



Intel[®] Omni-Path Fabric Software in Red Hat* Enterprise Linux* 7.7

Release Notes

Rev. 2.0

April 2020



You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

All product plans and roadmaps are subject to change without notice.

The products described 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.

Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Copyright © 2019–2020, Intel Corporation. All rights reserved.



Contents

1.0 Overview of the Release	5
1.1 Audience.....	5
1.2 Document Versions.....	5
1.3 Software License Agreement.....	6
1.4 If You Need Help.....	6
1.5 Packages in This Release.....	6
1.6 Supported Features.....	7
1.7 Supported MPI Libraries.....	8
1.8 Intel Hardware.....	8
1.9 Intel® OPA Compatibility Matrix.....	9
1.10 Installation Requirements.....	9
1.10.1 Best Practices.....	9
1.10.2 Installation Instructions.....	9
1.11 Product Constraints.....	11
1.12 Product Limitations.....	11
2.0 Issues	12
2.1 Resolved Issues.....	12
2.2 Open Issues.....	13



Tables

1	Supported Document Versions.....	5
2	Supported Hardware.....	8
3	Intel® OPA Compatibility Matrix.....	9
4	Issues Resolved in this Release.....	12
5	Open Issues.....	13



1.0 Overview of the Release

These Release Notes are intended for Intel® Omni-Path Fabric software provided in box with the OS release. This document provides a brief overview of the changes introduced into the Intel® Omni-Path Software by this release. References to more detailed information are provided where necessary. The information contained in this document is intended as supplemental information only; it should be used in conjunction with the documentation provided for each component.

These Release Notes list the features supported in this software release, open issues, and issues that were resolved during release development.

1.1 Audience

The information provided in this document is intended for installers, software support engineers, service personnel, and system administrators.

1.2 Document Versions

Intel® Omni-Path publications are available at the following URLs. For documents compatible with this release, refer to the V10.9.0 documents listed in the table below.

- Intel® Omni-Path Switches Installation, User, Reference Guides, and Release Notes
<http://www.intel.com/omnipath/SwitchPublications>
- Intel® Omni-Path Software Installation, User, Reference Guides, and Release Notes (includes HFI documents)
<http://www.intel.com/omnipath/FabricSoftwarePublications>

The following table lists the end user document versions supported by this release.

Table 1. Supported Document Versions

Title	Doc. Number	Revision
<i>Intel® Omni-Path Fabric Quick Start Guide</i>	J57479	6.0
<i>Intel® Omni-Path Fabric Setup Guide</i>	J27600	10.0
<i>Intel® Omni-Path Fabric Switches Hardware Installation Guide</i>	H76456	7.0
<i>Intel® Omni-Path Host Fabric Interface Installation Guide</i>	H76466	5.0
<i>Intel® Omni-Path Fabric Software Installation Guide</i>	H76467	11.0
<i>Intel® Omni-Path Fabric Switches GUI User Guide</i>	H76457	10.0
<i>Intel® Omni-Path Fabric Switches Command Line Interface Reference Guide</i>	H76458	10.0
<i>Intel® Omni-Path Fabric Suite FastFabric User Guide</i>	H76469	11.0
<i>Intel® Omni-Path Fabric Suite Fabric Manager User Guide</i>	H76468	11.0
<i>Intel® Omni-Path Fabric Suite Fabric Manager GUI User Guide</i>	H76471	11.0
<i>continued...</i>		



