

Intel[®] Data Plane Development Kit (Intel[®] DPDK)

Release Notes

Package Version: 1.7

June 2014



Description of Release

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: <http://www.intel.com/design/literature.htm>

Any software source code reprinted in this document is furnished for informational purposes only and may only be used or copied and no license, express or implied, by estoppel or otherwise, to any of the reprinted source code is granted by this document.

Code Names are only for use by Intel to identify products, platforms, programs, services, etc. ("products") in development by Intel that have not been made commercially available to the public, i.e., announced, launched or shipped. They are never to be used as "commercial" names for products. Also, they are not intended to function as trademarks.

See Appendix A for an overview of the applicable software licenses.

Intel, the Intel logo, Intel Core and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2012-2014, Intel Corporation. All rights reserved.

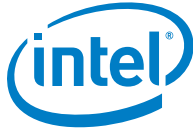


Revision History

| Date | Revision | Description |
|----------------|----------|----------------------------------------|
| May 2014 | -008 | Supports public software release 1.7 |
| January 2014 | -007 | Supports public software release 1.6 |
| November 2013 | -006 | Supports public software release 1.5.2 |
| October 2013 | -005 | Supports public software release 1.5.1 |
| September 2013 | -004 | Supports public software release 1.5.0 |
| August 2013 | -003 | Supports public software release 1.4.1 |
| June 2013 | -002 | Supports public software release 1.3.1 |
| November 2012 | -001 | Supports public software release 1.2.3 |

Contents

| | | |
|---|----------------------------------------------------------------------------------------------------|----|
| 1 | Description of Release | 7 |
| | 1.1 Important Note..... | 7 |
| | 1.2 Using Intel® DPDK Upgrade Patches | 7 |
| | 1.3 Documentation Roadmap | 7 |
| 2 | New Features | 10 |
| 3 | Supported Features | 11 |
| 4 | Supported Operating Systems | 16 |
| 5 | Updating Applications from Previous Versions..... | 17 |
| | 5.1 Intel® DPDK 1.6 to Intel® DPDK 1.7 | 17 |
| | 5.2 Intel® DPDK 1.5 to Intel® DPDK 1.6 | 17 |
| | 5.3 Intel® DPDK 1.4 to Intel® DPDK 1.5 | 17 |
| | 5.4 Intel® DPDK 1.3 to Intel® DPDK 1.4.x | 17 |
| | 5.5 Intel® DPDK 1.2 to Intel® DPDK 1.3 | 18 |
| | 5.6 Intel® DPDK 1.1 to Intel® DPDK 1.2 | 18 |
| 6 | Known Issues and Limitations | 20 |
| | 6.1 Pause Frame Forwarding does not work properly on igb..... | 20 |
| | 6.2 Running TestPMD with SRIOV in Domain U may cause it to hang when XENVIRT switch is on | 20 |
| | 6.3 Vhost-xen cannot detect Domain U application exit on Xen version 4.0.1 | 21 |



Description of Release

| | | |
|------|----------------------------------------------------------------------------------------------------------------------------------------|----|
| 6.4 | Virtio incorrect header length used if MSI-X is disabled by kernel driver..... | 21 |
| 6.5 | gmake clean may silently fail for some example applications | 22 |
| 6.6 | In packets provided by the PMD, some flags are missing | 22 |
| 6.7 | The rte_malloc library is not fully implemented..... | 23 |
| 6.8 | HPET reading is slow | 23 |
| 6.9 | HPET timers do not work on the Osage customer reference platform | 24 |
| 6.10 | Not all variants of supported NIC types have been used in testing | 24 |
| 6.11 | Multi-process sample app requires exact memory mapping | 25 |
| 6.12 | Unstable system performance across application executions with 2MB pages. | 25 |
| 6.13 | Packets are not sent by the 1 GbE/10 GbE SR-IOV driver when the source MAC address is not the MAC address assigned to the VF NIC | 26 |
| 6.14 | SR-IOV drivers do not fully implement the rte_ethdev API | 26 |
| 6.15 | PMD does not work with --no-huge EAL command line parameter | 27 |
| 6.16 | Some hardware off-load functions are not supported by the VF Driver | 28 |
| 6.17 | Kernel crash on IGB port unbinding..... | 28 |
| 6.18 | Link status change not working with MSI interrupts..... | 28 |
| 6.19 | Twinpond and Ironpond NICs do not report link status correctly | 29 |
| 6.20 | Discrepancies between statistics reported by different NICs | 29 |
| 6.21 | Error reported opening files on Intel® DPDK initialization | 30 |
| 6.22 | Intel® QuickAssist Technology sample application does not work on a 32-bit OS on Shumway | 30 |
| 6.23 | IEEE1588 support possibly not working with an Intel® Ethernet Controller I210 NIC..... | 31 |
| 6.24 | Differences in how different Intel NICs handle maximum packet length for jumbo frame | 31 |
| 6.25 | Link status interrupt not working in VF drivers..... | 32 |
| 6.26 | Binding PCI devices to igb_uio fails on Linux* kernel 3.9 when more than one device is used..... | 32 |
| 6.27 | GCC might generate Intel® AVX instructions for processors without Intel® AVX support..... | 32 |
| 6.28 | Ethertype filter could receive other packets (non-assigned) in Niantic..... | 33 |
| 6.29 | Double VLAN does not work on Intel® 40G ethernet controller..... | 33 |
| 6.30 | Cannot set link speed on Intel® 40G ethernet controller..... | 34 |
| 6.31 | Stopping the port does not down the link on Intel® 40G ethernet controller . | 34 |
| 7 | Resolved Issues..... | 36 |
| 7.1 | KNI does not provide Ethtool support for all NICs supported by the Poll-Mode Drivers | 36 |
| 7.2 | Linux IPv4 forwarding is not stable with vhost-switch on high packet rate..... | 36 |
| 7.3 | PCAP library overwrites mbuf data before data is used | 37 |
| 7.4 | MP Client Example app - flushing part of TX is not working for some ports if set specific port mask with skipped ports..... | 37 |
| 7.5 | Packet truncation with Intel® I350 Gigabit Ethernet Controller..... | 38 |
| 7.6 | Device initialization failure with Intel® Ethernet Server Adapter X520-T2 | 38 |
| 7.7 | Intel® DPDK kernel module is incompatible with Linux kernel version 3.3..... | 39 |
| 7.8 | Initialization failure with Intel® Ethernet Controller X540- T2..... | 39 |
| 7.9 | rte_eth_dev_stop() function does not bring down the link for 1 GB NIC ports | 40 |
| 7.10 | It is not possible to adjust the duplex setting for 1GB NIC ports | 40 |



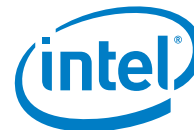
| | | |
|------|---------------------------------------------------------------------------------------------------------------------------------------------|----|
| 7.11 | Calling <code>rte_eth_dev_stop()</code> on a port does not free all the mbufs in use by that port | 41 |
| 7.12 | PMD does not always create rings that are properly aligned in memory | 41 |
| 7.13 | Checksum offload might not work correctly when mixing VLAN-tagged and ordinary packets | 42 |
| 7.14 | Port not found issue with Intel® 82580 Gigabit Ethernet Controller | 42 |
| 7.15 | Packet mbufs may be leaked from mempool if <code>rte_eth_dev_start()</code> function fails..... | 42 |
| 7.16 | Promiscuous mode for 82580 NICs can only be enabled after a call to <code>rte_eth_dev_start</code> for a port..... | 43 |
| 7.17 | Incorrect CPU socket information reported in <code>/proc/ cpuinfo</code> can prevent the Intel® DPDK from running..... | 43 |
| 7.18 | L3FWD sample application may fail to transmit packets under extreme conditions | 44 |
| 7.19 | L3FWD-VF might lose CRC bytes..... | 44 |
| 7.20 | 32-bit Intel® DPDK sample applications fails when using more than one 1 GB hugepage | 45 |
| 7.21 | I2fwd fails to launch if the NIC is the Intel® 82571EB Gigabit Ethernet Controller | 45 |
| 7.22 | 32-bit Intel® DPDK applications may fail to initialize on 64-bit OS | 46 |
| 7.23 | Lpm issue when using prefixes >24 | 46 |
| 7.24 | IXGBE PMD hangs on port shutdown when not all packets have been sent | 46 |
| 7.25 | Config file change can cause build to fail..... | 47 |
| 7.26 | <code>rte_cmdline</code> library should not be used in production code due to limited testing | 47 |
| 7.27 | Some <code>*_INITIALIZER</code> macros are not compatible with C++ | 48 |
| 7.28 | No traffic through bridge when using <code>exception_path</code> sample application | 48 |
| 7.29 | Segmentation Fault in <code>testpmd</code> after config fails | 49 |
| 7.30 | Linux kernel <code>pci_cfg_access_lock()</code> API can be prone to deadlock..... | 49 |
| 7.31 | When running multi-process applications, “ <code>rte_malloc</code> ” functions cannot be used in secondary processes | 50 |
| 7.32 | Configuring maximum packet length for IGB with VLAN enabled may not take into account the length of VLAN tag | 50 |
| 7.33 | Intel® I210 Ethernet controller always strips CRC of incoming packets..... | 51 |
| 7.34 | EAL can silently reserve less memory than requested..... | 51 |
| 7.35 | SSH connectivity with the board may be lost when starting an Intel® DPDK application | 51 |
| 7.36 | Remote network connections lost when running autotests or sample applications..... | 52 |
| 7.37 | KNI may not work properly in a multi-process environment | 52 |
| 7.38 | Hash library cannot be used in multi-process applications with multiple binaries | 53 |
| 7.39 | Unused hugepage files are not cleared after initialization | 53 |
| 7.40 | Packet reception issues when virtualization is enabled | 54 |
| 8 | Frequently Asked Questions (FAQ)..... | 55 |
| 8.1 | When running the test application, I get “EAL: <code>map_all_hugepages()</code> : open failed: Permission denied Cannot init memory”? | 55 |
| 8.2 | If I want to change the number of TLB Hugepages allocated, how do I remove the original pages allocated? | 55 |



Description of Release

| | | |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 8.3 | I have set up a total of 1024 Hugepages (that is, allocated 512 2M pages to each NUMA node). However, if I execute "l2fwd -c f -m 64 -n 3 -- -p 3", I get the following output, indicating that there are no socket 0 hugepages to allocate the mbuf and ring structures to? | 55 |
| 8.4 | I am running a 32-bit Intel® DPDK application on a NUMA system, and sometimes the application initializes fine but cannot allocate memory. Why is that happening?..... | 56 |
| 8.5 | On application startup, there is a lot of EAL information printed. Is there any way to reduce this? | 56 |
| 8.6 | How can I tune my network application to achieve lower latency? | 57 |
| 8.7 | Without NUMA enabled, my network throughput is low, why? I have a dual Intel® Xeon® E5645 processors @2.40 GHz with four Intel® 82599 10 Gigabit Ethernet NICs. Using eight logical cores on each processor with RSS set to distribute network load from two 10 GbE interfaces to the cores on each processor..... | 57 |
| 8.8 | I am getting errors about not being able to open files. Why? | 58 |
| 8.9 | Does my kernel require patching to run the Intel® DPDK? | 58 |
| 8.10 | VF driver for IXGBE devices cannot be initialized..... | 59 |
| Appendix A | Intel® DPDK License Overview..... | 60 |
| A.1 | BSD License | 62 |
| A.2 | GPLv2 License Information..... | 62 |
| A.3 | Dual BSD/LGPLv2 License Information..... | 64 |
| A.4 | Dual BSD/GPL License Information..... | 65 |

§



1 Description of Release

These release notes cover the new features, fixed bugs and known issues for Intel® Data Plane Development Kit (Intel® DPDK) release version 1.7.0.

For instructions on compiling and running the release, see the *Intel® DPDK Getting Started Guide*.

1.1 Important Note

All Intel® DPDK questions and technical problems including those regarding the Ethernet* Controllers for the Poll Mode Driver should be reported through the Intel® Premier Support site <http://premier.intel.com/premier> or access your IBL account and click the **Intel® Premier Support** link to enter issues under the Product Name "Data Plane Development Kit (DPDK)", which are then routed to our support team.

1.2 Using Intel® DPDK Upgrade Patches

For minor updates to the main Intel® DPDK releases, the software may be made available both as a new full package and as a patch file to be applied to the previously released package. In the latter case, the following commands should be used to apply the patch on top of the already-installed package for the previous release:

```
# cd $RTE_SDK
# patch -p1 < /path/to/patch/file
```

Once the patch has been applied cleanly, the Intel® DPDK can be recompiled and used as before (described in the *Intel® DPDK Getting Started Guide*).

