

Solving the Hypervisor Network I/O Bottleneck

Solarflare Virtualization Acceleration

White Paper

By: Gary Gumanow

9 October 2007

SF-101233-TM

Introduction

With increased pressure on IT departments to *do more with less*, IT is turning to server virtualization to get the most out of their server platforms. The adoption of virtualization technology to address these business conditions in enterprise data centers has led to improvements in scalability, manageability, and availability.

Solarflare’s SolarStorm™ 10 Gigabit vNIC controllers are built from the ground up to alleviate the I/O bottleneck in a virtualized environment and can provide up to three times the network I/O performance when compared with a non-accelerated virtualized environment.

This paper will highlight the improvements that SolarStorm’s virtualization acceleration can provide, not only in networking I/O performance, but also in server efficiency.

The Problem

As servers become more powerful, IT departments are consolidating workloads on fewer and fewer servers utilizing virtualization software techniques. Over the past ten years, the scale-out server model based upon standards-based high-volume components has been a very successful tool for delivering higher performance. Along the way applications have been tuned to maximize the benefits of this scale-out data center model. However, this scale-out or server sprawl has added to the complexity and cost of managing data centers, and IT managers are now moving to consolidate the number of physical servers. IT shops can now achieve higher performance from the same or smaller data center footprint.

Before virtualization, each server and application had dedicated bandwidth and hardware access to the network through its network adapter and switch port. Consolidating applications onto fewer servers requires applications to share that host’s physical hardware (network adapter and bandwidth).

Virtualized or shared network I/O cannot keep up with the number of applications consolidated. This is often compounded even further by the need to provide even greater I/O bandwidth for storage networking, negating the benefit of server consolidation. In effect, IT resource concerns of data center power and cooling, is now replaced with concerns over network I/O performance (Figure 1). A new approach is needed to breakthrough this I/O bottleneck on a virtualized server platform.

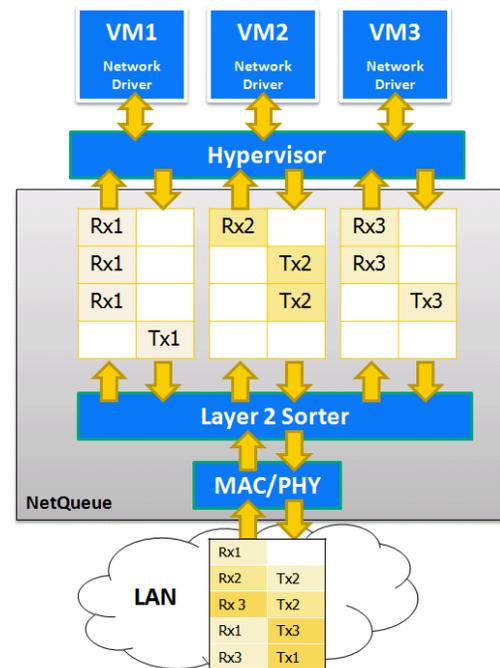


Figure 1 – All network I/O goes through the hypervisor and becomes the bottleneck.

SolarStorm™ virtualization acceleration

The virtualization acceleration provided by Solarflare’s SolarStorm™ 10GbE vNIC controller offloads the I/O burden of the hypervisor. SolarStorm™ virtualization acceleration (SVA) is an integrated hardware and software solution of the SolarStorm vNIC controller, along with modifications made by the hypervisor OS vendors (Figure 2). SVA provides a bypass from the virtual machine network driver directly to the SolarStorm 10GbE vNIC controller, bypassing the hypervisor. This moves network I/O processor resources from the hypervisor to the VM’s, more effectively utilizes today’s multi-core architectures, and thus allows for further consolidation of physical servers.

This also frees the hypervisor to be more efficient and responsive, as well as, to perform other tasks, such as, management tasks, monitoring the health of virtual machines, migrating virtual machines, and providing security functions.

The PCI-SIG specification for PCI-IOV* is currently under development and Solarflare plans to support this industry standard when it becomes ratified. PCI-IOV is complementary to SVA and is not expected to impact the SVA architecture. Acceleration from PCI-IOV is expected to deliver the same virtualized performance and efficiency as SVA. See Figure 3.