Title	Doc. Number	Revision
Intel® Omni-Path Fabric Host Software User Guide	H76470	11.0
Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide	H76473	11.0
Intel® Omni-Path Fabric Performance Tuning User Guide	H93143	13.0
Intel® Omni-Path IP and LNet Router Design Guide (Old title: Intel® Omni-Path IP and Storage Router Design Guide)	H99668	8.0
Building Containers for Intel® Omni-Path Fabrics using Docker* and Singularity* Application Note	J57474	6.0
Intel® Omni-Path Management API Programmer's Guide	J68876	4.0
Configuring Non-Volatile Memory Express* (NVMe*) over Fabrics on Intel® Omni-Path Architecture Application Note	J78967	1.0
Intel® Omni-Path Fabric Software Release Notes	K38338	1.0
Intel® Omni-Path Fabric Manager GUI Release Notes	K38339	1.0
Intel® Omni-Path Fabric Switches Release Notes (includes managed and externally-managed switches)	K38337	1.0
Intel® Omni-Path Fabric Unified Extensible Firmware Interface (UEFI) Release Notes	K21145	1.0
Intel® Omni-Path Fabric Thermal Management Microchip (TMM) Release Notes	K21147	1.0
Intel® Omni-Path Fabric Firmware Tools Release Notes	K21148	1.0

1.3 Software License Agreement

This software is provided under license agreements and may contain third-party software under separate third-party licensing. Please refer to the license files provided with the software for specific details.

1.4 If You Need Help

Technical support for Intel® Omni-Path products is available 24 hours a day, 365 days a year. Please contact Intel Customer Support or visit <http://www.intel.com/omnipath/support> for additional detail.

1.5 Packages in This Release

Intel® Omni-Path Software Packages
Packages created by Intel
opa-address-resolution-10.9.0.0.204-1.el7.x86_64
opa-basic-tools-10.9.0.0.204-1.el7.x86_64
opa-fastfabric-10.9.0.0.204-1.el7.x86_64
opa-fm-10.9.0.0.204-1.el7.x86_64
opa-libopamgt-10.9.0.0.204-1.el7.x86_64
libfabric-1.7.0-1.el7.x86_64
libpsm2-11.2.78-1.el7.x86_64
<i>continued...</i>



Intel® Omni-Path Software Packages
Firmware binaries delivered by Intel
8051 firmware version 1.27.0
SBus Master firmware version 0x10130001
PCIe SerDes firmware version 0x4755
Fabric SerDes firmware version 0x1055
Packages used by Intel
rdma-core-22.1-3.el7.x86_64 (libhfi1)
openmpi-1.10.7-5.el7.x86_64
openmpi3-3.1.3-2.el7.x86_64
mpitests-openmpi-5.4.2-1.el7.x86_64
mpitests-openmpi3-5.4.2-1.el7.x86_64
mpitests-mvapich222-5.4.2-1.el7.x86_64
mpitests-mvapich23-5.4.2-1.el7.x86_64
mvapich2-2.2-psm2-2.2-4.el7.x86_64
mvapich23-psm2-2.3-4.el7.x86_64
mpitests-mvapich222-psm2-5.4.2-1.el7.x86_64
mpitests-mvapich23-psm2-5.4.2-1.el7.x86_64

HFI Programmable Firmware

To download Intel programmable firmware for HFIs, refer to the following:

- [Unified Extensible Firmware Interface \(UEFI\)](#)
- [Thermal Management Module \(TMM\)](#)
- [Firmware Tools](#)

NOTE

Refer to the [Intel® OPA Compatibility Matrix](#) on page 9 for the firmware versions compatible with this release.

1.6 Supported Features

- The list of supported hardware is in [Table 2](#) on page 8.
- Product constraints are described in [Product Constraints](#) on page 11.
- New cable data collection tool (AOC Health Monitoring via PM).
- Intel® OPA support for cgroups.
- Support for multiple virtual fabric security.
- Active Optical Cables. For details, see the Cable Matrix at: <https://www.intel.com/content/www/us/en/products/network-io/high-performance-fabrics/omni-path-cables.html>



- Support for active optical cables (AOC) on server platforms using integrated HFI for OPA (commonly known as "-F").
- Support for Power Class 2 active optical cables (AOC). See [Product Constraints](#) on page 11 for more information.
- Legacy BIOS Boot Mode Enhancements to support boot over fabric, custom board descriptions, and pre-boot platform configuration data for AOC support.
- Multi-endpoint functionality. See the *Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide* for details.
- Support for OpenFabrics Interfaces (OFI), a framework that includes libraries (including libfabric) and applications used to export fabric communication services to applications.
- Support for NVMe over Fabric Protocol
- Virtual Fabric creation has been enhanced to better support advanced topologies, including the ability to place multicast traffic on a separate SL from unicast traffic. For details, see the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide*, section 2.