Note: If the patch does not apply cleanly, perhaps because of modifications made locally to the software, it is recommended to use the full release package for the minor update, instead of using the patch.

1.3 Documentation Roadmap

The following is a list of Intel® DPDK documents in the suggested reading order:

- **Release Notes** (this document): Provides release-specific information, including supported features, limitations, fixed issues, known issues and so on. Also, provides the answers to frequently asked questions in FAQ format.
- **Getting Started Guide:** Describes how to install and configure the Intel® DPDK software; designed to get users up and running quickly with the software.
- **FreeBSD* Getting Started Guide:** A document describing the use of the Intel® DPDK with FreeBSD* has been added in Intel® DPDK Release 1.6.0. Refer to this



Description of Release

guide for installation and configuration instructions to get started using the Intel® DPDK with FreeBSD*.

- **Programmer's Guide:** Describes:
 - The software architecture and how to use it (through examples), specifically in a Linux* application (linuxapp) environment
 - The content of the Intel® DPDK, the build system (including the commands that can be used in the root Intel® DPDK Makefile to build the development kit and an application) and guidelines for porting an application
 - Optimizations used in the software and those that should be considered for new development

A glossary of terms is also provided.

- **API Reference:** Provides detailed information about Intel® DPDK functions, data structures and other programming constructs.
- **Sample Applications User Guide:** Describes a set of sample applications. Each chapter describes a sample application that showcases specific functionality and provides instructions on how to compile, run and use the sample application.

The following sample applications are included:

- Command Line
- Exception Path (into Linux* for packets using the Linux TUN/TAP driver)
- Hello World
- Integration with Intel® QuickAssist Technology
- Link Status Interrupt (Ethernet* Link Status Detection)
- IP Reassembly
- IP Fragmentation
- IPv4 Multicast
- L2 Forwarding (supports virtualized and non-virtualized environments)
- L2 Forwarding IVSHMEM
- L3 Forwarding
- L3 Forwarding with Power Management
- L3 Forwarding in a Virtualized Environment
- Load Balancing
- Multi-process
- QoS Scheduler + Dropper
- QoS Metering
- Quota & Watermarks
- Timer
- VMDQ and DCB L2 Forwarding
- VMDQ L2 Forwarding
- Userspace vhost
- Userspace vhost switch
- Netmap
- Kernel NIC Interface (KNI)



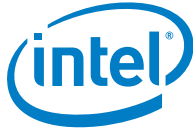
In addition, there are some other applications that are built when the libraries are created. The source for these applications is in the DPDK/app directory and are called:

- `test`
- `testpmd`

Once the libraries are created, they can be found in the `build/app` directory.

- The `test` application provides a variety of specific tests for the various functions in the Intel® DPDK.
- The `testpmd` application provides a number of different packet throughput tests and examples of features such as how to use the Flow Director found in the Intel® 82599 10 Gigabit Ethernet Controller.

The `testpmd` application is documented in the *Intel® DPDK Testpmd Application Note* (525362). The `test` application is not currently documented. However, you should be able to run and use `test` application with the command line help that is provided in the application.

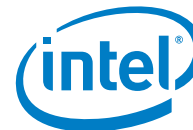


2 *New Features*

- Packet Distributor library for dynamic, single-packet at a time, load balancing
- IP fragmentation and reassembly library
- Support for IPv6 in IP fragmentation and reassembly sample applications
- Support for VFIO for mapping BARs and setting up interrupts
- Link Bonding PMD Library supporting round-robin, active backup, balance (layer 2, layer 2+3, and layer 3+4) and broadcast bonding modes
- Support zero copy mode RX/TX in user space vhost sample
- Support multiple queues in virtio-net PMD
- Support for Intel® 40GbE Controllers
- Support NIC filters in addition to flow director for Intel® 1GbE and 10GbE Controllers

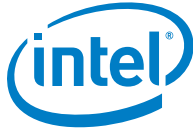
For further features supported in this release, see [Chapter 3 Supported Features](#).

§



3 Supported Features

- Virtualization (KVM)
 - Userspace vhost switch:
New sample application to support userspace virtio back-end in host and packet switching between guests.
- Virtualization (Xen)
 - Support for Intel® DPDK application running on Xen Domain0 without hugepages.
 - Para-virtualization
Support front-end Poll Mode Driver in guest domain
Support userspace packet switching back-end example in host domain
- FreeBSD* 9.2 support for librte_pmd_e1000, librte_pmd_ixgbe and Virtual Function variants. Please refer to the *Intel® DPDK for FreeBSD* Getting Started Guide*. Application support has been added for the following:
 - multiprocess/symmetric_mp
 - multiprocess/simple_mp
 - l2fwd
 - l3fwd
- Support for sharing data over QEMU IVSHMEM
- Support for Intel® Communications Chipset 8925 to 8955 Series in the DPDK-QAT Sample Application
- New VMXNET3 driver for the paravirtual device presented to a VM by the VMware* ESXi Hypervisor.
- BETA: example support for basic Netmap applications on Intel® DPDK
- Support for the wireless KASUMI algorithm in the dpdk_qat sample application
- Hierarchical scheduler implementing 5-level scheduling hierarchy (port, sub-port, pipe, traffic class, queue) with 64K leaf nodes (packet queues).
- Packet dropper based on Random Early Detection (RED) congestion control mechanism.
- Traffic Metering based on Single Rate Three Color Marker (srTCM) and Two Rate Three Color Marker (trTCM).
- An API for configuring RSS redirection table on the fly
- An API to support KNI in a multi-process environment
- IPv6 LPM forwarding
- Power management library and sample application using CPU frequency scaling
- IPv4 reassembly sample application
- Quota & Watermarks sample application

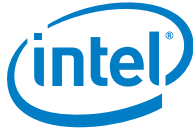


Supported Features

- PCIe Multi-BAR Mapping Support
- Support for Physical Functions in Poll Mode Driver for the following devices:
 - Intel® 82576 Gigabit Ethernet Controller
 - Intel® i350 Gigabit Ethernet Controller
 - Intel® 82599 10-Gigabit Ethernet Controller
 - Intel® XL710/X710 40-Gigabit Ethernet Controller
- Quality of Service (QoS) Hierarchical Scheduler: Sub-port Traffic Class Oversubscription
- Multi-thread Kernel NIC Interface (KNI) for performance improvement
- Virtualization (KVM)
 - Para-virtualization
 - Support virtio front-end poll mode driver in guest virtual machine
 - Support vHost raw socket interface as virtio back-end via KNI
 - SR-IOV Switching for the 10G Ethernet Controller
 - Support Physical Function to start/stop Virtual Function Traffic
 - Support Traffic Mirroring (Pool, VLAN, Uplink and Downlink)
 - Support VF multiple MAC addresses (Exact/Hash match), VLAN filtering
 - Support VF receive mode configuration
- Support VMDq for 1 GbE and 10 GbE NICs
- Extension for the Quality of Service (QoS) sample application to allow statistics polling
- New `libpcap`-based poll-mode driver, including support for reading from 3rd Party NICs using Linux kernel drivers
- New multi-process example using `fork()` to demonstrate application resiliency and recovery, including reattachment to and re-initialization of shared data structures where necessary
- New example (`vmdq`) to demonstrate VLAN-based packet filtering
- Improved scalability for scheduling large numbers of timers using the `rte_timer` library
- Support for building the Intel® DPDK as a shared library
- Support for Intel® Ethernet Server Bypass Adapter X520-SR2
- Poll Mode Driver support for the Intel® Ethernet Connection I354 on the Intel® Atom™ Processor C2000 Product Family SoCs
- IPv6 exact match flow classification in the `l3fwd` sample application
- Support for multiple instances of the Intel® DPDK
- Support for Intel® 82574L Gigabit Ethernet Controller - Intel® Gigabit CT Desktop Adapter (previously code named "Hartwell")
- Support for Intel® Ethernet Controller I210 (previously code named "Springville")
- Early access support for the Quad-port Intel® Ethernet Server Adapter X520-4 and X520-DA2 (code named "Spring Fountain")



- Support for Intel® X710/XL710 40 Gigabit Ethernet Controller (code named "Fortville")
- Core components:
 - `rte_mempool`: allocator for fixed-sized objects
 - `rte_ring`: single- or multi- consumer/producer queue implementation
 - `rte_timer`: implementation of timers
 - `rte_malloc`: malloc-like allocator
 - `rte_mbuf`: network packet buffers, including fragmented buffers
 - `rte_hash`: support for exact-match flow classification in software
 - `rte_lpm`: support for longest prefix match in software for IPv4 and IPv6
 - `rte_sched`: support for QoS scheduling
 - `rte_meter`: support for QoS traffic metering
 - `rte_power`: support for power management
 - `rte_ip_frag`: support for IP fragmentation and reassembly
- Poll Mode Driver - Common (`rte_ether`)
 - VLAN support
 - Support for Receive Side Scaling (RSS)
 - IEEE1588
 - Buffer chaining; Jumbo frames
 - TX checksum calculation
 - Configuration of promiscuous mode, and multicast packet receive filtering
 - L2 Mac address filtering
 - Statistics recording
- IGB Poll Mode Driver - 1 GbE Controllers (`librte_pmd_e1000`)
 - Support for Intel® 82576 Gigabit Ethernet Controller (previously code named "Kawela")
 - Support for Intel® 82580 Gigabit Ethernet Controller (previously code named "Barton Hills")
 - Support for Intel® I350 Gigabit Ethernet Controller (previously code named "Powerville")
 - Support for Intel® 82574L Gigabit Ethernet Controller - Intel® Gigabit CT Desktop Adapter (previously code named "Hartwell")
 - Support for Intel® Ethernet Controller I210 (previously code named "Springville")
 - Support for L2 Ethertype filters, SYN filters, 2-tuple filters and Flex filters for 82580 and i350
 - Support for L2 Ethertype filters, SYN filters and L3/L4 5-tuple filters for 82576
- Poll Mode Driver - 10 GbE Controllers (`librte_pmd_ixgbe`)
 - Support for Intel® 82599 10 Gigabit Ethernet Controller (previously code named "Niantic")
 - Support for Intel® Ethernet Server Adapter X520-T2 (previously code named "Iron Pond")



Supported Features

- Support for Intel® Ethernet Controller X540-T2 (previously code named “Twin Pond”)
- Support for Virtual Machine Device Queues (VMDq) and Data Center Bridging (DCB) to divide incoming traffic into 128 RX queues. DCB is also supported for transmitting packets.
- Support for auto negotiation down to 1 Gb
- Support for Flow Director
- Support for L2 Ethertype filters, SYN filters and L3/L4 5-tuple filters for 82599EB
- Poll Mode Driver - 40 GbE Controllers (librte_pmd_i40e)
 - Support for Intel® XL710 40 Gigabit Ethernet Controller
 - Support for Intel® X710 40 Gigabit Ethernet Controller
- Environment Abstraction Layer (librte_eal)
 - Multi-process support
 - Multi-thread support
 - 1 GB and 2 MB page support
 - Atomic integer operations
 - Querying CPU support of specific features
 - High Precision Event Timer support (HPET)
 - PCI device enumeration and blacklisting
 - Spin locks and R/W locks
- Test PMD application
 - Support for PMD driver testing
- Test application
 - Support for core component tests
- Sample applications
 - Command Line
 - Exception Path (into Linux* for packets using the Linux TUN/TAP driver)
 - Hello World
 - Integration with Intel® Quick Assist Technology drivers 1.0.0, 1.0.1 and 1.1.0 on Intel® Communications Chipset 89xx Series C0 and C1 silicon.
 - Link Status Interrupt (Ethernet* Link Status Detection)
 - IPv4 Fragmentation
 - IPv4 Multicast
 - IPv4 Reassembly
 - L2 Forwarding (supports virtualized and non-virtualized environments)
 - L3 Forwarding (IPv4 and IPv6)
 - L3 Forwarding in a Virtualized Environment
 - L3 Forwarding with Power Management
 - QoS Scheduling
 - QoS Metering + Dropper
 - Quota & Watermarks



- Load Balancing
- Multi-process
- Timer
- VMDQ and DCB L2 Forwarding
- Kernel NIC Interface (with ethtool support)
- Userspace vhost switch
- Interactive command line interface (`rte_cmdline`)
- Updated 10 GbE Poll Mode Driver (PMD) to the latest BSD code base providing support of newer ixgbe 10 GbE devices such as the Intel® X520-T2 server Ethernet adapter
- An API for configuring Ethernet flow control
- Support for interrupt-based Ethernet link status change detection
- Support for SR-IOV functions on the Intel® 82599, Intel® 82576 and Intel® i350 Ethernet Controllers in a virtualized environment
- Improvements to SR-IOV switch configurability on the Intel® 82599 Ethernet Controllers in a virtualized environment.
- An API for L2 Ethernet Address “whitelist” filtering
- An API for resetting statistics counters
- Support for RX L4 (UDP/TCP/SCTP) checksum validation by NIC
- Support for TX L3 (IPv4/IPv6) and L4 (UDP/TCP/SCTP) checksum calculation offloading
- Support for IPv4 packet fragmentation and reassembly
- Support for zero-copy Multicast
- New APIs to allow the “blacklisting” of specific NIC ports.
- Header files for common protocols (IP, SCTP, TCP, UDP)
- Improved multi-process application support, allowing multiple co-operating Intel® DPDK processes to access the NIC port queues directly.
- CPU-specific compiler optimization
- Improvements to the Load Balancing sample application
- The addition of a PAUSE instruction to tight loops for energy-usage and performance improvements
- Updated 10 GbE Transmit architecture incorporating new upstream PCIe* optimizations.
- IPv6 support:
 - Support in Flow Director Signature Filters and masks
 - RSS support in sample application that use RSS
 - Exact match flow classification in the L3 Forwarding sample application
 - Support in LPM for IPv6 addresses

S



4 *Supported Operating Systems*

The following Linux* distributions were successfully used to generate or run the Intel® DPDK.

