

Mellanox NIC's Performance Report with DPDK 17.05

Rev 1.0



NOTE:

THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT ("PRODUCT(S)") AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES "AS-IS" WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCT(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES: LOSS OF USE, DATA, OR PROFITS: OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.



Mellanox Technologies 350 Oakmead Parkway Suite 100 Sunnyvale, CA 94085 U.S.A. www.mellanox.com

Tel: (408) 970-3400 Fax: (408) 970-3403

© Copyright 2017. Mellanox Technologies Ltd. All Rights Reserved.

Mellanox®, Mellanox logo, Accelio®, BridgeX®, CloudX logo, CompustorX®, Connect-IB®, ConnectX®, CoolBox®, CORE-Direct®, EZchip®, EZchip logo, EZappliance®, EZdesign®, EZdriver®, EZsystem®, GPUDirect®, InfiniHost®, InfiniBridge®, InfiniScale®, Kotura®, Kotura logo, Mellanox CloudRack®, Mellanox CloudXMellanox®, Mellanox Federal Systems®, Mellanox HostDirect®, Mellanox Multi-Host®, Mellanox Open Ethernet®, Mellanox OpenCloud®, Mellanox OpenCloud Logo®, Mellanox PeerDirect®, Mellanox ScalableHPC®, Mellanox StorageX®, Mellanox TuneX®, Mellanox Connect Accelerate Outperform logo, Mellanox Virtual Modular Switch®, MetroDX®, MetroX®, MLNX-OS®, NP-1c®, NP-2®, NP-3®, Open Ethernet logo, PhyX®, PlatformX®, PSIPHY®, SiPhy®, StoreX®, SwitchX®, Tilera®, Tilera logo, TestX®, TuneX®, The Generation of Open Ethernet logo, UFM®, Unbreakable Link®, Virtual Protocol Interconnect®, Voltaire® and Voltaire logo are registered trademarks of Mellanox Technologies, Ltd.

All other trademarks are property of their respective owners.

For the most updated list of Mellanox trademarks, visit http://www.mellanox.com/page/trademarks

Intel® and the Intel logo are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries Xeon® is a trademark of Intel Corporation or its subsidiaries in the U.S. and/or other countries. HPE® is registered trademark owned by Hewlett-Packard Development Company, L.P. HPQ Holdings, LLC IXIA® is registered trademark owned by Ixia CORPORATION CALIFORNIA



Table of Contents

Do	cume	nt Revision History	6
Αb	out th	nis Report	7
1	Test	Description	8
	1.1	General	8
	1.2	Test Procedure	8
2	Test	#1 Mellanox ConnectX-4 Lx 10GbE Throughput at Zero Packet Loss (4x 10GbE)	9
	2.1	Test Settings	10
	2.2	Test Results	10
3	Test	#2 Mellanox ConnectX-4 Lx 25GbE Throughput at Zero Packet Loss (2x 25GbE)	11
	3.1	Test Settings	12
	3.2	Test Results	12
4	Test	#3 Mellanox ConnectX-4 Lx 40GbE Throughput at Zero Packet Loss (2x 40GbE)	13
	4.1	Test Settings	14
	4.2	Test Results	14
5	Test	#4 Mellanox ConnectX-4 100GbE Throughput at Zero Packet Loss (1x 100GbE)	15
	5.1	Test Settings	16
	5.2	Test Results	16
6	Test	#5 Mellanox ConnectX-5 Ex 100GbE Throughput at Zero Packet Loss (1x 100GbE)	17
	6.1	Test Settings	18
	6.2	Test Results	18



List of Figures

Figure 1: Test #1 Setup – Mellanox ConnectX-4 Lx 10GbE connected to IXIA	9
Figure 2: Test #1 Results – Mellanox ConnectX-4 Lx 10GbE Throughput at Zero Packet Loss	10
Figure 3: Test #2 Setup – Mellanox ConnectX-4 Lx 25GbE connected to IXIA	11
Figure 4: Test #2 Results – Mellanox ConnectX-4 Lx 25GbE Throughput at Zero Packet Loss	12
Figure 5: Test #3 Setup – Mellanox ConnectX-4 Lx 40GbE connected to IXIA	13
Figure 6: Test #3 Results – Mellanox ConnectX-4 Lx 40GbE Throughput at Zero Packet Loss	14
Figure 7: Test #4 Setup – Mellanox ConnectX-4 100GbE connected to IXIA	15
Figure 8: Test #4 Results – Mellanox ConnectX-4 100GbE Throughput at Zero Packet Loss	16
Figure 9: Test #5 Setup – Mellanox ConnectX-5 Ex 100GbE connected to IXIA	17
Figure 10: Test #5 Results – Mellanox ConnectX-5 Ex 100GhF Throughput at Zero Packet Loss	18



