iSCSI at 40Gbps

Line Rate Throughput and 3M+ IOPS with Terminator 5

Executive Summary

This paper presents iSCSI performance results for Chelsio's latest Terminator 5 (T5) ASIC running over a 40Gbps Ethernet interface.

Using a single iSCSI target, the results show 40Gb line rate performance using standard Ethernet frames with I/O sizes as small as 2KB, and more than 3M IOPS for 512B and 1KB I/O sizes. These results clearly show iSCSI to be an ideal fit for very high performance storage networking solutions. Furthermore, thanks to using the routable and reliable TCP/IP as a foundation, iSCSI allows highly scalable and cost effective installations using regular Ethernet switches.

Overview

The Terminator 5 (T5) ASIC from Chelsio Communications, Inc. is a fifth generation, high-performance 2x40Gbps/4x10Gbps, unified wire engine which offers PDU iSCSI offload capability in protocol acceleration for both file and block-level storage (iSCSI) traffic. Furthermore, iSCSI support is part of a complete, fully virtualized unified wire offload suite that includes FCoE, RDMA over Ethernet, TCP and UDP sockets and user space I/O.

By leveraging Chelsio's proven TCP Offload Engine (TOE), offloaded iSCSI over T5 enjoys a distinct performance advantage over regular NIC. The entire solution, which includes Chelsio's iSCSI Offload software, the T5 adapter, and an off-the-shelf computer system – including a high end disk subsystem, provides industry leading performance, with both IOPS and bandwidth the highest available today. The resulting solution is highly competitive with special purpose systems and storage infrastructure currently on the market in both performance and cost.

Unlike FC and FCoE, iSCSI runs over regular Ethernet infrastructure, without the need for specialized FC fabrics, or expensive DCB enabled switches and FibreChannel Forwarder switches. By using IP, it is also routable over the WAN and scalable beyond a local area environment. Finally, the TCP transport allows reliable operation over any link types, including naturally lossy media such as wireless.

Finally, T5 includes enhanced data integrity protection for all protocols, and particularly so for storage traffic, including full end-to-end T10-DIX support for both iSCSI and FCoE, as well as internal datapath CRC and ECC-protected memory.

Test Results

The following table summarizes the READ, WRITE and READ/WRITE throughput and IOPS numbers obtained varying the I/O sizes using the **iometer** tool. The maximum IOPS numbers reaches approximately 3M at small I/O sizes, one way performance reaches line rate at sizes close to 2KB, and bidirectional performance reaches line rate at 4KB.

IO Size (B)	T580-CR iSCSI								
	READ			**WRITE**			**READ/WRITE**		
	MB/s	IOPS	CPU (%)	MB/s	IOPS	CPU (%)	MB/s	IOPS	CPU (%)
512	1412.769	2893351	100	1118.985	2291681	100	1255.579	2571425	100
1024	2834.05	2902067	100	2225.926	2279349	100	2504.368	2564473	100
2048	4335.602	2219828	49	3424.745	1753469	60	4957.682	2538333	100
4096	4460.599	1141913	20	4082.393	1045093	29	6878.897	1760998	53
8192	4485.334	574122.7	10	4444.914	568948.9	17	8018.514	1026370	28
16384	4485.229	287054.6	7	4410.54	282274.6	13	8360.684	535083.8	20
32768	4491.347	143723.1	5	4498.942	143966.1	10	8686.52	277968.7	14

The following graph plots the performance results, showing how line rate 40Gbps is achieved even at I/O sizes as small as 2KB for unidirectional transfer and 4KB for bidirectional transfer.

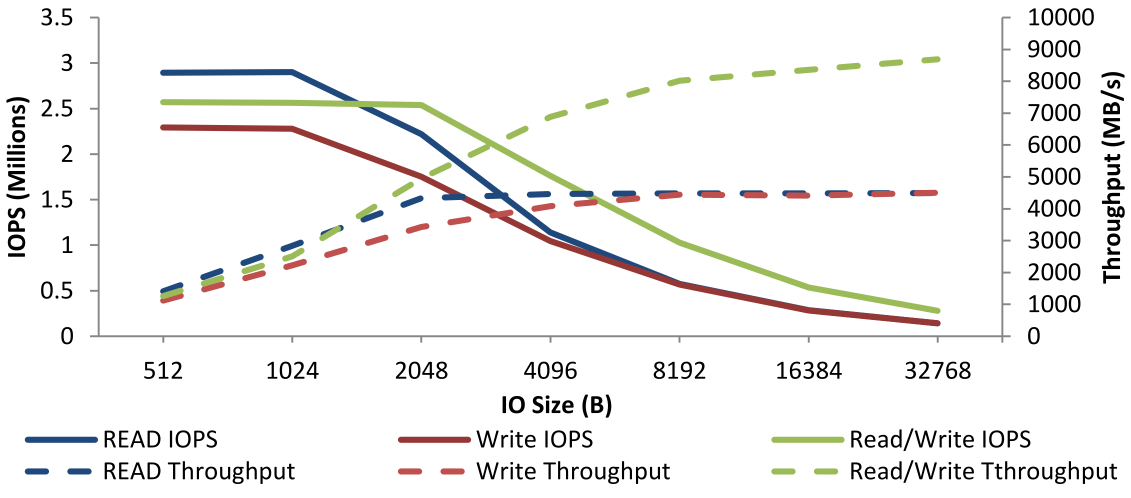


Figure 1 – Throughput and IOPS

The following graphs compares READ benchmarking data at 40Gbps and 10Gbps, showing near perfect 4x scaling of the performance. Similar results were observed with WRITE and READ/WRITE.