- FreeBSD* 9.2
- Fedora release 18
- Ubuntu* 12.04 LTS
- Wind River* Linux* 5
- Red Hat* Enterprise Linux 6.3
- SUSE Enterprise Linux* 11 SP2

These distributions may need additional packages that are not installed by default, or a specific kernel. Refer to the *Intel® DPDK Getting Started Guide* for details.

§



5 Updating Applications from Previous Versions

Although backward compatibility is being maintained across Intel® DPDK releases, code written for previous versions of the Intel® DPDK may require some code updates to benefit from performance and user experience enhancements provided in later Intel® DPDK releases.

5.1 Intel® DPDK 1.6 to Intel® DPDK 1.7

Note the following difference between 1.6 and 1.7:

- The “default” target has been renamed to “native”

5.2 Intel® DPDK 1.5 to Intel® DPDK 1.6

Note the following difference between 1.5 and 1.6:

- The CONFIG_RTE_EAL_UNBIND_PORTS configuration option, which was deprecated in Intel® DPDK 1.4.x, has been removed in Intel® DPDK 1.6.x. Applications using the Intel® DPDK must be explicitly unbound to the `igb_uio` driver using the `dpdk_nic_bind.py` script included in the Intel® DPDK release and documented in the *Intel® DPDK Getting Started Guide*.

5.3 Intel® DPDK 1.4 to Intel® DPDK 1.5

Note the following difference between 1.4 and 1.5:

- Starting with version 1.5, the top-level directory created from unzipping the release package will now contain the release version number, that is, `DPDK-1.5.2/` rather than just `DPDK/`.

5.4 Intel® DPDK 1.3 to Intel® DPDK 1.4.x

Note the following difference between releases 1.3 and 1.4.x:

- In Release 1.4.x, Intel® DPDK applications will no longer unbind the network ports from the Linux* kernel driver when the application initializes. Instead, any ports to be used by Intel® DPDK must be unbound from the Linux driver and bound to the `igb_uio` driver before the application starts. This can be done using the `pci_unbind.py` script included with the Intel® DPDK release and documented in the *Intel® DPDK Getting Started Guide*.

If the port unbinding behavior present in previous Intel® DPDK releases is



required, this can be re-enabled using the `CONFIG_RTE_EAL_UNBIND_PORTS` setting in the appropriate Intel® DPDK compile-time configuration file.

- In Release 1.4.x, HPET support is disabled in the Intel® DPDK build configuration files, which means that the existing `rte_eal_get_hpet_hz()` and `rte_eal_get_hpet_cycles()` APIs are not available by default. For applications that require timing APIs, but not the HPET timer specifically, it is recommended that the API calls `rte_get_timer_cycles()` and `rte_get_timer_hz()` be used instead of the HPET-specific APIs. These generic APIs can work with either TSC or HPET time sources, depending on what is requested by an application, and on what is available on the system at runtime.

For more details on this and how to re-enable the HPET if it is needed, please consult the *Intel® DPDK Getting Started Guide*.

5.5 Intel® DPDK 1.2 to Intel® DPDK 1.3

Note the following difference between releases 1.2 and 1.3:

- In release 1.3, the Intel® DPDK supports two different 1 GBe drivers: `igb` and `em`. Both of them are located in the same library: `lib_pmd_e1000.a`. Therefore, the name of the library to link with for the `igb` PMD has changed from `librte_pmd_igb.a` to `librte_pmd_e1000.a`.
- The `rte_common.h` macros, `RTE_ALIGN`, `RTE_ALIGN_FLOOR` and `RTE_ALIGN_CEIL` were renamed to, `RTE_PTR_ALIGN`, `RTE_PTR_ALIGN_FLOOR` and `RTE_PTR_ALIGN_CEIL`. The original macros are still available but they have different behavior. Not updating the macros results in strange compilation errors.
- The `rte_tailq` is now defined statically. The `rte_tailq` APIs have also been changed from being public to internal use only. The old public APIs are maintained for backward compatibility reasons. Details can be found in the *Intel® DPDK API Reference*.
- The method for managing mbufs on the NIC RX rings has been modified to improve performance. To allow applications to use the newer, more optimized, code path, it is recommended that the `rx_free_thresh` field in the `rte_eth_conf` structure, which is passed to the Poll Mode Driver when initializing a network port, be set to a value of 32.

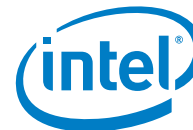
5.6 Intel® DPDK 1.1 to Intel® DPDK 1.2

Note the following difference between release 1.1 and release 1.2:

- The names of the 1G and 10G Ethernet drivers have changed between releases 1.1 and 1.2. While the old driver names still work, it is recommended that code be updated to the new names, since the old names are deprecated and may be removed in a future release.

The items affected are as follows:

- Any macros referring to `RTE_LIBRTE_82576_PMD` should be updated to refer to `RTE_LIBRTE_IGB_PMD`.



- Any macros referring to `RTE_LIBRTE_82599_PMD` should be updated to refer to `RTE_LIBRTE_IXGBE_PMD`.
- Any calls to the `rte_82576_pmd_init()` function should be replaced by calls to `rte_igb_pmd_init()`.
- Any calls to the `rte_82599_pmd_init()` function should be replaced by calls to `rte_ixgbe_pmd_init()`.
- The method used for managing mbufs on the NIC TX rings for the 10 GbE driver has been modified to improve performance. As a result, different parameter values should be passed to the `rte_eth_tx_queue_setup()` function. The recommended default values are to have `tx_thresh.tx_wthresh`, `tx_free_thresh`, as well as the new parameter `tx_rs_thresh` (all in the `struct rte_eth_txconf` datatype) set to zero. See the "Configuration of Transmit and Receive Queues" section in the *Intel® DPDK Programmer's Guide* for more details.

Note: If the `tx_free_thresh` field is set to `TX_RING_SIZE+1`, as was previously used in some cases to disable free threshold check, then an error is generated at port initialization time. To avoid this error, configure the TX threshold values as suggested above.

S



6 Known Issues and Limitations

This section describes known issues with the Intel® DPDK software, Release 1.6.0.

6.1 Pause Frame Forwarding does not work properly on igb

| Title | Pause Frame forwarding does not work properly on igb |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00384637 |
| Description | For igb devices <code>rte_eth_flow_ctrl_set</code> is not working as expected. Pause frames are always forwarded on igb, regardless of the RFCE, MPMCF and DPF registers. |
| Implication | Pause frames will never be rejected by the host on 1G NICs and they will always be forwarded. |
| Resolution/ Workaround | There is no workaround available. |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

6.2 Running TestPMD with SRIOV in Domain U may cause it to hang when XENVIRT switch is on

| Title | Running TestPMD with SRIOV in Domain U may cause it to hang when XENVIRT switch is on |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00168949 |
| Description | When TestPMD is run with only SRIOV port <code>./testpmd -c f -n 4 -- -i</code> , the following error occurs: PMD: gntalloc: ioctl error EAL: Error - exiting with code: 1 Cause: Creation of mbuf pool for socket 0 failed Then, alternately run SRIOV port and virtIO with testpmd: <code>testpmd -c f -n 4 -- -i</code> <code>testpmd -c f -n 4 --use-dev="eth_xenvirt0" -- -i</code> |
| Implication | DomU will not be accessible after you repeat this action some times |
| Resolution/ Workaround | Run testpmd with a <code>"--total-num-mbufs=N(N<=3500)"</code> |
| Affected Environment/ | Fedora 16, 64 bits + Xen hypervisor 4.2.3 + Domain 0 kernel 3.10.0 |



| | |
|---------------|----------------------------------------------------------------------------------------------|
| Title | Running TestPMD with SRIOV in Domain U may cause it to hang when XENVIRT switch is on |
| Platform | +Domain U kernel 3.6.11 |
| Driver/Module | TestPMD Sample Application |

6.3 Vhost-xen cannot detect Domain U application exit on Xen version 4.0.1

| | |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Vhost-xen cannot detect Domain U application exit on Xen 4.0.1. |
| Reference # | IXA00168947 |
| Description | When using Intel® DPDK applications on Xen 4.0.1, e.g. TestPMD Sample Application, on killing the application (e.g. killall testmd) vhost-switch cannot detect the domain U exited and does not free the Virtio device. |
| Implication | Virtio device not freed after application is killed when using vhost-switch on Xen 4.0.1 |
| Resolution | |
| Affected Environment/Platform | Xen 4.0.1 |
| Driver/Module | Vhost-switch |

6.4 Virtio incorrect header length used if MSI-X is disabled by kernel driver

| | |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Virtio incorrect header length used if MSI-X is disabled by kernel driver or if VIRTIO_NET_F_MAC is not negotiated. |
| Reference # | IXA00384256 |
| Description | <p>The Virtio header for host-guest communication is of variable length and is dependent on whether MSI-X has been enabled by the kernel driver for the network device.</p> <p>The base header length of 20 bytes will be extended by 4 bytes to accommodate MSI-X vectors and the Virtio Network Device header will appear at byte offset 24.</p> <p>The Userspace Virtio Poll Mode Driver tests the guest feature bits for the presence of VIRTIO_PCI_FLAG_MISIX, however this bit field is not part of the Virtio specification and resolves to the VIRTIO_NET_F_MAC feature instead.</p> |



Known Issues and Limitations

| Title | Virtio incorrect header length used if MSI-X is disabled by kernel driver or if VIRTIO_NET_F_MAC is not negotiated. |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Implication | The Intel® DPDK kernel driver will enable MSI-X by default, however if loaded with "intr_mode=legacy" on a guest with a Virtio Network Device, a KVM-Qemu guest may crash with the following error: "virtio-net header not in first element". If VIRTIO_NET_F_MAC feature has not been negotiated, then the Userspace Poll Mode Driver will assume that MSI-X has been disabled and will prevent the proper functioning of the driver. |
| Resolution | Ensure #define VIRTIO_PCI_CONFIG(hw) returns the correct offset (20 or 24 bytes) for the devices used, where in rare cases MSI-X is disabled or VIRTIO_NET_F_MAC has not been negotiated. |
| Affected Environment/ Platform | Virtio devices where MSI-X is disabled or VIRTIO_NET_F_MAC feature has not been negotiated. |
| Driver/Module | librte_pmd_virtio |

6.5 gmake clean may silently fail for some example applications

| Title | When using FreeBSD* 9.2 gmake clean may silently fail for some sample applications |
|-----------------------------------|-------------------------------------------------------------------------------------------|
| Reference # | IXA00834605 |
| Description | gmake clean may silently fail leaving the source object files intact. |
| Implication | Application object files are not removed. |
| Resolution | Manually remove ./build folders or rebuild application source after editing. |
| Affected Environment/ Platform | FreeBSD* 9.2 and below |
| Driver/Module | Example Applications |