1.7 Supported MPI Libraries

The list below shows the different MPI libraries tested with RHEL* 7.7 for Intel® Omni-Path Fabric Software.

- Open MPI3
- MVAPICH23

1.8 Intel Hardware

The following table lists the Intel hardware supported in this release.

NOTE

The Intel® PSM2 implementation has a limit of four (4) HFIs.

Table 2. Supported Hardware

Hardware	Description
Intel® Xeon® Processor E5-2600 v3 product family	Haswell CPU-based servers
Intel® Xeon® Processor E5-2600 v4 product family	Broadwell CPU-based servers
Intel® Xeon® Scalable Processors	Skylake CPU-based servers
2nd Generation Intel® Xeon® Scalable Processors	Cascade Lake CPU-based servers
Intel® Xeon Phi™ x200 Product Family	Knights Landing CPU-based servers
Intel® Xeon Phi™ 72x5 Processor Family	Knights Mill CPU-based servers
Intel® Omni-Path Host Fabric Interface 100HFA016 (x16)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Host Fabric Interface 100HFA018 (x8)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Switch 100SWE48Q	Managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48U	Externally-managed 48-port Edge Switch
<i>continued...</i>	



Hardware	Description
Intel® Omni-Path Switch 100SWE48UFH	Externally-managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE48QFH	Managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE24Q	Managed 24-port Edge Switch
Intel® Omni-Path Switch 100SWE24U	Externally-managed 24-port Edge Switch
Intel® Omni-Path Director Class Switch 100SWD24	Director Class Switch 100 Series, up to 768 ports
Intel® Omni-Path Director Class Switch 100SWD06	Director Class Switch 100 Series, up to 192 ports

1.9 Intel® OPA Compatibility Matrix

The following component versions are compatible with Intel® Omni-Path software in RHEL* 7.7. The bold text below represents the base component versions for this release.

Table 3. Intel® OPA Compatibility Matrix

UEFI	TMM	Managed Switch	Externally-Managed Switch	FM GUI
1.9.2.0.0	10.9.0.0.208	10.8.0.0.186	10.8.0.0.186	10.9.3.0.22
				10.9.2.0.7
1.9.0.1.0				10.9.1.0.14
				10.9.0.0.208
1.8.1.0.0	10.8.0.0.214	10.8.0.0.186	10.8.0.0.186	10.8.0.0.206

1.10 Installation Requirements

This section provides instructions and information on installing the software.

1.10.1 Best Practices

- Intel recommends that users update to the latest versions of Intel® Omni-Path firmware and software to obtain the most recent functional and security updates.
- To improve security, the administrator should log out users and disable multi-user logins prior to performing provisioning and similar tasks.
- To improve security, Intel recommends configuring the `MgmtAllowed` setting and consider limiting access to port configuration changes by limiting access to Userspace Management Datagrams (UMADs). Refer to the *Intel® Omni-Path Fabric Software Installation Guide*, About User Queries Settings for more information.

1.10.2 Installation Instructions

Perform the steps in this section to install the default Intel® Omni-Path Software configuration.



Assumptions

- You are logged in as root or with root privileges.
- You have a list of IPv4 addresses and netmasks for each IPoIB interface you are going to set up.
- RHEL* packages are available in a yum repository.

References

- Refer to the *Intel® Omni-Path Fabric Software Installation Guide* for related software requirements and next steps.
- Refer to the *Intel® Omni-Path Fabric Switches Hardware Installation Guide* for related firmware requirements.

