The Common Communication Interface (CCI)

Presented by: Galen Shipman
Technology Integration Lead
Oak Ridge National Laboratory

Collaborators: Scott Atchley, George Bosilca, Peter Braam, David Dillow, Patrick Geoffray, Brice Goglin, Ken Matney, Ron Minnich, Jeff Squyres, Geoffroy Vallee
Sockets in Data Centers

- Sockets is the de-facto standard Application Programming Interface (API) in networking
  - Portable, robust, simple

- Commonly uses TCP or UDP on the wire

- Designed in the 1980s
  - Relatively slow and lossy networks
  - Limited host concurrency
The Sockets API Has Problems

- Difficult to leverage networking innovations:
  - Semantics incompatible with zero-copy techniques
  - No portable support for asynchronous operations
  - Poor scalability with per-peer buffering and polling

- A bottleneck on application performance
  - Bad at 10GbE, worse at 40GbE or 100GbE
Need an alternative programming interface to reap the benefits of high-speed Ethernet

Experiences from high performance interconnects:
- Techniques: OS-bypass, zero-copy, scalability
- Vendor-neutral ecosystem through an open API
A Modern Network API

- **Common Communication Interface (CCI)**
  - Performance: low latency, high throughput, low CPU overhead, efficient multi-thread and NUMA
  - Scalability: no per-peer resources
  - Robustness: connection-oriented model
  - Portability: network and vendor neutral
  - Simplicity: compact API, event-driven

- **A modern paradigm for modern Ethernet**
  - A simple, flexible and logical API
CCI Basics

- **Endpoints**
  - Virtualized instance of a device

- **Connections**
  - Allows granular control of reliability and ordering attributes

- **Communication**
  - Small Messages
  - Remote Memory Access
Endpoints and Connections

- **Endpoints**
  - Complete container of resources
  - An event driven model
    - Application may poll or block
    - Events include send, recv, connection establishment, etc.
    - Events may contain resources (buffers for small messages)

- **Connections**
  - Per peer - a single endpoint can handle many connections
  - Scalable, no per-peer send/recv buffers or event queues
Communication

- **Small Messages**
  - Always buffered on both send and receive side
  - Library manages buffers, not the application
  - Message may be processed in-place

- **Bulk Data**
  - RMA communication for bulk-data transfer
  - Zero-copy when available
  - No implicit order for efficient link aggregation
    - explicit fence
  - May be combined with delivery of a remote Event
CCI Unleashes Modern Ethernet Performance

- 10GbE: UDP-CCI
- 10GbE: Linux-Direct-CCI
- 10GbE: Vendor-CCI
- IB QDR: CCI

Latency (µsec) vs. Message Size (bytes)

San Jose, CA USA
February 2012
CCI will not replace Sockets overnight
- Both are complementary in data centers
- Migrate performance-sensitive, intra-application communication to CCI

<table>
<thead>
<tr>
<th>CCI</th>
<th>Sockets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application controls both sides of the communication</td>
<td>Application controls only one side of the communication</td>
</tr>
<tr>
<td>Performance gain worth the porting effort</td>
<td>Existing implementation is good enough</td>
</tr>
</tbody>
</table>

*East-West traffic*  
*North-South traffic*
Competition: Verbs API

- Designed and driven by InfiniBand
- Incredibly Complex API
- Portability issues between IB and iWARP
- Limited scalability
  - per-QP resources, memory footprint
- Vendor specific semantics
  - Limits portability
  - Raises the bar for breaking into the market
Our Approach

- CCI defines the API not the software stack
  - Free to innovate under a common API
- BSD-style license
  - Easy to commercialize your derivative work
  - Easy to leverage existing code base
  - Protects your IP
- Apache-style contributor agreement
  - Protects the entire CCI community
Current Partners

- Cisco
- Inria
- Myricom
- Oak Ridge National Laboratory
- Sandia National Laboratories
- University of Tennessee, Knoxville

San Jose, CA USA
February 2012
Conclusion

- Sockets API cannot leverage modern Ethernet NICs capabilities
- We propose CCI, a novel communication interface built on over a decade of high performance networking experience
- CCI allows application to fully benefit from modern Ethernet networks
- CCI enables an open, vendor-neutral high performance Ethernet ecosystem
Visit http://cci-forum.com

Galen Shipman
gshipman@ornl.gov

This work is sponsored in part by the Office of Advanced Scientific Computing Research (ASCR); U.S. Department of Energy. The work was performed in part at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. De-AC05-00OR22725.