PERFORMANCE AWARE VIRTUALIZATION IN I/O WORKLOADS

Muwaabe Nassir, Dr. R.Suchithra,
MS(IT) Department, Jain University
JCRoad. Bangalore
muwaabe@gmail.com
suchithrasuriya@gmail.com

***Abstract* – Cloud computing has taken on the I.T industry through a fast technological revolution. And undeniably the underlying technology preferred lately is virtualization, categorized differently by names such as OS virtualization, server virtualization, kernel virtualization, Network virtualization, hypervisors etc. These technologies are becoming more inexpensive and less complex every day, making them attractive technologies and a requirement to the fast growing IT industry. Factors such as consolidation, cost savings, dynamic provisioning, fluid migration and the cost of administration are driving everything IT to experiment with some form of virtual machine product today.**

The fanciness of this technology has seen the average number of VMs per physical server rising. All of these VMs must share the available I/O bandwidth, which becomes further strained. 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) thus causing a bottleneck in I/O performance. 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) and storage performance, the two key I/O areas for most enterprise applications. It also represents 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.

Focusing on the performance of the I/O, we take a study on how to assign the VM directly to the I/O devices without emulation through software virtualization.

Keywords - Virtualization, Network Overlays, VMware, Virtual Networks, Hypervisor

I. INTRODUCTION

The advent of microcomputers in the late 80’s and their widespread adoption during the 90’s along with ubiquitous networking brought the distribution of computing to new grounds. Large number of client machines connected to numerous servers of various types gave rise to new computational paradigms such as client-server and peer-to-peer systems. These new environments brought with them several challenges and problems including reliability, security, increased administration cost and complexity, increased floor space, power consumption, and thermal dissipation requirements.

The recent rebirth of the use of virtualization techniques in commodity, inexpensive servers and client machines is poised to address these problems as Daniel A. Manasce sights [1] in a very elegant way. This paper describes the basic concepts in virtualization and explains how virtualization can be used to support server consolidation efforts with respect to I/O performance as a result 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 mid 2000s, virtualization technology began impacting the x86 server markets, 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 [2]. 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 i.e. virtualization as a test bed. But things started to change by the end of 2005, shifting 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 center of attraction towards virtualization within 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. 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 internal cloud architecture and adopt external clouds, things will become even more dynamic, bringing higher utilization to the I/O fabric and server sub systems

II. RELATED WORKS

The rising cost and complexity of building and operating modern data centers have led organizations make more innovations into making the data center infrastructure simpler and up to the mark with the current technology improvements. Although the cost of data center networking equipment is relatively small compared to the cost of server hardware and software, the underlying network fabric is the core that connects all mission critical resources. As sighted in the study by Juniper Networks [3], a simpler, more streamlined data center fabric means greater efficiency and productivity and lower operating costs. In addition, shared (centralized or distributed) storage, be it file-based (Network Access Storage (NAS), or block-based (storage area network (SAN) using Internet Small Computer System Interface (iSCSI), Fibre Channel (FC), and Fibre Channel over Ethernet (FCoE)) are essential elements of an effective compute and storage solution for data centers and the cloud. They can be used in concert to support advanced virtual systems and the overall virtual networking infrastructure.

Traditionally, servers are deployed with multiple I/O cards to connect to multiple separate physical network segments or even completely separate network infrastructures: dual SAN for disk access, another SAN or LAN for backup, dual LAN for client/server or campus LAN connection, out-of-band management, vMotion and cluster traffic. I/O convergence helps to reduce the number of such interfaces and networks. It has been promoted along with Ethernet or IP-based storage technologies such as iSCSI (Internet small computer system interface) NAS and more recently FCoE.

With the increased affordability and rapid adoption of 10GbE in the data center, Ethernet is poised to take on the connectivity tasks formerly relegated to InfiniBand and fibre-channel to become the dominant data center networking technology. Reducing the number of I/O cards and network ports drives many potential savings.

Virtualization comes at a cost of reduced performance due to hypervisor architecture. Following studies by Broadcom [4], today's virtualization architecture includes VM with device driver, I/O stack and applications, layered on top of a Virtualization layer that includes device emulation, I/O stack and physical device driver managing the Ethernet network controller. This additional virtualization layer adds overhead and degrades system performance including higher CPU utilization and lower bandwidth.

Broadcom Ethernet network controllers support stateless offloads such as TCP Check Sum Offload (CSO), which enables network adapters to compute TCP checksum on transmit and receive, and TCP Large Send Offload (LSO), which allows TCP layer to build a TCP message up to 64 KB long and send it in one call down the stack through IP and the Ethernet device driver, saving the host CPU from having to compute the checksum in a virtual environment. Jumbo frame support in virtual environments also saves CPU utilization due to interrupt reduction and increases throughput by allowing the system to concentrate on the data in the frames, instead of the frames around the data. However, performance is still limited due to the single threaded nature of hypervisor in processing I/O and duplicate I/O copies in the virtualization layer. Netqueue support in VMware and VMQ support on Microsoft Hyper-V removes single queue bottlenecks and use of state-full offloads such as TCP offload. The iSCSI hardware based acceleration in virtual environments is proven to provide excellent performance on VM.

Quite sizeable amount of research has also been conducted along the lines of trying to avert the issues on how to optimize the network performance in such scenario.

