“Interference-aware incoming message detection for MPI threaded progression”

Masahiro Miwa, Kohta Nakashima, Akira Naruse
Fujitsu Laboratories Ltd.
masahiro.miwa@jp.fujitsu.com

Abstract:

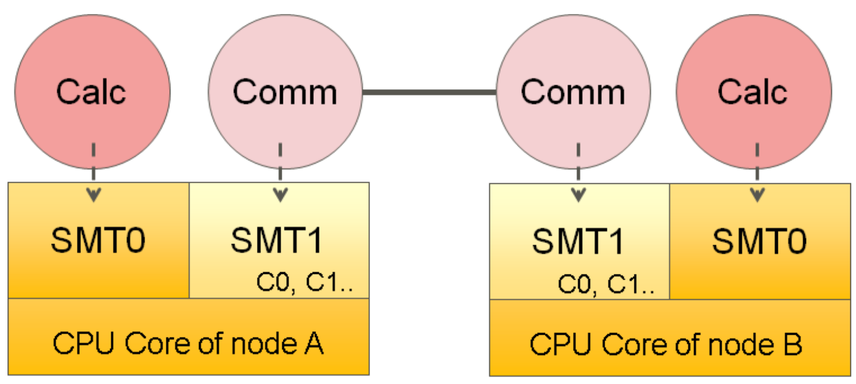
For computation-communication overlap, communications need to be progressed asynchronously.
We use Simultaneous MultiThread (SMT) for communication and apply MONITOR/MWAIT instruction for incoming message detection to avoid performance degradation of computation thread.
26% performance improvement of computation thread is achieved with only slightly increased latency.

1. Background:

- Computation time is shortened by using large scale PC cluster systems, accelerators; while communication time is relatively larger
- Computation-communication overlapping is used to reduce communication time and to achieve higher performance
  - ✓ Communications need to be asynchronously progressed
- Major progression methods in MPI:

Manual	・operate progression (call MPI_Test () repeatedly) in main calculation ・not practicable
Threaded	・use a separate thread ・communications are easily operated asynchronously, but performance problem is considered

2. Proposal:



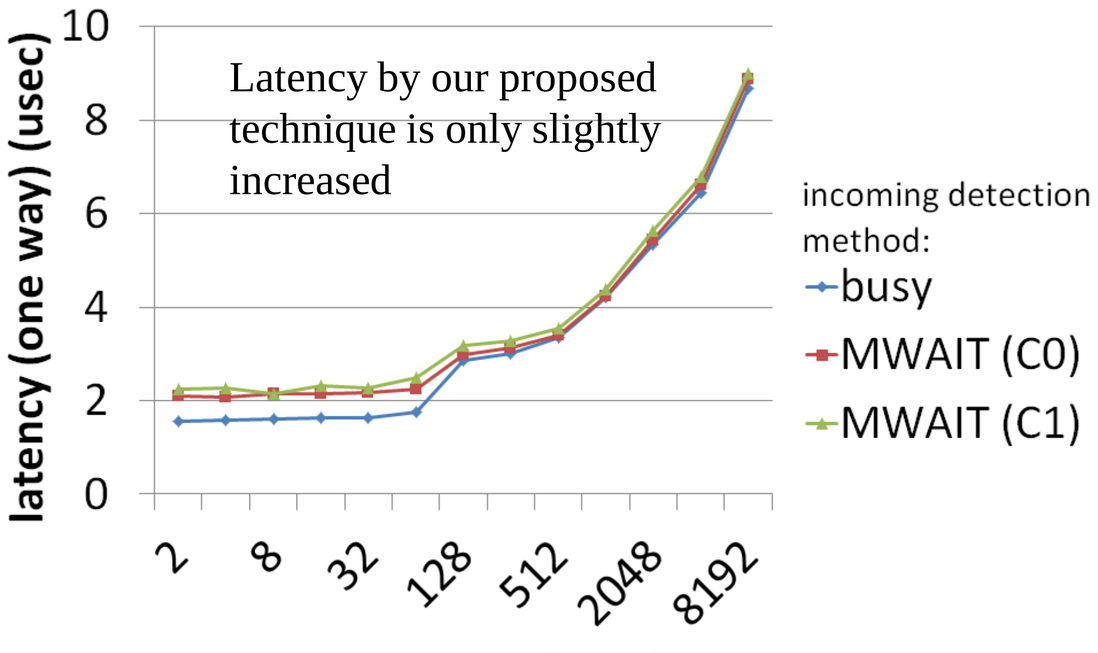
- Using SMT for communication thread with applying MONITOR/MWAIT instructions for incoming message detection
- SMT is used to avoid context switches.
- MONITOR/MWAIT instructions are applied
  - ✓ to avoid performance degradation due to commonly-used busy polling method
  - ✓ with processor low power state (C-state) consideration

Experiments are conducted in 2-node InfiniBand (ConnectX QDR) connected IA-servers system, Software setup includes RHEL5.4 for OS and openmpi 1.6.0 (gcc compiled) for MPI.

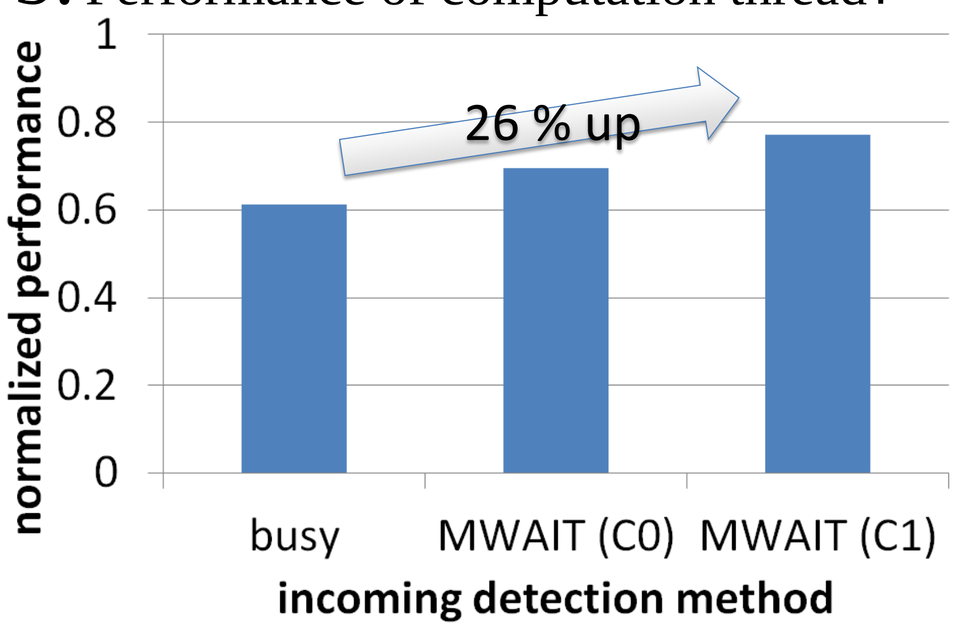
3. Experimental results:

- 26% performance improvement of computation thread with only slightly increased latency.

a. MPI pingpong Latency:



b. Performance of computation thread:



- MWAIT instruction with processor halt state (C1) largely improved performance of computation thread

Evaluation method:
Himeno Benchmark (sequential) performance is measured while the other thread is pingpong-ing.

4. Future plan

- Real benchmark application evaluation
- Comparison to other methods (e.g., dedicated physical core for communication)

FUJITSU | Fujitsu Laboratories Limited

shaping tomorrow with you