



Connect. Accelerate. Outperform.™

# **Mellanox DPDK Release Notes**

Rev 2.2\_2.7

[www.mellanox.com](http://www.mellanox.com)

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

|  |           |
|--|-----------|
| <b>1 Overview .....</b>                                      | <b>5</b>  |
| 1.1 System Requirements .....                                | 5         |
| <b>2 Changes and Major New Features in Rev 2.2_2.7 .....</b> | <b>6</b>  |
| <b>3 mlx4 and mlx5 PMD Drivers Features.....</b>             | <b>7</b>  |
| <b>4 Known Issues .....</b>                                  | <b>8</b>  |
| 4.1 General Known Issues .....                               | 8         |
| 4.2 mlx4 PMD Known Issues.....                               | 8         |
| 4.3 mlx5 PMD Known Issue .....                               | 9         |
| <b>5 Bug Fixes .....</b>                                     | <b>12</b> |
| <b>6 Change Log History .....</b>                            | <b>14</b> |

## List of Tables

|  |    |
|--|----|
| Table 1: System Requirements .....           | 5  |
| Table 2: Changes and Major New Features..... | 6  |
| Table 3: General Known Issues .....          | 8  |
| Table 4: mlx4 PMD Known Issues .....         | 8  |
| Table 5: ConnectX-4 PMD Bug Fixes .....      | 12 |
| Table 6: Change Log History .....            | 14 |

# 1 Overview

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

## 1.1 System Requirements

**Table 1: System Requirements**

| Specification                 | Value   |
|-------------------------------|---|
| Network Adapter Cards         | ConnectX®-3 / ConnectX®-3 Pro / ConnectX®-4 /<br>ConnectX®-4 Lx network adapter card.<br>(These must be configured to work in ETH mode.)  |
| Firmware                      | <ul style="list-style-type: none"> <li>ConnectX 4: 12.16.1006 / 12.16.1010</li> <li>ConnectX-4 Lx: 14.16.1006 / 12.16.1010</li> <li>ConnectX 3/ ConnectX-3 Pro: 2.36.5000</li> </ul> <p><b>Note:</b> To receive firmware 12/14.16.1010, please contact Mellanox Support: support@mellanox.com</p> |
| Linux Driver Stack            | MLNX_OFED- 3.3-1.0.0.0  |
| ESXi 6.0 Driver               | ConnectX-4 / ConnectX-4 Lx: 4.15.5-10   |
| ESXi 5.5 U1 Driver            | ConnectX-4 / ConnectX-4 Lx: 4.5.5-10  |
| Operating Systems and Kernels | All OSs supported by MLNX_OFED  |
| Minimum memory requirements   | 16 GB RAM   |
| Transport                     | Ethernet  |
| CPU Arch                      | x86   |

## 2 Changes and Major New Features in Rev 2.2\_2.7

**Table 2: Changes and Major New Features**

| Driver               | Changes  |
|----------------------|--|
| ConnectX-4 PMD, mlx5 | <ul style="list-style-type: none"><li>• Major CPU usage performance optimizations</li><li>• Performance improvements for messages &gt;=256B</li><li>• Automatic scatter gather setting:<br/>RX according to MTU and TX is set to HW maximum</li><li>• Added support for flow director drop queue</li><li>• Replaced compilation options with command line parameters</li><li>• Added support for ESX 6.0 and 5.5 SR-IOV for ConnectX-4</li><li>• HW TX VLAN insertion only (removed SW VLAN insertion)</li><li>• Bug fixes</li></ul> |

### 3 mlx4 and mlx5 PMD Drivers Features

| Feature                                    | mlx4 PMD                       | mlx5 PMD                      |
|--|--------------------------------|-------------------------------|
| Supported NICs                             | ConnectX®-3<br>ConnectX®-3 Pro | ConnectX®-4<br>ConnectX®-4 Lx |
| PCI mapping                                | Function per device            | Function per port             |
| KVM SR-IOV                                 | Yes                            | Yes                           |
| ESX 6.0 SR-IOV                             | Yes                            | Yes                           |
| ESX 5.5 SR-IOV                             | No                             | 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                            | No                            |
| Flow Director                              | No                             | Yes                           |
| RX VLAN stripping                          | No                             | Yes                           |
| TX VLAN insertion                          | No                             | Yes                           |

## 4 Known Issues

### 4.1 General Known Issues

**Table 3: General Known Issues**

| Subject                                       | Description  | Workaround  |
|---|--|---|
| InfiniBand                                    | Mellanox PMDs does not support InfiniBand.   | N/A   |
| Support for DPDK integrated shared library    | Mellanox PMDs does not support DPDK integrated shared library  | PMD compiled as shared library can be used  |
| Hardware counters                             | Hardware counters are not implemented.   | N/A   |
| The Primary and Secondary multi process model | The Primary and Secondary multi process model is currently supported in TX only.   | Use multithreaded model instead of multi process in case mutli process RX is needed |
| Active - Passive Bonding mode                 | 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. |   |
| Bond LAG interface is down                    | Bond PMD does not notice when Mellanox PMD's ports are up  | Run ifconfig down and up on the bond slaves of the Mellanox interfaces              |

