Scali MPI - ScaMPI

Overview

Scali's MPI implementation, ScaMPI™, is fully compliant with the MPI 1.1 specification, ensuring application compatibility. ScaMPI is designed to support scalable systems, focusing on sustained performance of systems up to hundreds of nodes. We are proud to offer a complete solution with extremely good latency numbers and excellent bandwidth in a multi-thread-safe implementation. This is achieved by exploiting the *shared address space architecture* of SCI, Scalable Coherent Interface.

Scali's commitment to performance is combined with extensive RAS (Reliability, Availability and Serviceability) features required for large systems; ScaMPI continues to operate in an error free manner during cable exchanges, dynamic change of interconnect routing tables, interconnect reset etc.

Scalable Systems

Scaling an application to a large number of MPI processes, while keeping the problem size constant, requires more frequent exchange of messages which becomes smaller and smaller. This is the situation when the goal of the scaling is to reduce the execution time. ScaMPI has been developed with these applications in mind and provides extremely low latency for short messages as depicted in figure 1 and 2. Furthermore, ScaMPI supports multi-threaded applications. Thus, it is possible to reduce the number of MPI processes to the number of nodes in the system, instead of using one process per CPU.

ScaMPI has also included a highly optimized implementation of MPI's collective operations. These are ideally suited for solving large problems, where high sustained bandwidth is an important requirement. ScaMPI's collective operations show close to linear performance scaling with growing system size.

Features and benefits

- **Highly Optimized Implementation** - ScaMPI's message latency, measured as half the round-trip delay of a zero length MPI message, is less than 5 µsec. Sustained bandwidth exceeds 80 MBytes/s.
- **Multi-Thread-Safe and Hot** - Multithreaded applications can fully exploit ScaMPI and multiple threads can simultaneously request services and conduct communication using ScaMPI.
- **Fault Tolerance** - ScaMPI operates transparently to transient networks errors, interconnect reset or change of routing tables.

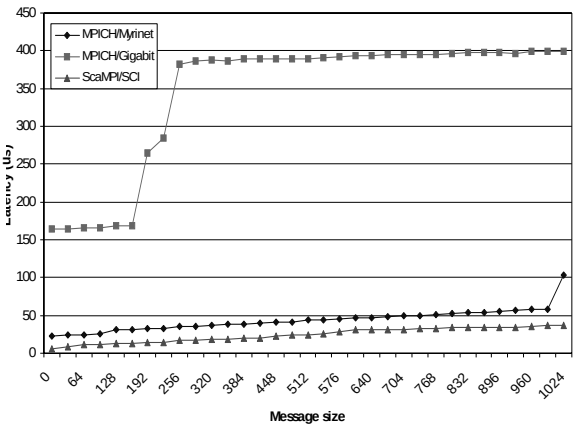

Figure 1: ScaMPI Ping/Pong latency[1]

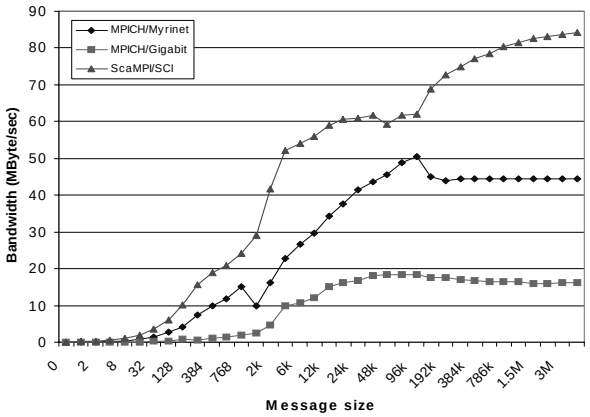

Figure 2: ScaMPI Ping/Pong bandwidth[1]

