



Mellanox DPDK

Release Notes

Rev 16.11_3.0

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 2016. Mellanox Technologies LTD. All Rights Reserved.

Mellanox®, Mellanox logo, BridgeX®, CloudX logo, Connect-IB®, ConnectX®, CoolBox®, CORE-Direct®, EZchip®, EZchip logo, EZappliance®, EZdesign®, EZdriver®, EZsystem®, GPUDirect®, InfiniHost®, InfiniScale®, Kotura®, Kotura logo, Mellanox Federal Systems®, Mellanox Open Ethernet®, Mellanox ScalableHPC®, Mellanox Connect Accelerate Outperform logo, Mellanox Virtual Modular Switch®, MetroDX®, MetroX®, MLNX-OS®, NP-1c®, NP-2®, NP-3®, Open Ethernet logo, PhyX®, SwitchX®, Tiler®, Tiler logo, TestX®, The Generation of Open Ethernet logo, UFM®, 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>

Contents

Document Revision History	5
1 Overview	6
1.1 System Requirements	6
2 Changes and Major New Features in Rev 16.11_3.0	7
3 mlx4 and mlx5 PMD Drivers Features	8
4 Known Issues	10
5 Bug Fixes	15
6 Changes and New Features History	18

List of Tables

Table 1: Document Revision History	5
Table 2: System Requirements.....	6
Table 3: Changes and Major New Features.....	7
Table 4: Known Issues	10
Table 5: Bug Fixes	15
Table 6: Changes and New Features History.....	18

Document Revision History

Table 1: Document Revision History

Release	Date	Description
Rev 16.11_3.0	June 30, 2017	Initial release of this DPDK version.

1 Overview

These are the release notes for mlx4 and mlx5 DPDK Poll-Mode Driver (PMD) for Mellanox ConnectX®-3 Pro and ConnectX®-4/ConnectX®-4 Lx/ConnectX®-5/ConnectX®-5 Ex Ethernet adapters.

1.1 System Requirements

Table 2: System Requirements

Specification	Value
Network Adapter Cards	ConnectX®-3 Pro / ConnectX®-4 / ConnectX®-4 Lx / ConnectX®-5 / ConnectX®-5 Ex network adapter card. (These must be configured to work in ETH mode.)
Firmware	<ul style="list-style-type: none"> • ConnectX-5/ConnectX-5 Ex: v16.20.1010 • ConnectX-4: v12.20.1010 • ConnectX-4 Lx: v14.20.1010 • ConnectX-3 Pro: v2.40.7000
Linux Driver Stack	MLNX_OFED_LINUX-4.1-1.x.x.x
Tested ESXi Driver	ESXi 6.0 for ConnectX-4 / ConnectX-4 Lx: v4.16.9.2
Tested Operating Systems and Kernels	<ul style="list-style-type: none"> • Debian8.3 • RHEL/CentOS7.2 • RHEL/CentOS7.3 • SLES12SP2 • Ubuntu 16.04 • Ubuntu 16.10 • Ubuntu 14.04 <p>Note: For additional OSes, please contact Mellanox Support (support@mellanox.com).</p>
Minimum memory requirements	16 GB RAM
Transport	Ethernet
CPU Arch	x86, PPC (Power8)

2 Changes and Major New Features in Rev 16.11_3.0

Table 3: Changes and Major New Features

Driver	Changes
ConnectX-4/5 PMD, mlx5	<ul style="list-style-type: none">• Added Delay Drop Rx queue support• Added HW TSO support for VXLAN packets without the inner Ethernet header. Supported only for PFs.• Added HW checksum offload for VXLAN packets without the inner Ethernet header. Supported only for PFs.• Performance Optimizations.• Bug fixes