### 4.2 mlx4 PMD Known Issues

**Table 4: mlx4 PMD Known Issues**

| Subject  | Description  | Workaround  |
|--|--|---|
| Multicast / Broadcast Self-loopback on SR-IOV VF | When a Multicast or Broadcast packet is sent to the SR-IOV VF, it is returned back to the port.  | N/A   |
| Performance degradation when SGE_NUM =4          | Performance degradation might occur for small packets when PMD is compiled with SGE_NUM = 4 compared to the performance when SGE_NUM = 1 | If scattered packets are not used, compile PMD with SGE_NUM = 1 |
| Hardware checksum offloading on ConnectX-3       | Hardware checksum offloading is not supported in ConnectX-3  | Use ConnectX3-Pro if hardware checksum offloading is needed     |
| Sending abnormal packets                         | Sending abnormal packets, smaller then 16B and bigger than the configured MTU causes PMD queues to enter an error state.                 | run dev_stop and dev_start                                      |
| ESX 6.0 and DMFS A0 mode                         | DMFS A0 mode is not supported in ESX 6.0 VM  | N/A   |



| Subject  | Description  | Workaround                             |
|--|--|--|
| ESX 6.0 and HW checksum offloading                   | RX hardware checksum offloads are not supported in ESX 6.0 VM  | N/A                                    |
| ESX 6.0 and “add MAC”                                | Only one additional MAC can be added in additional to the VF’s MAC in ESX 6.0 VM   | N/A                                    |
| VXLAN hardware checksum offloading                   | VXLAN hardware checksum offloading is supported only if the used kernel supports it  | N/A                                    |
| L4 checksum report with hardware checksum offloading | When the packet does not include TCP/UDP header and the hardware checksum offloading is enabled, L4 checksum will be reported as bad | N/A                                    |
| RSS  | RSS hash key and options cannot be modified.   | N/A                                    |
| Promiscuous and SR-IOV                               | Promiscuous mode does not work when SR-IOV is enabled.   | N/A                                    |
| Promiscuous and SR-IOV                               | In testpmd: Although promiscuous mode fails in SR-IOV, it still shows as enabled<br>This is DPDK implementation bug.                 | N/A                                    |
| VLAN filtering                                       | VLAN filtering is supported only with non-optimized steering mode.   | See QSG, RX VLAN Filter section        |
| Number of configured RSS queues must be power of 2   | Number of configured RSS queues must be power of 2.  | Use only power of 2 RSS queues         |
| Broadcast packets and VLAN filter on KVM VM          | Broadcast packets with VLAN are not received when VLAN filter is configured on KVM VM  | N/A                                    |
| MLNX OFED 2.4  | MLNX_OFED 2.4 and below is not supported   | It is recommended to use MLNX_OFED 3.3 |

### 4.3 mlx5 PMD Known Issue

**Table 4: mlx5 PMD Known Issues**

| Subject                        | Description   | Workaround  |
|--------------------------------|---|---|
| RSS IP performance degradation | Performance degradation might occur when running DPDK application with RSS IP and using firmware 14.16.1006 / 12.16.1006. | Either use RSS UDP instead, or upgrade to firmware v12/14.16.1010 |
| Power8 ARCH                    | Power8 Arch is not supported  | Will be added in next release                                     |
| VXLAN HW Checksum offloading   | VXLAN HW Checksum offloading is not supported   | Will be added in next releases                                    |
| VXLAN RSS                      | VXLAN traffic is spread between RSS queues according to the outer header  | Will be added in next releases                                    |
| Flow director VXLAN filter     | Flow director VXLAN filter is not supported   | Will be added in next releases                                    |

