High-speed I/O devices, such as solid state disks (SSDs) and 10Gb Ethernet (10GbE) links, will bring new levels of I/O performance but will further pressure existing I/O limitations. The evolution of converged fabrics in the datacenter will increasingly merge cluster, storage, and network traffic on a single physical fabric, requiring a high-performance I/O framework to support high throughput levels. This I/O fabric will also underpin the new internal cloud datacenter architecture that is an emerging, and maturing, model for "private clouds" within the enterprise.

The industry is looking for standards that will address these I/O requirements by providing new mechanisms for VMs to effectively share high-speed devices as a means to deliver peak throughput and lower latency to the VMs and to the applications they host. One new standard that has been published by the Peripheral Component Interconnect Special Interest Group (PCI-SIG) is called SR-IOV (short for single-root I/O virtualization). Adoption of this standard would "fit" with existing and widely deployed PCI Express (PCIe) I/O interfaces on an array of server products. Importantly, it would support the different I/O links within an enterprise network — such as 1GbE, 10GbE, and Fibre Channel. With the trend toward "converged" networking fabrics, support for all of these networking links is important to enterprise customers.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

Intel is working with other vendors to support the SR-IOV standard and to see that this standard is adopted across the industry. Wide adoption would accelerate the pace at which virtualization is proceeding — hastening the rate of transformation within the datacenter and enabling enterprises to move more computing resources into the virtualized x86 server space while allowing more mission-critical workloads to perform well within the virtualized infrastructure.

SITUATION OVERVIEW

The Rise of Virtualization

Virtualization technology has been deployed for many years, beginning with mainframes in the datacenter and proceeding to Unix servers and distributed systems over time. In the early to mid-2000s, virtualization technology began impacting the x86 server market, with the deployment of virtualization software from VMware.

By 2003, the market was evolving toward a new way of using virtualization software on x86 servers. About 70% of all virtualization software deployments in 2003 were related to software development and testing — applying the technology inside a sandbox of large organizations' test and development labs for consolidation purposes. But by the end of 2005, IDC saw spending shift from the consolidation of software development and testing environments toward the consolidation of applications within the production part of the IT infrastructure.

Since then, the industry has transitioned to focus more heavily on production-level consolidation, which continues to be a primary motivator for customers to bring virtualization within their organizations. In the interim, a variety of competitive solutions have entered the market, including multiple implementations of the open source Xen hypervisor technology and Kernel-based Virtual Machine (KVM) on Linux and Microsoft's Hyper-V.

In 2009 and 2010, several trends are expected. Production-level consolidation will continue and will begin to extend to the most mission-critical and performance-sensitive applications. Virtualizing these applications will demand virtualization-optimized hardware to reduce hypervisor overhead and, in particular, address I/O latency and throughput. In addition, as CPU core counts and memory sizes increase, VM density will continue to rise, putting a greater burden on individual servers to be highly available and fault tolerant.

Beyond consolidation, enterprises are beginning to leverage the hypervisor for extended use cases such as high availability, disaster recovery, and resource optimization. Virtualization offers a cheaper and sometimes better way to perform these tasks, taking advantage of the instant provisioning and live migration features. However, this also results in increased I/O demand as these "dynamic" virtualization features can regularly shuffle around VMs and data. As enterprises migrate to an internal cloud architecture and adopt external clouds, things will become even more dynamic, bringing higher utilization to the I/O fabric and server subsystems.

Despite the high interest in virtualization, IDC notes that not all servers are virtualized — yet. According to an IDC virtualization study, companies have virtualized only about 20% of their servers as of the beginning of 2009 — with plans to get to 50% in 2010. One big factor that is "gating" even faster adoption of virtualization is the requirement that I/O be improved in this virtualized environment. With so many VMs competing for available bandwidth, the aggregate I/O and the utilization and routing of that I/O "highway" of bandwidth must be increased. New initiatives, including those of the PCI-SIG, have brought about new standards for virtualized I/O, such as the SR-IOV standard, which is beginning to be adopted by system OEMs and ODMs and to be leveraged by customers who find that virtualized I/O will support next-generation applications better than conventional I/O components.

I/O Problems Associated with Virtualization

A hypervisor virtualizes all parts of a server, including I/O. Thus, any virtualized server is already utilizing some form of I/O virtualization by using software to share I/O devices such as an Ethernet network interface card (NIC). Although this function is convenient, much like other virtualized functions of the server, it is done in software and carries some overhead. The overhead of this software I/O virtualization reduces the overall I/O throughput of the system and increases latency. This I/O limitation primarily affects network (Ethernet) performance and storage (DAS, IP SAN, FC SAN, NAS) performance, the two key I/O areas for most enterprise applications. Software I/O virtualization is also a significant computational task because it requires additional CPU cycles. This may represent a significant overhead to the server (CPU utilization), and it has the result of significantly reducing available resources for VM workload processing, affecting consolidating ratio potential.

Trends in Virtualization Deployments That Are Straining Current I/O Limits

Virtualization by Default

The market is moving toward virtualizing the majority of infrastructure, including even tier 1 applications. The capex savings, and increasingly the opex savings, realized using virtualization has proved to be extremely compelling, motivating enterprises to increase the scale of their virtualization deployments. According to IDC's annual virtualization study, at most organizations, a virtual server is now the default build, with special exceptions required for a physical build. As a result, the number of VMs is growing at an exponential rate. As VMs begin to dominate physical servers, the demand for virtualization-optimized servers is growing as enterprises seek to achieve near native CPU, memory, and I/O performance from virtual servers.

Pushing for Ever Higher VM Density

The average number of VMs per physical server is rising — having grown from an average of two to four VMs per physical server several years ago to a new average of six VMs per physical server in 2008. It is not unusual to see 20–30 VMs per server for advanced deployments on new server hardware. The hardware platform is enabling higher VM density through the use of increased (processor) cores (now quad-core, or more), virtualization hardware assists built into newer CPUs (reduces virtualization

software overhead), and larger amounts of attached memory. This in turn has made it possible to host more demanding enterprise workloads on x86 server platforms as well as increase the number hosted per server. In these types of "rich" systems, I/O becomes further strained. All of these VMs must share the available I/O bandwidth, which has not scaled relative to the advances in virtualized CPU and memory resources.

Deployment of Enterprise Workloads in a Virtualized Environment

IDC research indicates that two-thirds of all virtual servers run production-level applications, including mission-critical workloads such as databases. Customers surveyed by IDC indicated that the decision to virtualize has less to do with the type of application than the characteristics of a workload. Applications that already have high utilization rates and those with high I/O requirements were the least likely to be virtualized.

However, several recent factors are now enabling the virtualization of the most demanding tier 1 applications:

- Powerful new server hardware that accelerates virtualization and reduces hypervisor overhead involves increased core count and improved native performance, allowing for the consolidation of applications even with high utilization rates.
- Increasing functionality in hypervisors, such as growing levels of virtual multithreading ("vSMP-wayness") and addressable memory available to individual VMs, enables more challenging applications to be virtualized.
- Increasingly compelling virtualization features are being added that have benefits beyond consolidation, such as mobility, flexibility, high availability, and agile management.

As enterprises push to virtualize nearly all their workloads, including demanding tier 1 applications, I/O subsystems will feel the strain more than ever.

Virtualization-Optimized Hardware

Virtualization has become the new design point for x86 servers, with servers built from the chip level and up for virtualization. In addition to the usual increases over time in native processing power and increased core count, server systems now actually accelerate virtualization with hardware assists. These improving hardware assists are continuing to reduce the performance penalties due to virtualization. With more and more workloads being virtualized and virtualization becoming the standard build at enterprises, server buyers are now expecting higher and higher levels of virtualization performance. While hardware assists have greatly impacted CPU and memory performance, much of I/O is still software based, leading to a performance mismatch that can affect many virtualization use cases.

Scaling I/O Capacity with More I/O Adapters

As the VM density increases, most customers are scaling I/O capacity by installing more adapters. For example, it is not uncommon to see six or more physical 1GbE NICs in a single server, with two (one primary, one backup) dedicated to live

migration, two to the service console, and two to the actual VMs. Customers install more NICs to increase the overall bandwidth of the system and isolate various network functions and for redundancy. However, this technique is reaching its practical limits as customers run out of expansion slots and as costs, cabling, power consumption, and management of all these devices become too much.

Emergence of Higher-Bandwidth Pipes and I/O Devices

Today's server I/O is dominated by 1GbE LAN, 2–4GBps FC SAN, and locally attached hard disks. The advent of 10GbE, Data Center Bridging (DCB, an enhanced Ethernet for datacenters), 8Gbps FC, as well as high-throughput SSDs and the maturation of their respective ecosystems brings "thicker pipes," which demand more efficient sharing methods by the virtualized server.

Enabling the "Dynamic Datacenter"

The next generation of virtualization has the potential to bring unprecedented efficiency and service levels to the datacenter. Server virtualization will have tighter integration and coordination with network, I/O, and storage, eliminating stovepipes by creating the fully virtualized datacenter. Operations will be policy based and service oriented/driven. An automated and intelligent management capability on top of the hypervisor will be required to achieve this dynamic datacenter. Given these characteristics, this kind of infrastructure will naturally lend itself to an internal cloud delivery model, where users can self-service and provision resources instantly, scale their applications, and be charged a utilitylike rate. Virtualization will also enable easier sourcing from external clouds, as abstraction is a key step in achieving internal-to-external cloud interoperability. High-performance, flexible I/O will be key to virtualizing the full range of workloads as well as enabling dynamic management, which requires constant, high-performance I/O between servers and to the external cloud.

THE BUSINESS NEED FOR IMPROVED I/O IN THE VIRTUALIZED ENTERPRISE

As enterprises push for better returns through higher consolidation ratios and the virtualization of all their applications, software-emulated I/O is rapidly becoming a limiting factor for virtualization. The demand to virtualize I/O-intensive tier 1 applications such as database and technical/compute-intensive applications and move to a fully virtualized, dynamic datacenter will require an I/O architecture that can deliver near native performance, increased throughput, and flexibility.

SR-IOV as a Solution for Virtualized I/O

IOV is simply the abstraction of the logical details of I/O from the physical, essentially to separate the upper-layer protocols from the physical connection or transport. There are several forms of IOV, which are explained in the IOV taxonomy sidebar. This discussion focuses specifically on IOV as it relates to server virtualization, in particular the SR-IOV standard and NICs.

Past Approaches to Addressing Virtualized I/O

- **Software.** As previously discussed, a hypervisor intrinsically virtualizes I/O by creating software emulations of I/O devices, such as a NIC, which allows multiple VMs to share a single server's hardware. The software overhead of this approach is significant. Later software techniques used paravirtualization, which exposed certain hypervisor APIs to the VM, which, in turn, ran a modified guest operating system (OS) or drivers that would utilize these APIs to speed up certain operations such as I/O. Though faster than the fully emulated approach, paravirtualization was still software based and incurred overhead.

- **Direct assignment.** To achieve native I/O performance, direct assignment of I/O devices to VMs can be accomplished. In this scenario, a VM is directly assigned and interfaces to a hardware I/O device (e.g., NIC or RAID controller), bypassing the hypervisor altogether, but requires one adapter or one adapter port for every VM on the server, which creates a scaling problem. In addition, the benefits of abstraction are lost because VMs are now directly tied to hardware, making any kind of VM migration difficult.

How SR-IOV Works

SR-IOV is a PCI-SIG standard that allows for efficient sharing of PCI devices among virtual machines and is implemented in the hardware to achieve near native I/O performance. SR-IOV creates a number of virtual function interfaces in the hardware of a physical PCI device. These virtual functions, which are essentially virtualized instances of the physical device, are then directly assigned to VMs and allow them to share this physical device and perform I/O without hypervisor software overhead.

COMPONENTS OF AN SR-IOV SOLUTION

SR-IOV requires many components of a server to work together to enable IOV:

- A system chipset that supports I/O isolation and direct assignment such as Intel Virtualization Technology for Directed I/O (Intel® VT-d)

- An SR-IOV–enabled NIC (or other I/O device) that has virtual functions built into the hardware

- A Basic Input/Output System (BIOS) that recognizes the enabling chipset and SR-IOV adapters

- A hypervisor, which must be modified to enable SR-IOV capabilities and for which users need to install a driver allowing the hypervisor to recognize the particular SR-IOV device

- A guest OS driver installed in the guest OS that allows it to perform I/O directly with the SR-IOV device

The recommended route for most users initially will be to attain SR-IOV technology through their server vendor because vendors will be expected to integrate and qualify all the required components as a more unified I/O solution, making it easier to deploy

and adopt this new technology. While it may be possible to upgrade some existing servers that have the prerequisite chipset, it is much more complex than simply installing a new NIC. Users will have to update the BIOS (if one is available) as well as their hypervisor and drivers. As the technology becomes more embedded into the majority of server systems and hypervisors over time, users will then be able to simply insert an SR-IOV I/O device and expect it to work. Because SR-IOV was designed to be compatible with existing PCIe technology, it is not expected to be disruptive to the IT infrastructure that is already in place.

Benefits of SR-IOV

The SR-IOV platform offers a number of technical benefits that improve I/O performance and consolidate system footprint and management while offering cost-effective, increased scalability and data protection and security.

- Performance is improved by:
  - Achieving near native I/O performance due to direct assignment of the device's virtual functions to the VM (Overall I/O throughput increases dramatically and latency is reduced by eliminating the software overhead.)
  - Reducing the system overhead incurred by the hypervisor for I/O, freeing the CPU to be used for productive tasks
  - Enabling hardware-assisted, efficient sharing of I/O devices
- I/O consolidation allows system managers to:
  - Use virtual functions instead of multiple physical I/O devices to separate I/O functions, including those related to storage and those related to networking technologies
  - Reduce hardware costs, simplify cabling, lower power consumption, reduce switch port usage, and reduce the number of devices to be managed
  - Increase the utilization of I/O devices for higher efficiency
- Increased scalability built into the hardware allows system managers to use a single higher-bandwidth I/O device (such as a 10GbE NIC) instead of multiple lower-bandwidth cards (such as 1GbE NICs) to meet bandwidth requirements while using virtual functions to isolate and provision network resources like separate physical NICs. In addition, this conserves valuable expansion slot space for other types of devices.
- Data protection and security is enhanced by using hardware instead of software to create and enforce data and I/O stream isolation between virtual machines.

BUSINESS VALUE OF I/O VIRTUALIZATION

Enhanced I/O through the use of IOV can increase the well-proven returns and savings from virtualization in that it enables system administrators to:

- Achieve higher consolidation ratios for increased server hardware savings
- Extend virtualization capex savings to previously unvirtualizable applications
- Extend virtualization operational efficiency savings to a new class of applications
- Improve response time and quality of service for applications
- Realize network/storage hardware capex savings from reduced adapter count, less cabling, and fewer switch ports
- Realize network/storage opex savings from simplified cabling, power savings, and improved management because fewer adapters and ports need to be managed

IDC's IOV Taxonomy

This sidebar presents a brief overview of IDC's IOV taxonomy to help clear up market confusion over the various contexts in which I/O virtualization is often mentioned:

- **Software IOV** (hypervisors, software-emulated I/O, paravirtualized I/O). This is the default configuration for the current generation of virtualization. Hypervisors virtualize an entire server, including the I/O, using software techniques. While very compatible and flexible, software emulation generally carries significant overhead.
- **Intraserver IOV.** This refers to a range of hardware-based technologies to virtualize the I/O to VMs hosted on a server. Most solutions are based around the PCI-SIG SR-IOV standard. Components can include:
  - Platform technology (Intel VT-d, ATS capability)
  - I/O adapter technology (multiqueue, SR-IOV virtual functions)
  - Enabling BIOS, drivers
- **IOV switches.** This refers to proprietary or PCI-SIG multi-root I/O virtualization (MR-IOV)–compliant I/O directors that virtualize I/O across multiple servers in a rack or a blade chassis; generally used for consolidation, bandwidth sharing, flexible provisioning, and simplified cabling.
- **Converged fabrics.** These technologies run multiple I/O protocols over a common fabric such as Fibre Channel over Ethernet (FCoE) or the upcoming DCB standard.