6.6 In packets provided by the PMD, some flags are missing

| Title | In packets provided by the PMD, some flags are missing |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 3 |
| Description | In packets provided by the PMD, some flags are missing. The application does not have access to information provided by the hardware (packet is broadcast, packet is multicast, packet is IPv4 and so on). |
| Implication | The "ol_flags" field in the "rte_mbuf" structure is not correct and |



| | |
|-----------------------------------|--------------------------------------------------------------------------------------------------|
| Title | In packets provided by the PMD, some flags are missing |
| | should not be used. |
| Resolution | The application has to parse the Ethernet header itself to get the information, which is slower. |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

6.7 The rte_malloc library is not fully implemented

| | |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------|
| Title | The rte_malloc library is not fully implemented |
| Reference # | 6 |
| Description | The rte_malloc library is not fully implemented. |
| Implication | All debugging features of rte_malloc library described in architecture documentation are not yet implemented. |
| Resolution | No workaround available. |
| Affected Environment/ Platform | All |
| Driver/Module | rte_malloc |

6.8 HPET reading is slow

| | |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | HPET reading is slow |
| Reference # | 7 |
| Description | Reading the HPET chip is slow. |
| Implication | An application that calls "rte_get_hpet_cycles()" or "rte_timer_manage()" runs slower. |
| Resolution | The application should not call these functions too often in the main loop. An alternative is to use the TSC register through "rte_rdtsc()" which is faster, but specific to an lcore and is a cycle reference, not a time reference. |
| Affected Environment/ Platform | All |
| Driver/Module | Environment Abstraction Layer (EAL) |



6.9 HPET timers do not work on the Osage customer reference platform

| Title | HPET timers do not work on the Osage customer reference platform |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 17 |
| Description | HPET timers do not work on the Osage customer reference platform (which includes an Intel® Xeon® processor 5500 series processor) using the released BIOS from Intel. |
| Implication | On Osage boards, the implementation of the "rte_delay_us()" function must be changed to not use the HPET timer. |
| Resolution | This can be addressed by building the system with the "CONFIG_RTE_LIBEAL_USE_HPET=n" configuration option or by using the --no-hpet EAL option. |
| Affected Environment/Platform | The Osage customer reference platform. Other vendor platforms with Intel® Xeon® processor 5500 series processors should work correctly, provided the BIOS supports HPET. |
| Driver/Module | lib/librte_eal/common/include/rte_cycles.h |

6.10 Not all variants of supported NIC types have been used in testing

| Title | Not all variants of supported NIC types have been used in testing |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 28 |
| Description | <p>The supported network interface cards can come in a number of variants with different device ID's. Not all of these variants have been tested with the Intel® DPDK.</p> <p>The NIC device identifiers used during testing:</p> <ul style="list-style-type: none">• Intel® 82576 Gigabit Ethernet Controller [8086:10c9]• Intel® 82576 Quad Copper Gigabit Ethernet Controller [8086:10e8]• Intel® 82580 Dual Copper Gigabit Ethernet Controller [8086:150e]• Intel® I350 Quad Copper Gigabit Ethernet Controller [8086:1521]• Intel® 82599 Dual Fibre 10 Gigabit Ethernet Controller [8086:10fb]• Intel® Ethernet Server Adapter X520-T2 [8086: 151c]• Intel® Ethernet Controller X540-T2 [8086:1528]• Intel® 82574L Gigabit Network Connection [8086:10d3]• Emulated Intel® 82540EM Gigabit Ethernet Controller |



| | |
|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Not all variants of supported NIC types have been used in testing |
| | [8086:100e] <ul style="list-style-type: none"> Emulated Intel® 82545EM Gigabit Ethernet Controller [8086:100f] Intel® Ethernet Server Adapter X520-4 [8086:154a] Intel® Ethernet Controller I210 [8086:1533] |
| Implication | Risk of issues with untested variants. |
| Resolution | Use tested NIC variants. For those supported Ethernet controllers, additional device IDs may be added to the software if required. |
| Affected Environment/Platform | All |
| Driver/Module | Poll-mode drivers |

6.11 Multi-process sample app requires exact memory mapping

| | |
|-------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Multi-process sample app requires exact memory mapping |
| Reference # | 30 |
| Description | The multi-process example application assumes that it is possible to map the hugepage memory to the same virtual addresses in client and server applications. Occasionally, very rarely with 64-bit, this does not occur and a client application will fail on startup. The Linux "address-space layout randomization" security feature can sometimes cause this to occur. |
| Implication | A multi-process client application fails to initialize. |
| Resolution | See the "Multi-process Limitations" section in the Intel® DPDK Programmer's Guide for more information. |
| Affected Environment/Platform | All |
| Driver/Module | Multi-process example application |

6.12 Unstable system performance across application executions with 2MB pages

| | |
|--------------|---------------------------------------------------------------------------------|
| Title | Unstable system performance across application executions with 2MB pages |
| Reference # | IXA00372346 |



Known Issues and Limitations

| | |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Unstable system performance across application executions with 2MB pages |
| Description | The performance of an Intel® DPDK application may vary across executions of an application due to a varying number of TLB misses depending on the location of accessed structures in memory. This situation occurs on rare occasions. |
| Implication | Occasionally, relatively poor performance of Intel® DPDK applications is encountered. |
| Resolution/ Workaround | Using 1 GB pages results in lower usage of TLB entries, resolving this issue. |
| Affected Environment/ Platform | Systems using 2 MB pages |
| Driver/Module | All |

6.13 Packets are not sent by the 1 GbE/10 GbE SR-IOV driver when the source MAC address is not the MAC address assigned to the VF NIC

| | |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Packets are not sent by the 1 GbE/10 GbE SR-IOV driver when the source MAC address is not the MAC address assigned to the VF NIC |
| Reference # | IXA00168379 |
| Description | The 1 GbE/10 GbE SR-IOV driver can only send packets when the Ethernet header's source MAC address is the same as that of the VF NIC. The reason for this is that the Linux "ixgbe" driver module in the host OS has its anti-spoofing feature enabled. |
| Implication | Packets sent using the 1 GbE/10 GbE SR-IOV driver must have the source MAC address correctly set to that of the VF NIC. Packets with other source address values are dropped by the NIC if the application attempts to transmit them. |
| Resolution/ Workaround | Configure the Ethernet source address in each packet to match that of the VF NIC. |
| Affected Environment/ Platform | All |
| Driver/Module | 1 GbE/10 GbE VF Poll Mode Driver (PMD) |

6.14 SR-IOV drivers do not fully implement the rte_ethdev API

| | |
|--------------|-----------------------------------------------------------------|
| Title | SR-IOV drivers do not fully implement the rte_ethdev API |
| Reference # | 59 |



| Title | SR-IOV drivers do not fully implement the rte_ethdev API |
|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Description | <p>The SR-IOV drivers only supports the following rte_ethdev API functions:</p> <ul style="list-style-type: none"> • rte_eth_dev_configure() • rte_eth_tx_queue_setup() • rte_eth_rx_queue_setup() • rte_eth_dev_info_get() • rte_eth_dev_start() • rte_eth_tx_burst() • rte_eth_rx_burst() • rte_eth_dev_stop() • rte_eth_stats_get() • rte_eth_stats_reset() • rte_eth_link_get() • rte_eth_link_get_no_wait() |
| Implication | Calling an unsupported function will result in an application error. |
| Resolution/ Workaround | Do not use other rte_ethdev API functions in applications that use the SR-IOV drivers. |
| Affected Environment/ Platform | All |
| Driver/Module | VF Poll Mode Driver (PMD) |

6.15 PMD does not work with --no-huge EAL command line parameter

| Title | PMD does not work with --no-huge EAL command line parameter |
|--------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00373461 |
| Description | Currently, the Intel® DPDK does not store any information about memory allocated by malloc() (for example, NUMA node, physical address), hence PMD drivers do not work when the --no-huge command line parameter is supplied to EAL. |
| Implication | Sending and receiving data with PMD will not work. |
| Resolution/ Workaround | Use huge page memory or use VFIO to map devices. |
| Affected Environment/ Platform | Systems running the Intel® DPDK on Linux |
| Driver/Module | Poll Mode Driver (PMD) |



6.16 Some hardware off-load functions are not supported by the VF Driver

| Title | Some hardware off-load functions are not supported by the VF Driver |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00378813 |
| Description | Currently, configuration of the following items is not supported by the VF driver: <ul style="list-style-type: none">• IP/UDP/TCP checksum offload• Jumbo Frame Receipt• HW Strip CRC |
| Implication | Any configuration for these items in the VF register will be ignored. The behavior is dependant on the current PF setting. |
| Resolution/ Workaround | For the PF (Physical Function) status on which the VF driver depends, there is an option item under PMD in the config file. For others, the VF will keep the same behavior as PF setting. |
| Affected Environment/ Platform | All |
| Driver/Module | VF (SR-IOV) Poll Mode Driver (PMD) |

6.17 Kernel crash on IGB port unbinding

| Title | Kernel crash on IGB port unbinding |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| Reference # | 74 |
| Description | Kernel crash may occur when unbinding 1G ports from the igb_uio driver, on 2.6.3x kernels such as shipped with Fedora 14. |
| Implication | Kernel crash occurs. |
| Resolution/ Workaround | Use newer kernels or do not unbind ports. |
| Affected Environment/ Platform | 2.6.3x kernels such as shipped with Fedora 14 |
| Driver/Module | IGB Poll Mode Driver (PMD) |

6.18 Link status change not working with MSI interrupts

| Title | Link status change not working with MSI interrupts |
|-------------|----------------------------------------------------|
| Reference # | IXA00378191 |



| Title | Link status change not working with MSI interrupts |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Description | MSI interrupts are not supported by the PMD. |
| Implication | Link status change will only work with legacy or MSI-X interrupts. |
| Resolution/ Workaround | The igb_uio driver can now be loaded with either legacy or MSI-X interrupt support. However, this configuration is not tested. |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

6.19 Twinpond and Ironpond NICs do not report link status correctly

| Title | Twinpond and Ironpond NICs do not report link status correctly |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00378800 |
| Description | Twin Pond/Iron Pond NICs do not bring the physical link down when shutting down the port. |
| Implication | The link is reported as up even after issuing "shutdown" command unless the cable is physically disconnected. |
| Resolution/ Workaround | None. |
| Affected Environment/ Platform | Twin Pond and Iron Pond NICs |
| Driver/Module | Poll Mode Driver (PMD) |

6.20 Discrepancies between statistics reported by different NICs

| Title | Discrepancies between statistics reported by different NICs |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00378113 |
| Description | Gigabit Ethernet devices from Intel include CRC bytes when calculating packet reception statistics regardless of hardware CRC stripping state, while 10-Gigabit Ethernet devices from Intel do so only when hardware CRC stripping is disabled. |
| Implication | There may be a discrepancy in how different NICs display packet reception statistics. |
| Resolution/ Workaround | None |
| Affected Environment/ Platform | All |



Known Issues and Limitations

| Title | Discrepancies between statistics reported by different NICs |
|---------------|-------------------------------------------------------------|
| Driver/Module | Poll Mode Driver (PMD) |

6.21 Error reported opening files on Intel® DPDK initialization

| Title | Error reported opening files on Intel® DPDK initialization |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 91 |
| Description | On Intel® DPDK application startup, errors may be reported when opening files as part of the initialization process. This occurs if a large number, for example, 500 or more, or if hugepages are used, due to the per-process limit on the number of open files. |
| Implication | The Intel® DPDK application may fail to run. |
| Resolution/ Workaround | If using 2 MB hugepages, consider switching to a fewer number of 1 GB pages. Alternatively, use the "ulimit" command to increase the number of files which can be opened by a process. |
| Affected Environment/ Platform | All |
| Driver/Module | Environment Abstraction Layer (EAL) |

6.22 Intel® QuickAssist Technology sample application does not work on a 32-bit OS on Shumway

| Title | Intel® QuickAssist Technology sample applications does not work on a 32-bit OS on Shumway |
|--------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 93 |
| Description | The Intel® Communications Chipset 89xx Series device does not fully support NUMA on a 32-bit OS. Consequently, the sample application cannot work properly on Shumway, since it requires NUMA on both nodes. |
| Implication | The sample application cannot work in 32-bit mode with emulated NUMA, on multi-socket boards. |
| Resolution/ Workaround | There is no workaround available. |
| Affected Environment/ Platform | Shumway |
| Driver/Module | All |



