**OFI WG Data Storage / Data Access Subteam Weekly telecom – 12/16/2014**

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

**Agenda**

* role call, agenda bashing
* recording DS/DA calls
* Next steps
* Requirements gathering? finished?
* Meetings over the next few weeks (holidays)

**Next Steps**

-what drove the user level stuff was that people had actual applications in mind.

- how does this turn into a kernel driver that gets loaded?

- go to solution space? use cases? e.g. SRP

- if there is nobody who wants to write something to it, there is no reason to have the API…

- possible use cases:

- block storage (SRP)

- byte addressable memory

- persistent block storage

- RPC style messaging-based I/O i.e. Lustre

- have we identified anything missing in base libfabrics?

- no mechanism for a callback in the case of e.g. a completion queue entry. No way today to pass in a callback pointer.

- in getinfo call there is a string representation of a port number, but converting strings to integers inside the kernel is problematic.

- will need additional calls for:

- interrupt management API, eg. coalescing values, enable/disable interrupts

- memory registration

- an ability to rate limit traffic e.g. MPI and storage traffic mixed on the same fabric. Is this really more of a management function versus something that would be controlled via the API? e.g., if you have a b/w reservation for 10Gb using OSCARS for example, how do you rate limit a 40Gb NIC? How fine-grained does that need to be?

- an example: checkpoints are written to node local memory, which eventually needs to be drained. How to do so without impacting MPI?

- congestion. Would it be helpful to provide feedback on congestion indications to a higher level?

- multi-homed storage devices: for a local storage device which is hosted by a node, is there any way to access that storage device from the fabric to access the stored data if the node that is hosting it fails? This would require two interfaces to the device, for example an NVMe device with a local port and a fabric-attached port.

AI: Reese – look through SRP code to uncover other things that may be missing.

By introducing a second class of traffic (e.g. storage) to the existing libfabrics, we uncover a fair number of new/interesting issues, mainly related to QoS and/or congestion.

- are we limited to storage? No.

- are we limited to kernel applications? No.

**Next Steps**

**Agenda for next meeting**

Summary of Requirements gathered to date

**Next regular telecom**

Next meeting: Tuesday, 1/6/15.

8am-9am Pacific daylight time

**NOTE:** We have switched over to using Webex (courtesy of Cisco). The URL for joining meetings is:

<https://cisco.webex.com/cisco/j.php?J=200935598&PW=67935ad6df07030d5f05044a5b0f>