OPENFABRICS ALLIANCE

iWARP Update

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP

#OFADevWorkshop

Increasing interest in iWARP

- RDMA Use Cases
  - High Performance Computing
  - File and Block Storage
  - NVM access
  - Virtual Machine migration
  - Low-latency messaging middleware
  - Virtualization and Cloud deployments place important requirements on these use cases

- Key iWARP value propositions for these use cases
  - Engineered for “typical” Ethernet (best effort, no DCB, no QCN, etc)
  - Natively Routable
  - Multi-pathing supported at Layer 3 (as well as Layer 2)
  - Reliable and proven TCP Transport
    - Mature and efficient retransmission algorithms
    - Dynamic and verified congestion algorithms

iWARP Standards

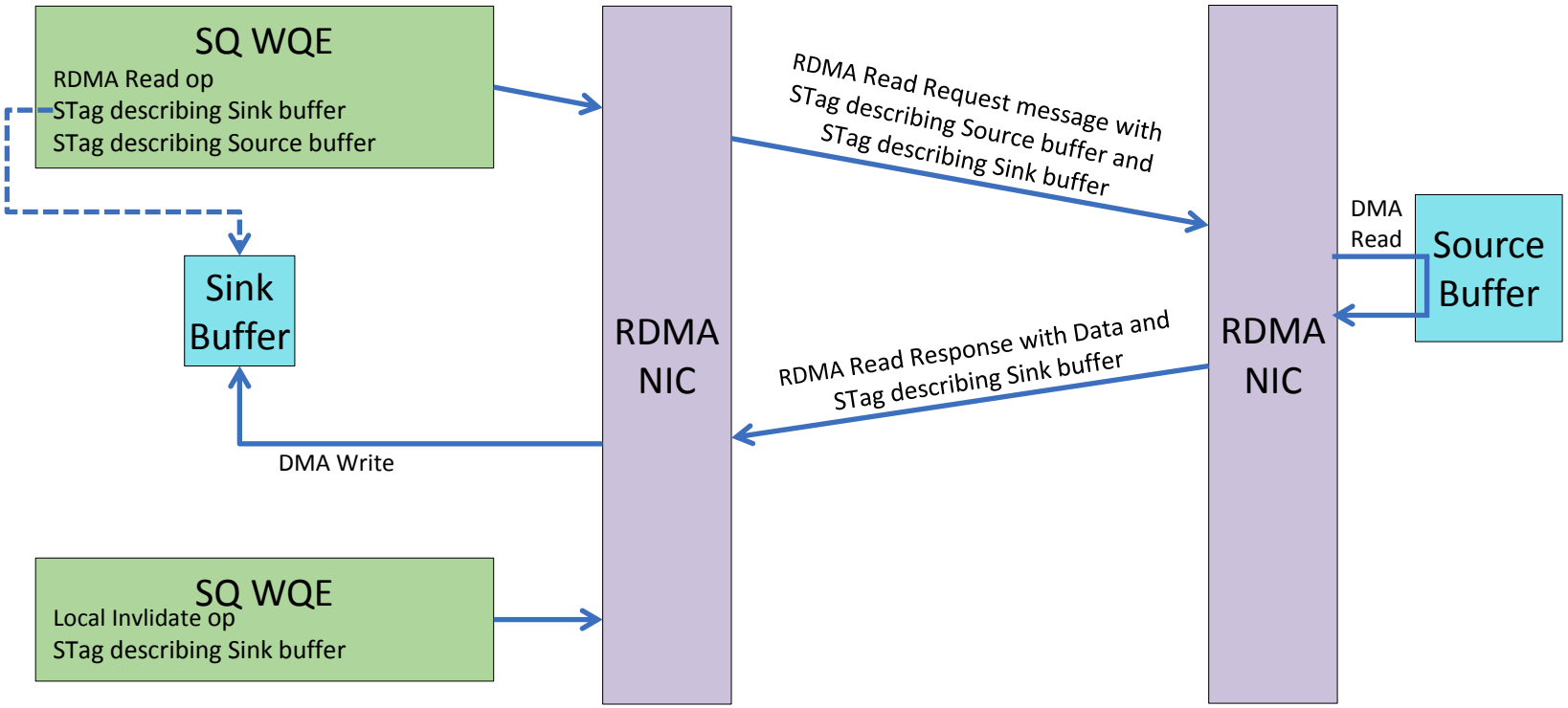
- iWARP updates and enhancements are handled by the IETF STORM (Storage Maintenance) working group
- Finalized RFCs

RFC 5040	A Remote Direct Memory Access Protocol Specification
RFC 5041	Direct Data Placement over Reliable Transports
RFC 5044	Marker PDU Aligned Framing for TCP Specification
RFC 6580	IANA Registries for the RDDP Protocols
RFC 6581	Enhanced RDMA Connection Establishment