How SVA Works

The SolarStorm controller establishes a vNIC (virtual NIC) for each virtual machine in hardware, thus communicating with that virtual machine network driver directly. Each vNIC in this case could be considered a network queue or stream, with its own memory protection and security. The SolarStorm vNIC controller supports up to 4096 vNICs while other 10GbE controllers on the market today support only 16 virtual connections.

- In non-accelerated mode the vNIC communicates via the hypervisor to the virtual machines, overloading the hypervisor, limiting the CPU bandwidth for the virtual machine’s virtualized applications.

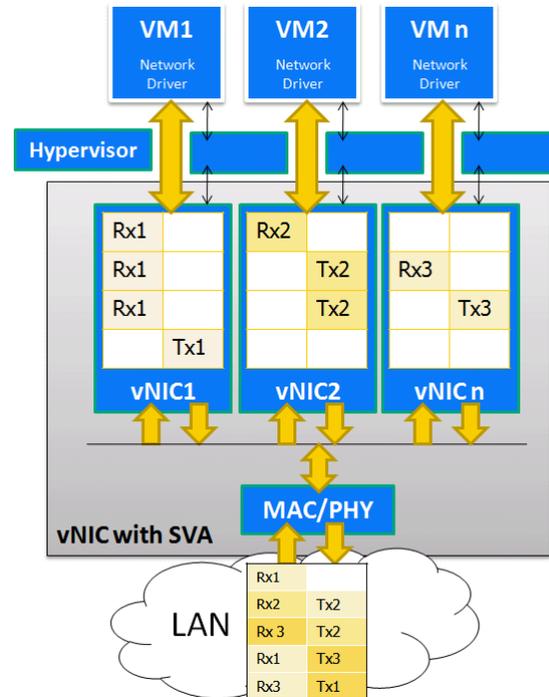


Figure 2 – SVA offloads network I/O traffic and preserves hypervisor management resources.

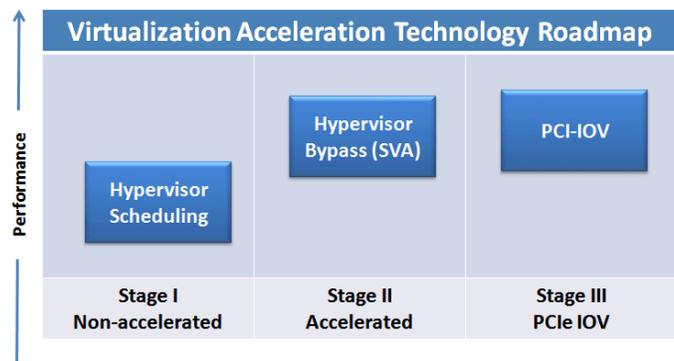


Figure 3 – Virtualization acceleration stages and relative performance

- In accelerated mode the vNIC still maintains a virtual connection through the hypervisor for management of the virtual machines. However, for network I/O the vNIC shifts network traffic to a vNIC for bypass mode. This greatly improves network I/O, virtual machine application performance and hypervisor responsiveness (Figure 2).

SVA Performance

In a performance study, Solarflare was able to measure the benefits of running in both an accelerated and non-accelerated virtualized environment. When comparing the same hardware environment, with the same applications virtualized, Solarflare observed a dramatic increase in hypervisor responsiveness, as well as, a 3x performance gain in application performance. See the Chart 1. This configuration included two Dell PowerEdge 2950 Quad-Core Intel® Xeon® processor based servers running a modified XenEnterprise 4.0. Each physical machine was running three virtual machines (RHEL5), with NetPerf running on each virtual machine. The physical machines were connected by a pair of SolarStorm™ 10GBASE-T adapters.

Results

Running in non-accelerated mode, the throughput observed was 3.1Gbps and the hypervisor CPU utilization was 95%. With bypass mode enabled throughput climbed 3x to 9.2Gbps and the hypervisor CPU utilization dropped to 8%.