3 mlx4 and mlx5 PMD Drivers Features

Feature	mlx4 PMD	mlx5 PMD
Supported NICs	ConnectX®-3 Pro	ConnectX®-4 ConnectX®-4 Lx ConnectX®-5 ConnectX®-5 Ex
PCI mapping	Function per device	Function per port
KVM SR-IOV	Yes	Yes
ESX 6.0 SR-IOV	Yes	Yes
Scattering/gathering RX/TX packets	Yes	Yes
Multiple RX (with RSS/RCA) and TX queues	Yes	Yes
IPv4, TCP IPv4, UDP IPv4 RSS	Yes	Yes
IPv6 RSS	Yes	Yes
VXLAN RSS	Yes	According to the outer header
Number of RSS queues	Power of 2	Any
Get and Set RSS key per flow type (rss_hf)	No	Yes
Multiple MAC addresses	Yes	Yes
VLAN filtering	Yes	Yes
Link state information	Yes	Yes
Software counters/statistics	Yes	Yes
Start/stop/close operations	Yes	Yes
Multiple physical ports host adapter	Yes	Yes
Promiscuous mode	Yes	Yes
Multicast Promiscuous	Yes	Yes
Checksums hardware offloading	Yes	Yes
Checksum VXLAN hardware offloading	Yes	Yes
Flow Director	No	Yes
RX VLAN stripping	No	Yes
TX VLAN insertion	No	Yes
Port extended statistics	No	Yes

TSO	No	Yes
VXLAN TSO	No	Yes
VXLAN TSO with packets without Ethernet header	No	Yes
Delay Drop Rx queue	No	Yes

4 Known Issues

The following are DPDK Known Issues.

Table 4: Known Issues

Internal Ref.	Issues
General Known Issues	
-	<p>Description: Mellanox PMDs does not support InfiniBand.</p> <p>Workaround: N/A</p> <p>Keywords: InfiniBand</p>
-	<p>Description: Mellanox PMDs does not support DPDK integrated shared library.</p> <p>Workaround: Use the PMD compiled as a shared library.</p> <p>Keywords: DPDK integrated shared library</p>
-	<p>Description: Hardware queue counters are not implemented.</p> <p>Workaround: N/A</p> <p>Keywords: Hardware queue counters</p>
-	<p>Description: The Primary and Secondary multi process model is currently supported in TX only.</p> <p>Workaround: Use multithreaded model instead of multi process in case a mutli process RX is needed</p> <p>Keywords: The Primary and Secondary multi process model</p>
-	<p>Description: Bond PMD does not configure the needed MAC address to Bond slaves in case of failover. Active-Passive Bonding mode is available in case of KVM SR-IOV when 2 VMs are configured with the same MAC.</p> <p>Workaround: N/A</p> <p>Keywords: Active - Passive Bonding mode</p>
-	<p>Description: When the packet does not include TCP/UDP header and the hardware checksum offloading is enabled, L4 checksum is reported as bad.</p> <p>Workaround: N/A</p> <p>Keywords: L4 checksum report with hardware checksum offloading</p>
-	<p>Description: Promiscuous mode does not work when SR-IOV is enabled.</p> <p>Workaround: N/A</p> <p>Keywords: Promiscuous mode and SR-IOV</p>
-	<p>Description: In testpmd: Although promiscuous mode fails in SR-IOV, it is still shown as enabled This is DPDK implementation bug.</p> <p>Workaround: N/A</p> <p>Keywords: Promiscuous mode and SR-IOV</p>
-	<p>Description: Using OVS 2.5.0 requires several additional patches for it to function properly</p>