CHALLENGES/OPPORTUNITIES FOR SR-IOV

Challenges

As with any new technology, there are bound to be some challenges for adoption of SR-IOV. This section outlines some of the issues that are likely to surface relative to this change in I/O technology. Technical issues include the following:

- **Direct assignment.** SR-IOV achieves near native performance by direct assignment, but direct assignment also loses some of the flexibility and compatibility of software emulation, requiring adaptation of virtualization software.

- **Live migration.** VMs that use SR-IOV may be live migrated to a system without such capability or a system with different SR-IOV hardware. A mechanism needs to be developed to where a VM can "renegotiate" with the hypervisor as to whether the new target has SR-IOV capability and either fall back to software emulation if it does not or possibly change hardware profiles to accommodate different SR-IOV hardware.

- **New driver architecture.** With each VM directly interfacing to an I/O device using virtual functions, each with its own driver, it becomes vague as to where ultimate device control resides. A central location for processing events and resolving control conflicts must be established through a master driver in the hypervisor. Guest drivers in the VM would primarily interface directly with the hardware but coordinate with the master hypervisor drivers in certain instances. This will require hypervisor vendors to modify their software for SR-IOV.

At the same time, there are also likely to be business and market challenges ahead, as would be the case for adoption of any new technology within the datacenter. Some of these issues include the following.

- **Market awareness.** At this point, customers do not have a full understanding of IOV in terms of the different types of IOV available and the technical and business benefits that IOV offers. Many do not know that I/O is a limiting factor in their virtualization deployments. Part of the challenge is translating the technical benefit messages into business benefit messages understood by business unit managers who ultimately approve IT purchases.

- **Industry cooperation.** SR-IOV requires many IHVs, OEMs, and ISVs to coordinate and build support for this technology in order to deliver a qualified and integrated solution to the end user. The fact that the PCI-SIG has published the SR-IOV standard (as first introduced in September 2007) should help to overcome this obstacle. The industry has rolled out key components of the solution, with enabling chipsets and SR-IOV adapters already available.

Opportunities

It is important to note that there are many opportunities for the use of SR-IOV as it is introduced into the broad IT marketplace. Some of the leading opportunities for this new technology are as follows:

- **Extending virtualization to tier 1 applications.** The sheer number and performance requirements of applications being virtualized are straining existing I/O limits. By deploying systems with IOV technology, customers can improve throughput and allow more enterprise applications to be run in a virtualized environment, fitting in well with the trend of virtualizing the entire infrastructure.

- **Extending I/O consolidation.** Virtualization is creating "I/O sprawl" as users seek to keep up with I/O requirements, leading to consolidation opportunities with IOV-enabled systems.

- **Enabling the dynamic datacenter.** Customers are moving toward the development and deployment of dynamic datacenters (internal clouds), which will be fully virtualized and include virtualization technologies beyond just server technologies. Vendors supplying IOV-enabled systems will be able to better support dynamic datacenter initiatives, aimed at improving IT flexibility and business agility.

CONCLUSION

Adoption of virtualization is entering a new phase. After a first wave of virtualizing servers to gain greater efficiency and to consolidate workloads onto fewer server "footprints," a second wave of business-critical and mission-critical workloads will be deployed into this computing environment. Because these workloads are more demanding, they will require greater I/O rates than earlier deployments.

In recognition of this change in IT deployments, the aggregate amount of I/O that can be supported within any given server will need to rise dramatically. All of this must happen within the confines of a highly dense computing environment. The advent of converged network fabrics makes this all the more timely and important for next-generation virtualization. The PCI-SIG has created its SR-IOV standard to answer these IT processing requirements for virtualized IT infrastructure. But this change in the datacenter will also bring business benefits, as computing will be more adaptive — allowing applications to move more easily to alternate computing resources and allowing IT resources to become more closely aligned with changing business requirements.