Solarflare will continue to roll out SVA instantiated in controller silicon and software drivers in 2007 and 2008. The company has already announced its co-development partnerships with VMware, and publicly demonstrated SVA on a modified XenEnterprise 4.0 release. In 2008, Solarflare will make available to its server OEM and NIC OEM customers the

full suite of support for all the major operating systems, including both non-accelerated and accelerated support for Xen OpenSource, Xen Enterprise, VMWare 3.5, and Microsoft Viridian.

Summary

Solarflare's SolarStorm™ 10GbE vNIC controller provides the powerful acceleration to relieve network I/O bottlenecks hidden in virtualized environments, allows IT managers to direct server resources to the applications that need performance, and *do more with less*. SolarStorm's virtualization acceleration boosts network throughput by 3x, and reduces CPU utilization of the hypervisor dramatically, from 95% to 5%. Virtual machine applications get back the I/O and server resources they need, especially in a consolidated, virtualized environment.

Solarflare continues to deliver on the promise of the converged network, with acceleration for iSCSI and virtualization. When utilized together, the SolarStorm 10GbE vNIC provides the bandwidth and server

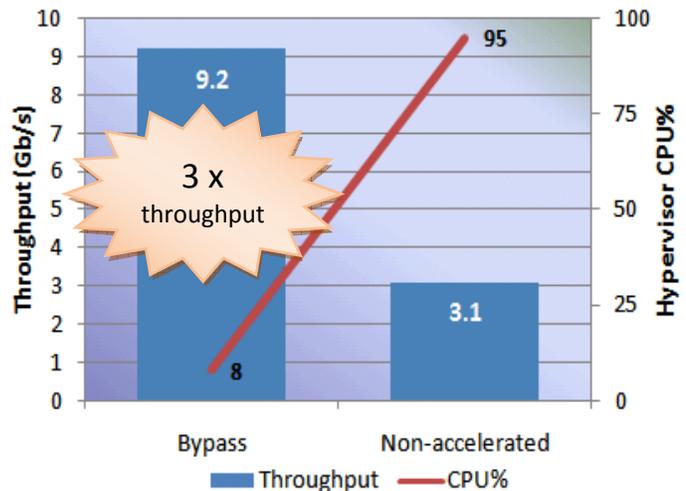


Chart 1 – Shows the dramatic 3x throughput improvement, and CPU utilization reduction from using Solarflare's 10GbE vNIC controller's virtualization acceleration (SVA). CPU % shows utilization of dom0 or the hypervisor

resources required for both data and storage to coexist on the same network. These are just some of the inherent advanced features of the SolarStorm 10GbE vNIC controller. In the future, Solarflare will support PCI Express IOV for even further advancements in converged network performance.

Moore's Law continues to move the computer industry forward with improved processor and platform performance, and more IT shops will continue to look for ways to get the most out of their hardware and software. Virtualization and the SolarStorm™ 10GbE vNIC controller are an excellent method for accomplishing this goal.

About 10Xpress® and Solarstorm™

Solarflare Communications designed the Solarstorm 10 Gigabit Ethernet vNIC controller to deliver the industry's highest performance, lowest-power 10GbE solution and leverages the host server platform resources as efficiently as possible. Solarstorm server adapter reference designs (10GBASE-T, 10GBASE-CX4, and XFP optical) are available now. The Solarflare 10Xpress 10GBASE-T PHY supports distances of at least 100 meters on UTP Category 6A cabling. The PHY is also available now. Pricing is available upon request. For more information about these products, contact: productinfo@solarflare.com

About Solarflare Communications, Inc.

Solarflare Communications is a semiconductor company delivering the next level of high performance Ethernet. Leading provider of standards-based 10 Gigabit Ethernet silicon and reference designs, Solarflare delivers robust and power-efficient solutions that are cost effective and easy to deploy. Now real and primetime, Solarflare 10 Gigabit Ethernet makes possible next generation applications such as streaming rich media, Web 2.0, data center virtualization and network convergence. Solarflare is working with key industry partners and customers to ensure interoperability and drive a complete ecosystem. The privately held company is headquartered in Irvine, California with a development center in Cambridge, UK. For more information, visit <http://www.solarflare.com>.

* All product and company names herein may be trademarks of their registered owners.