

IETF iWARP Update

Brian Hausauer
IETF STORM
Nov 7, 2013

Increasing interest in iWARP

- **Public/private cloud uses:**
 - File and Block Storage
 - Virtual Machine migration
 - Low-latency messaging middleware
 - HPC-as-a-Service
- **Key iWARP value propositions for these usage cases:**
 - Engineered for best-effort Ethernet
 - Neither lossless Ethernet nor DCB required
 - 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

Recent STORM work on iWARP

- Driving iWARP Extensions into the iWARP specification
 - Focused on eliminating the application-visible differences between iWARP and InfiniBand
- draft-ietf-storm-rdmap-ext-08
 - Authors from multiple iWARP providers
 - Adds missing iWARP operations:
 - Atomic Operations
 - RDMA Write with Immediate Data
 - Last Call closed on Oct 15, 2013
 - Ready to submission to STORM AD and IESG for initial request for publication as an RFC
- Next phase ready for discussion

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

Future iWARP Expansions

- Remaining InfiniBand/iWARP differences
 - RDMA Read semantics
 - Send with Immediate Operations
 - New draft coming soon
- Storage
 - Acknowledged Writes
 - Reliable multicast
- HPCaaS
 - Address known RDMA resource scaling deficiencies when used in very high core count clusters