

RDMA Status Update

IETF 111

NFSv4 Working Group

July 29-30, 2021

Tom Talpey

RDMA charter and work item

- The NFSv4 working group is also responsible for maintenance and extension of the RDMA protocols originally developed by the now-concluded RDDP WG (RFCs 5040-5045, e.g., RDMAP, DDP and MPA) , whose maintenance and extension were previously handled by the now-concluded STORM WG.
- “The WG will extend RDMA to enhance the Memory Placement operations such as Flush, Atomic Write and Validation using Integrity Signatures.”

Mar 2021

Request publication of RDMA Extensions for Enhanced Memory Placement

Status of Draft

- <https://datatracker.ietf.org/doc/html/draft-talpey-rdma-commit-01>
- Individual draft, industry authors
- Problem statement, discussion, and protocol specification for iWARP extension
- Most recent update March 2020

```
Versions: 00 01
NFSv4 (provisionally)
Internet-Draft
Updates: 5040 7306 (if approved)
Intended status: Standards Track
Expires: September 10, 2020

T. Talpey
Microsoft
T. Hurson
Intel
G. Agarwal
Marvell
T. Reu
Chelsio
March 9, 2020
```

RDMA Extensions for Enhanced Memory Placement
draft-talpey-rdma-commit-01

Abstract

This document specifies extensions to RDMA (Remote Direct Memory Access) protocols to provide capabilities in support of enhanced remotely-directed data placement on persistent memory-addressable devices. The extensions include new operations supporting remote commitment to persistence of remotely-managed buffers, which can provide enhanced guarantees and improve performance for low-latency storage applications. In addition to, and in support of these, extensions to local behaviors are described, which may be used to guide implementation, and to ease adoption. This document updates [RFC5040](#) (Remote Direct Memory Access Protocol (RDMA)) and updates [RFC7306](#) (RDMA Protocol Extensions).

Why the delay?

- Author company change (Tom now “Unaffiliated”)
- IBTA extension activity
 - “Memory Placement Extensions” in upcoming IB/RoCE 1.5 update
 - Flush and Atomic Write semantics fully aligned
 - Verify still under discussion
 - Paused IETF work to ensure alignment – now ready to proceed
- Interest increasing from ULP consumers
 - Persistent memory-aware applications (e.g. storage, database, pub/sub)
 - Emerging platform interconnect extensions (e.g. CXL, PCIe, NVMe)
 - Definition and use of peer-to-peer bus protocols, steering tags, etc
- Strong agreement on Flush, Atomic Write and Verify as fundamental

Plan for Update to Draft

- Publish by September as NFSv4 WG document, per charter
- Refresh references and other text (“commit” => “flush”)
- Corrections to:
 - Missing MO field in operations
 - Revise VERIFY proposed definition
 - Add ordering rules
 - Incorporate other comments received
- Update to new XML RFC v3 schema and tools
- Revise NFSv4 WG work item schedule dates

Other RDMA work item possibility

- RDMA (iWARP) over QUIC
- Discussed previously, but deferred until QUIC core protocol complete
- With RFC9000 published, is it time to consider?
- Ideas from QUIC inheritance:
 - RDMA secure by default, protecting direct placement
 - Multichannel or multistream RDMA, with certain ordering guarantees
 - RDMA datagram or RDM (reliably delivered messaging) mode
 - N.B. IB and RoCE currently support these