High Performance iSCSI for Virtual Machines

Software Defined Storage with Terminator 5

Executive Summary

Server virtualization has become an essential tool in building agile and efficient IT infrastructures. However, the flexibility it provides often comes at the expense of performance, particularly in terms of network I/O. This paper shows how virtual machines can experience large improvements in storage I/O performance through access to the fully offloaded iSCSI in Chelsio's latest Terminator 5 (T5) ASIC. With both iSCSI initiator and target running in a VM space, performance is shown to exceed 300K IOPS. This breakthrough enables the virtualization of storage systems without loss of performance, and paves the way for high performance Software Defined Storage.

Overview

The Terminator 5 (T5) from Chelsio Communications, Inc. is a fifth generation, hyper-virtualized, high-performance unified wire device with full iSCSI offload support, among other networking, clustering and storage protocols. All of the iSCSI protocol processing is implemented in hardware, thus freeing the CPU for other computing needs. Furthermore, the adapter hardware handles the expensive byte touching operations, such as CRC computation and verification, and direct DMA to the final host memory destination.

T5 supports 128 independent virtual interfaces in hardware, with access control and filtering, as well as an integrated virtual switch. The virtual interfaces provide separate configuration and statistics capabilities, and can be used along with PCI SR-IOV support, or independently thereof. The virtual switch implements broadcast and multicast support, and is capable of switching between virtual machines as well as between external ports.

While virtualization brings in many benefits in improved flexibility, agility, reliability, manageability and resource efficiency, it imposes additional overheads that particularly impact network performance. T5's unique capability to allow protocol offload access within virtual machines, delivers improved performance and efficiency to the overall virtualized infrastructure.

This paper demonstrates how high storage networking performance can be achieved for initiator and target software running within virtual machines. By eliminating the most significant performance bottleneck in virtualizing storage systems, this T5 capability removes a major hurdle in delivering competitive software defined storage solutions, in addition to considerable savings in power and resources utilization.

Test Results

The following graph illustrates benchmarking data obtained for random READ, WRITE and READ/WRITE IOPS with 4K-64K I/O size using **fio** tool.

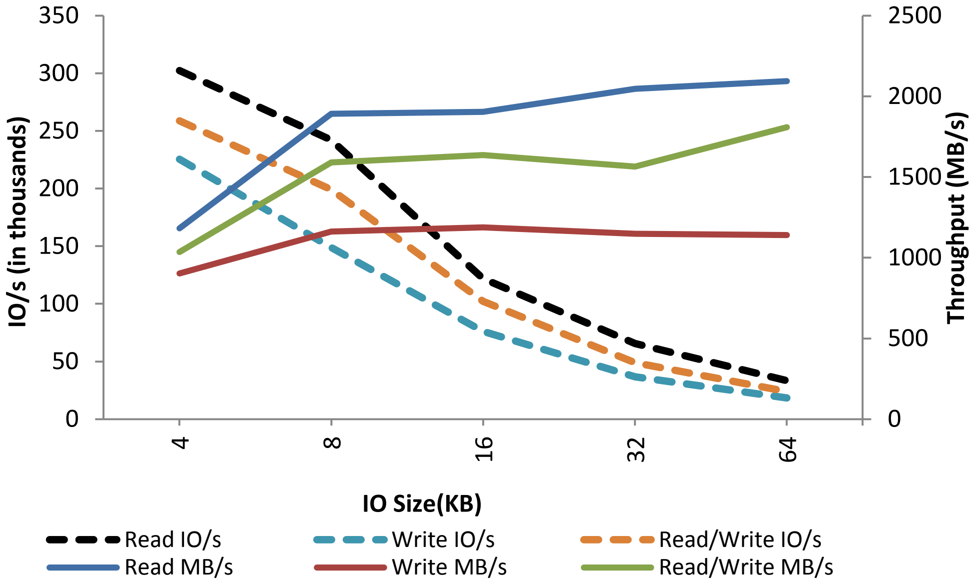

Figure 1 – iSCSI IOPS and Throughput vs. I/O Size

Test Setup

The test setup consists of test VM and a target VM running on the same system, and communicating using the T5 on-chip embedded virtual switch. The target VM communicates through VM Direct Path to the T5 adapter, whereas the initiator side runs a para-virtualized driver to utilize the fully offloaded T5 initiator.

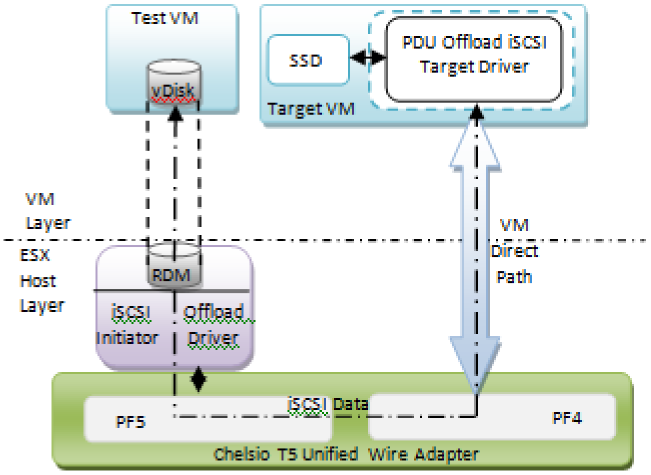

Figure 2 – Test Setup

Test Configuration

The following sections provide the test setup configuration details:

Network configuration

The network configuration consists of two components: a 10G switch and an ESXi host machine configured as FOiSCSI Initiator. Two VMs were configured: one as an iSCSI Target (Target VM) and the discovered LUN was attached to the other (Test VM).