| Subject  | Description   | Workaround   |
|--|---|--|
| The dest MAC of the sent packet                    | When destination MAC of a sent packet is the same as the port's MAC, the packet is returned to the port and is not sent to the network.<br>Note: On ConnectX-4 Lx, the packet is sent when <code>txq_mpw_en</code> is set to 1 (the default configuration).   | N/A  |
| RSS with non-power of 2 RX queues performance      | When the number of RX queues is not power of 2, the RETA table size is automatically configured to 256 to achieve the best flows spreading. Still in some cases traffic spreading between the queues can be non-equal   | Use power of 2 number of queues if possible  |
| Maximum supported size of TX/RX queue              | The maximum supported number of descriptors in TX/RX queues is equal to 32K. When <code>txq_inline</code> or <code>txq_inline_new</code> parameter is used, the number is smaller and is equal to $32K/(txq\_inline/32)$ .<br>If the number of descriptors of TX or RX queues is higher than the above maximum, segmentation fault will occur     | Use maximum number of TX and RX descriptors as explained.  |
| Performance of small messages when MTU > mbuf size | Performance degradation of small messages might occur when MTU is higher than the mbuf size   | N/A  |
| RX VLAN Stripping has unexpected behavior          | 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.<br>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. | Use only one API to configure RX stripping   |
| TX VLAN insertion offloading with ConnectX-4 LX    | 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   | Set <code>txq_mpw_en=0</code> .<br>Please check the Quick Start Guide for more information regarding command line options. |
| <code>txq_inline_new=128</code>                    | When <code>txq_inline_new</code> is set to 128B, packets with size of 512 are not being sent.   | Set <code>txq_inline_new=64</code>   |
| Flow Director: Ports & VLAN masks format           | Flow Director Port and VLAN mask have to be in little endian format   |  |
| Changing MTU during traffic                        | Changing MTU during traffic results in segmentation fault.  | Stop the ports before changing the MTU   |
| Flow Control settings on KVM VM                    | Flow Control settings are not allowed on VM   | Set Flow Control settings on the Hypervisor  |

| Subject                                       | Description  | Workaround |
|---|--|------------|
| Allmulticast on KVM and ESX VM                | Allmulticast on VM is not supported  | N/A        |
| Adding MAC on KVM VM                          | To add MAC on a VM in addition to the DPDK API that should be used ( <code>rte_eth_dev_mac_addr_add</code> ), the following command must be run:<br><pre>bridge fdb add dev &lt;interface name&gt; &lt;MAC ADDRESS&gt;</pre> | N/A        |
| Adding unicast MAC on ESX VM is not supported | ESX VM supports only a single unicast MAC  | N/A        |
| MTU configuration on KVM VM                   | When configuring the MTU on the VM make sure the MTU on the hypervisor is the same as on the VM  | N/A        |
| Flow director on ESX VM                       | Flow director is not supported on ESX VM   | N/A        |

## 5 Bug Fixes

**Table 5: ConnectX-4 PMD Bug Fixes**

| Subject                               | Description   | Found in          | Fixed in          |
|---------------------------------------|---|-------------------|-------------------|
| Flow director return error code       | Fixed inconsistent return value in Flow Director  | MLNX_DPDK 2.2_1.6 | MLNX_DPDK 2.2_2.7 |
| testpmd application and CRC stripping | By default, testpmd configures CRC to be scattered by the HW, but it does not strip the CRC header.<br>ConnectX-4 adds CRC header for every packet that is sent. In case of IO forwarding on a loopback setup, the forwarded packet increases each send by 4B | MLNX_DPDK 2.2_1.6 | MLNX_DPDK 2.2_2.7 |
| Jumbo frames                          | Fixed an issue that resulted in the following error when working with MTU > 2048 and the PMD is compiled with SGE_NUM=4:<br>mlx5: got completion with error:  | MLNX_DPDK 2.2_1.4 | MLNX_DPDK 2.2_1.6 |
| Flow Director: zero values            | Avoid packet duplication with overlapping Flow Director rules. For example the case where Flow Director mask is zero and the value is not   | MLNX_DPDK 2.2_1.4 | MLNX_DPDK 2.2_1.6 |
| HW capabilities check                 | Fixed an error which occurred when the software configured FCS stripping or HW padding which were not supported by the firmware.  | MLNX_DPDK 2.2_1.4 | MLNX_DPDK 2.2_1.6 |
| L4 checksum of L3 packet              | Fixed a faulty L4 checksum which was reported when the packet is L3   | MLNX_DPDK 2.1_1   | MLNX_DPDK 2.2_1   |
| L3 and L4 checksum of L2 packet       | Fixed a faulty L3 and L4 checksums which were reported as bad in case packet is L2  | MLNX_DPDK 2.1_1   | MLNX_DPDK 2.2_1   |
| Parallel builds compilation           | Fixed an issue that showed some features as not supported although they were when DPDK was compiled with -j option  | MLNX_DPDK 2.1_1   | MLNX_DPDK 2.2_1   |
| IPv6 resolution protocol              | Fixed an issue causing IPv6 resolution protocol to not work in case all-multicast options were not configured   | MLNX_DPDK 2.1_1   | MLNX_DPDK 2.2_1   |

