



OSGC
May 13, 2008



Open MPI and Sun HPC ClusterTools A Technical Overview

Len Wisniewski
Engineering Manager
Software Developer Tools and Services
Sun Microsystems

Open MPI open-source community

- Currently 15 Members, 9 Contributors, 1 Partner
- Plus individual contributors



Open MPI goals

- Create a free, open source, peer-reviewed, production-quality complete MPI-2 implementation.
- Provide extremely high, competitive performance (latency, bandwidth, ...pick your favorite metric).
- Directly involve the HPC community with external development and feedback (vendors, 3rd party researchers, users, etc.).
- Provide a stable platform for 3rd party research and commercial development.
- Help prevent the "forking problem" common to other MPI projects.
- Support a wide variety of HPC platforms and environments.



www.open-mpi.org



Open MPI architecture

MPI API Layer

PML = Pt2Pt Messaging Layer

BML = BTL Management Layer

**TCP
BTL**

**Shared Memory
BTL**

**Open IB
(User Verbs)
BTL**

**uDAPL
BTL**

BTL = Byte Transfer Layer

Open MPI Supported Platforms

Resource Managers

Sun Grid Engine

PBS Pro / Torque / Open PBS

rsh / ssh

SLURM

LoadLeveler

Xgrid

Yod

Compilers

Sun Studio

gcc

PGI

Intel

Pathscale

Operating Systems

Solaris

Linux

Mac OS X

Windows

Interconnects

TCP / ethernet

Shared memory

Infiniband

Myrinet (GM and MX)

Portals

Loopback

Sun HPC ClusterTools

- Previously based on proprietary source code derived from Thinking Machines GlobalWorks technology
- Starting with Sun HPC ClusterTools 7 (CT 7), CT is a binary distribution of Open MPI
 - CT 7 based on Open MPI 1.2
 - CT 7.1 based on Open MPI 1.2.4
 - CT 8 to be based on Open MPI 1.3
- www.sun.com/clustertools
 - CT 7.1 is available
 - Early access version of CT 8 available now!



Sun's Contributions to Open MPI

- Features contributed (or collaborated on) by Sun
 - Sun Grid Engine plug-in
 - Sun Studio compiler support
 - Infiniband support on Solaris
 - Myrinet MX support on Solaris
 - Totalview / Allinea DDT support on Solaris
 - Processor affinity support on Solaris
 - Sun packaging
 - Dtrace examples
 - Parallel job utilities
 - MPI Test Tool development
 - <https://svn.open-mpi.org/trac/mtt>



Open MPI 1.3

- Improved job startup scalability
- Improved MPI_THREAD_MULTIPLE support
- Checkpoint / restart support
- Support for Platform LSF
- OpenIB BTL improvements
 - iWARP support
 - XRC / ConnectX support
- Processor affinity improvements
- Message logging
- VampirTrace support
- Many more new features and improvements
 - <https://svn.open-mpi.org/trac/ompi/wiki>
 - See Release document for 1.3 series



Sun HPC ClusterTools 8

- Based on Open MPI 1.3
- Features
 - Linux support
 - Mellanox ConnectX support
 - Increased scalability
 - Support for 1024 nodes / 4096 processes
 - And more...TACC-sized clusters
 - Profiling support
 - Dtrace providers
 - Sun Studio Analyzer tight integration
 - MPI PERUSE
 - VampirTrace
 - Infiniband multi-rail support on Solaris





Len Wisniewski
leonard.wisniewski@sun.com