Internal Ref.	Issues
	<p>Workaround: Use OVS 2.6.0 branch instead. For further information, refer to the Mellanox Community.</p> <p>Keywords: OVS support</p>
-	<p>Description: Working with multiple containers where separate network namespaces are created is not supported</p> <p>Workaround: N/A</p> <p>Keywords: Containers</p>
-	<p>Description: Enabling Auto-inline feature might cause host to crash,</p> <p>Workaround: N/A</p> <p>Keywords: Auto-inline, inline</p>
mlx4 PMD Known Issues	
-	<p>Description: When a Multicast or Broadcast packet is sent to the SR-IOV VF, it is returned back to the port.</p> <p>Workaround: Disable mc_loopback using <code>ethtool --set-priv-flags <netdevice> disable_mc_loopback on</code>.</p> <p>Note: Disabling MC loopback will disable steering between local VFs.</p> <p>Keywords: Multicast / Broadcast Self-loopback on SR-IOV VF</p>
-	<p>Description: Performance degradation might occur for small packets when PMD is compiled with SGE_NUM = 4 compared to the performance when SGE_NUM = 1.</p> <p>Workaround: If scattered packets are not used, compile PMD with SGE_NUM = 1</p> <p>Keywords: Performance degradation</p>
-	<p>Description: Sending abnormal packets, smaller than 16B and larger than the configured MTU causes PMD queues to enter an error state.</p> <p>Workaround: Run <code>dev_stop</code> and <code>dev_start</code></p> <p>Keywords: Abnormal packets' size</p>
-	<p>Description: RSS hash key and options cannot be modified.</p> <p>Workaround: N/A</p> <p>Keywords: RSS</p>
-	<p>Description: VLAN filtering is supported only with non-optimized steering mode.</p> <p>Workaround: See QSG, RX VLAN Filter section</p> <p>Keywords: VLAN filtering</p>
-	<p>Description: Number of configured RSS queues must be power of 2.</p> <p>Workaround: Use only power of 2 RSS queues</p> <p>Keywords: Number of configured RSS queues must be power of 2</p>
-	<p>Description: Broadcast packets with VLAN are not received when VLAN filter is configured on KVM VM</p> <p>Workaround: N/A</p> <p>Keywords: Broadcast packets and VLAN filter on KVM VM</p>
-	<p>Description: VXLAN traffic is not being spread among the RSS queues</p>

Internal Ref.	Issues
	<p>Workaround: N/A</p> <p>Keywords: RSS and VXLAN traffic</p>
mlx5 PMD Known Issue	
-	<p>Description: When a Multicast or Broadcast packet is sent to the SR-IOV VF, it is returned back to the port.</p> <p>Workaround: N/A</p> <p>Keywords: Multicast / Broadcast Self-loopback on SR-IOV VF</p>
-	<p>Description: Enabling auto-inline feature might cause host to crush.</p> <p>Workaround: N/A</p> <p>Keywords: Auto-inline, inline</p>
-	<p>Description: System hang during heavy traffic and stopping DPDK process</p> <p>Workaround: Disable ethernet driver steering rules with ethtool.</p> <p>example: <code>ethtool -U eth2 flow-type ether dst E4:1D:2D:EA:CC:30 action -1 loc 0</code></p> <p>Keywords: System hang during heavy traffic</p>
-	<p>Description: Broadcast packets are not being received on the KVM VM.</p> <p>Workaround: N/A</p> <p>Keywords: Broadcast packets on KVM VM</p>
-	<p>Description: When using default firmware configuration, LCS events are not received when <code>rte_eth_dev_set_link_down/up</code> are called.</p> <p>Workaround:</p> <ol style="list-style-type: none"> Run: <code>mlxconfig -d /dev/mst/mt4115_pciconf0 set KEEP_ETH_LINK_UP_P1=0 KEEP_ETH_LINK_UP_P2=0</code> Reboot the server. <p>Keywords: LCS events</p>
-	<p>Description: Flow pattern without any specific VLAN will match for VLAN packets as well. Meaning, the flow rule: <code>"flow create 0 ingress pattern eth / vlan vid is 3 / ipv4 / end ..."</code> will only match VLAN packets with vid=3.</p> <p>The flow rules: <code>"flow create 0 ingress pattern eth / ipv4 / end ..."</code> or <code>"flow create 0 ingress pattern eth / vlan / ipv4 / end ..."</code> will match any IPv4 packet (VLAN included).</p> <p>Workaround: Use specific VLAN for filtering.</p> <p>Keywords: VLAN, flow</p>
-	<p>Description: Packets that have UDP destination port 4791 will be silently dropped by the NIC which may cause to low performance of 0, packet loss tests. This port is a reserved port and should not be used.</p> <p>Workaround: Use a different port as the UDP destination port</p>