**Table 6: ConnectX-3 PMD Bug Fixes**

| Subject  | Description   | Found in       | Fixed in                |
|--|---|----------------|-------------------------|
| Windriver OS support                                     | Fixed DPDK on Windriver   | MLNX_OFED 3.2  | MLNX_OFED 3.3           |
| TX checksum offloading on KVM                            | Fixed TX checksum offloading on KVM VM.   | MLNX_OFED 3.1  | MLNX_OFED 3.3           |
| Multicast Promiscuous mode and SR-IOV                    | Fixed Virtual Machines (VM) with enabled Multicast Promiscuous mode do not receive the multicast traffic that other VMs on the same hypervisor are registered to.                   | MLNX_OFED 2.4  | MLNX_OFED 3.0 and above |
| QP setup failure   | Fixed QP setup failure on Debian 3.16.7-ckt7-1  | DPDK 2.1 inbox | 2.1_1.1                 |
| Memory corruption when fork() is used in the application | Fixed an unpredicted behavior due to “cow” mechanism when an application used fork() or system call that were not related to the networking.  | DPDK 2.0 inbox | 2.0_2.8.4               |
| 1Gb/s port's link reported as 10Gb/s                     | Fixed a wrong speed reporting when a 10Gb/s port was set as 1Gb/s (10Gb/s instead of 1Gb/s)   | DPDK 2.0 inbox | 2.0_2.8.4               |
| Multiple RX VLAN filters                                 | Fixed a multiple RX VLAN filters issue preventing multiple RX VLAN filters to function properly although all were configured. In this case, only the first one functioned properly. | DPDK 2.0 inbox | 2.0_2.8.4               |

**Table 7: DPDK v2.2 Bug Fixes and Changes**

| Subject                                   | Description   | Found in       | Fixed in           |
|---|---|----------------|--------------------|
| ConnectX-4 100G/b port link speed         | Added support for 100 Gb/s link rate                            | DPDK 2.2       | MLNX_DPD K 2.2_1.6 |
| Compilation on Power8                     | Fixed compilation failure with <code>-pedantic</code> on Power8 | DPDK 2.2 inbox | MLNX_DPD K 2.2_1.4 |
| Add support for 512 entries in RETA table |   | DPDK 2.2 inbox | MLNX_DPD K 2.2_1.4 |

## 6 Change Log History

**Table 6: Change Log History**



**NOTE:** The changes are from the previous DPDK version on dpdk.org and not between two MLNX\_DPDK versions. DPDK (dpdk.org) Release Notes can be found at: [http://dpdk.org/doc/guides/rel\\_notes/](http://dpdk.org/doc/guides/rel_notes/)

| Driver                                      | Changes  |
|---|--|
| <b>Rev 2.2_1.6 from Rev 2.2_1.4</b>         |  |
| ConnectX-4 PMD, mlx5                        | <ul style="list-style-type: none"> <li>• DPDK bug fixes</li> </ul>   |
| ConnectX-3 PMD, mlx4                        | <ul style="list-style-type: none"> <li>• Bug fixes</li> </ul>  |
| <b>Rev 2.2_1.4 from DPDK 2.2</b>            |  |
| ConnectX-4 PMD, mlx5                        | <ul style="list-style-type: none"> <li>• Added support for flow director filters (RTE_FDIR_MODE_PERFECT and RTE_FDIR_MODE_PERFECT_MAC_VLAN modes)</li> <li>• Single core and max performance optimizations</li> <li>• Added support for RX VLAN stripping</li> <li>• Added support for TX VLAN insertion (supported only with MLNX_OFED-3.2-1.5.0.1 and above)</li> <li>• Added callbacks to support link (up / down) changes</li> <li>• Added support for HW packet padding (supported only with MLNX_OFED-3.2-1.5.0.1 and above)</li> <li>• Added support for scattering FCS (supported only with MLNX_OFED-3.2-1.5.0.1 and above)</li> <li>• Added TX support for secondary processes</li> <li>• Added support for MLNX_OFED 3.2-1.0.1.1 and higher</li> <li>• DPDK fixes (see section <a href="#">Table 7: DPDK v2.2 Bug Fixes</a>)</li> </ul> |
| ConnectX-4 PMD, mlx4                        | <ul style="list-style-type: none"> <li>• Added support for VMware ESX 6.0 U1</li> </ul>  |
| <b>Rev 2.1_1.1 from DPDK 2.1 inbox</b>      |  |
| ConnectX-4 PMD, mlx5                        | <ul style="list-style-type: none"> <li>• Added support for ConnectX®-4 and ConnectX®-4 LX NICs</li> <li>• Added support for MLNX_OFED 3.1-1.0.0</li> <li>• DPDK fixes (see section <a href="#">Table 7: DPDK v2.2 Bug Fixes</a>)</li> </ul>  |
| <b>Rev 2.1_1.1 from MLNX DPDK 2.0_2.8.4</b> |  |
| ConnectX-3 PMD, mlx4                        | <ul style="list-style-type: none"> <li>• Removed support for VMware ESX 5.5</li> <li>• Added support for Accelerated verbs – PMD does not include libibverbs and libmlx4</li> </ul>  |