iWARP In-Progress RFCs

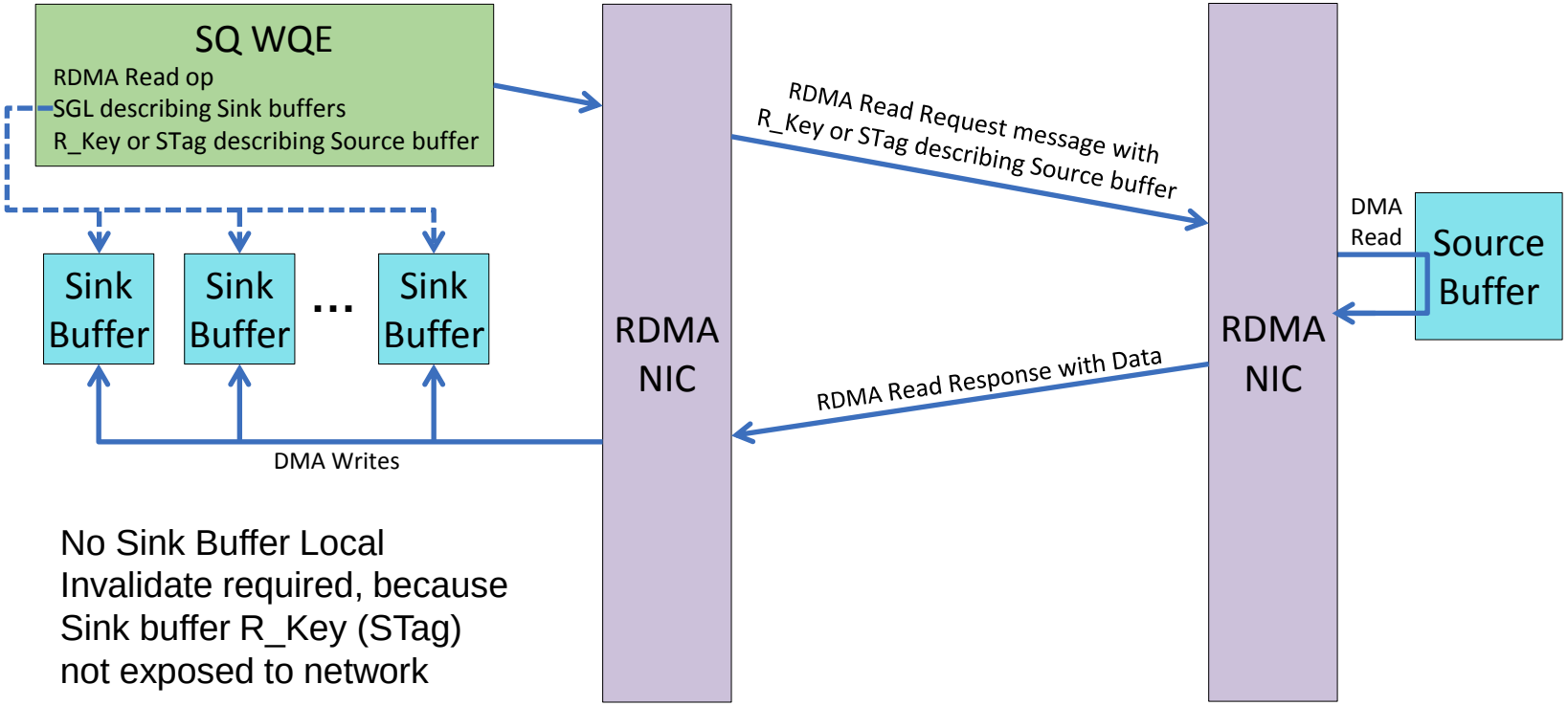
- draft-ietf-storm-rdmap-ext-09.txt
  - Extends RFC 5040
    - Adds Atomic Operations and Immediate Data
  - Authors from Intel, Broadcom, Chelsio
  - IESG approved. Next step RFC Editor Queue
- draft-ietf-wood-rdmap-ext-v2-00.txt
  - Extends RFC 5040
    - Add Send with Immediate Data
    - Add IB-style RDMA Read
  - Authors from Intel
  - Submitted for initial review by STORM working group

Traditional iWARP RDMA Read



This slide has animations – view in Presentation mode

IB-style RDMA Read



No Sink Buffer Local Invalidate required, because Sink buffer R_Key (STag) not exposed to network

In-progress RFCs enable this flow on iWARP

This slide has animations – view in Presentation mode

Goal of In-Progress RFCs

- Common Application capabilities across all flavors of RDMA
- These RFCs remove all known application differences between iWARP and InfiniBand

IETF Alignment/Synergy with iWARP

- iWARP currently leverages:
  - TCP
    - Reliable transport and congestion management
  - Explicit Congestion Notification
    - Inherited from TCP/IP layers
- iWARP will naturally adopt/use:
  - Tunneling/Network Overlays
    - iWARP works with (but does not require) existing tunnel protocols (ie Generic Routing Encapsulation) and NVO3 technology investigations
- Connectionless messaging to complement iWARP RDMA
  - Typically realized with unreliable datagrams (unicast and multicast)
  - InfiniBand has Unreliable Datagram (UD)
  - UDP may be used in place of UD for Ethernet implementations
    - No new wire protocol standards required

iWARP Ecosystem

- Strong industry support to evolve iWARP
- Good alignment with IETF, and support in STORM to evolve the standards
- OFED 3.5-2 stable drivers from multiple vendors: cxgb3, cxgb4, nes
- Intel is implementing iWARP RDMA as a key capability in *Fort Park*. *Fort Park* is an Ethernet IP block that will be integrated into future Intel server chipsets.

Call to Action

- Participate in STORM standards reviews
- iWARP RNIC vendors and system software vendors consider supporting the in-progress RFCs as soon as possible
- Develop future RDMA extensions with a goal to enable them across all flavors of RDMA

Thank You

OPENFABRICS ALLIANCE

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP