

FreeBSD 40GbE Netmap Performance

Disruptive Performance with Terminator 5

Executive Summary

netmap is a framework for fast and safe access to network devices, that streamlines network I/O resulting in significant performance improvements compared to normal operating system stack operation. netmap requires explicit support in device drivers, and the FreeBSD kernel drivers now includes native support for Chelsio's T5 based adapters.

This report presents performance benchmark results for Chelsio's solution at 40Gbps, demonstrating record breaking performance with up to 70+ MPPS using a single 40GbE port.

Overview

The Terminator 5 (T5) ASIC from Chelsio Communications, Inc. is a fifth generation, high-performance 2x40Gbps/4x10Gbps fully virtualized server adapter engine with Unified Wire capability, allowing storage, compute and networking traffic to run simultaneously. Thanks to integrated, standards-based FCoE/iSCSI and RDMA offload, T5 adapters are high performance drop in replacements for Fibre Channel storage adapters and InfiniBand RDMA adapters. Furthermore, they also excel at normal server adapter functionality, providing high packet processing rate, high throughput and low latency for common network applications. T5 provides extensive support for stateless offload operation for both IPv4 and IPv6 (IP, TCP and UDP checksum offload, Large Send Offload, Large Receive Offload, Receive Side Steering/Load Balancing, and flexible line rate Filtering).

netmap is an efficient framework for high speed packet I/O, implemented as a kernel module for FreeBSD. **netmap** follows a simple data transfer model, supports multi-queue adapters and uses standard system calls, facilitating porting of existing applications or writing new ones.

Chelsio is a strong supporter of FreeBSD and Chelsio's in-boxed drivers are recognized as performance and quality leaders. Recent support for netmap over Chelsio devices has been integrated within the FreeBSD kernel and allows plug-and-play operation, facilitating access to the performance levels shown in this report.

Test Results

The following table summarizes the performance results in million packets/sec (MPPS):

netmap	Unidirectional		Bidirectional
	Tx	Rx	Tx+Rx
MPPS	44	40	71

The following section displays the output packets/sec numbers using the **pkt-gen** tool, with the device under test (DUT) connected to two peers.

Unidirectional Workload

Tx:

```
# ./pkt-gen -f tx -S 00:07:43:ab:cd:f0 -D 00:07:43:29:b1:d1 -i ncxl0
586.471754 main [1715] interface is ncxl0
586.471772 extract_ip_range [291] range is 10.0.0.1:0 to 10.0.0.1:0
586.471776 extract_ip_range [291] range is 10.1.0.1:0 to 10.1.0.1:0
586.492305 main [1910] mapped 334980KB at 0x801c00000
Sending on netmap:ncxl0: 4 queues, 1 threads and 1 cpus.
10.0.0.1 -> 10.1.0.1 (00:07:43:ab:cd:f0 -> 00:07:43:29:b1:d1)
586.492324 main [1994] Sending 512 packets every 0.000000000 s
586.492326 main [1996] Wait 2 secs for phy reset
588.618295 main [1998] Ready...
588.618326 nm_open [456] overriding ifname ncxl0 ringid 0x0 flags 0x1
588.618412 sender_body [1073] start, fd 4 main_fd 3
588.622685 sender_body [1142] drop copy
589.678727 main_thread [1512] 44474020 pps (47156737 pkts in 1060321 usec)
590.680227 main_thread [1512] 44545938 pps (44612757 pkts in 1001500 usec)
591.743727 main_thread [1512] 44559740 pps (47389284 pkts in 1063500 usec)
592.745225 main_thread [1512] 44564236 pps (44630993 pkts in 1001498 usec)
593.747225 main_thread [1512] 44564584 pps (44653713 pkts in 1002000 usec)
594.749227 main_thread [1512] 44552457 pps (44641651 pkts in 1002002 usec)
595.812730 main_thread [1512] 44426400 pps (47247610 pkts in 1063503 usec)
```

