**OFI WG Data Storage / Data Access Subteam Weekly telecom – 10/28/2014**

**OFIWG Download Site:** [www.openfabrics.org](http://www.openfabrics.org) 🡪OFED/OFA Resources 🡪 OpenFabrics Interfaces WG

**Agenda**

* role call,
* agenda bashing
* early thoughts on requirements

Q: Are we focused on networking, or the VFS layer?

A: This is about networking, when viewed from a storage perspective. e.g. it’s not so much about Lustre as it is about LNET/LND.

Jim Ryan- bring in someone from SNIA? SNIA is pursuing this from the NVM side

A: Yes please.

Start with requirements gathering - Volunteers?

- Bernard volunteers for requirements for byte addressable memory

- Venkat V – volunteers for file I/O, (LNET and also GPFS)

- Frank Y – has some write-ups on requirements and will present next week.

- Stan Smith – also has some perspective from a network standpoint, less so for filesystem, but more about what a filesystem would want from a network interface, specifically LNET/LND.

-Paul – some thoughts about storage over networks in general. Hard to commit to writing it down, but will try to do so ASAP.

Stan Smith: “Ideal Lustre IF.pptx”

Collected from former Whamclouders and others, blue sky thinking.

1. Reliable datagrams – scalable, no connection footprint, useful in large systems scaling

- no order preservation requirement

- small size

- senders should adhere to injection limits – i.e. congestion. There has to be a concept of a limit, which implies that there is some sort of injection rate limit.

- “quasi-reliable”, meaning datagrams only get dropped on transmission failure, not on congestion.

2. Remote Memory Access

- PUT/GET type operations

- passive side buffers guaranteed to be pre-registered

- should be sensibly-sized 1K 🡪 1MB

- Datagrams are the only unsolicited messages – everything else is moved by agreement between the endpoints (RMA put/get).

3. Two priorities for both communication styles (RD & RMA)

- high priority comms complete with max guaranteed latency, but with vastly reduced injection rates (up to 100 high priority communications per second injected at any endpoint with a maximum latency to delivery of 100mS. Used for communications to maintain global cluster membership state. Historically, this has been done with some kind of heartbeat, which tends to get stacked up behind other, larger messages.

- standard priority communications.

-QoS on a per transaction basis – does this include latency, B/W…?

Stan will seek the appropriate approvals to post these to the reflector.

AR Paul – circulate the link for posting presentations

Frank asks about the status of a potential kernel maintainer. (Bart had volunteered to send the name of a potential storage maintainer). All agreed that we want to interact with the kernel community both from the perspective of release scheduling and from the perspective of what is generally acceptable to the kernel maintainers. Also agreed that it is a bit early in the process for us just yet. Jeff B (NASA) points out that the kernel community generally wants to talk in fairly specific terms.

**Agenda for next meeting**

- NetApp to present some initial thoughts on requirements

**Next regular telecom**

Next meeting: Tuesday, 11/4/14

8am-9am Pacific daylight time