List of Tables

Table 1: Document Revision History	6
Table 2: Test #1 Setup	9
Table 3 : Test #1 Settings	10
Table 4: Test #1 Results – Mellanox ConnectX-4 Lx 10GbE Throughput at Zero Packet Loss	10
Table 5: Test #2 Setup	11
Table 6: Test #2 Settings	12
Table 7: Test #2 Results – Mellanox ConnectX-4 Lx 25GbE Throughput at Zero Packet Loss	12
Table 8: Test #3 Setup	13
Table 9: Test #3 Settings	14
Table 10: Test #3 Results – Mellanox ConnectX-4 Lx 40GbE Throughput at Zero Packet Loss	14
Table 11: Test #4 Setup	15
Table 12: Test #4 Settings	16
Table 13: Test #4 Results – Mellanox ConnectX-4 100GbE Throughput at Zero Packet Loss	16
Table 14: Test #5 Setup	17
Table 15: Test #5 Settings	18
Table 16: Test #5 Results – Mellanox ConnectX-5 Ex 100GbE Throughput at Zero Packet Loss	18



Document Revision History

Table 1: Document Revision History

Revision	Date	Description
1.0	18-July-2017	Initial report release



About this Report

The purpose of this report is to provide Packet Rate performance data for Mellanox ConnectX®-4, ConnectX®-4 Lx and ConnectX-5 Network Interface Cards (NICs) achieved with the specified Data Plane Development Kit (DPDK) release. The report provides both the measured Packet Rate performance and the procedures and configurations to replicate the results. This document does not cover all network speeds available with the ConnectX® family of NICs and is intended as a general reference of achievable performance for the specified DPDK release.

Target Audience

This document is intended for engineers implementing applications with DPDK to guide and help achieving optimal performance.



1 Test Description

1.1 General

Setup is made up of the following components:

- HPE® ProLiant DL380 Gen9 Server
- Mellanox ConnectX® NIC
- IXIA® XM12 packet generator

Tests utilize l3fwd (https://www.dpdk.org/doc/guides/sample-app-ug/l3-forward.html) as the test application for maximum throughput with zero packet loss at various frame sizes based on RFC2544 https://tools.ietf.org/html/rfc2544.

1.2 Test Procedure

The packet generator transmits a specified frame rate towards the DUT and counts the received frame rate sent back from the DUT. Throughput is determined with the maximum achievable transmit frame rate and is equal to the received frame rate i.e. zero packet loss.

- Duration for each test is 60 seconds
- Traffic of 8192 IP flows is generated per port
- IxNetwork (Version 8.00EA) is used with the IXIA packet generator



Mellanox ConnectX-4 Lx 10GbE Throughput at Zero Packet Loss (4x 10GbE)

Table 2: Test #1 Setup

Item	Description	
Test	Test #1 – Mellanox ConnectX-4 Lx 10GbE Throughput at zero packet loss	
Server	HPE ProLiant DL380 Gen 9	
СРИ	Intel® Xeon® CPU E5-2697A v4 (Broadwell) @ 2.60GHz 16 CPU cores * 2 NUMA nodes	
RAM	256GB: 4 * 32GB DIMMs * 2 NUMA nodes @ 2400MHz	
BIOS	P89 v2.00 (12/27/2015)	
NIC	Two of MCX4121A-XCA - ConnectX-4 Lx network interface card; 10GbE dualport SFP28; PCIe3.0 x8; ROHS R6	
Operating System	Red Hat Enterprise Linux Server 7.2 (Maipo)	
Kernel Version	3.10.0-327.el7.x86_64	
GCC version	4.8.5 20150623 (Red Hat 4.8.5-4) (GCC)	
Mellanox NIC firmware version	14.18.2000	
Mellanox OFED driver version	MLNX_OFED_LINUX-4.0-2.0.0.1	
DPDK version	17.05.0	
Test Configuration	2 NICs, 2 ports used on each NIC. Each port has 1 queue assigned for a total of 4 queues. 1 queue assigned per logical core for a total of 4 logical cores for 4 ports. Each port receives a stream of 8192 IP flows from the IXIA	

Device Under Test (DUT) is made up of the HPE server and two Mellanox ConnectX®-4 Lx NICs with two 10GbE ports each (total 4 ports). The DUT is connected to the IXIA packet generator which generates traffic towards each of the ConnectX®-4 Lx NIC ports. The ConnectX®-4 Lx received data traffic is passed through DPDK to the test application I3fwd and is redirected to the opposite port on the same NIC. IXIA measures throughput with zero packet loss.