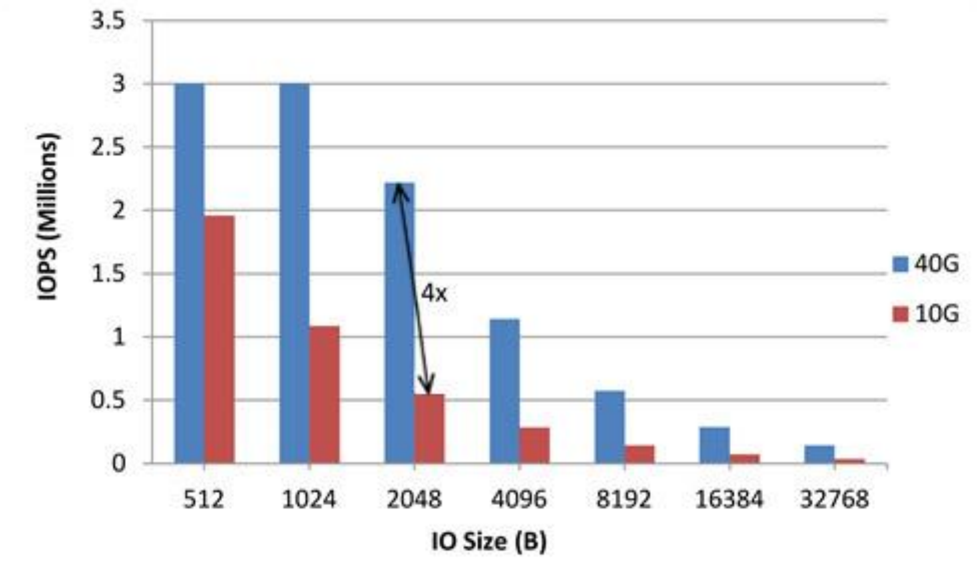


Figure 2 – READ IOPS at 40/10Gbps

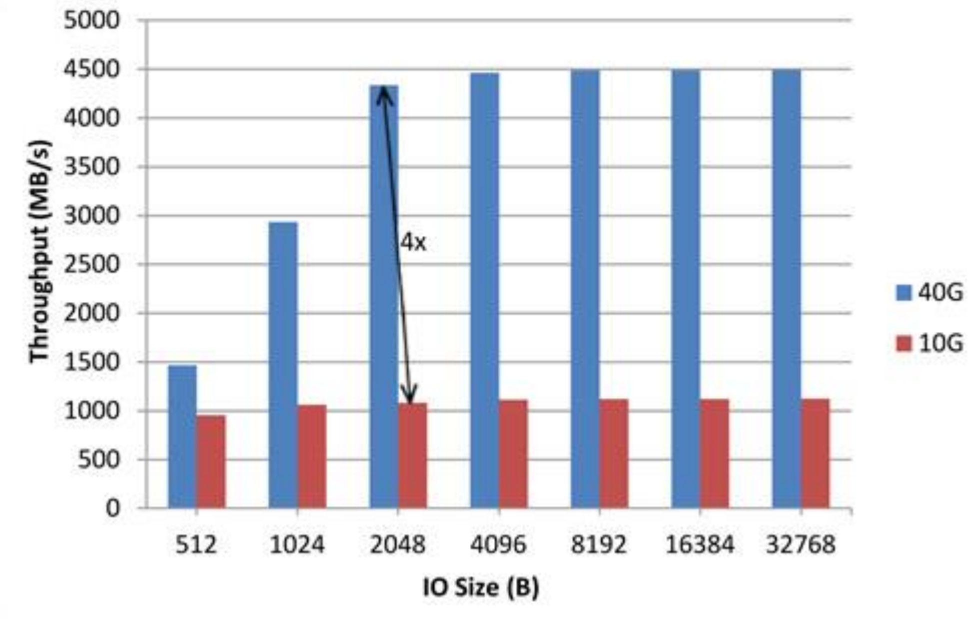
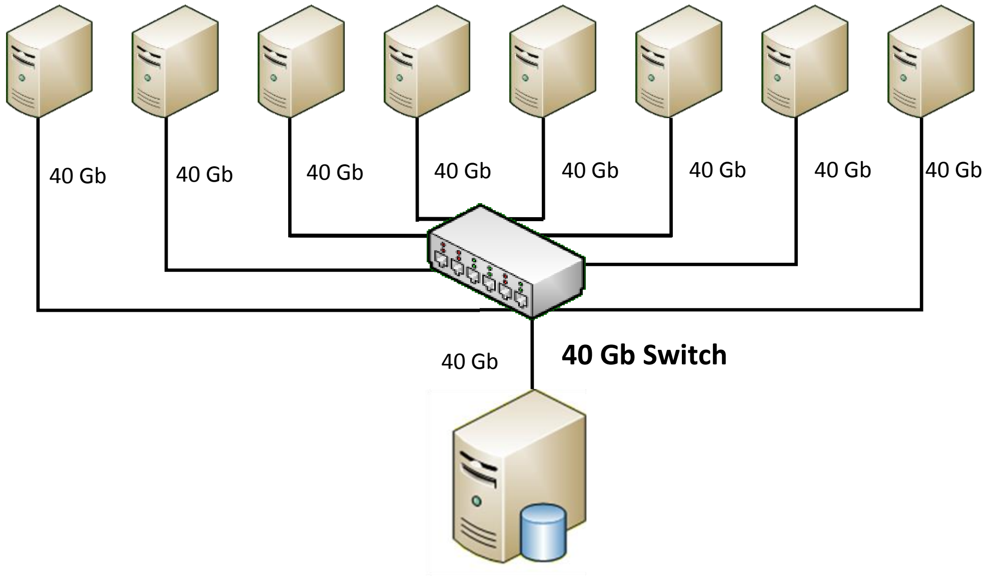


