Mellanox OFED for FreeBSD for ConnectX-4/ConnectX-4 Lx/ ConnectX-5/ConnectX-6 Release Note

Rev 3.5.2

NOTE:

THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT □PRODUCT(S)ᴺ AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES AS-ISꟷ 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

Table of Contents

List of Tables

Release Update History

Table 1 - Release Update History

Release	Date	Description
Rev 3.5.2	September 29, 2019	Initial release of this version.

1 Introduction

These are the release notes for Mellanox Technologies' driver for FreeBSD Rev 3.5.2 driver kit for Mellanox ConnectX®-4, ConnectX®-4 Lx, ConnectX®-5, ConnectX®-5 Ex adapter cards supporting the following uplinks to servers:

Uplink/HCAs	Driver Name	Uplink Speed
ConnectX®-4	mlx5	• InfiniBand: SDR, QDR, FDR, FDR10, EDR • Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, 56GigE[a], and 100GigE
ConnectX®-4 Lx		• Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, and 50GigE
ConnectX®-5/ ConnectX®-5 Ex		• InfiniBand: SDR, QDR, FDR, FDR10, EDR • Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, and 100GigE
ConnectX®-6		• InfiniBand: SDR, EDR, HDR • Ethernet: 10GigE, 25GigE, 40GigE, 50GigE, 100GigE[b], and 200GigE (**alpha**)

    a. 56 GbE is a Mellanox propriety link speed and can be achieved while connecting a Mellanox adapter cards to Mellanox SX10XX switch series or connecting a Mellanox adapter card to another Mellanox adapter card.

    b. ConnectX-6 Ethernet adapter cards currently support Force mode only. Auto-Negotiation mode is **not** supported.

1.1 Supported Platforms and Operating Systems

The following are the supported OSs in Mellanox OFED for FreeBSD for ConnectX-4/ConnectX-4 Lx/ConnectX-5/ConnectX-6 Rev 3.5.2:

Table 2 - Supported Platforms and Operating Systems

Operating System	Platform
FreeBSD 13	AMD64/x86_64

1.2 Supported Adapters Firmware Versions

Mellanox OFED for FreeBSD Rev 3.5.2 supports the following Mellanox network adapter cards:

Supported Adapters	Current Firmware Rev.
ConnectX®-4	12.26.1040
ConnectX®-4 Lx	14.26.1040
ConnectX®-5/ConnectX-5 Ex	16.26.1040
ConnectX®-6	20.26.1040

2   Changes and New Features in Rev 3.5.2

Table 3 - Changes and New Features in Rev 3.5.2

Category	Description
SR-IOV over ESXi	Added support for SR-IOV Guest over ESXi. **NOTE**: Certain earlier versions of VMware prevented proper function of MS/MSI-X emulation on FreeBSD Guests. Since these versions are not accurately accounted for, MSI is blacklisted by default, preventing mlx5(4) driver from loading properly. If MS/MSI-X emulation does not function properly on your current hypervisor, you may set the loader tunables `hw.pci.enable_msi` and `hw.pci.enable_msix` to 1, and `hw.pci.honor_msi_blacklist` to 0.
Forward Error Correction (FEC) Configuration	Added support for FEC configuration via the FreeBSD driver.
IRQ Labeling	IRQ IDs are now labeled according to their Mellanox device functionality.
Priority Flow Control (PFC) Hardware Buffer Configuration	[**Beta**] Added support for configuring HW buffers for priority flow control (PFC).
Port Module Events	Added counters for various port module events.
ConnectX-6 Firmware Dump	Expanded firmware dump support for ConnectX-6 adapter cards.
Bug Fixes	See Section 4, "Bug Fixes", on page 10.

For additional information on the new features, please refer to the User Manual.

3 Known Issues

The following is a list of general limitations and known issues of the various components of this Mellanox OFED for FreeBSD release.

Table 4 - Known Issues

Internal Ref.	Issue
-	**Description**: When reloading the module during heavy traffic, FreeBSD driver may become unstable.
	Workaround: N/A
	Keywords: Reload, stress, heavy traffic
	Discovered in Release: 3.5.2
1320335	**Description:** When Witness is enabled, the following message may appear in logs: "`lock order reversal in mlx5_en_rx and in_pcb/tcp_input`".
	Workaround: N/A
	Keywords: Witness, LOR
	Discovered in Release: 3.5.0
1578093	**Description:** ibstat tool shows the wrong value of "rate" after unplugging the cable from the HCA.
	Workaround: N/A
	Keywords: ibstate, rate
	Discovered in Release: 3.5.0
1439351	**Description:** Link local GIDs are dysfunctional when IPv6 address is configured for the first time.
	Workaround: Set the net device state to "up". For example: `# ifconfig mce0 up`
	Keywords: RoCE, IPv6
	Discovered in Release: 3.4.2
1435021	**Description:** All Rx priority pause counters values increase when Rx global pause is enabled.
	Workaround: Ignore Rx priority pause counters when Rx global pause is enabled.
	Keywords: Rx pause counters, priority
	Discovered in Release: 3.4.2
1434034	**Description:** RDMA-CM applications do not work when PCP is configured on one side of the connection.
	Workaround: Make sure PCP is configured on both sides of the connection.
	Keywords: RDMA-CM, PCP
	Discovered in Release: 3.4.2

Table 4 - Known Issues

Internal Ref.	Issue
1428828	**Description:** Extended join multicast API is not supported.
	Workaround: N/A
	Keywords: RDMA, Multicast
	Discovered in Release: 3.4.2
1313461	**Description:** When Packet Pacing is enabled in firmware, only one traffic class will be supported by the firmware.

	Workaround: Disable Packet Pacing in the firmware configuration. For example:
	
	`# cat /tmp/disable_pp.txt`
	`MLNX_RAW_TLV_FILE`
	`0x00000004 0x0000010c 0x00000000 0x00000000`
	`# mlxconfig -d pci0:4:0:0 -f /tmp/disable_pp.txt set_raw`

	Keywords: Firmware, Packet Pacing
	Discovered in Release: 3.4.2
1227471	**Description:** When loading and unloading linuxkpi module, the following error message will appear in the dmesg, indicating that a memory leak has occurred: "`Warning: memory type linux leaked memory on destroy (2 allocations, 64 bytes leaked). Warning: memory type linuxcurrent leaked memory on destroy (7 allocations, 896 bytes leaked).`"
	Workaround: N/A
	Keywords: linuxkpi
	Discovered in Release: 3.4.1
-	**Description:** The following error message may be printed to dmesg when using static configuration via rc.conf: "`loopback_route: deletion failed`" This is a kernel-related issue.
	Workaround: N/A
	Keywords: Static Configuration
-	**Description:** Choosing a wrong interface media type will cause a "no carrier" status and the physical port will not be active.
	Workaround: N/A
	Keywords: Media Type
-	**Description:** There is no TCP traffic when configuring MTU in the range of 72-100 bytes in ConnectX®-4 Lx.
	Workaround: N/A
	Keywords: MTU

4 Bug Fixes

The table below lists the bugs fixed in this release.

Table 5 - Bug Fixes

Internal Ref.	Issue
1243940	**Description:** Fixed the issue where RDMA applications (user space and Kernel space) might hang when restarting the driver during traffic.
	Keywords: RDMA, driver restart
	Discovered in Release: 3.4.1
	Fixed in Release: 3.5.1
1402958	**Description:** Fixed the issue where interfaces were not loaded after firmware software reset while RDMA traffic was running in the background.
	Keywords: Self healing, RDMA
	Discovered in Release: 3.4.2
	Fixed in Release: 3.5.1
1581628	**Description:** Fixed the issue were driver unload used to hang while RDMA user space application was running.
	Keywords: RDMA, driver unload
	Discovered in Release: 3.5.0
	Fixed in Release: 3.5.1
1554671	**Description:** Fixed the issue where mlx5ib unload used to fail while OpenSM was running in the background.
	Keywords: mlx5ib, OpenSM, RDMA
	Discovered in Release: 3.5.0
	Fixed in Release: 3.5.1
1498467	**Description:** Added support for 10G-ER and 10G-LR modules recognition.
	Keywords: SFP module
	Discovered in Release: 3.4.2
	Fixed in Release: 3.5.0
1175757	**Description:** Added support for running RDMA CM with IPoIB.
	Keywords: RDMA CM, IPoIB
	Discovered in Release: 3.4.1
	Fixed in Release: 3.5.0

