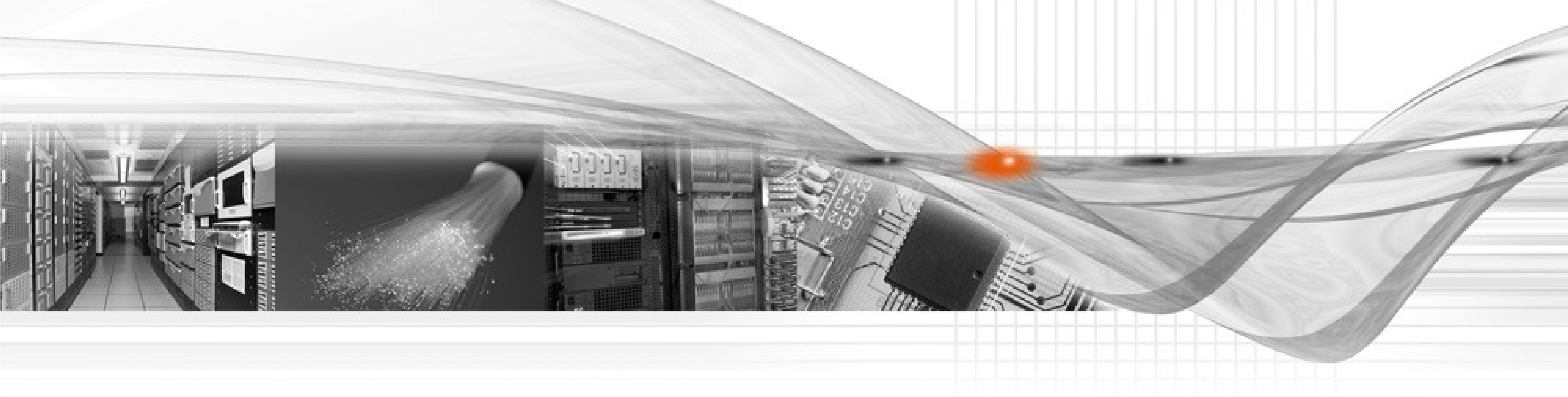

OPENFABRICS ALLIANCE

iWARP Enhancements

Authors: Sharp/Meigs Intel CNG
Date: March 28, 2012

iWARP in OFED

- iWARP support first discussed at 2005 workshop
- iWARP development based on 1.2 release in 2007
  - This was not part of the distribution yet..1.3
- Three different iWARP providers implemented today
  - Ammasso 1100
  - Chelsio T3/T4
  - Intel NetEffect NE020
- Common fast path verbs operations for IB and iWARP
  - Some application visible differences
- iWARP connection management strictly IP based
- Applications can be IB/iWARP agnostic
  - Use RDMA CM, avoid unique operations

Application Visible Differences

- iWARP does not support the IB connection manager
  - RDMA Connection Manager hides IB specifics
- iWARP missing some operations
  - Immediate data operations
  - Atomic operations
- Some operations different
  - RDMA Read has single local buffer element
  - RDMA Read LKey requires remote access
  - Local Invalidate Operations added
- Some differences in error semantics
  - “Empty” Receive Queue MAY cause connection to fail
  - Remote errors MAY be detected after local completion

iWARP Standards Activity

- IETF chosen for iWARP standards
  - IETF handles TCP/IP based standards
- Original iWARP standards published by IETF
- New drafts in process:
  - “Peer to Peer” draft close to being published
    - Enhanced Connection Setup
    - Already implemented in OFED iWARP providers
  - Draft for additional operations under review
    - Authored by Broadcom, Chelsio and Intel
    - Adds Atomic operations and Immediate data
- Other standards work
  - Verbs updates (more on next slide)

iWARP Verbs Enhancements

- Verbs standard drives interoperable implementations and API definitions
  - Not an OS specific API definition
  - A basis for API definition instead
- Forming new iWARP verbs extension consortium
  - Original verbs developed in the RDMA Consortium
- Required verbs work:
  - Cover new operation codes
  - Address minor inconsistencies in original specifications
  - Add multicast support
  - Add ability to query more optional behaviors

Application Recommendations

- Use the IP based RDMA Connection Manager
- Avoid use of unique IB or iWARP operations
- Plan ahead for remaining differences
  - Use single local buffer on RDMA Read
  - Remote rights for LKey
- Let us know about differences not addressed by standards work

Next Steps

- Continue shepherding RDMA Protocol Extensions draft through IETF process
    - Please review drafts at http://tools.ietf.org/wg/storm/
- Form iWARP verbs specification consortium
- Discuss need for new IETF standards work
- Contact David Fair (david.l.fair@intel.com) for further information regarding standards work

Backup

Peer Connect IETF Draft Status

- Extends RFCs 5043 and 5044
- Original iWARP did not have a concept of RTR state exit
  - Assumed active side of connection sent first RDMA message
- Application was responsible for ORD/IRD negotiation
  - Typically used application messages or private data
- iWARP connection establishment enhancements:
  - Ready to Receive (RTR) Message Negotiation
  - Standardized ORD/IRD Negotiation
- "Peer connect" draft is about to become an RFC
  - RFC numbers assigned
  - Final edits in progress

RDMAP IETF Draft Status

- Extends RFC 5040
- Adds atomic operations and immediate data
- RDMA Protocol Extensions Draft in active state
- Currently on revision 2
  - Addressed comments received from IETF community
  - More input encouraged
- Draft should move to “Last Call” status soon
  - Possibly in April
- Submission to IESG as Standards Track follows
  - Possibly in July