Figure 1: Test #1 Setup - Mellanox ConnectX-4 Lx 10GbE connected to IXIA

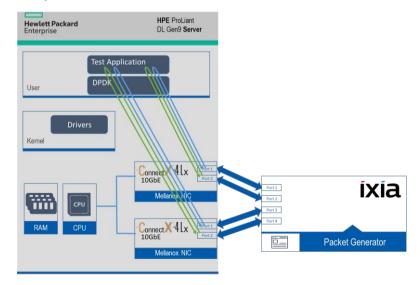




Table 3: Test #1 Settings

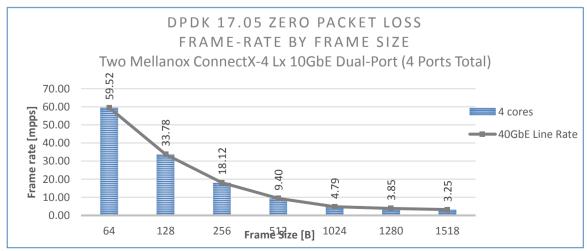
Item	Description
BIOS	Boot in "Legacy BIOS mode" Power Profile PERFORMANCE; C-states OFF; P-states OFF; TurboBoost ON; HyperThreading OFF; Virt OFF; VT-d OFF; SR-IOV OFF; SMI OFF; IONonPostedPrefetching OFF; See "Configuring and tuning HPE ProLiant Servers for low-latency applications": hpe.com > Search "DL380 low latency"
BOOT Settings	isolcpus=0-7,16-23 intel_idle.max_cstate=0 processor.max_cstate=0 intel_pstate=disable nohz_full=0-7,16-23 rcu_nocbs=0-7,16-23 rcu_novb_poll default_hugepagesz=1G hugepagesz=1G hugepages=64 audit=0 nosoftlockup
DPDK Settings	Enable mlx5 PMD before compiling DPDK: In .config file generated by "make config", set: "CONFIG_RTE_LIBRTE_MLX5_PMD=y" During testing, I3fwd was given real-time scheduling priority.
L3fwd settings	Added /I3fwd/main.c:85: #define RTE_TEST_RX_DESC_DEFAULT 4096 #define RTE_TEST_TX_DESC_DEFAULT 4096 Added /I3fwd/I3fwd.h:47: #define MAX_PKT_BURST 64
Command Line	/root/dpdk/examples/l3fwd/build/app/l3fwd -c 0xf00000 -n 4 -w 05:00.0 -w 05:00.1 -w 0b:00.0 -w 0b:00.1 socket-mem=8192,256p 0xf -Pconfig='(0,0,23),(1,0,22),(2,0,21),(3,0,20)'
Other optimizations	a) Flow Control OFF: "ethtool -A \$netdev rx off tx off" b) Memory optimizations: "sysctl -w vm.zone_reclaim_mode=0"; "sysctl -w vm.swappiness=0" c) Move all IRQs to far NUMA node: "IRQBALANCE_BANNED_CPUS=\$LOCAL_NUMA_CPUMAP irqbalance oneshot" d) Disable irqbalance: "systemctl stop irqbalance" e) Change PCI MaxReadReq to 1024B for each port of each NIC: Run "setpci -s \$PORT_PCI_ADDRESS 68.w", it will return 4 digits ABCD> Run "setpci -s \$PORT_PCI_ADDRESS 68.w=3BCD"

2.2 Test Results

Table 4: Test #1 Results - Mellanox ConnectX-4 Lx 10GbE Throughput at Zero Packet Loss

Frame Size (Bytes)	Frame Rate (Mpps)	Line Rate [40G] (Mpps)	% Line Rate
64	59.52	59.52	100
128	33.78	33.78	100
256	18.12	18.12	100
512	9.40	9.40	100
1024	4.79	4.79	100
1280	3.85	3.85	100
1518	3.25	3.25	100

Figure 2: Test #1 Results - Mellanox ConnectX-4 Lx 10GbE Throughput at Zero Packet Loss





Mellanox ConnectX-4 Lx 25GbE Throughput at Zero Packet Loss (2x 25GbE)