- **The host machine** was configured with an Intel Xeon CPU E5-1660 v2 processor running at 3.70GHz. Chelsio T520-LL-CR dual-port adapter was installed in the system with Chelsio's FOiSCSI initiator driver (*ima-csiostor-2.3.0-1, scsi-csiostor-1.0.40*) and ESXi 5.5 hypervisor.
- **The Target VM** was configured with 4 vCPUs and 32 GB of RAM with Fedora 17 [Kernel - 3.6.11-1.fc17.x86_64] installed as Guest OS. Chelsio's PDU Offload iSCSI target driver build 734 was installed and configured on PF4 of Chelsio adapter.
- **The Test VM** was configured with 8 vCPUs and 4GB of RAM with Fedora 17 [Kernel - 3.6.11-1.fc17.x86_64] installed as Guest OS.

Storage Topology and Configuration

The target VM consisted of a single SSD configured in pass-through mode.

I/O Benchmarking Configuration

fio was used to assess the I/O capacity of the configuration. The I/O size used was 4K Bytes to 64K Bytes with an I/O access pattern of random READs, WRITEs and READ/WRITEs.

fio config file used in the test:

```
[global]
ioengine=libaio
iodepth=64
direct=1
numjobs=32
filesize=1g
size=32g
time_based
runtime=60
group_reporting=1
cpus_allowed=0,1
ioscheduler=noop
#################
[random_read_4k]
rw=randread
blocksize=4k
filename=/dev/sdb
stonewall
```

Setup Configuration

The setup was configured using the following procedure:

- **On the Host**

i. Port0 and Port1 of Chelsio T5 CNA were connected to the Cisco switch for the purpose of link-up.
ii. The SSD and PF4 of the Chelsio adapter were configured in pass-through mode.
iii. FOiSCSI Initiator driver was installed and host machine was rebooted:

```
[root@host]# esxcli software vib install -v /path-to-driver/ima-
csiostor-2.3.0-1OEM.550.0.0.136.1.i386.vib

[root@host]# esxcli software vib install -v /path-to-driver/scsi-
csiostor-1.0.40-1OEM.550.0.0.136.1.x86_64.vib

[root@host]# reboot
```

iv. After the host was up, driver was enabled to load with "csio_loopback=1" option and rebooted:

```
[root@host]# esxcfg-module -s csio_loopback=1 csiostor
[root@host]# reboot
```

v. After the host was up, the SSD and PF4 of the Chelsio adapter were attached to the target VM using vSphere Client.

- **On the Target VM**

i. The target VM was then powered up and the following sequence of commands were executed:

```
[root@host]# modprobe cxgb4 tx_vm=1
[root@host]# modprobe t4_tom cop_managed_offloading=1
[root@host]# modprobe chiscsi_t4 ddp_reg_modify=0
```

ii. *chinfootool* was used to generate iSCSI license information file.
iii. The file obtained in the previous step was sent to Chelsio Support team for license.
iv. The license file received was installed to unlock the iSCSI Target software:

```
[root@host]# cp -f chiscsi.key /etc/chelsio-iscsi/chiscsi.key
```

v. Chelsio PDU Offload iSCSI target was configured using the following *chiscsi* conf file with the SSD as a LUN :

```
# sample iSCSI configuration file with 1 target
# iSCSI Global Settings
global:
# change iscsi_offload_mode to TOE or ULP if desired
       iscsi_offload_mode=TOE
       iscsi_auth_order=CHAP
# change the target vendor ID to "Chelsio"
#      iscsi_target_vendor_id=Chelsio
target:
```

```
TargetName=iqn.2004-05.com.chelsio.target0
TargetAlias=iscsitarget0
MaxRecvDataSegmentLength=8192
HeaderDigest=None,CRC32C
DataDigest=None,CRC32C
ImmediateData=Yes
InitialR2T=Yes
MaxOutstandingR2T=1
MaxConnections=4
TargetSessionMaxCmd=2048
TargetDevice=/dev/ssdb1
PortalGroup=1@102.11.11.32:3260
```

vi. *cop* tool was used to distribute connections to different offload queues:

```
[root@host]# echo "any => offload ddp bind random" >  ~/cop.conf
[root@host]# cop -d -o ~/cop.out ~/cop.conf
[root@host]# cxgbtool ethX policy ~/cop.out
```

vii. Chelsio iSCSI worker threads were mapped to different cores using taskset.

viii. Interrupts of offload queues and SSD were mapped to separate cores from that of iSCSI worker threads.

ix. Next, the Chelsio PDU iSCSI target was started:

```
[root@host] # iscsictl -S target=ALL
```

x. Logged in to Target VM using ESXi FOiSCSI initiator through vSphere Client. The SSD was discovered on the Initiator machine.

- **On the Test VM**

i. Para-Virtualized driver was used for attaching the discovered LUN to the Test VM as Raw Device Mapping (RDM) disk, which appears as local disk.

ii. Finally, LUN on the Test VM was formatted and traffic was run using the **fio** tool:

```
[root@host]# rmmod vmw_pvscsi
[root@host]# modprobe vmw_pvscsi cmd_per_lun=255 ring_pages=4
[root@host]# mkfs.ext3 /dev/sdb
[root@host]# fio fio_config_file
```

Conclusion

This paper demonstrates the benefits of providing direct access to the offloaded iSCSI implementation in Chelsio’s T5 adapter, showing how T5 can deliver more than 300K IOPS performance in a virtualized environment, where both target and initiator run within a VM. This eliminates a significant hurdle in the path of fully virtualizing storage systems, a key enabler of storage defined networking.

Related Links

The Chelsio Terminator 5 ASIC