100G iSCSI Offload Performance for AMD EPYC

Using AMD EPYC 7551 Platform & Chelsio T6 Adapter

Executive Summary

AMD EPYC, industry's first hardware-embedded x86 server security solution, is a system on chip (SoC) which provides exceptional processing power coupled with high-end memory and I/O resources to meet workload demands of any scale, from virtualized infrastructures to cloud-era datacenters. The combination of the AMD EPYC 7551 server with Chelsio's industry-leading Unified Wire adapter solution delivers compelling performance, power and total cost of ownership (TCO) advantages. This enables innovative topologies and networked computing models to address the most demanding processing needs. This paper shows Chelsio 100G iSCSI offload solution delivering 98 Gbps line-rate throughput and more than 2.7M IOPS with minimal CPU utilization for a cost-effective enterprise-class storage target solution built with volume, off-the-shelf hardware and software components.

Test Results

The following graph presents READ, WRITE IOPS and throughput performance of Chelsio T6 iSCSI solution using null block device as storage array. The results are collected using the **fio** tool with I/O size varying from 4k to 512k bytes with an access pattern of random READs and WRITEs.

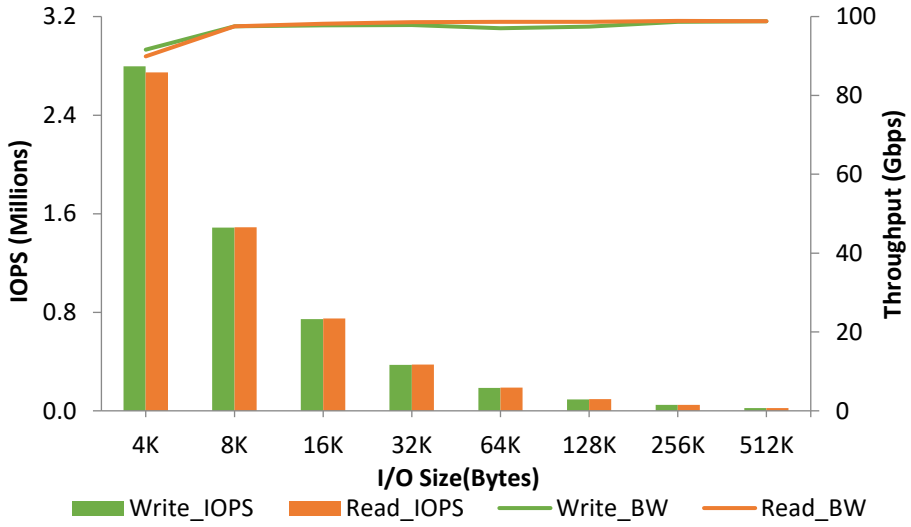

Figure 1 – READ, WRITE Throughput & IOPS vs. I/O size

As seen from the above results, T6 iSCSI Offload solution delivers an exceptional line-rate throughput performance of 98 Gbps for both READ and WRITE. In addition, the solution scores a commendable IOPS performance of 2.7M. With a consistent and scalable performance, Chelsio iSCSI Offload solution provides an all-round SAN solution for exceptional I/O performance and efficiency.

The below graph plots the iSCSI target CPU Utilization. It reduces from 50% for 4 Kbytes IO size to ~6% for 512 Kbytes IO size. This highlights the value of iSCSI offload, where high IOPs and high throughput can be achieved with low CPU utilization. This in turn frees up the target node for application use.

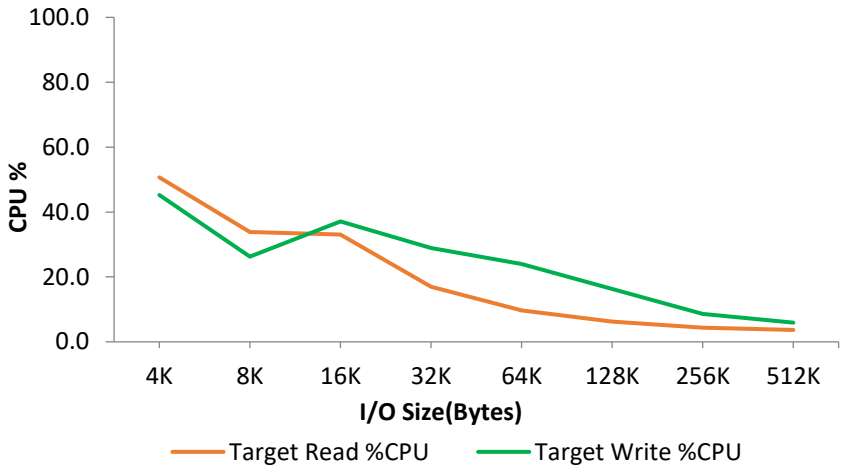

Figure 2 – iSCSI Target CPU Utilization vs. I/O size

The Demonstration

The setup consists of an LIO iSCSI target machine connected to 4 initiator machines through a 100GbE switch using single port on each system. MTU of 9000B was used. Latest Chelsio Unified Wire driver was installed on each machine.