Table 5: Test #2 Setup

Item	Description	
Test	Test #2 – Mellanox ConnectX-4 Lx 25GbE Throughput at zero packet loss	
Server	HPE ProLiant DL380 Gen 9	
CPU	Intel(R) Xeon(R) CPU E5-2680A v3 @ 2.50GHz 12 CPU cores * 2 NUMA nodes	
RAM	128GB: 4 * 16GB DIMMs * 2 NUMA nodes @ 2133MHz	
BIOS	P89 v2.00 (12/27/2015)	
NIC	One MCX4121A-ACAT - ConnectX-4 Lx network interface card 25GbE dual-port SFP28; PCIe3.0 x8; ROHS R6	
Operating System	Red Hat Enterprise Linux Server 7.3 (Maipo)	
Kernel Version	3.10.0-327.el7.x86_64	
GCC version	4.8.5 20150623 (Red Hat 4.8.5-11) (GCC)	
Mellanox NIC firmware version	14.18.2000	
Mellanox OFED driver version	MLNX_OFED_LINUX-4.0-2.0.0.1	
DPDK version	17.05.0	
Test Configuration	1 NIC, 2 ports used on the NIC. Each port receives a stream of 8192 IP flows from the IXIA Each port has 4 queues assigned for a total of 8 queues 1 queue assigned per logical core with a total of 8 logical cores	

Device Under Test (DUT) is made up of the HPE server and the Mellanox ConnectX®-4 Lx NIC with dual-port. The DUT is connected to the IXIA packet generator which generates traffic towards the ConnectX-4 NIC

The ConnectX $^{\circ}$ -4 Lx data traffic is passed through DPDK to the test application I3fwd and is redirected to the opposite direction on the same port. IXIA measures throughput and packet loss.

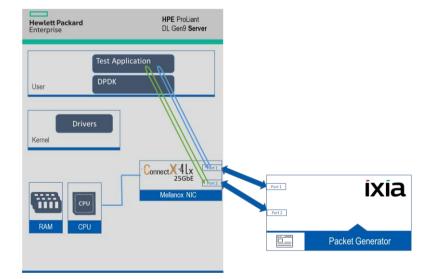


Figure 3: Test #2 Setup - Mellanox ConnectX-4 Lx 25GbE connected to IXIA



Table 6: Test #2 Settings

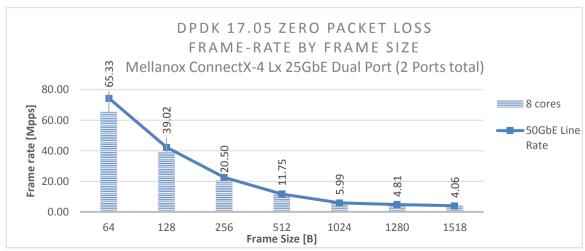
Item	Description
BIOS	Boot in "Legacy BIOS mode" Power Profile PERFORMANCE; C-states OFF; P-states OFF; TurboBoost ON; HyperThreading OFF; Virt OFF; VT-d OFF; SR-IOV OFF; SMI OFF IONonPostedPrefetching OFF; See "Configuring and tuning HPE ProLiant Servers for low-latency applications": hpe.com > Search "DL380 low latency"
BOOT Settings	isolcpus=0-7,16-23 intel_idle.max_cstate=0 processor.max_cstate=0 intel_pstate=disable nohz_full=0-7,16-23 rcu_nocbs=0-7,16-23 rcu_novb_poll default_hugepagesz=1G hugepagesz=1G hugepages=64 audit=0 nosoftlockup
DPDK Settings	Enable mlx5 PMD before compiling DPDK: In .config file generated by "make config", set: "CONFIG_RTE_LIBRTE_MLX5_PMD=y" During testing, l3fwd was given real-time scheduling priority.
L3fwd settings	Added /l3fwd/main.c:85: #define RTE_TEST_RX_DESC_DEFAULT 4096 #define RTE_TEST_TX_DESC_DEFAULT 4096 Added /l3fwd/l3fwd.h:47: #define MAX_PKT_BURST 64
Command Line	/root/dpdk/examples/l3fwd/build/app/l3fwd -c 0x3f030 -n 4 -w 05:00.0,txq_inline=200 -w 05:00.1,txq_inline=200socket-mem=8192p 0x3 -P config='(0,0,17),(0,1,16),(0,2,15),(0,3,14),(0,4,13),(0,5,12),(0,6,5),(0,7,4)'
Other optimizations	a) Flow Control OFF: "ethtool -A \$netdev rx off tx off" b) Memory optimizations: "sysctl -w vm.zone_reclaim_mode=0"; "sysctl -w vm.swappiness=0" c) Move all IRQs to far NUMA node: "IRQBALANCE_BANNED_CPUS=\$LOCAL_NUMA_CPUMAP irqbalance oneshot" d) Disable irqbalance: "systemctl stop irqbalance" e) Change PCI MaxReadReq to 1024B for each port of each NIC: Run "setpci -s \$PORT_PCI_ADDRESS 68.w", it will return 4 digits ABCD> Run "setpci -s \$PORT_PCI_ADDRESS 68.w=3BCD"

