OPEN IB
ALLIANCE

Transport Neutral OpenIB
Key iWARP Vendor Requirements

Uri Elzur, Caitlin Bestler
Broadcom

Tom Tucker
Open Grid Computing

Definitions

- Transport Neutral

    – Lacking preference perceived or actual between RDMA capable transports of different types

- Transport Independent

    – Not dependent on transport specific features or conventions

- Vendor Neutral

    – Lacking preference perceived or actual between HCA providers from different vendors

- Vendor Independent

    – Not dependent on vendor specific features or implementation

- OpenSTAC

    – This is a place holder for the new name for OpenIB

- Application

    – A user mode program written to industry standard API e.g. MPI library or Sockets

- ULP (Upper Layer Protocol)

    – A library or middleware between the application and OpenSTAC

Key Goals

- Applications run unmodified on all supported RDMA transports

- Drivers and ULP are open source BSD/GPL

- ULP are adapted to become transport neutral

- Add iWARP support to OpenIB gen2 with minimal changes

- Special transport features remain available for applications that want them

- Applications run unmodified on new versions of the stack

- Expanded development of support for RDMA in key applications (NFS, iSER, etc...)

  - Not required for the purpose of integrating IB and iWARP

  - All future Transport Neutral ULP/Applications could benefit from these new development – "write the ULP Once!"
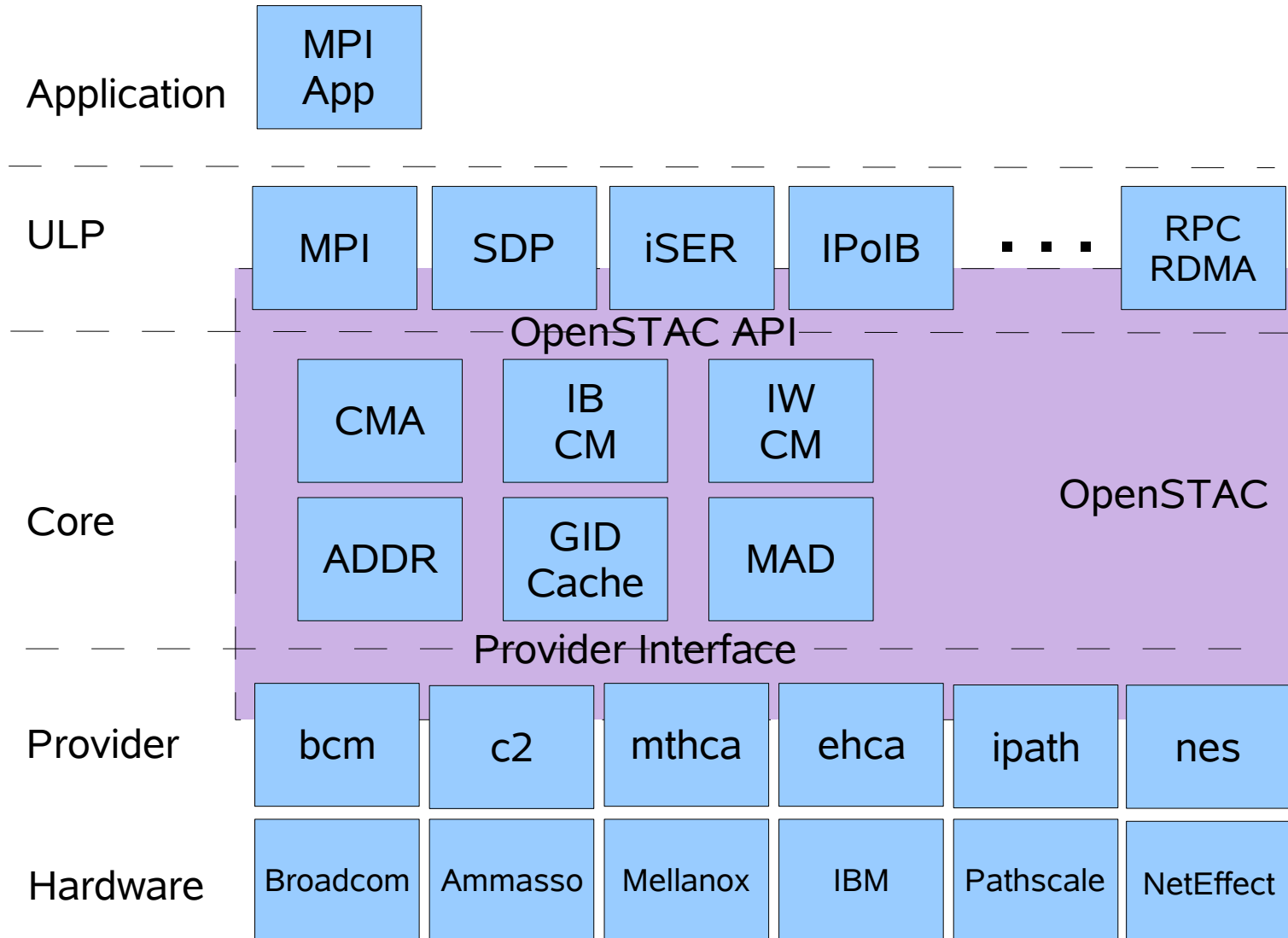
3

Specific Non-Goals

- Slow the flow of new features into the kernel

- Destabilize the code base

- Enable proprietary drivers and ULP

- Break application backwards compatibility

- Hide or disable transport specific features

- Degrade overall stack performance

- Diminish the number or scope of the ULP development projects

- Require either transport to emulate the behaviour of another

Current Status

- Functional changes required are 90% complete, e.g.
    - Transport independent connection manager
    - Transport independent device definition (ib_verbs.h)
    - Generic address resolution handling
    - A real-live iWARP device driver
- Cosmetic changes are 0% complete
    - Transport neutral names
    - Documentation
- 0% performance degradation
    - There is no transport switch logic in the data path
- Get it at : https://openib.org/svn/gen2/branches/iwarp
- Patch has been submitted for integration

OpenSTAC Scope

Application — MPI App

ULP — MPI | SDP | iSER | IPoIB | . . . | RPC RDMA

OpenSTAC API

Core — CMA | IB CM | IW CM | ADDR | GID Cache | MAD | OpenSTAC

Provider Interface

Provider — bcm | c2 | mthca | ehca | ipath | nes

Hardware — Broadcom | Ammasso | Mellanox | IBM | Pathscale | NetEffect

OpenSTAC API Scope

- OpenSTAC API specifies interfaces between applications and RDMA transports

- OpenSTAC API specifies interfaces between the OpenSTAC stack and a device driver

- OpenSTAC API does not define a hardware interface

OpenSTAC API

- OpenSTAC API draws from IBTA 1.2, RDMAC and RNIC-PI as needed

- OpenSTAC will define and document a set of Transport and Vendor Independent API that is the minimum any provider will support

  – Which API/features are "common" (i.e. required for all transports)

  – Which API/features are "specific" (i.e. unique to a particular implementation or transport)

  – Method for application to determine features supported by a provider

- OpenSTAC API must not rely on vendor/transport dependent behaviour

- Verbs must allow compliance with standards

  – e.g. currently no way to pass up TERMINATE message

OpenSTAC API & Verbs

- Applications can be written that are unaware of the underlying RDMA transport

- Transport specific and optional features are accessible

- Optional feature support may be queried at run-time

- OpenSTAC API & Verbs will not unfavourably impact the performance of any transport

OpenSTAC API & Conn Mgmt

- Transport Independent
  - Peer addressing is IP based
  - Applications may optionally still access transport dependent CM features (e.g. MAD, QoS, etc...)
- Peer address verification is no less strong than for native stack
- Longer term requirement
  - Support for iWARP connection establishment in streaming mode
    - My be necessary to support ISER per IETF RFC
  - Drivers need notification when the network changes
    - Static route modified by the user
    - Neighbour update
    - ICMP redirect
    - Path MTU change

OpenSTAC API - Naming

- Transport neutral naming convention for all symbols exported by the transport neutral interfaces
  - rdma_
- Uniform naming convention for transport specific symbols
  - ib_
  - iw_
  - etc...

ULP Guidelines

- ULP should utilize vendor and transport independent API and connection management:

  - MPI

  - uDAPL

  - SDP

  - iSER/iSCSI

- ULP will not require transport specific behaviour

  - e.g. in order data placement

Provider (Device Driver)

- Driver must be included in the OpenIB source repository as GPL/BSD.

- Firmware can be distributed separately from driver, i.e. firmware is part of the hardware.

Organizational Requirements

- New Name – OpenSTAC
  - Open Source Transport Agnostic Consortium
- Charter must be modified to be Transport Neutral
- Charter must specify that at least IB and iWARP will be supported
  - IB 1.2
  - RDMA Consortium
  - IETF RDDP 1.0