Procedures

Perform the following steps to install the default Intel® Omni-Path Software configuration using RHEL* OS:

Step	Task/Prompt	Action
Install OPA-Basic Software		
1.	At the command prompt, enter the installation command for <code>opa-basic-tools</code> .	Type <code>yum install -y opa-basic-tools</code> and press Enter .
2.	At the command prompt, reboot the server.	Type <code>reboot</code> and press Enter .
3.	Check your link using <code>opainfo</code> .	Type <code>opainfo</code> and press Enter . Example output: <pre>hfil_0:1 PortGID: 0xfe80000000000000:001175010163f931 PortState: Active LinkSpeed Act: 25Gb En: 25Gb LinkWidth Act: 4 En: 4 LinkWidthDnGrd ActTx: 4 Rx: 4 En: 3,4 LCRC Act: 14-bit En: 14-bit,16-bit, 48-bit Mgmt: True LID: 0x00000010-0x00000010 SM LID: 0x0000000c SL: 0 QSFP: AOC , 5m FINISAR CORP P/N FCBN425QB1C05 Rev A Xmit Data: 0 MB Pkts: 251 Recv Data: 0 MB Pkts: 251 Link Quality: 5 (Excellent)</pre>
4.	Install the <code>rdma-core</code> rpm.	Type <code>yum install -y rdma-core</code> and press Enter .
5.	On all compute nodes: install the PSM2 library.	Type <code>yum install -y libpsm2</code> and press Enter .
Install Intel® Omni-Path Fabric Suite Components on the Management Node		
6.	Install FastFabric.	Type <code>yum install -y opa-fastfabric</code> and press Enter .
7.	Install the <code>opa-address-resolution</code> rpm on all nodes.	Type <code>yum install -y opa-address-resolution</code> and press Enter .
8.	Install Fabric Manager.	Type <code>yum install -y opa-fm</code> and press Enter .
9.	Start the Fabric Manager.	Type <code>systemctl start opafm</code> and press Enter .
Set up IPoIB IPV4 Configuration		
<i>continued...</i>		



Step	Task/Prompt	Action
10.	Manually edit or create the <code>ifcfg-ibX</code> file.	<p><i>Note:</i> Use the OS distribution-supplied instructions for setting up network interfaces.</p> <p>Type <code>cat /etc/network-scripts/ifcfg-ib0</code> and press Enter.</p> <p>Example output:</p> <pre>DEVICE=ib0 TYPE=infiniband BOOTPROTO=static IPADDR=10.228.200.173 BROADCAST=10.228.203.255 NETWORK=10.228.200.0 NETMASK=255.255.252.0 ONBOOT=yes CONNECTED_MODE=yes MTU=65520</pre> <p>NOTE: To configure datagram mode for AIP, change <code>CONNECTED_MODE=no</code> and remove (comment out) <code>MTU=</code> of the <code>ifcfg-ib0</code> file. Further details can be found in the <i>Intel® Omni-Path Fabric Performance Tuning User Guide</i>.</p>
11.	Bring up the <code>ib0</code> interface.	Type <code>ifup ib0</code> and press Enter .
12.	Perform a test ping.	<p>Type <code>ping <remote IPoIB address></code> and press Enter.</p> <p>For example:</p> <pre>ping 10.228.200.161 PING 10.228.200.161 (10.228.200.161) 56(84) bytes of data. 64 bytes from 10.228.200.161: icmp_seq=1 ttl=64 time=0.863 ms</pre>
End Task		

1.11 Product Constraints

- Power class 2 AOC are supported. You must use UEFI version 1.5 or newer for proper operation. Servers using integrated HFI (-F) requires a specific BIOS level to support power class 2 AOC; contact your BIOS vendor for more information.

1.12 Product Limitations

This release has the following product limitations:

- Performance Administration (PA) Failover should not be enabled with FMs running on differing software versions.
To disable PA failover, edit the `/etc/opa-fm/opa_fm.xml` file and in the `<Pm>` section, change `<ImageUpdateInterval>` to 0.
- Enabling UEFI Optimized Boot on some platforms can prevent the HFI UEFI driver from loading during boot. To prevent this, do not enable UEFI Optimized Boot.