3.2 Test Results

Table 7: Test #2 Results - Mellanox ConnectX-4 Lx 25GbE Throughput at Zero Packet Loss

Frame Size (Bytes)	Frame Rate (Mpps)	Line Rate [50G] (Mpps)	% Line Rate
64	65.33	74.4	87.81
128	39.02	42.23	92.40
256	20.50	22.64	90.55
512	11.75	11.75	100.00
1024	5.99	5.99	100.00
1280	4.81	4.81	100.00
1518	4.06	4.06	100.00

Figure 4: Test #2 Results - Mellanox ConnectX-4 Lx 25GbE Throughput at Zero Packet Loss



12



Mellanox ConnectX-4 Lx 40GbE Throughput at Zero Packet Loss (2x 40GbE)

Table 8: Test #3 Setup

Item	Description
Test	Test #3 – Mellanox ConnectX-4 Lx 40GbE Throughput at zero packet loss
Server	HPE ProLiant DL380 Gen 9
CPU	Intel(R) Xeon(R) CPU E5-2697A v4 (Broadwell) @ 2.60GHz 16 CPU cores * 2 NUMA nodes
RAM	256GB: 4 * 32GB DIMMs * 2 NUMA nodes @ 2400MHz
BIOS	P89 v2.00 (12/27/2015)
NIC	Two of MCX4131A-BCA - ConnectX-4 Lx network interface card; 40GbE single-port QSFP28; PCIe3.0 x8; ROHS R6
Operating System	Red Hat Enterprise Linux Server 7.2 (Maipo)
Kernel Version	3.10.0-327.el7.x86_64
GCC version	4.8.5 20150623 (Red Hat 4.8.5-4) (GCC)
Mellanox NIC firmware version	14.18.2000
Mellanox OFED driver version	MLNX_OFED_LINUX-4.0-2.0.0.1
DPDK version	17.05.0
Test Configuration	 2 NICs, 1 port used on each NIC. Each port receives a stream of 8192 IP flows from the IXIA Subtest #1: Each port has 2 queues assigned for a total of 4 queues 1 queue assigned per logical core with a total of 4 logical cores and 4 queues for 2 ports Subtest #2: Each port has 4 queues assigned for a total of 8 queues 1 queue assigned per logical core with a total of 8 logical cores and 8 queues

Device Under Test (DUT) is made up of the HPE server and the two Mellanox ConnectX®-4 Lx NICs with one 40GbE port each (total of 2 ports). The DUT is connected to the IXIA packet generator which generates traffic towards each of the ConnectX-4 Lx NICs.

The ConnectX®-4 Lx data traffic is passed through DPDK to the test application I3fwd and is redirected to the opposite card's port. IXIA measures throughput and packet loss.

Hewlett Packard Enterprise HPE ProLiant DL Gen9 Server User ixia Connect X-4Lx Packet Generator

