What’s new – Volume 1 Release 1.7

Overview
Specification update overview

- Volume 1, Release 1.7, published July 11, 2023
- The specification defines InfiniBand and RoCE
- Available to IBTA Members

- 2091 pages
- 60 comments submitted and included
- New features added by both the LWG and the MgtWG
What’s new in Vol1 Release 1.7

IBTA - Management Working Group
Support For Large Radix Switches

• 1.7 Spec Version
  • Finalize backward compatible support for large radix switches and directed route (DR) MADs
  • Support XDR speeds

• Next Steps
  • Add support for XDR speed FEC modes
  • Review and enhance various sections of the specification to incorporate user feedback
Update DR For Large Radix Switches

• Directed route algorithm now supports large radix switches as endpoints and as intermediate devices
Next Generation Speed

• Spec 1.7 supports XDR speed ~200Gb/s per lane.
  • QSFP ➔ 800 Gb/s
  • QSFP-DD and OSFP ➔ 1600 Gb/s

• Update the PortInfo MAD with new extended speeds to support the future generation

• Updates were made to chapters – 14 and 15
What’s new in Vol1 Release 1.7

IBTA - Link Working Group
Network Probing Problem Statement

• RDMA congestion control is evolving
  • Timely
  • HPCC
  • Swift

• A simple in-band RTT measurement primitive is not available in RDMA transport
  • E.g. No response on RDMA READ
  • E.g. ACK coalescing on RDMA WRITE / SEND

• New primitives are required for efficient congestion control e.g.:
  • End to end round trip measurements
  • End to end telemetry collection

• Network Probing extensions (Annex 20) are addressing this requirement
  • End to end measurement collection primitives between reaction point and notification point
  • No RDMA transport level changes, independent of the transport service and link layer (IB / RoCE)
Network Probing Design Guidelines

• No impact for data path packets
  • E.g. RDMA / SEND / ACK

• No changes to transport service / link / network layers
  • E.g. RC, UC, RD, UD
  • E.g. IB / RoCE

• Interoperability and support
  • Ability to work on any RoCE/IB platform

• Network routing robustness
  • Network probe packet should follow the flow

• Ability to hold payload & relay back

• Robustness to network configuration
  • Ability to work with & without PFC / ECN / etc.
Network Probing Architecture

- Reaction Point
- Network
- Notification Point

- RDMA CM
- App
- QP
Network Probe Overview

- Network Probes are a generalized mechanism for probing the state of the network.
- Probes are sent from one end point to another and may interact with network entities along the way.
- Network Probes can be used to collect information about the network without the need to have a specific process running on the remote node.
- Network Probes utilize the basic MAD format and appear as standard MAD packets in the network.
Others

• Multicast congestion control recommendation
• Ordering & error flows clarification for MPE Verify Check / Verify Compute
• Memory windows interoperability with MPE
• APM clarification for RoCE
For more information

https://www.infinibandta.org/ibta-specification/

• RDMA vendors:
  • Implement Network Probing in your InfiniBand and RoCE adapter(s)
  • Implement Large Radix Switches

• RDMA users:
  • Enhance your application(s) and ULP(s) to leverage Network Probing