Fine Grain MPI

Earl J. Dodd

Humaira Kamal, Alan Wagner
@ University of British Columbia
Agenda

- Motivation
- Fine-Grain MPI
- Key System Features
- Novel Program Design.
• Introduction of multicore has changed the architecture of modern processors dramatically.
• Plethora of languages and frameworks have emerged to express fine-grain concurrency on multicore systems.
New Languages and Frameworks

concurrency  threads/processes  parallel

Axum
C++ Concurrency in Action
Erlang
Programming in Scala
HPX
Intel Threading Building Blocks
Pthreads
OpenMP
Java Concurrency in Practice
golang
Haskell

multicore  cluster
How to take advantage of multicore with seamless execution across a cluster?
Let $X = \text{MPI}$
FG-MPI: FINE-GRAIN MPI

- FG-MPI extends the execution model of the Message Passing Interface (MPI) to expose large-scale, fine-grain concurrency.
Decoupling an MPI process from an OS-level process.
FG-MPI System

- Has light-weight, scalable design integrated into MPICH middleware which leverages its architecture.
- Implements location-aware communication inside OS-processes and nodes.
- Allows the user to scale to millions of MPI processes without needing the corresponding number of processor cores.
- Allows granularity of MPI programs to be adjusted through the command-line to better fit the cache leading to improved performance.
- Enables design of novel algorithms and vary the number of MPI processes to match the problem rather than the hardware.
- Enables task oriented program design due to decoupling from hardware and support for function-level concurrency.
Executing FG-MPI Programs

Example of SPMD MPI program

- with 16 MPI processes,
- assuming two nodes with quad-core.

8 pairs of processes executing in parallel, where each pair interleaves execution.
Decoupled from Hardware

- Fit the number of processes to the problem rather than the number of cores.

```
mpiexec -nfg 250 -n 4 myprog
```
Flexible Process Mapping

- Flexibly move the boundary of MPI processes mapped to OS-processes, cores and machines.

\[
\text{mpiexec } -\text{nfg 1000 } -\text{n 4 myprog}
\]

\[
\text{mpiexec } -\text{nfg 500 } -\text{n 8 myprog}
\]

\[
\text{mpiexec } -\text{nfg 750 } -\text{n 4 myprog: } -\text{nfg 250 } -\text{n 4 myprog}
\]
Scalability

- Can have hundreds and thousands of MPI processes on a laptop or cluster.

```
mpiexec -nfg 30000 -n 8 myprog
```

- 100 Million processes on 6500 cores.

```
mpiexec -nfg 16000 -n 6500 myprog
```
Novel Program Design

- Modelling of emergent systems
  - Bird flocking.

- Distributed data structures
  - Every data item is an MPI process.
Dynamic Graph Applications

How to query large amounts of real-time data to extract relationship information?

- Twitter feeds
- Sensor data feeds
- Financial data

Companies with an Executive in common: Every dot represents a executive/director from a publicly listed company; People are connected to one another if they served the company at the same time.

FG-MPI
Distributed Skip-list with support for Range-querying

Scalable, using thousands of processors executing on over 200 cores
FG-MPI: A Finer Grain Concurrency Model for MPI

March 19, 2014 at 3:00 PM - 4:00 PM CT

Society of HPC Professionals (SHPCP)
http://hpcsociety.org/events?eventId=849789&EventViewMode=EventDetails
Thank You …

http://www.cs.ubc.ca/~humaira/fgmpi.html

or google “FG-MPI”

Dr. Alan Wagner
UBC
+1-604-822-6450
wagner@cs.ubc.ca

Dr. Humaira Kamal
UBC
+1-604-822-6450
kamal@cs.ubc.ca

Sarwar Alam
UBC
+1-604-827-3985
sarwar@cs.ubc.ca

Earl J. Dodd
Scalable Analytics Inc.
+1-713-446-4963
Earl.Dodd@scalableanalytics.com
Publications