Figure 5: Test #3 Setup - Mellanox ConnectX-4 Lx 40GbE connected to IXIA



Table 9: Test #3 Settings

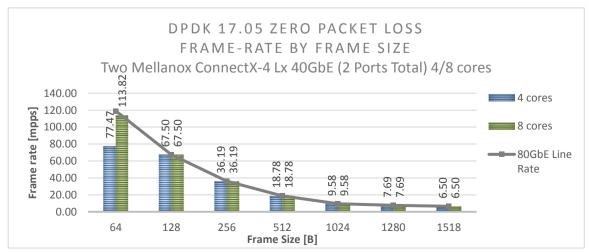
Item	Description
BIOS	Boot in "Legacy BIOS mode" Power Profile PERFORMANCE; C-states OFF; P-states OFF; TurboBoost ON; HyperThreading OFF; Virt OFF; VT-d OFF; SR-IOV OFF; SMI OFF IONonPostedPrefetching OFF; See "Configuring and tuning HPE ProLiant Servers for low-latency applications": hpe.com > Search "DL380 low latency"
BOOT Settings	isolcpus=0-7,16-23 intel_idle.max_cstate=0 processor.max_cstate=0 intel_pstate=disable nohz_full=0-7,16-23 rcu_nocbs=0-7,16-23 rcu_novb_poll default_hugepagesz=1G hugepagesz=1G hugepages=64 audit=0 nosoftlockup
DPDK Settings	Enable mlx5 PMD before compiling DPDK: In .config file generated by "make config", set: "CONFIG_RTE_LIBRTE_MLX5_PMD=y" During testing, l3fwd was given real-time scheduling priority.
L3fwd settings	Added /l3fwd/main.c:85: #define RTE_TEST_RX_DESC_DEFAULT 4096 #define RTE_TEST_TX_DESC_DEFAULT 4096 Added /l3fwd/l3fwd.h:47: #define MAX_PKT_BURST 64
Command Line	/root/dpdk/examples/l3fwd/build/app/l3fwd -c 0xf00 -n 4 -w 84:00.0 -w 81:00.0socket-mem=0,8192p 0x3 -Pconfig='(0,0,8),(0,1,9),(1,0,10),(1,1,11)'
Other optimizations	a) Flow Control OFF: "ethtool -A \$netdev rx off tx off" b) Memory optimizations: "sysctl -w vm.zone_reclaim_mode=0"; "sysctl -w vm.swappiness=0" c) Move all IRQs to far NUMA node: "IRQBALANCE_BANNED_CPUS=\$LOCAL_NUMA_CPUMAP irqbalance oneshot" d) Disable irqbalance: "systemctl stop irqbalance" e) Change PCI MaxReadReq to 1024B for each port of each NIC: Run "setpci -s \$PORT_PCI_ADDRESS 68.w", it will return 4 digits ABCD> Run "setpci -s \$PORT_PCI_ADDRESS 68.w=3BCD"

4.2 Test Results

Table 10: Test #3 Results - Mellanox ConnectX-4 Lx 40GbE Throughput at Zero Packet Loss

Frame Size (Bytes)	4 Queues per port (total 8 cores)		2 Queues per port (total 4 cores)		Line Date [000]
	Frame Rate 8 Queues (Mpps)	% Line Rate 8 Queues	Frame Rate 4 Queues (Mpps)	% Line Rate 4 Queues	Line Rate [80G] (Mpps)
64	113.82	95.60	77.47	65.07	119.05
128	67.50	99.90	67.50	99.90	67.57
256	36.19	99.89	36.19	99.89	36.23
512	18.78	99.89	18.78	99.89	18.80
1024	9.58	100	9.58	100	9.58
1280	7.69	100	7.69	100	7.69
1518	6.50	100	6.50	100	6.50

Figure 6: Test #3 Results - Mellanox ConnectX-4 Lx 40GbE Throughput at Zero Packet Loss





Mellanox ConnectX-4 100GbE Throughput at Zero Packet Loss (1x 100GbE)

Table 11: Test #4 Setup

Item	Description	
Test	Test #4 – Mellanox ConnectX-4 100GbE Throughput at zero packet loss	
Server	HPE ProLiant DL380 Gen 9	
CPU	Intel(R) Xeon(R) CPU E5-2697A v4 (Broadwell) @ 2.60GHz 16 CPU cores * 2 NUMA nodes	
RAM	256GB: 4 * 32GB DIMMs * 2 NUMA nodes @ 2400MHz	
BIOS	P89 v2.00 (12/27/2015)	
NIC	One MCX415A-CCAT- ConnectX-4 network interface card 100GbE single-port QSFP28; PCIe3.0 x16; ROHS R6	
Operating System	Red Hat Enterprise Linux Server 7.2 (Maipo)	
Kernel Version	3.10.0-327.el7.x86_64	
GCC version	4.8.5 20150623 (Red Hat 4.8.5-4) (GCC)	
Mellanox NIC firmware version	12.18.2000	
Mellanox OFED driver version	MLNX_OFED_LINUX-4.0-2.0.0.1	
DPDK version	17.05.0	
Test Configuration 1 NIC, 1 port used on NIC, The port has 8 queues assigned to it, logical core for a total of 8 logical cores. Each port receives a stream of 8192 IP flows from the IXIA		

Device Under Test (DUT) is made up of the HPE server and the Mellanox ConnectX®-4 NIC with a single port. The DUT is connected to the IXIA packet generator which generates traffic towards the ConnectX®-4 NIC.

The ConnectX-4 data traffic is passed through DPDK to the test application l3fwd and is redirected to the opposite direction on the same port. IXIA measures throughput and packet loss.