Table 5 - Bug Fixes

Internal Ref.	Issue
1337448/ 1485155/ 1470374	**Description:** Fixed the issue of when rebooting a virtual machine (VM), the following log message may appear: `warning: event(0) on port 0`
	Keywords: Virtualization, RDMA
	Discovered in Release: 3.4.2
	Fixed in Release: 3.5.0
1297834	**Description:** Fixed the issue of when running over VLAN, RDMA loopback traffic used to fail.
	Keywords: RDMA, loopback, VLAN
	Discovered in Release: 3.4.1
	Fixed in Release: 3.4.2
1258718	**Description:** Fixed the issue of when working in RoCE mode using ConnectX-4 HCAs only, a bandwidth performance degradation used to occur when sending/ receiving a message of any size larger than 16K.
	Keywords: RoCE, performance, ConnectX-4
	Discovered in Release: 3.4.1
	Fixed in Release: 3.4.2
1273118/ 1399014	**Description:** Added support for RDMA multicast traffic.
	Keywords: RDMA, multicast
	Discovered in Release: 3.4.1
	Fixed in Release: 3.4.2
765775	**Description:** Suppressed EEPROM error message/s that used to be received when SFP cages were empty.
	Keywords: EEPROM, SFP
	Discovered in Release: 3.0.0
	Fixed in Release: 3.3.0
854565	**Description:** Allowed setting software MTU size below the value of 1500.
	Keywords: MTU
	Discovered in Release: 3.0.0
	Fixed in Release: 3.3.0

5 Change Log History

Table 6 - Change Log History

Release	Category	Description
3.5.1	Firmware Upgrade Using mlx5tool	Added the ability to burn firmware of MFA2 format using mlx5tool and Kernel module.
	Dynamic Interrupt Moderation (DIM)	Added the ability to adaptively configure interrupt moderation based on network traffic.
3.5.0	Relaxed Ordering	Added support for configuring PCIe packet write ordering via sysctl.
	Enhanced Transmission Selection (ETS)	Added support for setting the bandwidth limit as a ratio rather than in bits per second. The ratio must be an integer number between 1 and 100, inclusive. This feature also enables setting a minimal BW guarantee on traffic classes (TCs).
	Ethernet Counters	Added support for the following new counters: • tx_jumbo_packets • rxstat0.bytes • txstat0tc0.bytes
3.4.2	RoCE Packet Sniffing	Added support for RoCE packets sniffing using tcpdump tool.
	VLAN 0 Priority Tagging	Added support for 802.1Q Ethernet frames to be transmitted with VLAN ID set to zero in RoCE mode.
	Differentiated Service Code Point (DSCP)	Added support for classifying and managing network traffic and providing quality of service (QoS) on IP and RoCE networks.
	Trust State	Added support for prioritizing sent/received packets based on packet fields.
	Reset Flow	Added support for a reset mechanism to recover from fatal failures. Upon such failures, a firmware dump for all relevant registers will be triggered, followed by a firmware and driver reset.
	RDMA Mutlicast Support	Added support for sending and receiving RDMA multicast packets.
3.4.1	Explicit Congestion Notification (ECN)	Added support for ECN, which enables end-to-end congestion notifications between two end-points when a congestion occurs.
	Rate Limiting	Added support for users to rate limit a specific Traffic Class.
	Priority Flow Control (PFC)	Added the ability to apply pause functionality to specific classes of traffic on the Ethernet link. **Note**: Currently, only layer 2 PFC (PCP) is supported.
	Rx Hardware Time-Stamping	Added support for adding high-quality hardware time-stamping on incoming packets.
	Firmware Dump	Added the ability to dump hardware registered data upon demand.
3.3.0	Packet Pacing	Also known as "rate limit", this feature is now supported at a GA level. **Note**: This feature is supported in firmware v12.17.1016 and above.

Table 6 - Change Log History

Release	Category	Description
3.0.0	Hardware LRO	Added support for Large Receive Offload (LRO) in the hardware. It increases inbound throughput of high-bandwidth network connections by reducing CPU overhead. Hardware LRO is only supported in ConnectX®-4.
	Completion Based Moderation	Added the option to reset the timer for generating interrupts upon completion generation.
	EEPROM Cable Reading	Added support for EEPROM cable reading via ifconfig and sysctl. EEPROM is only supported in ConnectX®-4.
	Interface Name	Changed the interface name from mlx5en<X> to mce<X>.