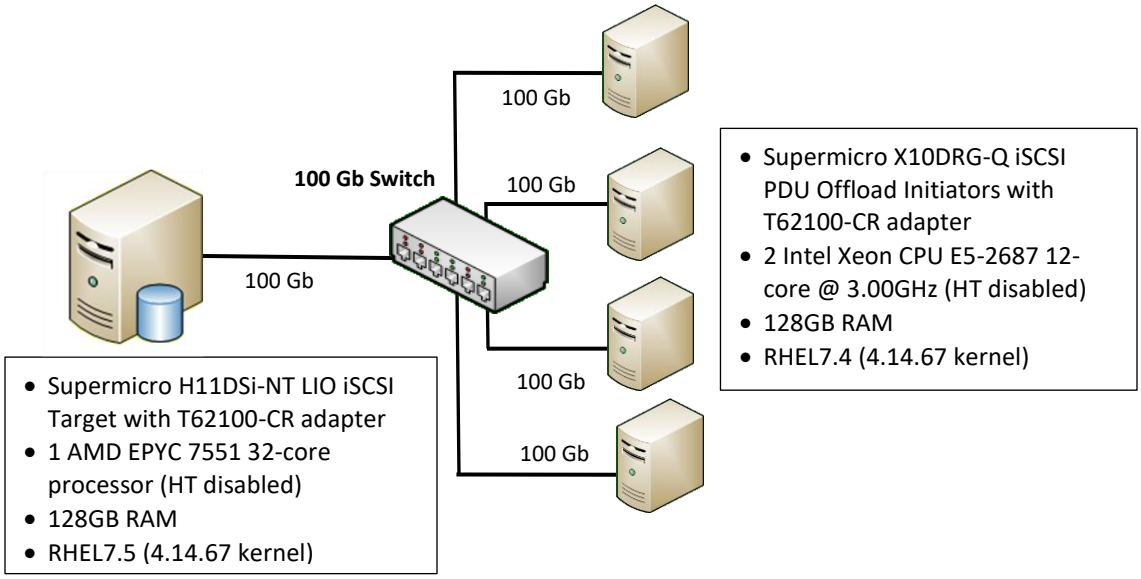

Figure 3 – Test Setup

Storage Configuration

The target was configured in offload mode with 32 Ramdisk (nullio) LUNs, each of 600MB size. Each initiator uses 8 connections.

Setup Configuration

Following performance tunings were done on AMD EPYC server:

i. Updated BIOS to latest version (v1.1b was used in this case).

ii. BIOS Settings:
SVM (Virtualization), Global C-state Control, Hyperthreading, IOMMU, SR-IOV and Core Performance Boost were *Disabled.*
Determinism Slider was set to *Performance*
Memory Interleaving was set to *Auto*

iii. All the memory channels were populated with maximum supported speed.

iv. Added *'iommu=pt cpuidle.off=1 processor.max_cstate=0'* to the kernel command line to disable c-states and rebooted the machine.

Target/Initiator Configuration

Following configuration was done on all the machines:

i. Following power saving profile was set.

```
[root@host~]# tuned-adm profile network-throughput
```

ii. Installed Chelsio Unified Wire v3.9.0.0.

```
[root@host~]# make install
```

Target Configuration

i. LIO iSCSI Offload target driver was loaded.

```
[root@host~]# modprobe cxgbit
```

ii. Chelsio interface was assigned with IPv4 address, MTU 9000 and brought-up.

```
[root@host~]# ifconfig ethX <IP address> mtu 9000 up
```