6.23 IEEE1588 support possibly not working with an Intel® Ethernet Controller I210 NIC

| | |
|--------------------------------|------------------------------------------------------------------------------------------|
| Title | IEEE1588 support may not work with an Intel® Ethernet Controller I210 NIC |
| Reference # | IXA00380285 |
| Description | IEEE1588 support is not working with an Intel® Ethernet Controller I210 NIC. |
| Implication | IEEE1588 packets are not forwarded correctly by the Intel® Ethernet Controller I210 NIC. |
| Resolution/ Workaround | There is no workaround available. |
| Affected Environment/ Platform | All |
| Driver/Module | IGB Poll Mode Driver |

6.24 Differences in how different Intel NICs handle maximum packet length for jumbo frame

| | |
|--------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Differences in how different Intel NICs handle maximum packet length for jumbo frame |
| Reference # | 96 |
| Description | 10 Gigabit Ethernet devices from Intel do not take VLAN tags into account when calculating packet size while Gigabit Ethernet devices do so for jumbo frames. |
| Implication | When receiving packets with VLAN tags, the actual maximum size of useful payload that Intel Gigabit Ethernet devices are able to receive is 4 bytes (or 8 bytes in the case of packets with extended VLAN tags) less than that of Intel 10 Gigabit Ethernet devices. |
| Resolution/ Workaround | Increase the configured maximum packet size when using Intel Gigabit Ethernet devices. |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |



6.25 Link status interrupt not working in VF drivers

| Title | Link status interrupts not working in the VF drivers |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00381312 |
| Description | Due to the driver not setting up interrupts for VF drivers, the NIC does not report link status change to VF devices. |
| Implication | Link status interrupts will not work in VM guests. |
| Resolution/ Workaround | There is no workaround available. |
| Affected Environment/ Platform | All |
| Driver/Module | VF (SR-IOV) Poll Mode Driver (PMD) |

6.26 Binding PCI devices to igb_uio fails on Linux* kernel 3.9 when more than one device is used

| Title | Binding PCI devices to igb_uio fails on Linux* kernel 3.9 when more than one device is used |
|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 97 |
| Description | A known bug in the uio driver included in Linux* kernel version 3.9 prevents more than one PCI device to be bound to the igb_uio driver. |
| Implication | The Poll Mode Driver (PMD) will crash on initialization. |
| Resolution/ Workaround | Use earlier or later kernel versions, or apply the following patch . |
| Affected Environment/ Platform | Linux* systems with kernel version 3.9 |
| Driver/Module | igb_uio module |

6.27 GCC might generate Intel® AVX instructions for processors without Intel® AVX support

| Title | Gcc might generate Intel® AVX instructions for processors without Intel® AVX support |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00382439 |
| Description | When compiling Intel® DPDK (and any Intel® DPDK app), gcc may generate Intel® AVX instructions, even when the processor does not support Intel® AVX. |
| Implication | Any Intel® DPDK app might crash while starting up. |



| | |
|--------------------------------|---------------------------------------------------------------------------------------------|
| Title | Gcc might generate Intel® AVX instructions for processors without Intel® AVX support |
| Resolution/ Workaround | Either compile using icc or set EXTRA_CFLAGS='-O3' prior to compilation. |
| Affected Environment/ Platform | Platforms which processor does not support Intel® AVX. |
| Driver/Module | Environment Abstraction Layer (EAL) |

6.28 Ethertype filter could receive other packets (non-assigned) in Niantic

| | |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Ethertype filter could receive other packets (non-assigned) in Niantic |
| Reference # | IXA00169017 |
| Description | <p>On Intel® Ethernet Controller 82599EB:</p> <p>When Ethertype filter (priority enable) was set, unmatched packets also could be received on the assigned queue, such as ARP packets without 802.1q tags or with the user priority not equal to set value.</p> <p>Launch the testpmd by disabling RSS and with multiply queues, then add the ethertype filter like: "add_ethertype_filter 0 ethertype 0x0806 priority enable 3 queue 2 index 1", and then start forwarding.</p> <p>When sending ARP packets without 802.1q tag and with user priority as non-3 by tester, all the ARP packets can be received on the assigned queue.</p> |
| Implication | The user priority comparing in Ethertype filter cannot work probably. It is the NIC's issue due to the response from PAE: "In fact, ETQF.UP is not functional, and the information will be added in errata of 82599 and X540." |
| Resolution/ Workaround | None |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

6.29 Double VLAN does not work on Intel® 40G ethernet controller

| | |
|--------------|--------------------------------------------------------------------|
| Title | Double VLAN does not work on Intel® 40G ethernet controller |
| Reference # | IXA00386480 |



Known Issues and Limitations

| Title | Double VLAN does not work on Intel® 40G ethernet controller |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Description | On Intel® 40G Ethernet Controller: Double VLAN does not work. This was confirmed a firmware issue which will be fixed in later versions of firmware. |
| Implication | After setting double vlan to be enabled on a port, no packets can be transmitted out on that port. |
| Resolution/ Workaround | None |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

6.30 Cannot set link speed on Intel® 40G ethernet controller

| Title | Cannot set link speed on Intel® 40G ethernet controller |
|--------------------------------|----------------------------------------------------------------------------------------|
| Reference # | IXA00386379 |
| Description | On Intel® 40G Ethernet Controller: It cannot set the link to specific speed. |
| Implication | The link speed cannot be changed forcedly, though it can be configured by application. |
| Resolution/ Workaround | None |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

6.31 Stopping the port does not down the link on Intel® 40G ethernet controller

| Title | Stopping the port does not down the link on Intel® 40G ethernet controller |
|------------------------|---------------------------------------------------------------------------------------------|
| Reference # | IXA00386380 |
| Description | On Intel® 40G Ethernet Controller: Stopping the port does not really down the port link. |
| Implication | The port link will be still up after stopping the port. |
| Resolution/ Workaround | None |
| Affected Environment/ | All |



| | |
|---------------|-----------------------------------------------------------------------------------|
| Title | Stopping the port does not down the link on Intel® 40G ethernet controller |
| Platform | |
| Driver/Module | Poll Mode Driver (PMD) |

§



7 Resolved Issues

This section describes previously known issues that have been resolved since release version 1.2.

7.1 KNI does not provide Ethtool support for all NICs supported by the Poll-Mode Drivers

| | |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | KNI does not provide ethtool support for all NICs supported by the Poll Mode Drivers |
| Reference # | IXA00383835 |
| Description | To support ethtool functionality using the KNI, the KNI library includes separate driver code based of the linux kernel drivers. Because this driver code is separate from the driver code for the poll-mode drivers, the set of supported NICs for these two components may differ. Because of this, in this release, the KNI driver does not provide "ethtool" support for the Intel® Ethernet Connection I354 on the Intel Atom Processor C2000 Product Family SoCs. |
| Implication | Ethtool support with KNI will not work for NICs such as the Intel® Ethernet Connection I354. Other KNI functionality, such as injecting packets into the Linux kernel is unaffected. |
| Resolution/ Workaround | Updated for Intel® Ethernet Connection I354. |
| Affected Environment/ Platform | Platforms using the Intel® Ethernet Connection I354 or other NICs unsupported by KNI ethtool |
| Driver/Module | KNI |

7.2 Linux IPv4 forwarding is not stable with vhost-switch on high packet rate

| | |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Linux IPv4 forwarding is not stable with vhost-switch on high packet rate. |
| Reference # | IXA00384430 |
| Description | Linux IPv4 forwarding is not stable in Guest when Tx traffic is high from traffic generator using two virtio devices in VM with 10G in host. |
| Implication | Packets cannot be forwarded by userspace vhost-switch and Linux IPv4 forwarding if the rate of incoming packets is greater than 1Mpps. |



| | |
|--------------------------------------|-----------------------------------------------------------------------------------|
| Title | Linux IPv4 forwarding is not stable with vhost-switch on high packet rate. |
| Resolution/ Workaround | N/A |
| Affected Environment/ Platform | All |
| Driver/Module | Sample application |

7.3 PCAP library overwrites mbuf data before data is used

| | |
|--------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | PCAP library overwrites mbuf data before data is used |
| Reference # | IXA00383976 |
| Description | PCAP library allocates 64 mbufs for reading packets from PCAP file, but declares them as static and reuses the same mbufs repeatedly rather than handing off to the ring for allocation of new mbuf for each read from the PCAP file. |
| Implication | In multi-threaded applications data in the mbuf is overwritten. |
| Resolution/ Workaround | Fixed in eth_pcap_rx() in rte_eth_pcap.c |
| Affected Environment/ Platform | All |
| Driver/Module | Multi-threaded applications using PCAP library |

7.4 MP Client Example app - flushing part of TX is not working for some ports if set specific port mask with skipped ports

| | |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | MP Client Example app - flushing part of TX is not working for some ports if set specific port mask with skipped ports |
| Reference # | 52 |
| Description | When ports not in a consecutive set, for example, ports other than ports 0,1 or 0,1,2,3, are used with the client-server sample app, when no further packets are received by a client, the application may not flush correctly any unsent packets already buffered inside it. |
| Implication | Not all buffered packets are transmitted if traffic to the client application is stopped. While traffic is continually received for transmission on a port by a client, buffer flushing happens normally. |



Resolved Issues

| | |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Title | MP Client Example app - flushing part of TX is not working for some ports if set specific port mask with skipped ports |
| Resolution/ Workaround | Changed line 284 of the client.c file: from "send_packets(port);" to "send_packets(ports->id[port]);" |
| Affected Environment/ Platform | All |
| Driver/Module | Client-Server Multi-process Sample application |

7.5 Packet truncation with Intel® I350 Gigabit Ethernet Controller

| | |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Packet truncation with Intel I350 Gigabit Ethernet Controller |
| Reference # | IXA00372461 |
| Description | The setting of the hw_strip_crc field in the rte_eth_conf structure passed to the rte_eth_dev_configure() function is not respected and hardware CRC stripping is always enabled. If the field is set to 0, then the software also tries to strip the CRC, resulting in packet truncation. |
| Implication | The last 4 bytes of the packets received will be missing. |
| Resolution/ Workaround | Fixed an omission in device initialization (setting the STRCRC bit in the DVMOLR register) to respect the CRC stripping selection correctly. |
| Affected Environment/ Platform | Systems using the Intel® I350 Gigabit Ethernet Controller |
| Driver/Module | 1 GbE Poll Mode Driver (PMD) |

7.6 Device initialization failure with Intel® Ethernet Server Adapter X520-T2

| | |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Device initialization failure with Intel® Ethernet Server Adapter X520-T2 |
| Reference # | 55 |
| Description | If this device is bound to the Linux kernel IXGBE driver when the Intel® DPDK is initialized, the device initialization fails with error code -17 "IXGBE_ERR_PHY_ADDR_INVALID". |
| Implication | The device is not initialized and cannot be used by an application. |
| Resolution/ Workaround | Introduced a small delay in device initialization to allow Intel® DPDK to always find the device. |
| Affected Environment/ | Systems using the Intel® Ethernet Server Adapter X520-T2 |



| | |
|---------------|----------------------------------------------------------------------------------|
| Title | Device initialization failure with Intel® Ethernet Server Adapter X520-T2 |
| Platform | |
| Driver/Module | 10 GbE Poll Mode Driver (PMD) |

7.7 Intel® DPDK kernel module is incompatible with Linux kernel version 3.3

| | |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Intel® DPDK kernel module is incompatible with Linux kernel version 3.3 |
| Reference # | IXA00373232 |
| Description | The igb_uio kernel module fails to compile on systems with Linux kernel version 3.3 due to API changes in kernel headers. |
| Implication | The compilation fails and Ethernet controllers fail to initialize without the igb_uio module. |
| Resolution/Workaround | Kernel functions pci_block_user_cfg_access()/pci_cfg_access_lock() and pci_unblock_user_cfg_access()/pci_cfg_access_unlock() are automatically selected at compile time as appropriate. |
| Affected Environment/Platform | Linux systems using kernel version 3.3 or later |
| Driver/Module | UIO module |

7.8 Initialization failure with Intel® Ethernet Controller X540- T2

| | |
|-------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Title | Initialization failure with Intel® Ethernet Controller X540-T2 |
| Reference # | 57 |
| Description | This device causes a failure during initialization when the software tries to read the part number from the device EEPROM. |
| Implication | Device cannot be used. |
| Resolution/Workaround | Removed unnecessary check of the PBA number for the device. |
| Affected Environment/Platform | Systems using the Intel® Ethernet Controller X540-T2 |
| Driver/Module | 10 GbE Poll Mode Driver (PMD) |