2.0 Issues

This section lists the resolved and open issues in the Intel® Omni-Path Software.

2.1 Resolved Issues

The following table lists issues that are resolved in this release.

Table 4. Issues Resolved in this Release

ID	Description	Resolved in Release
132207	Kernel crash caused by the ib_srpt module.	RHEL* 7.7
139743 143031 143115	Under a very heavy load through the IPoIB interface, the kernel warning <code>NETDEV WATCHDOG: ib0 (hfil): transmit queue 0 timed out</code> , followed by the messages <code>queue stopped 1, tx_head xxx, tx_tail xxx</code> and <code>transmit timeout: latency xxxx msec</code> s may be seen.	RHEL* 7.7
141793	Use of a static buffer could produce an incorrect device name (hfi1_x) in dmesg logging.	RHEL* 7.7
143449	PM will scroll LQI=0 and Integrity Exceeded Threshold logs when an additional VF with QoS enabled and a device group that is not "All". <i>Note:</i> This issue does not occur when running against the default opafm.xml configuration file.	RHEL* 7.7
144165	Nodes unable to ping on IPoIB. <i>Note:</i> This issue occurs when a host port disappears and reappears from the FM's topology (usually due discovery timeout or major fabric disruption), while the port remains ACTIVE the entire time. This results in the host port not being a member of the IP multicast groups. The primary symptom is the inability to resolve IP addresses via ARP.	RHEL* 7.7
144996	Running workloads with more than 78 ranks with the Open MPI OFI MTL over OFI Verbs;OFI_RXM provider may result in a hang with message sizes larger than 65 KB.	RHEL* 7.7
145474	OFI Verbs <code>mpi_stress</code> may cause verbs/MSG provider completion queue overrun that results in dropped completions. They show up as sequence errors in the test.	RHEL* 7.7
145855	If the Admin VF is not running on VL0, the HSM may get into a state where it is unable to talk to the fabric. The sweep will log the following errors: <pre>opamgt ERROR: [<pid>] omgmt_send_mad2: send failed; Invalid argument, agent id 2 MClass 0x81 method 0x1 attrId 0x11 attrM 0x0 WARN [topology]: SM: sm_send_stl_request_impl: Error Sending to Path:[1] Lid:[0xffffffff] [Can't find node in topology!]. AID:[NODEINFO] TID:[0x0000000000000031] Status:[OK (0x00000000)] WARN [topology]: SM: topology_main: TT: too many errors during sweep, will re-sweep in a few seconds rc: 108: unrecoverable error</pre>	RHEL* 7.7
146456	In a fabric with only one Edge switch using the fat tree routing algorithm, a port can get stuck in the <code>Init (LinkUp)</code> state after the port is bounced.	RHEL* 7.7

continued...



ID	Description	Resolved in Release
STL-46606 STL-47956 STL-48661	Bouncing a link or rebooting a device under certain fabric conditions may cause a switch in the fabric to be removed from the Fabric Manager's internal view of the topology leading to fabric disruptions and instability.	RHEL* 7.7
STL-46790	In cases where GSI services are active and the FM is receiving capability change traps (common after node reboots), FM responsiveness may be impacted. This could result in data traffic disruption or unexpected FM failovers. GSI traffic would include the PM, SA, and DBSync (FM failover).	RHEL* 7.7
STL-47546	When an ISL goes down in the middle of an FM sweep (due to a disruption in the fabric such as a reboot), the SA copy of topology becomes invalid when the Fattree routing algorithm is used. SA queries that use this topology (e.g., path record query) fail. <i>Note:</i> A path record query failure can be seen in FM log as "INVALID TOPOLOGY" messages. The issue will resolve after the FM's next successful sweep.	RHEL* 7.7

2.2 Open Issues

The following table lists the open issues for this release.

