



Open MPI and InfiniBand



Jeff Squyres
Open MPI Architect

Presentation_ID © 2006 Cisco Systems, Inc. All rights reserved. Cisco Confidential

1

Open MPI is Community

- Originally founded by 4 organizations
- Now has 13 members
 - Industry and academic partners
 - More organization in the process of being added
- **Bleeding edge research and industry hardening**
- Collaboration is key
 - The sum is greater than its parts
- Current status
 - Version 1.2b1 released at SC'06

SC2006 Mellanox Booth © 2006 Cisco Systems, Inc. All rights reserved. Cisco Public

2

Top 500: #6 Sandia Thunderbird cluster

- Linpack result
 - 4347 dual processor nodes
 - 53 teraflops
 - 84.66% network efficiency
- Powered by Open Fabrics / Open MPI
 - Mellanox HCAs
 - Cisco switches

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

3

Community Members

Academia / Research

- HLRS
- Indiana University
- Sandia National Laboratory
- Los Alamos National Laboratory
- University of Dresden
- University of Houston
- University of Tennessee

Industry

- Cisco
- IBM
- Mellanox
- Myricom
- QLogic
- Sun
- Voltaire

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

4

Open MPI works everywhere

- Linux, Solaris, OS X, Windows
- Run-time systems
 - BProc
 - POE / LoadLeveler
 - rsh/ssh
 - SLURM
 - Torque / PBS Pro / OpenPBS
 - XGrid
- InfiniBand (OpenFabrics and VAPI)
- InfiniPath
- Myrinet (MX and GM)
- Portals
- Shared memory
- TCP
- uDAPL

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

5

Open MPI works everywhere

- Linux, Solaris, OS X, Windows
- Run-time systems
 - BProc
 - POE / LoadLeveler
 - rsh/ssh
 - SLURM
 - Torque / PBS Pro / OpenPBS
 - XGrid
- **InfiniBand (OpenFabrics and VAPI)**
- InfiniPath
- Myrinet (MX and GM)
- Portals
- Shared memory
- TCP
- uDAPL

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

6

InfiniBand support in Open MPI

- Full support for OpenFabrics
 - Previously known as “OpenIB”
 - Distributed in the OpenFabrics Enterprise Distribution (OFED)
 - Also available from Open MPI web site
- Powered #6 entry in the Top 500 list
 - Sandia thunderbird cluster
- VAPI is also supported, but being phased out
 - No new work except critical bug fixes

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

7

OpenFabrics support

- Developed by multiple IB vendors, research organizations, academics
- Repeatable, reliable high performance
 - Low latency
 - High bandwidth (pipelined registration, etc.)
- Native support for multi-port
 - Stripe large messages across available IB ports
- Multi-LID support for congestion avoidance
- Dozens of run-time tunable parameters for power users

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

8

Open MPI emerging research / work

- Many forms of fault tolerance
 - Checkpoint / restart
 - Logging, distributed, ...etc.
- Thread safety and asynchronous progress
- Host heterogeneity
- Data reliability, failover
- More networks, schedulers

- ...and more

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

9

Don't see what you want? Come join us!

- Open MPI reflects the union of the priorities of its members
 - Come be part of the process
- Open MPI is a production-quality platform for research

<http://www.open-mpi.org/community/contribute>

SC2006 Mellanox Booth

© 2006 Cisco Systems, Inc. All rights reserved.

Cisco Public

10

Questions?



Make today an Open MPI day!

<http://www.open-mpi.org/>