7.9 **rte_eth_dev_stop() function does not bring down the link for 1 GB NIC ports**

| | |
|--------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Title | rte_eth_dev_stop() function does not bring down the link for 1 GB NIC ports |
| Reference # | IXA00373183 |
| Description | When the rte_eth_dev_stop() function is used to stop a NIC port, the link is not brought down for that port. |
| Implication | Links are still reported as up, even though the NIC device has been stopped and cannot perform TX or RX operations on that port. |
| Resolution | The rte_eth_dev_stop() function now brings down the link when called. |
| Affected Environment/ Platform | All |
| Driver/Module | 1 GbE Poll Mode Driver (PMD) |

7.10 **It is not possible to adjust the duplex setting for 1GB NIC ports**

| | |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | It is not possible to adjust the duplex setting for 1 GB NIC ports |
| Reference # | 66 |
| Description | The rte_eth_conf structure does not have a parameter that allows a port to be set to half-duplex instead of full-duplex mode, therefore, 1 GB NICs cannot be configured explicitly to a full- or half-duplex value. |
| Implication | 1 GB port duplex capability cannot be set manually. |
| Resolution | The PMD now uses a new field added to the rte_eth_conf structure to allow 1 GB ports to be configured explicitly as half- or full-duplex. |
| Affected Environment/ Platform | All |
| Driver/Module | 1 GbE Poll Mode Driver (PMD) |



7.11 Calling `rte_eth_dev_stop()` on a port does not free all the mbufs in use by that port

| | |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Calling <code>rte_eth_dev_stop()</code> on a port does not free all the mbufs in use by that port |
| Reference # | 67 |
| Description | The <code>rte_eth_dev_stop()</code> function initially frees all mbufs used by that port's RX and TX rings, but subsequently repopulates the RX ring again later in the function. |
| Implication | Not all mbufs used by a port are freed when the port is stopped. |
| Resolution | The driver no longer re-populates the RX ring in the <code>rte_eth_dev_stop()</code> function. |
| Affected Environment/ Platform | All |
| Driver/Module | IGB and IXGBE Poll Mode Drivers (PMDs) |

7.12 PMD does not always create rings that are properly aligned in memory

| | |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | PMD does not always create rings that are properly aligned in memory |
| Reference # | IXA00373158 |
| Description | The NIC hardware used by the PMD requires that the RX and TX rings used must be aligned in memory on a 128-byte boundary. The memzone reservation function used inside the PMD only guarantees that the rings are aligned on a 64-byte boundary, so errors can occur if the rings are not aligned on a 128-byte boundary. |
| Implication | Unintended overwriting of memory can occur and PMD behavior may also be effected. |
| Resolution | A new <code>rte_memzone_reserve_aligned()</code> API has been added to allow memory reservations from hugepage memory at alignments other than 64-bytes. The PMD has been modified so that the rings are allocated using this API with minimum alignment of 128-bytes. |
| Affected Environment/ Platform | All |
| Driver/Module | IGB and IXGBE Poll Mode Drivers (PMDs) |



7.13 Checksum offload might not work correctly when mixing VLAN-tagged and ordinary packets

| Title | Checksum offload might not work correctly when mixing VLAN-tagged and ordinary packets |
|---------------------------|----------------------------------------------------------------------------------------|
| Reference # | IXA00378372 |
| Description | Incorrect handling of protocol header lengths in the PMD driver |
| Implication | The checksum for one of the packets may be incorrect. |
| Resolution/ Workaround | Corrected the offset calculation. |
| Affected Environment/ | All |
| Driver/Module | Poll Mode Driver (PMD) |

7.14 Port not found issue with Intel® 82580 Gigabit Ethernet Controller

| Title | Port not found issue with Intel® 82580 Gigabit Ethernet Controller |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | 50 |
| Description | After going through multiple driver unbind/bind cycles, an Intel® 82580 Ethernet Controller port may no longer be found and initialized by the Intel® DPDK. |
| Implication | The port will be unusable. |
| Resolution/ Workaround | Issue was not reproducible and therefore no longer considered an issue. |
| Affected Environment/ Platform | All |
| Driver/Module | 1 GbE Poll Mode Driver (PMD) |

7.15 Packet mbufs may be leaked from mempool if rte_eth_dev_start() function fails

| Title | Packet mbufs may be leaked from mempool if rte_eth_dev_start() function fails |
|-------------|-------------------------------------------------------------------------------|
| Reference # | IXA00373373 |
| Description | The rte_eth_dev_start() function allocates mbufs to populate the NIC |



| | |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Title | Packet mbufs may be leaked from mempool if rte_eth_dev_start() function fails |
| | RX rings. If the start function subsequently fails, these mbufs are not freed back to the memory pool from which they came. |
| Implication | mbufs may be lost to the system if rte_eth_dev_start() fails and the application does not terminate. |
| Resolution/ Workaround | mbufs are correctly deallocated if a call to rte_eth_dev_start() fails. |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

7.16 Promiscuous mode for 82580 NICs can only be enabled after a call to rte_eth_dev_start for a port

| | |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Promiscuous mode for 82580 NICs can only be enabled after a call to rte_eth_dev_start for a port |
| Reference # | IXA00373833 |
| Description | For 82580-based network ports, the rte_eth_dev_start() function can overwrite the setting of the promiscuous mode for the device. Therefore, the rte_eth_promiscuous_enable() API call should be called after rte_eth_dev_start() for these devices. |
| Implication | Promiscuous mode can only be enabled if API calls are in a specific order. |
| Resolution/ Workaround | The NIC now restores most of its configuration after a call to rte_eth_dev_start(). |
| Affected Environment/ Platform | All |
| Driver/Module | Poll Mode Driver (PMD) |

7.17 Incorrect CPU socket information reported in /proc/cpuinfo can prevent the Intel® DPDK from running

| | |
|--------------|------------------------------------------------------------------------------------------------------------|
| Title | Incorrect CPU socket information reported in /proc/cpuinfo can prevent the Intel® DPDK from running |
| Reference # | 63 |



Resolved Issues

| | |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Description | The Intel® DPDK uses information supplied by the Linux kernel to determine the hardware properties of the system being used. On rare occasions, information supplied by /proc/cpuinfo does not match that reported elsewhere. In some cases, it has been observed that the CPU socket numbering given in /proc/cpuinfo is incorrect and this can prevent Intel® DPDK from operating. |
| Implication | The Intel® DPDK cannot run on systems where /proc/cpuinfo does not report the correct CPU socket topology. |
| Resolution/ Workaround | CPU socket information is now read from /sys/devices/cpu/cpuN/topology |
| Affected Environment/ Platform | All |
| Driver/Module | Environment Abstraction Layer (EAL) |

7.18 L3FWD sample application may fail to transmit packets under extreme conditions

| | |
|--------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| Title | L3FWD sample application may fail to transmit packets under extreme conditions |
| Reference # | IXA00372919 |
| Description | Under very heavy load, the L3 Forwarding sample application may fail to transmit packets due to the system running out of free mbufs. |
| Implication | Sending and receiving data with the PMD may fail. |
| Resolution/ Workaround | The number of mbufs is now calculated based on application parameters. |
| Affected Environment/ Platform | All |
| Driver/Module | L3 Forwarding sample application |

7.19 L3FWD-VF might lose CRC bytes

| | |
|--------------|---------------------------------------------------------------------------------------|
| Title | L3FWD-VF might lose CRC bytes |
| Reference # | IXA00373424 |
| Description | Currently, the CRC stripping configuration does not affect the VF driver. |
| Implication | Packets transmitted by the Intel® DPDK in the VM may be lacking 4 bytes (packet CRC). |
| Resolution/ | Set "strip_crc" to 1 in the sample applications that use the VF PMD. |



| | |
|-----------------------------------|-------------------------------------------|
| Title | L3FWD-VF might lose CRC bytes |
| Workaround | |
| Affected Environment/ Platform | All |
| Driver/Module | IGB and IXGBE VF Poll Mode Drivers (PMDs) |

7.20 32-bit Intel® DPDK sample applications fails when using more than one 1 GB hugepage

| | |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | 32-bit Intel® DPDK sample applications fails when using more than one 1 GB hugepage |
| Reference # | 31 |
| Description | 32-bit applications may have problems when running with multiple 1 GB pages on a 64-bit OS. This is due to the limited address space available to 32-bit processes. |
| Implication | 32-bit processes need to use either 2 MB pages or have their memory use constrained to 1 GB if using 1 GB pages. |
| Resolution | EAL now limits virtual memory to 1 GB per page size. |
| Affected Environment/ Platform | 64-bit systems running 32-bit Intel® DPDK with 1 GB hugepages |
| Driver/Module | Environment Abstraction Layer (EAL) |

7.21 I2fwd fails to launch if the NIC is the Intel® 82571EB Gigabit Ethernet Controller

| | |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| Title | I2fwd fails to launch if the NIC is the Intel® 82571EB Gigabit Ethernet Controller |
| Reference # | IXA00373340 |
| Description | The 82571EB NIC can handle only one TX per port. The original implementation allowed for a more complex handling of multiple queues per port. |
| Implication | The I2fwd application fails to launch if the NIC is 82571EB. |
| Resolution | I2fwd now uses only one TX queue. |
| Affected Environment/ Platform | All |
| Driver/Module | Sample Application |



7.22 32-bit Intel® DPDK applications may fail to initialize on 64-bit OS

| | |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | 32-bit Intel® DPDK applications may fail to initialize on 64-bit OS |
| Reference # | IXA00378513 |
| Description | The EAL used a 32-bit pointer to deal with physical addresses. This could create problems when the physical address of a hugepage exceeds the 4 GB limit. |
| Implication | 32-bit applications may not initialize on a 64-bit OS. |
| Resolution/ Workaround | The physical address pointer is now 64-bit. |
| Affected Environment/ Platform | 32-bit applications in a 64-bit Linux* environment |
| Driver/Module | Environment Abstraction Layer (EAL) |

7.23 Lpm issue when using prefixes >24

| | |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Title | Lpm issue when using prefixes >24 |
| Reference # | IXA00378395 |
| Description | Extended tbl8's are overwritten by multiple lpm rule entries when the depth is greater than 24. |
| Implication | LPM tbl8 entries removed by additional rules. |
| Resolution/ Workaround | Adding tbl8 entries to a valid group to avoid making the entire table invalid and subsequently overwritten. |
| Affected Environment/ Platform | All |
| Driver/Module | Sample applications |

7.24 IXGBE PMD hangs on port shutdown when not all packets have been sent

| | |
|-------------|----------------------------------------------------------------------|
| Title | IXGBE PMD hangs on port shutdown when not all packets have been sent |
| Reference # | IXA00373492 |
| Description | When the PMD is forwarding packets, and the link goes down, and port |



| | |
|--------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Title | IXGBE PMD hangs on port shutdown when not all packets have been sent |
| | shutdown is called, the port cannot shutdown. Instead, it hangs due to the IXGBE driver incorrectly performing the port shutdown procedure. |
| Implication | The port cannot shutdown and does not come back up until re-initialized. |
| Resolution/ Workaround | The port shutdown procedure has been rewritten. |
| Affected Environment/ Platform | All |
| Driver/Module | IXGBE Poll Mode Driver (PMD) |

7.25 Config file change can cause build to fail

| | |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Config file change can cause build to fail |
| Reference # | IXA00369247 |
| Description | If a change in a config file results in some Intel® DPDK files that were needed no longer being needed, the build will fail. This is because the *.o file will still exist, and the linker will try to link it. |
| Implication | Intel® DPDK compilation failure |
| Resolution | The Makefile now provides instructions to clean out old kernel module object files. |
| Affected Environment/ Platform | All |
| Driver/Module | Load balance sample application |

7.26 rte_cmdline library should not be used in production code due to limited testing

| | |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | rte_cmdline library should not be used in production code due to limited testing |
| Reference # | 34 |
| Description | The rte_cmdline library provides a command line interface for use in sample applications and test applications distributed as part of Intel® DPDK. However, it is not validated to the same standard as other Intel® DPDK libraries. |
| Implication | It may contain bugs or errors that could cause issues in production applications. |



Resolved Issues

| | |
|-----------------------------------|-----------------------------------------------------------------------------------------|
| Title | rte_cmdline library should not be used in production code due to limited testing |
| Resolution | The rte_cmdline library is now tested correctly. |
| Affected Environment/ Platform | All |
| Driver/Module | rte_cmdline |

7.27 Some *_INITIALIZER macros are not compatible with C++

| | |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Some *_INITIALIZER macros are not compatible with C++ |
| Reference # | IXA00371699 |
| Description | These macros do not work with C++ compilers, since they use the C99 method of named field initialization. The TOKEN_*_INITIALIZER macros in librte_cmdline have this problem. |
| Implication | C++ application using these macros will fail to compile. |
| Resolution/ Workaround | Macros are now compatible with C++ code. |
| Affected Environment/ Platform | All |
| Driver/Module | rte_timer, rte_cmdline |

7.28 No traffic through bridge when using exception_path sample application

| | |
|-------------|-------------------------------------------------------------------------------------------------------------------|
| Title | No traffic through bridge when using exception_path sample application |
| Reference # | IXA00168356 |
| Description | On some systems, packets are sent from the exception_path to the tap device, but are not forwarded by the bridge. |
| Implication | The sample application does not work as described in its sample application guide. |
| Resolution/ | If you cannot get packets through the bridge, it might be because IP |



| | |
|-----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Workaround | packet filtering rules are up by default on the bridge. In that case you can disable it using the following: # for i in /proc/sys/net/bridge/bridge_nf-*; do echo 0 > \$i; done |
| Affected Environment/ Platform | Linux |
| Driver/Module | Exception path sample application |

7.29 Segmentation Fault in testpmd after config fails

| | |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Segmentation Fault in testpmd after config fails |
| Reference # | IXA00378638 |
| Description | Starting testpmd with a parameter that causes port queue setup to fail, for example, set TX WTHRESH to non 0 when tx_rs_thresh is greater than 1, then doing "port start all". |
| Implication | Seg fault in testpmd |
| Resolution/ Workaround | Testpmd now forces port reconfiguration if the initial configuration failed. |
| Affected Environment/ Platform | All |
| Driver/Module | Testpmd Sample Application |

7.30 Linux kernel pci_cfg_access_lock() API can be prone to deadlock

| | |
|-----------------------------------|-----------------------------------------------------------------------------------------------------|
| Title | Linux kernel pci_cfg_access_lock() API can be prone to deadlock |
| Reference # | IXA00373232 |
| Description | The kernel APIs used for locking in the igb_uio driver can cause a deadlock in certain situations. |
| Implication | Unknown at this time; depends on the application. |
| Resolution/ Workaround | The igb_uio driver now uses the pci_cfg_access_trylock() function instead of pci_cfg_access_lock(). |
| Affected Environment/ Platform | All |
| Driver/Module | IGB UIO Driver |



7.31 When running multi-process applications, “rte_malloc” functions cannot be used in secondary processes

| | |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | When running multi-process applications, “rte_malloc” functions cannot be used in secondary processes |
| Reference # | 35 |
| Description | The rte_malloc library provides a set of malloc-type functions that reserve memory from hugepage shared memory. Since secondary processes cannot reserve memory directly from hugepage memory, rte_malloc functions cannot be used reliably. |
| Implication | The librte_malloc functions, for example, rte_malloc(), rte_zmalloc() and rte_realloc() cannot be used reliably in secondary processes. |
| Resolution/ Workaround | In addition to re-entrancy support, the Intel® DPDK now supports the reservation of a memzone from the primary thread or secondary threads. This is achieved by putting the reservation-related control data structure of the memzone into shared memory. Since rte_malloc functions request memory directly from the memzone, the limitation for secondary threads no longer applies. |
| Affected Environment/ Platform | All |
| Driver/Module | rte_malloc |

7.32 Configuring maximum packet length for IGB with VLAN enabled may not take into account the length of VLAN tag

| | |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------|
| Title | Configuring maximum packet length for IGB with VLAN enabled may not take into account the length of VLAN tag |
| Reference # | IXA00379880 |
| Description | For IGB, the maximum packet length configured may not include the length of the VLAN tag even if VLAN is enabled. |
| Implication | Packets with a VLAN tag with a size close to the maximum may be dropped. |
| Resolution/ Workaround | NIC registers are now correctly initialized. |
| Affected Environment/ Platform | All with IGB NICs |
| Driver/Module | IGB Poll Mode Driver (PMD) |



7.33 Intel® I210 Ethernet controller always strips CRC of incoming packets

| | |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Intel® I210 Ethernet controller always strips CRC of incoming packets |
| Reference # | IXA00380265 |
| Description | The Intel® I210 Ethernet controller (NIC) removes 4 bytes from the end of the packet regardless of whether it was configured to do so or not. |
| Implication | Packets will be missing 4 bytes if the NIC is not configured to strip CRC. |
| Resolution/ Workaround | NIC registers are now correctly initialized. |
| Affected Environment/ Platform | All |
| Driver/Module | IGB Poll Mode Driver (PMD) |

7.34 EAL can silently reserve less memory than requested

| | |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Title | EAL can silently reserve less memory than requested |
| Reference # | IXA00380689 |
| Description | During application initialization, the EAL can silently reserve less memory than requested by the user through the -m application option. |
| Implication | The application fails to start. |
| Resolution | EAL will detect if this condition occurs and will give an appropriate error message describing steps to fix the problem. |
| Affected Environment/ Platform | All |
| Driver/Module | Environmental Abstraction Layer (EAL) |

7.35 SSH connectivity with the board may be lost when starting an Intel® DPDK application



Resolved Issues

| | |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | SSH connectivity with the board may be lost when starting an Intel® DPDK application |
| Reference # | 26 |
| Description | Currently, the Intel® DPDK takes over all the NICs found on the board that are supported by the Intel® DPDK. This results in these NICs being removed from the NIC set handled by the kernel, which has the side effect of any SSH connection being terminated. See also issue #27. |
| Implication | Loss of network connectivity to board. |
| Resolution | Intel® DPDK now no longer binds ports on startup. Please refer to the Getting Started Guide for information on how to bind/unbind ports from Intel® DPDK. |
| Affected Environment/ Platform | Systems using a Intel® DPDK supported NIC for remote system access |
| Driver/Module | Environment Abstraction Layer (EAL) |

7.36 Remote network connections lost when running autotests or sample applications

| | |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Remote network connections lost when running autotests or sample applications |
| Reference # | 27 |
| Description | The PCI autotest and sample applications will scan for PCI devices and will remove from Linux* control those recognized by it. This may result in the loss of network connections to the system. |
| Implication | Loss of network connectivity to board when connected remotely. |
| Resolution | Intel® DPDK now no longer binds ports on startup. Please refer to the Getting Started Guide for information on how to bind/unbind ports from Intel® DPDK. |
| Affected Environment/ Platform | Systems using a Intel® DPDK supported NIC for remote system access |
| Driver/Module | Sample applications |

7.37 KNI may not work properly in a multi-process environment



| | |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | KNI may not work properly in a multi-process environment |
| Reference # | IXA00380475 |
| Description | Some of the network interface operations such as, MTU change or link UP/DOWN, when executed on KNI interface, might fail in a multi-process environment, although they are normally successful in the Intel® DPDK single process environment. |
| Implication | Some network interface operations on KNI cannot be used in an Intel® DPDK multi-process environment. |
| Resolution | The ifconfig callbacks are now explicitly set in either master or secondary process. |
| Affected Environment/ Platform | All |
| Driver/Module | Kernel Network Interface (KNI) |

7.38 Hash library cannot be used in multi-process applications with multiple binaries

| | |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Hash library cannot be used in multi-process applications with multiple binaries |
| Reference # | IXA00168658 |
| Description | The hash function used by a given hash-table implementation is referenced in the code by way of a function pointer. This means that it cannot work in cases where the hash function is at a different location in the code segment in different processes, as is the case where an Intel® DPDK multi-process application uses a number of different binaries, for example, the client-server multi-process example. |
| Implication | The Hash library will not work if shared by multiple processes. |
| Resolution/ Workaround | New API was added for multiprocess scenario. Please refer to Intel® DPDK Programmer's Guide for more information. |
| Affected Environment/ Platform | All |
| Driver/Module | librte_hash library |

7.39 Unused hugepage files are not cleared after initialization

| | |
|--------------|------------------------------------------------------------|
| Title | Hugepage files are not cleared after initialization |
| Reference # | IXA00383462 |



Resolved Issues

| Title | Hugepage files are not cleared after initialization |
|--------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Description | EAL leaves hugepages allocated at initialization in the hugetlbfs even if they are not used. |
| Implication | Reserved hugepages are not freed back to the system, preventing other applications that use hugepages from running. |
| Resolution/ Workaround | Reserved and unused hugepages are now freed back to the system. |
| Affected Environment/ Platform | All |
| Driver/Module | EAL |

7.40 Packet reception issues when virtualization is enabled

| Title | Packet reception issues when virtualization is enabled |
|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Reference # | IXA00369908 |
| Description | Packets are not transmitted or received on when VT-d is enabled in the BIOS and Intel IOMMU is used. More recent kernels do not exhibit this issue. |
| Implication | An application requiring packet transmission or reception will not function. |
| Resolution/ Workaround | Intel® DPDK Poll Mode Driver now has the ability to map correct physical addresses to the device structures. |
| Affected Environment/ Platform | All |
| Driver/Module | Poll mode drivers |

§



8 Frequently Asked Questions (FAQ)

8.1 When running the test application, I get “EAL: map_all_hugepages(): open failed: Permission denied Cannot init memory”?

This is most likely due to the test application not being run with `sudo` to promote the user to a superuser. Alternatively, applications can also be run as regular user. For more information, please refer to *Intel® DPDK Getting Started Guide*.

8.2 If I want to change the number of TLB Hugepages allocated, how do I remove the original pages allocated?

The number of pages allocated can be seen by executing the `cat /proc/meminfo | grep Huge` command. Once all the pages are mmaped by an application, they stay that way. If you start a test application with less than the maximum, then you have free pages. When you stop and restart the test application, it looks to see if the pages are available in the `/dev/huge` directory and `mmaps` them. If you look in the directory, you will see a number of 2M pages files. If you specified 1024, you will see 1024 files. These are then placed in memory segments to get contiguous memory.

If you need to change the number of pages, it is easier to first remove the pages. The `tools/setup.sh` script provides an option to do this. See the “Quick Start Setup Script” section in the *Intel® DPDK Getting Started Guide* for more information.

8.3 I have set up a total of 1024 Hugepages (that is, allocated 512 2M pages to each NUMA node). However, if I execute “`l2fwd -c f -m 64 -n 3 -- -p 3`”, I get the following output, indicating that there are no socket 0 hugepages to allocate the mbuf and ring structures to?

The `-m` command line parameter does not guarantee that huge pages will be reserved on specific sockets. Therefore, allocated huge pages may not be on socket 0. To



request memory to be reserved on a specific socket, please use the `--socket-mem` command-line parameter instead of `-m`.

8.4 I am running a 32-bit Intel® DPDK application on a NUMA system, and sometimes the application initializes fine but cannot allocate memory. Why is that happening?

32-bit applications have limitations in terms of how much virtual memory is available, hence the number of hugepages they are able to allocate is also limited (1 GB per page size). If your system has a lot (>1 GB per page size) of hugepage memory, not all of it will be allocated. Due to hugepages typically being allocated on a local NUMA node, the hugepages allocation the application gets during the initialization depends on which NUMA node it is running on (the EAL does not affinity cores until much later in the initialization process). Sometimes, the Linux OS runs the Intel® DPDK application on a core that is located on a different NUMA node from Intel® DPDK master core and therefore all the hugepages are allocated on the wrong socket.

To avoid this scenario, either lower the amount of hugepage memory available to 1 GB per page size (or less), or run the application with `taskset`, affinitying the application to a would-be master core. For example, if your EAL coremask is `0xff0`, the master core will usually be the first core in the coremask (`0x10`); this is what you have to supply to `taskset`, for example, `taskset 0x10 ./l2fwd -c 0xff0 -n 2`. In this way, the hugepages have a greater chance of being allocated to the correct socket. Additionally, a `--socket-mem` option could be used to ensure the availability of memory for each socket, so that if hugepages were allocated on the wrong socket, the application simply will not start.

8.5 On application startup, there is a lot of EAL information printed. Is there any way to reduce this?

Yes, each EAL has a configuration file that is located in the `/config` directory. Within each configuration file, you will find `CONFIG_RTE_LOG_LEVEL=8`. You can change this to a lower value, such as 6 to reduce this printout of debug information. The following is a list of LOG levels that can be found in the `rte_log.h` file. You must remove, then rebuild, the EAL directory for the change to become effective as the configuration file creates the `rte_config.h` file in the EAL directory.

```
#define RTE_LOG_EMERG 1U /* System is unusable. */
#define RTE_LOG_ALERT 2U /* Action must be taken immediately. */
#define RTE_LOG_CRIT 3U /* Critical conditions. */
#define RTE_LOG_ERR 4U /* Error conditions. */
#define RTE_LOG_WARNING 5U /* Warning conditions. */
#define RTE_LOG_NOTICE 6U /* Normal but significant condition. */
#define RTE_LOG_INFO 7U /* Informational. */
#define RTE_LOG_DEBUG 8U /* Debug-level messages. */
```




8.6 How can I tune my network application to achieve lower latency?

Traditionally, there is a trade-off between throughput and latency. An application can be tuned to achieve a high throughput, but the end-to-end latency of an average packet typically increases as a result. Similarly, the application can be tuned to have, on average, a low end-to-end latency at the cost of lower throughput.

To achieve higher throughput, the Intel® DPDK attempts to aggregate the cost of processing each packet individually by processing packets in bursts. Using the `testpmd` application as an example, the "burst" size can be set on the command line to a value of 16 (also the default value). This allows the application to request 16 packets at a time from the PMD. The `testpmd` application then immediately attempts to transmit all the packets that were received, in this case, all 16 packets. The packets are not transmitted until the tail pointer is updated on the corresponding TX queue of the network port. This behavior is desirable when tuning for high throughput because the cost of tail pointer updates to both the RX and TX queues can be spread across 16 packets, effectively hiding the relatively slow MMIO cost of writing to the PCIe* device.

However, this is not very desirable when tuning for low latency, because the first packet that was received must also wait for the other 15 packets to be received. It cannot be transmitted until the other 15 packets have also been processed because the NIC will not know to transmit the packets until the TX tail pointer has been updated, which is not done until all 16 packets have been processed for transmission.

To consistently achieve low latency even under heavy system load, the application developer should avoid processing packets in bunches. The `testpmd` application can be configured from the command line to use a burst value of 1. This allows a single packet to be processed at a time, providing lower latency, but with the added cost of lower throughput.

8.7 Without NUMA enabled, my network throughput is low, why? I have a dual Intel® Xeon® E5645 processors @2.40 GHz with four Intel® 82599 10 Gigabit Ethernet NICs. Using eight logical cores on each processor with RSS set to distribute network load from two 10 GbE interfaces to the cores on each processor.

Without NUMA enabled, memory is allocated from both sockets, since memory is interleaved. Therefore, each 64B chunk is interleaved across both memory domains.

The first 64B chunk is mapped to node 0, the second 64B chunk is mapped to node 1, the third to node 0, the fourth to node 1. If you allocated 256B, you would get memory that looks like this:



256B buffer

Offset 0x00 - Node 0
Offset 0x40 - Node 1
Offset 0x80 - Node 0
Offset 0xc0 - Node 1

Therefore, packet buffers and descriptor rings are allocated from both memory domains, thus incurring QPI bandwidth accessing the other memory and much higher latency. For best performance with NUMA disabled, only one socket should be populated.

8.8 I am getting errors about not being able to open files. Why?

As the Intel® DPDK operates, it opens a lot of files, which can result in reaching the open files limits, which is set using the `ulimit` command or in the `limits.conf` file. This is especially true when using a large number (>512) of 2 MB huge pages. Please increase the open file limit if your application is not able to open files. This can be done either by issuing a `ulimit` command or editing the `limits.conf` file. Please consult Linux* manpages for usage information.

8.9 Does my kernel require patching to run the Intel® DPDK?

Any kernel greater than version 2.6.33 can be used without any patches applied. The following kernels may require patches to provide hugepage support:

- kernel version 2.6.32 requires the following patches applied:
 - [add hugepage support to pagemap](#)
 - [fix hugepage memory leak](#)
 - [add nodemask arg to huge page alloc](#)
(not mandatory, but recommended on a NUMA system to support per-NUMA node hugepages allocation)
- kernel version 2.6.31, requires the following patches applied:
 - [fix hugepage memory leak](#)
 - [add hugepage support to pagemap](#)
 - [add uio name attributes and port regions](#)
 - [add nodemask arg to huge page alloc](#)
(not mandatory, but recommended on a NUMA system to support per-NUMA node hugepages allocation)

Note: Blue text in the lists above are direct links to the patch downloads.



8.10 VF driver for IXGBE devices cannot be initialized.

Some versions of Linux* IXGBE driver do not assign a random MAC address to VF devices at initialization. In this case, this has to be done manually on the VM host, using the following command:

```
ip link set <interface> vf <VF function> mac <MAC address>
```

where <interface> being the interface providing the virtual functions for example, eth0, <VF function> being the virtual function number, for example 0, and <MAC address> being the desired MAC address.





Appendix A Intel® DPDK License Overview

The following describes the various licenses used by the Intel® Data Plane Development Kit (Intel® DPDK). The purpose of the Intel® DPDK is to prove the abilities of the Intel® architecture processors and to provide users with a strong set of examples, libraries and proof points. By placing the majority of this software under the BSD License, users may choose to use the Intel® as is, parts of it, or just the ideas for their programs. All code may be modified by the user to suit their project needs and requirements.

Note: The license in each source file takes precedence over this document, and should be used as the definitive license for that file. All users should seek their legal team's guidance with respect to the licensing used by the Intel® DPDK.

The following table lists those files (or libraries) that are not under a BSD License. In some cases, these files are part of the standard Intel® DPDK release package, and in other cases may be a separate package that requires a separate download to be added to the Intel® DPDK. This document spells out those cases where possible.

The sections following the table provide the various licenses used. Please note that copyright notices may change overtime. It is the responsibility of all users to understand these licenses and seek their legal team's guidance.

The use of the GPLv2 License is confined to files in kernel loadable modules.

The use of the Dual BSD/LGPLv2 License and Dual BSD/GPL License allows use with either free/open source software or with proprietary software in userspace.

| File | Description | License |
|----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| igb_uio.c | 1st Released: 1.0 Location: DPDK/lib/librte_eal/linuxapp/igb_uio/ Description: This file is used for a kernel loadable module which is responsible for unbinding NICs from the Linux kernel and binding them to the Intel® DPDK | GPLv2 License Information |
| kni_dev.h kni_ethtool.c kni_misc.c kni_net.c | 1st Released: 1.3 Location: DPDK/lib/librte_eal/linuxapp/kni/ Description: The KNI kernel loadable module is a standard net driver which allows Intel DPDK Linux userspace applications to exchange packets/data with the Linux kernel. | GPLv2 License Information |



| File | Description | License |
|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| rte_kni_common.h | 1st Released: 1.3 Location: DPDK/lib/librte_eal/linuxapp/eal/include/exec-env/ Description: The KNI header files is utilized by both the Intel DPDK userspace application and the KNI kernel loadable module | Dual BSD/LGPLv2 License Information |
| All files under this directory are for the ethtool functionality. | 1st Released: 1.3 Location: DPDK/lib/librte_eal/linuxapp/kni/ethtool Description: The igb and ixgbe drivers for the Ethtool function available with the KNI kernel loadable. | GPLv2 License Information |
| rte_pci_dev_ids.h | 1st Released: 1.3 Location: DPDK/lib/librte_eal/common/include Description: Contains the PCI device ids for all devices. | Dual BSD/GPL License Information |

Added rte_dom0_mm.h file for Dual BSD/LGPLv2 License information:

rte_dom0_mm.h 1st Released 1.6.0
 Location: DPDK/lib/librte_eal/eal/include/exec-env/
 Description: The Dom0 header file is utilized by both the Intel® DPDK userspace application and the Dom0 kernel loadable module

Added dom0_mm_misc.h and dom0_mm)_misc.c file for Dual BSD/GPLv2 License Information:

dom0_mm_dev.h 1st Released 1.6.0
 dom0_mm_dev.c Location: DPDK/lib/librte_eal/linuxapp/xen_dom0
 Description: The Dom0 memory management kernel loadable module is a misc device driver which is used to facilitate allocating and mapping via an **IOCTL** (allocation) and **MMAP** (mapping).



Frequently Asked Questions (FAQ)

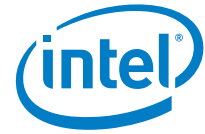
A.1 BSD License

```
/*-
 *
 *      BSD LICENSE
 *
 *      Copyright(c) 2010-2012 Intel Corporation. All rights reserved.
 *      All rights reserved.
 *
 *      Redistribution and use in source and binary forms, with or without
 *      modification, are permitted provided that the following conditions
 *      are met:
 *
 *      * Redistributions of source code must retain the above copyright
 *      notice, this list of conditions and the following disclaimer.
 *      * Redistributions in binary form must reproduce the above copyright
 *      notice, this list of conditions and the following disclaimer in
 *      the documentation and/or other materials provided with the
 *      distribution.
 *      * Neither the name of Intel Corporation nor the names of its
 *      contributors may be used to endorse or promote products derived
 *      from this software without specific prior written permission.
 *
 *
 *      THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS
 *      "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT
 *      LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR
 *      A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT
 *      OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL,
 *      SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT
 *      LIMITED TO, 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 OUT OF THE USE
 *      OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
 */
```

A.2 GPLv2 License Information

```
/*-
 *
 *      Copyright (c) 2010-2012, Intel Corporation
 *
 *      This program is free software; you can redistribute it and/or
 *      modify it under the terms of the GNU General Public License
 *      as published by the Free Software Foundation; either version 2
 *      of the License, or (at your option) any later version.
 *
 *      This program is distributed in the hope that it will be useful,
 *      but WITHOUT ANY WARRANTY; without even the implied warranty of
 *      MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
 *      GNU General Public License for more details.
 *
 *      You should have received a copy of the GNU General Public License
 *      along with this program; if not, write to the Free Software
 *      Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301,
 *      USA.
 *
 *      GNU GPL V2: http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
```

Frequently Asked Questions (FAQ)



*
* /



A.3 Dual BSD/LGPLv2 License Information

/*

This file is provided under a dual BSD/LGPLv2 license. When using or redistributing this file, you may do so under either license.

GNU LESSER GENERAL PUBLIC LICENSE

Copyright(c) 2007,2008,2009 Intel Corporation. All rights reserved.

This program is free software; you can redistribute it and/or modify it under the terms of version 2.1 of the GNU Lesser General Public License as published by the Free Software Foundation.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin St - Fifth Floor, Boston, MA 02110-1301 USA.

Contact Information: Intel Corporation

BSD LICENSE

Copyright(c) 2010-2012 Intel Corporation. All rights reserved. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

Neither the name of Intel Corporation nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, 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 OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

*/



A.4 Dual BSD/GPL License Information

```

/*-
 *      This file is provided under a dual BSD/GPLv2 license. When using or
 *      redistributing this file, you may do so under either license.
 *
 *      GPL LICENSE SUMMARY
 *
 *      Copyright(c) 2010-2012 Intel Corporation. All rights reserved.
 *
 *      This program is free software; you can redistribute it and/or modify
 *      it under the terms of version 2 of the GNU General Public License as
 *      published by the Free Software Foundation.
 *
 *      This program is distributed in the hope that it will be useful, but
 *      WITHOUT ANY WARRANTY; without even the implied warranty of
 *      MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU
 *      General Public License for more details.
 *
 *      You should have received a copy of the GNU General Public License
 *      along with this program; if not, write to the Free Software
 *      Foundation, Inc., 51 Franklin St - Fifth Floor, Boston, MA 02110-1301 USA.
 *      The full GNU General Public License is included in this distribution
 *      in the file called LICENSE.GPL.
 *
 *      Contact Information:
 *      Intel Corporation
 *
 *      BSD LICENSE
 *
 *      Copyright(c) 2010-2012 Intel Corporation. All rights reserved.
 *      All rights reserved.
 *
 *      Redistribution and use in source and binary forms, with or without
 *      modification, are permitted provided that the following conditions
 *      are met:
 *
 *      * Redistributions of source code must retain the above copyright
 *      notice, this list of conditions and the following disclaimer.
 *      * Redistributions in binary form must reproduce the above copyright
 *      notice, this list of conditions and the following disclaimer in
 *      the documentation and/or other materials provided with the
 *      distribution.
 *      * Neither the name of Intel Corporation nor the names of its
 *      contributors may be used to endorse or promote products derived
 *      from this software without specific prior written permission.
 *
 *      THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS
 *      "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT
 *      LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR
 *      A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT
 *      OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL,
 *      SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT
 *      LIMITED TO, 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 OUT OF THE USE
 *      OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

```



Frequently Asked Questions (FAQ)

*
* /

§