Hewlett Packard
Enterprise

Test Application
User

Drivers

Kernel

Connect 100GbE
Mellanox NIC

Packet Generator

Figure 7: Test #4 Setup - Mellanox ConnectX-4 100GbE connected to IXIA



Table 12: Test #4 Settings

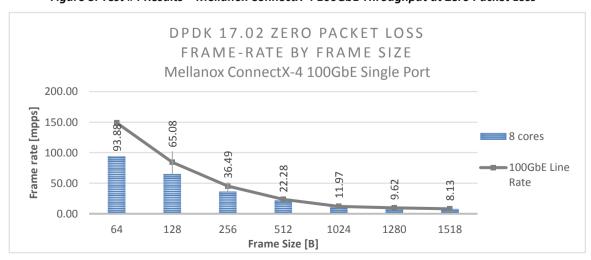
Item	Description
BIOS	Boot in "Legacy BIOS mode" Power Profile PERFORMANCE; C-states OFF; P-states OFF; TurboBoost ON; HyperThreading OFF; Virt OFF; VT-d OFF; SR-IOV OFF; SMI OFF IONonPostedPrefetching OFF; See "Configuring and tuning HPE ProLiant Servers for low-latency applications": hpe.com > Search "DL380 low latency"
BOOT Settings	isolcpus=0-7,16-23 intel_idle.max_cstate=0 processor.max_cstate=0 intel_pstate=disable nohz_full=0-7,16-23 rcu_nocbs=0-7,16-23 rcu_novb_poll default_hugepagesz=1G hugepagesz=1G hugepages=64 audit=0 nosoftlockup
DPDK Settings	Enable mlx5 PMD before compiling DPDK: In .config file generated by "make config", set: "CONFIG_RTE_LIBRTE_MLX5_PMD=y" During testing, l3fwd was given real-time scheduling priority.
L3fwd settings	Added /l3fwd/main.c:85: #define RTE_TEST_RX_DESC_DEFAULT 2048 #define RTE_TEST_TX_DESC_DEFAULT 2048 Added /l3fwd/l3fwd.h:47: #define MAX_PKT_BURST 64
Command Line	/root/dpdk/examples/l3fwd/build/app/l3fwd -c 0xff000000 -n 4 -w 88:00.0,txq_inline=128socket-mem=0,8192p 0x1 -Pconfig='(0,0,31),(0,1,30),(0,2,29),(0,3,28),(0,4,27),(0,5,26),(0,6,25),(0,7,24)'
Other optimizations	a) Flow Control OFF: "ethtool -A \$netdev rx off tx off" b) Memory optimizations: "sysctl -w vm.zone_reclaim_mode=0"; "sysctl -w vm.swappiness=0" c) Move all IRQs to far NUMA node: "IRQBALANCE_BANNED_CPUS=\$LOCAL_NUMA_CPUMAP irqbalance oneshot" d) Disable irqbalance: "systemctl stop irqbalance" e) Change PCI MaxReadReq to 1024B for each port of each NIC: Run "setpci -s \$PORT_PCI_ADDRESS 68.w", it will return 4 digits ABCD> Run "setpci -s \$PORT_PCI_ADDRESS 68.w=3BCD"

5.2 Test Results

Table 13: Test #4 Results - Mellanox ConnectX-4 100GbE Throughput at Zero Packet Loss

Frame Size (Bytes)	Frame Rate (Mpps)	Line Rate [100G] (Mpps)	% Line Rate
64	93.88	148.81	63.09
128	65.08	84.46	77.05
256	36.49	45.29	80.57
512	22.28	23.5	94.81
1024	11.97	11.97	100
1280	9.62	9.61	100
1518	8.13	8.13	100

Figure 8: Test #4 Results - Mellanox ConnectX-4 100GbE Throughput at Zero Packet Loss





Mellanox ConnectX-5 Ex 100GbE Throughput at Zero Packet Loss (1x 100GbE)

Table 14: Test #5 Setup

Item	Description	
Test	Test #5 – Mellanox ConnectX-5 Ex 100GbE Throughput at zero packet loss	
Server	HPE ProLiant DL380 Gen 9	
CPU	Intel(R) Xeon(R) CPU E5-2697A v4 (Broadwell) @ 2.60GHz 16 CPU cores * 2 NUMA nodes	
RAM	256GB: 4 * 32GB DIMMs * 2 NUMA nodes @ 2400MHz	
BIOS	P89 v2.00 (12/27/2015)	
NIC	One MCX516A-CDAT- ConnectX-5 Ex network interface card 100GbE dual-port QSFP28; PCIe3.0/PCIe4 x16; ROHS R6	
Operating System	Red Hat Enterprise Linux Server 7.3 (Maipo)	
Kernel Version	3.10.0-514.el7.x86_64	
GCC version	4.8.5 20150623 (Red Hat 4.8.5-11) (GCC)	
Mellanox NIC firmware version	16.19.1200	
Mellanox OFED driver version	MLNX_OFED_LINUX-4.0-2.0.0.1	
DPDK version	17.05.0	
Test Configuration 1 NIC, 1 port used on NIC, The port has 16 queues assigned to it, 1 clogical core for a total of 16 logical cores. Each port receives a stream of 8192 IP flows from the IXIA		

Device Under Test (DUT) is made up of the HPE server and the Mellanox ConnectX®-5 Ex NIC with a dual-port (only first port used in this test). The DUT is connected to the IXIA packet generator which generates traffic towards the ConnectX®-5 Ex NIC.

The ConnectX®-5 Ex data traffic is passed through DPDK to the test application l3fwd and is redirected to the opposite direction on the same port. IXIA measures throughput and packet loss.

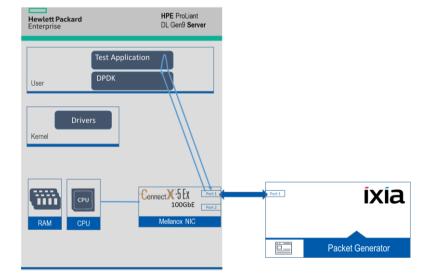


Figure 9: Test #5 Setup - Mellanox ConnectX-5 Ex 100GbE connected to IXIA



Table 15: Test #5 Settings

Item	Description
BIOS	Boot in "Legacy BIOS mode" Power Profile PERFORMANCE; C-states OFF; P-states OFF; TurboBoost ON; HyperThreading OFF; Virt OFF; VT-d OFF; SR-IOV OFF; SMI OFF IONonPostedPrefetching OFF; See "Configuring and tuning HPE ProLiant Servers for low-latency applications": hpe.com > Search "DL380 low latency"
BOOT Settings	isolcpus=0-7,16-23 intel_idle.max_cstate=0 processor.max_cstate=0 intel_pstate=disable nohz_full=0-7,16-23 rcu_nocbs=0-7,16-23 rcu_novb_poll default_hugepagesz=1G hugepagesz=1G hugepages=64 audit=0 nosoftlockup
DPDK Settings	Enable mlx5 PMD before compiling DPDK: In .config file generated by "make config", set: "CONFIG_RTE_LIBRTE_MLX5_PMD=y" During testing, I3fwd was given real-time scheduling priority.
L3fwd settings	Added /l3fwd/main.c:85: #define RTE_TEST_RX_DESC_DEFAULT 2048 #define RTE_TEST_TX_DESC_DEFAULT 2048 Added /l3fwd/l3fwd.h:47: #define MAX_PKT_BURST 64
Command Line	/l3fwd/build/app/l3fwd -c 0xffff0000 -n 4 -w 82:00.0socket-mem=0,16384p 0x1 -P config='(0,0,31),(0,1,30),(0,2,29),(0,3,28),(0,4,27),(0,5,26),(0,6,25),(0,7,24),(0,8,23),(0,9,22),(0,10,21),(0,11,20) ,(0,12,19),(0,13,18),(0,14,17),(0,15,16)'
Other optimizations	a) Flow Control OFF: "ethtool -A \$netdev rx off tx off" b) Memory optimizations: "sysctl -w vm.zone_reclaim_mode=0"; "sysctl -w vm.swappiness=0" c) Move all IRQs to far NUMA node: "IRQBALANCE_BANNED_CPUS=\$LOCAL_NUMA_CPUMAP irqbalance oneshot" d) Disable irqbalance: "systemctl stop irqbalance" e) Change PCI MaxReadReq to 1024B for each port of each NIC: Run "setpci -s \$PORT_PCI_ADDRESS 68.w", it will return 4 digits ABCD> Run "setpci -s \$PORT_PCI_ADDRESS 68.w=3BCD"

6.2 Test Results

Table 16: Test #5 Results - Mellanox ConnectX-5 Ex 100GbE Throughput at Zero Packet Loss

Frame Size (Bytes)	Frame Rate (Mpps)	Line Rate [100G] (Mpps)	% Line Rate
64	136.46	148.81	91.70
128	84.46	84.46	100
256	45.29	45.29	100
512	23.50	23.50	100
1024	11.97	11.97	100
1280	9.62	9.61	100
1518	8.13	8.13	100

Figure 10: Test #5 Results - Mellanox ConnectX-5 Ex 100GbE Throughput at Zero Packet Loss