Internal Ref.	Issues
	Keywords: UDP port 4791
-	<p>Description: Auto inline feature is supported only on a single port ConnectX-4 adapter card (is not supported on ConnectX-4 Lx) and cannot be used for 0 packet loss performance</p> <p>Workaround: Use auto-inline only with supported mode</p> <p>Keywords: Auto inline</p>
-	<p>Description: LCS events are not received on KVM/ESX VM when rte_eth_dev_set_link_down/up are called.</p> <p>Workaround: N/A</p> <p>Keywords: LCS events on VM</p>
-	<p>Description: VXLAN RX HW Checksum offloading is not supported</p> <p>Workaround: Will be added in next releases</p> <p>Keywords: VXLAN RX HW Checksum offloading</p>
-	<p>Description: VXLAN RH traffic is spread between RSS queues according to the outer header</p> <p>Workaround: Will be added in next releases</p> <p>Keywords: VXLAN RSS</p>
-	<p>Description: When destination MAC of a sent packet is the same as the ports MAC, the packet is returned to the port and is not sent to the network. Note: On ConnectX-4 Lx, the packet is sent when txq_mpw_en is set to 1 (the default configuration).</p> <p>Workaround: N/A</p> <p>Keywords: The dest MAC of the sent packet</p>
-	<p>Description: When the number of RX queues is not power of 2, the RETA table size is automatically configured to 512 to achieve the best flows spreading. Still in some cases traffic spreading between the queues can be non-equal</p> <p>Workaround: Use power of 2 number of queues if possible</p> <p>Keywords: RSS with non-power of 2 RX queues performance</p>
-	<p>Description: The maximum supported number of descriptors in TX/RX queues is equal to 32K. When txq_inline parameter is used the number is smaller and is equal to $32K/(txq_inline/32 + 1)$. When auto-inline is used, the maximum supported number of descriptor is 6K If the number of descriptors of TX or RX queues is higher than the above maximum, segmentation fault will occur</p> <p>Workaround: Use maximum number of TX and RX descriptors as explained.</p> <p>Keywords: Maximum supported size of TX/RX queue</p>
-	<p>Description: Performance degradation of small messages might occur when MTU is higher than the mbuf size</p> <p>Workaround: N/A</p> <p>Keywords: Performance of small messages when MTU > mbuf size</p>
-	Description: Adding MAC address on KVM VM is not supported

Internal Ref.	Issues
	<p>Workaround: N/A</p> <p>Keywords: Adding MAC address on KVM VM</p>
-	<p>Description: VLAN packets are received but are not sent when DPDK is running on KVM VM</p> <p>Workaround: N/A</p> <p>Keywords: VLAN packets on KVM VM</p>
-	<p>Description: DPDK has a flag per port to indicate the VLAN stripping state (On/Off), that updated when using <code>rte_eth_dev_set_vlan_offload()</code> API. This flag is not updated when calling the <code>rte_eth_dev_set_vlan_strip_on_queue()</code> API (VLAN stripping per queue), and can cause some confusion when mixing between those two APIs.</p> <p>Workaround: Use only one API to configure RX stripping</p> <p>Keywords: RX VLAN Stripping has unexpected behavior</p>
-	<p>Description: TX VLAN offloading insertion is supported only when <code>txq_mpw_en</code> is set to 0. By default is not supported since MPW is enabled</p> <p>Workaround: Set <code>txq_mpw_en=0</code>. Please check the Quick Start Guide for more information regarding command line options.</p> <p>Keywords: TX VLAN insertion offloading with ConnectX-4 LX</p>
-	<p>Description: Changing MTU during traffic results in segmentation fault.</p> <p>Workaround: Stop the ports before changing the MTU</p> <p>Keywords: Changing MTU during traffic</p>
-	<p>Description: Flow Control settings are not allowed on VM</p> <p>Workaround: Set Flow Control settings on the Hypervisor</p> <p>Keywords: Flow Control settings on KVM VM</p>
-	<p>Description: Allmulticast on VM is not supported</p> <p>Workaround: N/A</p> <p>Keywords: Allmulticast on KVM and ESX VM</p>
-	<p>Description: ESX VM supports only a single unicast MAC</p> <p>Workaround: N/A</p> <p>Keywords: Adding unicast MAC on ESX VM is not supported</p>
-	<p>Description: When configuring the MTU on the VM make sure the MTU on the hypervisor is the same as on the VM</p> <p>Workaround: N/A</p> <p>Keywords: MTU configuration on KVM VM</p>
-	<p>Description: Flow director is not supported on ESX VM</p> <p>Workaround: N/A</p> <p>Keywords: Flow director on ESX VM</p>