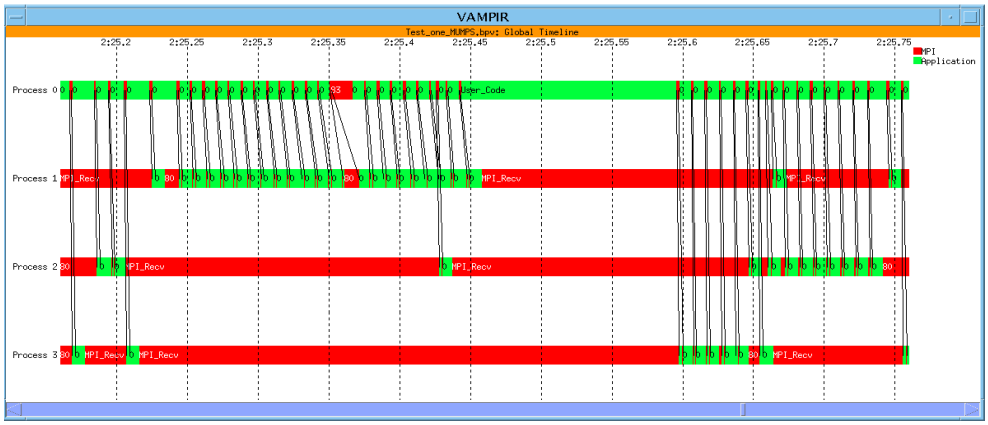

Figure 3: Vampir trace

- **Automatic selection of physical transport mechanism** - ScaMPI will automatically select the best suited transport medium for MPI messages; within a Symmetrical Multiprocessor node (SMP), shared memory will be used, whereas Scalable Coherent Interface -SCI - interconnect will be used between the nodes.
- **UNIX Command Line Replication** - The command line arguments to the application will automatically be provided to all MPI processes. This avoids tedious parsing and broadcast of parameters to the other MPI processes.
- **Exact Message Size Option** - Although not required by the MPI specification, ScaMPI has the option of verifying match between the effective received and transmitted message size.
- **MIMD Support** - ScaMPI supports the Multiple Instruction - Multiple Data model by having provision of launching different executables which constitute the whole MPI application.
- **Heterogeneous Clusters** - ScaMPI supports applications consisting of MPI processes hosted on nodes running different operating systems.
- **Graphical Frontend** - ScaMPI application can either be launched from *mpimon*, Scali's MPI start-up program, from the MPICH compatible *mpirun* scripts, or from the Scali graphical desktop.
- **Daemon launch** - ScaMPI uses a daemon to start the MPI processes. This enhances security and avoids use of the remote shell daemon (rshd).
- **Manual launch mode** - ScaMPI has the ability to start X terminals enabling manual start of selected MPI processes. This is useful if special trace-, profiling-tools, or debuggers are to be applied on selected portions of the MPI processes.
- **Support for TotalView** - ScaMPI fully supports Etnus' TotalView distributed debugger.
- **Support for Vampir** - Pallas Gmbh's Vampir trace system is supported by ScaMPI (figure 3).

Platforms supported

OS	Version	Architecture
Linux	RH 6.0	i86pc
Solaris	2.6 or 7	i86pc / UltraSPARC

Availability

Now

1) mpperf from MPICH, RedHat Linux 6.0 (2.2.7), Intel 440BX 82443BX AGP (rev 3) with Dual Pentium II 450 MHz

System	Network interface	Network topology	Com. library	Network driver
ScaMPI/SCI	Dolphin D311/D312 32 bit PCI-SCI card	6 nodes in a 2D mesh 2 x 4 config.	ScaMPI 1.8.0	ScaSCI 090 990716_124623
MPICH/Myrinet	AMCC Myrinet PCI M2-PCI-32 (rev 1)	6 nodes switched	MPICH/GM 1.1.2..7	gm-1.086
MPICH/Gigabit ethernet	Gigabit Ethernet card	6 nodes switched	MPICH 1.1.2	Hamachi 0.14