Figure 3 – READ Throughput at 40/10Gbps

Test Topology

The following diagram shows the test setup and topology.

iSCSI Initiators with T580-CR adapters running on Windows 2012 R2



iSCSI Target with T580-CR running on 3.6.11 kernel

Figure 4 – iSCSI Target connected to 8 Initiators using a 40Gb swtich

Test Configuration

The following sections provide the test configuration details.

Network Configuration

The network configuration consists of an iSCSI target storage array connected to 8 iSCSI initiator machines through a 40Gb switch. Standard MTU of 1500B was used.

- **The storage array** was configured with two Intel Xeon CPU E5-2687W v2 processors running at 3.40GHz with 64 GB of RAM. Chelsio T580-CR adapter was installed in the system with Chelsio’s iSCSI target driver v5.2.0-0734 and RHEL 6.4 (3.6.11 Kernel) operating system.

- **The initiator machines** were each setup with an Intel Xeon CPU E5-1660 v2 processor running at 3.70GHz with 64 GB of RAM each. Chelsio T580-CR adapter was installed in each system with Windows MS Initiator and Windows 2012 R2 operating system.

Storage Topology and Configuration

The storage array contains 64 iSCSI *ramdisk null-rw* targets. Each of the 8 initiators connects to 8 targets. The I/O sizes used varied from 512B to 32KB with an I/O access pattern of random READs and random WRITEs.

I/O Benchmarking Configuration

iometer is used to assess the I/O capacity of a configuration. This test used the following sample block sizes: 512B, 1KB, 2KB, 4KB, 8KB, 16KB and 32KB.

Parameters passed to Iometer

-
  ```
  dynamo.exe –l remote_iometer_iP –m localmachine ip //Add it for all
  initiators.
  ```
- 50 outstanding IO per Target.
- 8 worker threads.

Setup Configuration

The iSCSI target storage array was configured using the following procedure:

i. iSCSI Target drivers were loaded:

```
[root@host]# modprobe t4_tom cop_managed_offloading=1
[root@host]# modprobe chiscsi_t4
```

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. *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
```

vi. *t4_perftune.sh* script was run to map interrupts to different CPUs:

```
[root@host]# t4_perftune.sh
```

vii. *chiscsi_set_affinity.sh* script was run to map worker threads to different CPUs.

```
[root@host]# chiscsi_set_affinity.sh
```

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

```
global:
        iscsi_offload_mode=ULP
        iscsi_auth_order=CHAP
target:
        TargetName=iqn.2004-05.com.chelsio.target6
        TargetAlias=chiscsit6
        MaxRecvDataSegmentLength=262144
        HeaderDigest=None,CRC32C
        DataDigest=None,CRC32C
        ImmediateData=Yes
        InitialR2T=No
        MaxOutstandingR2T=1
        MaxConnections=4
        TargetSessionMaxCmd=256
        TargetDevice=ramdisk3,MEM,NULLRW,size=128MB
        PortalGroup=6@102.5.5.2:3260
```

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

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

Conclusion

This paper provided performance results for Chelsio's offloaded iSCSI solution running over Chelsio's T5 40Gbps Ethernet ASIC.

- Chelsio's T5 delivers line rate 40Gbps iSCSI SAN performance using its T580-CR Unified Wire Network adapter
- READ IOPS reaches approximately 3M for 512B IO size and reach the theoretical limit from 2KB IO Size.

Part of Chelsio's Unified Wire Ethernet solution, T5's iSCSI implementation provides the same combination of uncompromising performance and feature rich solution as the rest of the offloaded protocols.

Related Links

The Chelsio Terminator 5 ASIC
iSCSI over 40Gb Ethernet
TCP Offload at 40Gbps
High Performance iSCSI for Virtual Machines
Solaris/OpenIndiana at 40Gbps