1) XenLoop: XenLoop [5] is a full transparent and high performance inter-VM network loopback channel implemented in Xen. Guest virtual machine can switch between the standard network path and XenLoop channel seamlessly. Xenloop intercepts network packets under the network layer. If co resident communication is detected, packets would be sent to target VM through shared memory channel that bypasses the virtual network interface controller.

2) XenSocket: XenSocket [6] is high performance network channel designed for co-resident inter-VM communication with a static circular memory buffer shared between

VirtIO is designed to be a de-facto standard for virtual I/O devices, as comprehensively discussed by Ning [7] and Rusty Tussell [8], it is currently implemented in KVM and achieves excellent I/O performance. Like Xen, VirtIO is a para-virtualization solution that requires drivers to be installed in guest machines. However, VirtIO aims at a unified framework and interface for device drivers in different virtualization environment since the differences among hypervisors bring great difficulties when writing and maintaining device drivers. By now, VirtIO drivers for disk, network and PCI devices have been implemented for various operating systems. Performance is improved significantly when compared with original virtualization scenario. However, redundant data copy still exists.

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 such as NIC or RAID controller, bypassing the hypervisor altogether, but requires one adapter or one adapter port for every VM on 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.
High-throughput and low latency is especially important in a distributed system, where the I/O latency of each node impacts both cluster and overall application performance. Low latency is required in order to preserve data coherency in large database clusters implementing scalable SR-IOV [9] network

adapters. With VMware Direct Path network plug-in architecture and Broadcom SR-IOV device, a VF can be directly assigned to a VM. This yields native performance and eliminates additional I/O copy in the Hypervisor with the added advantage of being able to support all virtualization features including live migration or vMotion for VMware. Direct assignment of PCI devices to VM will be necessary for I/O appliances and high performance VMs.

Getting into the nitty-gritty of these different methods, we can easily see that they all have immense shortfalls in regards to performance. While they solve the problem on one hand they create a performance bottleneck on the other. F5 Networks Inc [10] puts emphasis on some of the issues that need not skip one's mind before thinking virtualization and performance.

While virtualizing with intent of optimized hardware may pay off having great impact on CPU and memory performance, much of I/O is still software based, leading to a performance mismatch that can affect many virtualization use cases since I/O is still constrained by the overwhelming workload.

Scaling I/O capacity with more I/O adapters doesn't prove to be any better either. As the VM density increases, the option of scaling I/O capacity by installing more adapters hits anyone's mind as a quick and straight solution. For example, it is not unique to see more than a single 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.

As previously discussed, addressing virtualized I/O software, 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. It is software based and incurs significant overhead.

III. DYNAMIC SINGLE ROOT IOV

Single Root I/O Virtualization (SR-IOV) is yet another approach that proposes a set of hardware enhancements for the PCI-Express (PCIe) device, which aims to remove major VMM intervention for performance data movement, such as the packet classification and address translation. SR-IOV inherits Direct I/O technology through using IOMMU to offload memory protection and address translation. An SR-IOV-capable device is able to create multiple "light-weight" instances of PCI function entities, known as Virtual-Functions (VFs). Each VF can be assigned to a guest for direct access, but still shares major device resources, and thus achieve both resource sharing and high performance.

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.

IV CONCLUSION

The tenacity of the innovations that are being bombarded on the IT industry each fall needs not to leave out any of the stake holders of the industry as all the aspects seem to be dependent on each other. Virtualization being one of the fast and fancied innovations in the industry should be the driving force of any vendor. The speed at which hardware and software vendors evolve the different products on market should be in line with the recent trends such that the hardware I/O devices do match the virtualization needs.

REFERENCES

[1] *Daniel A. Menasce: Virtualization: Concepts, Applications, And Performance Modeling.*

[2] *Gary P. Chen, Jean S. Bozman. Optimizing I/O "Virtualization: Preparing the Datacenter for Next-Generation Applications" IDC White Paper, September 2009*

[3] *"Cloud-Ready Data Center Reference Architecture" Juniper Networks White paper. February 2013*

[4] *"Broadcom Etherner Network Controller Enhanced Virtualization Functionality",Broadcom White Paper October 2009*

[5] *J. Wang, K.-L. Wright, and K. Gopalan. XenLoop: a transpar-ent high performance inter-vm network loopback. in HPDC '08: Proceedings of the 17th international symposium on High performance distributed computing. New York, NY, USA:ACM, June 2008*

[6] *X. Zhang, S. McIntosh, P. Rohatgi, and J. L. Griffin. XenSocket: A High-Throughput Interdomain Transport for Virtual Machines. in Mid-dleware '07: Proceedings of the ACM/IFIP/USENIX 2007 International Conference on Middleware. New York, NY, USA: Springer, November 2007*

[7] *Fengfeng Ning, Chllang Weng, Yuan Luo. "Virtualization I/O Optimization Based on Shared Memory" International Conference on Big Data IEEE 2013.*

[8] *R. Ressell, R. "Virtio:Towards a de-facto standard for virtual I/O Devices." SIGOPS Oper. Syst. Rev. 42, 5 (2008)*

[9] *Yaozu Dong, Xiaowei Yang etl " High Performance Network Virtualization with SR-IOV"*

[10] *F5 Networks Inc. "Keeping Your Head Above the cloud: Seven Data Center Challenges to Consider before Going Virtual" White Paper, 2008*