Rx:

```
# ./pkt-gen -f rx -S 00:07:43:ab:cd:f0 -i ncxl0
678.983451 main [1715] interface is ncxl0
678.983471 extract_ip_range [291] range is 10.0.0.1:0 to 10.0.0.1:0
678.983474 extract_ip_range [291] range is 10.1.0.1:0 to 10.1.0.1:0
679.004297 main [1910] mapped 334980KB at 0x801c00000 Receiving from netmap:ncxl0: 4
queues, 1 threads and 1 cpus.
679.004315 main [1996] Wait 2 secs for phy reset
681.130285 main [1998] Ready...
681.130315 nm_open [456] overriding ifname ncxl0 ringid 0x0 flags 0x1
681.130404 receiver_body [1247] reading from netmap:ncxl0 fd 4 main_fd 3
682.193786 receiver_body [1254] waiting for initial packets, poll returns 0 0
682.193787 main_thread [1512] 0 pps (0 pkts in 1063391 usec)
683.243225 main_thread [1512] 27205659 pps (28550680 pkts in 1049439 usec)
684.306724 main_thread [1512] 40268683 pps (42825704 pkts in 1063499 usec)
685.308224 main_thread [1512] 40273454 pps (40333864 pkts in 1001500 usec)
686.325224 main_thread [1512] 40262246 pps (40946704 pkts in 1017000 usec)
687.388722 main_thread [1512] 40267969 pps (42824904 pkts in 1063498 usec)
```

The results in either direction exceed 40MPPS, nearly saturating the PCI bus when overhead traffic is considered.

Bidirectional Workload

Tx:

```
# ./pkt-gen -i ncxl0 -f tx -S 00:07:43:ab:cd:f0 -D 00:07:43:29:b1:d1
686.723766 main [1715] interface is ncxl0
686.723782 extract_ip_range [291] range is 10.0.0.1:0 to 10.0.0.1:0
686.723786 extract_ip_range [291] range is 10.1.0.1:0 to 10.1.0.1:0
686.723830 main [1910] mapped 334980KB at 0x801c00000
Sending on netmap:ncxl0: 4 queues, 1 threads and 1 cpus.
10.0.0.1 -> 10.1.0.1 (00:07:43:ab:cd:f0 -> 00:07:43:29:b1:d1)
686.723844 main [1994] Sending 512 packets every 0.000000000 s
686.723846 main [1996] Wait 2 secs for phy reset
688.725157 main [1998] Ready...
688.725184 nm_open [456] overriding ifname ncxl0 ringid 0x0 flags 0x1
688.725267 sender_body [1073] start, fd 4 main_fd 3
688.729639 sender_body [1142] drop copy
689.726152 main_thread [1512] 43960811 pps (43999980 pkts in 1000891 usec)
693.731479 main_thread [1512] 44055343 pps (44113937 pkts in 1001330 usec)
694.733150 main_thread [1512] 44058869 pps (44132491 pkts in 1001671 usec)
695.734152 main_thread [1512] 43930514 pps (43974532 pkts in 1001002 usec)
698.737151 main_thread [1512] 42854403 pps (42897343 pkts in 1001002 usec)
699.739150 main_thread [1512] 42827901 pps (42913514 pkts in 1001999 usec)
700.740151 main_thread [1512] 42872826 pps (42915742 pkts in 1001001 usec)
701.741150 main_thread [1512] 42836265 pps (42879016 pkts in 1000998 usec)
705.759150 main_thread [1512] 42824846 pps (42910496 pkts in 1002000 usec)
```