5 Bug Fixes

Table 5: Bug Fixes

Internal Ref.	Issue
DPDK v16.11 Bug Fixes	
-	Description: Fixed redundant free of Tx buffer Keywords: mlx5 Discovered in Release: DPDK 16.07 Fixed in Release: MLNX_DPDK 16.11_3.0
-	Description: Fixed exception handling Keywords: mlx5 Discovered in Release: DPDK 16.07 Fixed in Release: MLNX_DPDK 16.11_3.0
-	Description: Fixed index handling for Tx ring Keywords: mlx5 Discovered in Release: DPDK 16.11_2.3 Fixed in Release: MLNX_DPDK 16.11_3.0
-	Description: Fixed Tx when first segment size is too short Keywords: mlx5 Discovered in Release: DPDK 16.11 Fixed in Release: MLNX_DPDK 16.11_3.0
-	Description: Fixed Completion buffer size Keywords: mlx5 Discovered in Release: DPDK 16.07 Fixed in Release: MLNX_DPDK 16.11_3.0
-	Description: Fixed PMD specific parameters defaults. Keywords: mlx5 Discovered in Release: DPDK 16.11_2.3 Fixed in Release: MLNX_DPDK 16.11_3.0
-	Description: Fixed updating RETA. Keywords: mlx5 Discovered in Release: DPDK 16.11 Fixed in Release: MLNX_DPDK 16.11_2.3
-	Description: Fixed extended statistics counters identification. Keywords: mlx5 Discovered in Release: DPDK 16.11 Fixed in Release: MLNX_DPDK 16.11_2.3

Internal Ref.	Issue
-	<p>Description: Fixed VLAN stripping indication.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_2.3</p>
-	<p>Description: Fixed extended statistics wrong number.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_2.3</p>
-	<p>Description: Fixed inline WQE consumption.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_2.3</p>
-	<p>Description: Fixed Tx WQE corruption caused by starvation.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_2.3</p>
-	<p>Description: Fixed Ethernet header re-writing.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_2.3</p>
-	<p>Description: Fixed Rx packet validation and type.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_1.5</p>
-	<p>Description: Changed the default RSS algorithm of the NIC to TOPLITZ.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_1.5/ MLNX_OFED 4.0-1.0.0.1</p>
-	<p>Description: Fixed unexpected TX drops.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_1.5</p>
-	<p>Description: Fixed data corruption with MPW on ConnectX-4 Lx.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_1.5</p>

Internal Ref.	Issue
-	<p>Description: Enlarged the RETA table to 512 instead of 256 in case of non-power of two to improve flows spreading.</p> <p>Keywords: mlx5</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_1.5</p>
-	<p>Description: Fixed IPv4 and IPv6 packet type report.</p> <p>Keywords: mlx5/mlx4</p> <p>Discovered in Release: DPDK 16.11</p> <p>Fixed in Release: MLNX_DPDK 16.11_1.5</p>

6 Changes and New Features History

Table 6: Changes and New Features History

Driver	Changes
Rev. 16.11_2.3	
ConnectX-4/5 PMD, mlx5	<ul style="list-style-type: none"> • Added HW TSO support for non-tunneled and tunneled packets • [Beta] Added HW TSO support for VXLAN packets without the inner Ethernet header. Such capability is supported only for PFs. • [Beta] Added HW checksum offload for VXLAN packets without the inner Ethernet header. Such capability is supported only for PFs. • Optimized Enhanced MPW for ConnectX®-5 • Added support for RETA table update • Added support for MPLS RSS (MLNX_OFED feature) • Added ‘out-of-buffer’ counter to extended statistics • Bug fixes
Rev. 16.11_1.5	
ConnectX-4/5 PMD, mlx5	<ul style="list-style-type: none"> • Added support for ConnectX®-5 NICs [at Beta level] • Added support for optimized MPW for ConnectX®-5 • Added support for MLNX_OFED 4.0-x.x.x.x • Vectorized logic on TX path to improve CPU utilization (backport from DPDK 17.02) • Improved RSS spreading in case of non-power of two number of queues • Added support for port extended statistics (backported from DPDK 17.02) • Added Auto-inline support for ConnectX-4 single port NIC (backported from MLNX_DPDK 2.2_4.8)