iii. CPU affinity was set.

```
[root@host~]# t4_perftune.sh -n -Q iSCSIT
```

iv. Configured 32 targets using *targetcli*.

```
[root@host~]# for i in `seq 1 32`; do targetcli /backstores/ramdisk/ create
nullio=true size=600M name=ramdisk$i ; done
[root@host~]# for i in `seq 1 32`; do targetcli /iscsi create iqn.2017-
09.org.linux-iscsi.target$i ; done
[root@host~]# for i in `seq 1 32`; do targetcli /iscsi/iqn.2017-
09.org.linux-iscsi.target$i/tpg1/ set attribute authentication=0
demo_mode_write_protect=0 generate_node_acls=1 cache_dynamic_acls=1  ; done
[root@host~]# for i in `seq 1 32`; do targetcli /iscsi/iqn.2017-
09.org.linux-iscsi.target$i/tpg1/luns create lun=0
storage_object=/backstores/ramdisk/ramdisk$i  ; done
[root@host~]# for i in `seq 1 32`; do targetcli /iscsi/iqn.2017-
09.org.linux-iscsi.target$i/tpg1/portals/ delete ip_address=0.0.0.0
ip_port=3260 ; done
[root@host~]# for i in `seq 1 32`; do targetcli /iscsi/iqn.2017-
09.org.linux-iscsi.target$i/tpg1/portals create ip_address=102.2.2.238
ip_port=3260 ; done
[root@host~]#for i in `seq 1 32`; do targetcli iscsi/iqn.2017-09.org.linux-
iscsi.target$i/tpg1/ set parameter InitialR2T=No; done
```

v. Enabled LIO Target Offload.

```
[root@host~]# for i in `seq 1 32  `; do targetcli /iscsi/iqn.2017-
09.org.linux-iscsi.target$i/tpg1/portals/10.1.1.8:3260 enable_offload
boolean=True ; done
```

Initiator Configuration

i. Added '*scsi_mod.use_blk_mq=Y processor.max_cstate=1 intel_pstate=disable intel_idle.max_cstate=0*' to the kernel command line and rebooted the machine.

ii. iSCSI Initiator driver was loaded:

```
[root@host~]# modprobe cxgb4i
```

iii. Chelsio interface was assigned with IPv4 address and brought-up.

```
[root@host~]# ifconfig ethX <IP address> mtu 9000 up
```

iv. CPU affinity was set:

```
[root@host~]# t4_perftune.sh -n -Q iSCSI
```

v. Target was discovered using *cxgb4i* iface

```
[root@host~]# iscsiadm -m discovery -t st -p 102.2.2.238:3260 -I
<cxgb4i_iface>
```

vi. Logged in to targets from the 4 initiators.

```
[root@host1~]# for i in `seq 1 8`; do iscsiadm -m node iqn.2017-
09.org.linux-iscsi.target$i -p 102.2.2.238:3260 -I <cxgb4i_iface> -l; done
[root@host2~]# for i in `seq 9 16`;do iscsiadm -m node -T iqn.2017-
09.org.linux-iscsi.target$i -p 102.2.2.238:3260 -I <cxgb4i_iface> -l ;done
[root@host3~]# for i in `seq 17 24`;do iscsiadm -m node -T iqn.2017-
09.org.linux-iscsi.target$i -p 102.2.2.238:3260 -I <cxgb4i_iface> -l ;done
[root@host4~]# for i in `seq 25 32`;do iscsiadm -m node -T iqn.2017-
09.org.linux-iscsi.target$i -p 102.2.2.238:3260 -I <cxgb4i_iface> -l ;done
```

vii. fio tool was run on all 4 initiators for Throughput and IOPS test.

```
[root@host~]# fio --rw=<randwrite/randwrite> --ioengine=libaio --name=random
--size=400m --invalidate=1 --direct=1 --runtime=30 --time_based --
fsync_on_close=1 --group_reporting --filename=/dev/sdb:…:/dev/sdi --
iodepth=32 --numjobs=8 --bs=<value>
```

Conclusion

This paper provided 100GbE iSCSI IOPS, throughput and CPU Utilization results for Chelsio's T62100-CR Unified Wire adapter in AMD EPYC 7551 based servers. Chelsio's iSCSI Offload solution delivers:

- Line-rate throughput of 98 Gbps for both READ and WRITE
- IOPS of more than 2.7M
- Minimal CPU Utilization

The entire solution, which includes Chelsio's iSCSI Offload software, the T6 adapter, and an off-the-shelf computer system including a high-end disk subsystem, provides industry leading performance with the highest IOPS and bandwidth available today. The resulting solution is highly competitive with special purpose systems and storage infrastructure currently on the market in both performance and cost.

Related Links

100G NVMe over Fabrics for AMD EPYC
100G Network Performance for AMD EPYC
100G OVS Kernel Datapath Offload for AMD EPYC
FreeBSD 100G TOE Performance for AMD EPYC
Demartek Evaluation: Chelsio Terminator 6 (T6) Unified Wire Adapter iSCSI Offload