Rx:

```
# ./pkt-gen -f rx -S 00:07:43:ab:cd:f0 -i ncxl0
681.891917 main [1715] interface is ncxl0
681.891937 extract_ip_range [291] range is 10.0.0.1:0 to 10.0.0.1:0
681.891940 extract_ip_range [291] range is 10.1.0.1:0 to 10.1.0.1:0
681.913229 main [1910] mapped 334980KB at 0x801c00000
Receiving from netmap:ncxl0: 4 queues, 1 threads and 1 cpus.
681.913247 main [1996] Wait 2 secs for phy reset
684.015404 main [1998] Ready...
684.015434 nm_open [456] overriding ifname ncxl0 ringid 0x0 flags 0x1
684.015522 receiver_body [1247] reading from netmap:ncxl0 fd 4 main_fd 3
685.043406 main_thread [1512] 0 pps (0 pkts in 1027893 usec)
685.043435 receiver_body [1254] waiting for initial packets, poll returns 0 0
686.106904 receiver_body [1254] waiting for initial packets, poll returns 0 0
686.106905 main_thread [1512] 0 pps (0 pkts in 1063499 usec)
687.108151 main_thread [1512] 6657705 pps (6666001 pkts in 1001246 usec)
688.109154 main_thread [1512] 39498077 pps (39537694 pkts in 1001003 usec)
689.111151 main_thread [1512] 36352694 pps (36425290 pkts in 1001997 usec)
690.112149 main_thread [1512] 30098385 pps (30128423 pkts in 1000998 usec)
693.117150 main_thread [1512] 30015829 pps (30075801 pkts in 1001998 usec)
696.123151 main_thread [1512] 29615715 pps (29674976 pkts in 1002001 usec)
697.124150 main_thread [1512] 29169834 pps (29198975 pkts in 1000999 usec)
701.129150 main_thread [1512] 29139157 pps (29197435 pkts in 1002000 usec)
```

Therefore, the Chelsio T580-CR adapter can process 70+ MPPS (42 TX+29 RX) while simultaneously transmitting and receiving.

Test Configuration

The following sections provide the test setup and configuration details.

Topology

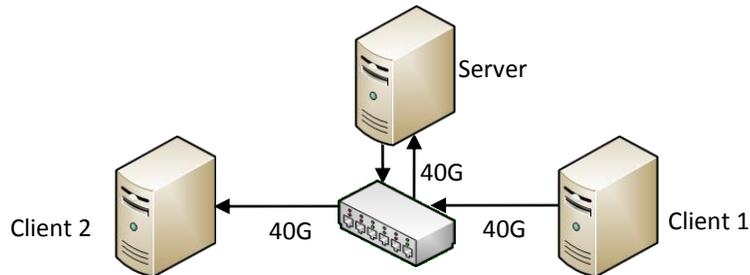


Figure 1 – Test Setup

Network Configuration

The test configuration consists of 3 machines: 1 Server and 2 Clients. The Server machine is setup with an Intel Xeon CPU E5-1660 v2 6-core processor running at 3.70GHz with 64GB of RAM. Each of the Client machines is setup with an Intel Xeon CPU E5-2687W v2 8-core processor clocked at 3.40GHz with 64GB of RAM. The Server is connected to the Client machines using a 40Gb switch. Standard MTU of 1500B is used. 1 Chelsio T580-CR adapter is installed in each system with inbox driver and FreeBSD 11.0-CURRENT operating system.

The Server transmits Netmap data to one client and concurrently receives from another client.

I/O Benchmarking Configuration

pkt-gen is used to measure network throughput.

Commands Used

```
root@host:~#cat /boot/loader.conf
hw.cxgbe.fl_pktshift=0
hw.cxgbe.toecaps_allowed=0
hw.cxgbe.nnmtxq10g=4
hw.cxgbe.nnmtxq10g=4
```

Conclusion

This report provides **netmap** performance results for Chelsio's T580-CR 2x40GbE server adapter in FreeBSD. The results show exceptional performance levels with up to 70+ MPPS on a single adapter. Such record breaking performance makes T5 ideal for high-performance applications, including wire analytics, traffic generation and monitoring applications that so far required specialized hardware.

Related Links

[The Chelsio Terminator 5 ASIC](#)

[FreeBSD NIC at 40Gbps](#)

[40Gb TOE vs NIC Performance](#)