libfabric: your new BFF

Sung-Eun Choi, Cray Inc.
Salishan Conference, Random Access Session
April 29, 2015
Background

The Open Fabrics Interface Working Group (OFI WG) formed in August 2013

• Co-chairs:
  – Sean Hefty, Intel
  – Paul Grun, Cray Inc.

Charter:

*Develop an extensible, open source framework and interface aligned with upper-layer protocols and applications needs for high-performance fabric services.*
The only network API you’ll ever need (we hope)
Why?

• Today middleware needs to be ported to a new (and sometimes more complicated) low-level network API every 3-5 years

• These hardware-specific APIs have to be supported for upwards of 10 years

• A common API gives you portability on day one*
How?

OFI WG is a community effort

• In no particular order: DOE, DOD, NASA, Intel, Cray, Cisco, Mellanox, IBM, UNH (plus storage vendors)...

• Expertise spans hardware and software
Charter: open source

Development on github

- [https://github.com/ofiwg](https://github.com/ofiwg)
- Dual licensing: GPL and BSD
Charter: ULP and apps

Engage the user community to define requirements

- MPI requirements
  - ETH Zurich, SNL, ORNL, ANL, Cisco, IBM, Intel, AMD, Cray, Microsoft, Mellanox, SGI, U Edinburgh/EPCC, U Alabama Birmingham

- PGAS and SHMEM requirements
  - LANL, ORNL, SNL, Intel, Mellanox, Cray
Charter: high performance fabric

Software leading hardware

- **Vendor involvement**
  - Can we influence HPC network vendors?
- **Extensible interface**
  - Can be vendor-specific
  - Good way to propose acceptance into main API
libfabric architecture

Middleware

libfabric APIs
- Control Interface
- Message Queue
- RDMA
- Atomics
- CM Services
- Event Queues
- Tag Matching
- Triggered Operations

Provider implementation
- I/O service
- I/O service
- I/O service
- ...
libfabric architecture: realized

available on github
libfabric in a visual nutshell

- **Control Services**
  - Discovery
  - fi_info

- **Communication Services**
  - Connection Management
  - Address Vectors

- **Completion Services**
  - Event Queues
  - Counters

- **Data Transfer Services**
  - Message Queues
  - RMA
  - Tag Matching
  - Atomics

**Triggered Operations**
Not shown: usage models

• Capabilities
  – Application desired *features* and *permissions*
  – e.g., RMA, Atomics, tag matching

• Attributes
  – Defines the *limits* and *behavior* of selected interfaces
  – e.g., thread safety, message ordering constraints

• Mode
  – Provider request on application
  – e.g., local memory registration, user-allocated context
“The Fine Print”

• It’s a large API with lots of options

• It’s a work-in-progress

• Have an opinion? Join us:
  – http://lists.openfabrics.org/mailman/listinfo/ofiwg
  – Weekly meetings: Tuesdays 9am PT
  – https://github.com/ofiwg/libfabric
Status

• Release 1.0 coming soon!
  – providers: sockets, verbs, PSM, usNIC
Thanks!

And thanks to Sean and Paul for slide content