Table 5. Open Issues

ID	Description	Workaround
129563	Memory allocation errors with MVAPICH2-2.1/Verbs.	<p><i>Note:</i> To avoid this issue, use MPIs over PSM. If you are using MPIs over verbs (not recommended), the following workaround is required:</p> <ul style="list-style-type: none"> When running MVAPICH2 jobs with a large number of ranks (for example, > 36 ranks but ≤ 72 ranks), you must set the following parameters in <code>/etc/security/limits.conf</code>: <ul style="list-style-type: none"> hard memlock unlimited soft memlock unlimited Also, you must increase the <code>lkey_table_size:LKEY</code> table size in bits (2^n, where $1 \leq n \leq 23$) from its default of 16 to 17. For instructions on setting module parameters, refer to the <i>Intel® Omni-Path Fabric Performance Tuning User Guide</i>, HFI1 Driver Module Parameters chapter.
135830	On Intel® Xeon Phi™ systems, failure observed during software upgrade when rebuilding the boot image. Error message contains: Rebuilding boot image with "/usr/bin/dracut -f"	<p>Due to the extended processing time of the dracut command on the Intel® Xeon Phi™ platform, Intel recommends the following:</p> <ul style="list-style-type: none"> Install and configure Intel® Xeon Phi™ systems separately. Change the <code>FF_TIMEOUT_MULT</code> value in <code>opafastfabric.conf</code> from 2 to 6 for Intel® Xeon Phi™ systems.
139368	Some applications compiled with older compilers may use a personality bit that signifies that READ should imply EXECUTE permissions. To improve system security, the hfi1 driver does not allow execute permissions on PSM memory maps. Therefore, applications that use READ implies EXECUTE will fail to run.	<p>As root, run the <code>execstack</code> tool to clear the executable bit on the binary:</p> <pre>execstack -c <binary></pre> <p>Alternatively, recompile the binary to not set this personality bit.</p>

continued...



ID	Description	Workaround
139613	The Subsystem Vendor and Subsystem Device ID in the PCI configuration space of Intel® Omni-Path discrete HFI cards may not indicate the correct OEM vendor and device. As a result, the <code>lspci</code> command may show incorrect Subsystem Vendor and Device ID information. This issue affects Intel server boards for Intel® Xeon® Processor v3 and v4 Product Family configured in Legacy OS boot mode.	Reconfigure the system from Legacy OS boot mode to UEFI boot mode.
141273	The in-distro version of <code>perftests</code> has bugs.	Use the upstream version of <code>perftest</code> from https://github.com/linux-rdma/perftest .
142330	MPI applications that leverage the PSM2 library's access to the HFI ASICs Memory Mapped IO and that access the MMIO directly (not via PSM2) can potentially cause an "unsupported opcode" error which some servers handle as a critical error.	Disable upstream error reporting using the AER mask register. <ul style="list-style-type: none">For discrete HFI ASICs (e.g., CHF PCIe card), use<pre>setpci -d 8086:24f0 ECAP_AER +8.l=00100000:00100000</pre>For integrated HFIs (e.g., KNL-F and SKX-F), use<pre>setpci -d 8086:24f1 ECAP_AER +8.l=00100000:00100000</pre>
STL-47571	Since <code>libfabric</code> 1.6, the <code>psm2</code> provider maps OFI endpoints directly to HFI contexts instead of multiplexing multiple OFI endpoints to a single HFI context. This relies on the multi-EP feature of the <code>PSM2</code> library and thus the provider automatically sets <code>PSM2_MULTI_EP=1</code> if it has not been set. However, enabling the multi-EP feature also disables context sharing. As the result, applications may experience the following runtime error when trying to oversubscribe CPU cores (which is usually the same as available HFI contexts). <pre>hfi_userinit: assign_context command failed: Device or resource busy PSM2 can't open hfi unit: -1 (err=23)</pre> <p><i>Note:</i> Applications that don't use <code>libfabric</code> are not affected.</p>	Set <code>PSM2_MULTI_EP=0</code> . <i>Note:</i> This only works for applications that open only one OFI endpoint per process.