HYDRA: Using Windows Desktop Systems in Distributed Parallel Computing

Arvind Gopu, Douglas Grover, David Hart, Richard Repasky, Joseph Rinkovsky, Steve Simms, Adam Sweeny, Peng Wang

University Information Technology Services
Indiana University
Problem Description

- Turn Windows desktop systems at IUB student labs into a scientific resource.
  - 2300 systems, 3 year replacement cycle
  - 1.5 Teraflops
  - Fast ethernet or better
- Harvest idle cycles.
Constraints

- Systems dedicated to students using desktop office applications — not parallel scientific computing
- Microsoft Windows environment
- Daily software rebuild
What could these systems be used for?

- Many small computations and a few small messages
  - Master-worker
  - Parameter studies
  - Monte Carlo
Assembling small ephemeral resources

- Different parallel libraries have constraints of some form or the other
  - MPI not designed to handle ephemeral resources
Solution

- Simple Message Brokering Library (SMBL)
  - Limited replacement for MPI
- Process and Port Manager (PPM)

... Plus ...

- Condor NT
  - Job management
- Web portal
  - Job submission
The Big Picture
We’ll discuss each part in more detail next…

The shaded box indicates components hosted on multiple desktop computers
Portal

- Creates and submits Condor files, handles data files
- Apache based
- PHP web interface

http://hydra.indiana.edu
Condor

- Condor for Windows NT/2000/XP
  - “Vanilla universe” -- no support for checkpointing or parallelism
  - Provides:
    - Security
    - Match-making
    - Fair sharing
    - File transfer
    - Job submission, suspension, preemption, restart
SMBL

- In charge of message delivery for each parallel session
- Client library implements selected MPI-like calls
- Both server and client library based on TCP socket abstraction
SMBL (Contd … )

Managing Temporary Workers

- SMBL server maintains a dynamic pool of client process connections
- Worker job manager hides details of ephemeral workers at the application level
- Porting from MPI is fairly straightforward
Process and Port Manager (PPM)

- Assigns port/host to each parallel session
- Starts the SMBL server and application processes on demand
- Directs workers to their servers
Once again … the big picture

The shaded box indicates components hosted on multiple desktop computers.
System Layout

- PPM, SMBL server and web portal running on Linux server -- Dual Intel Xeon 1.7 GHz, 2 GB memory and GigE inter-connect
- STC Windows worker machines -- Combination of different OS (Windows 2000/XP) and network inter-connect speeds (GigE/100 Mbps/10 Mbps)
Applications

- **FastDNAml-p**
  - Parallel application, master-worker model, small granularity of work
  - Provides generic interface for parallel communication library (MPI, PVM, SMBL)
  - Reliability built in: Foreman process copes with delayed or lost workers

- **Blast**
- **Meme**
Applications – FastDNAml

INDIANA UNIVERSITY
Hydra Portal
A service of the Research and Academic Computing Division of Indiana University

Submit a FastDNAml job
Submit a MEME job
Submit a BLAST job
Check on job status
Check on cluster status
Contact support
Help

Thursday, September 29 2005

Your E-mail address:
agopus@indiana.edu

File to upload: (max 100000 bytes)
/home/agopus/projects/datafile1

Number of bootstrap replicates (optional, limit 300):
100

Number of jumbles per repetition (optional, limit 3):

Number of CPUs to run on:
You are currently limited to using 512 CPUs per job. (total jobs is equal to the number of replicates multiplied by the number of jumbles)
12

Submit
FastDNAml-p Performance

![Graph showing performance of FastDNAml-p with different numbers of processors. The x-axis represents the number of processors, ranging from 0 to 300. The y-axis represents time, ranging from 1 to 1,000,000. Two lines are plotted: one for Research SP and one for Condor Cluster. The Research SP line shows a steeper decrease in time as the number of processors increases, while the Condor Cluster line shows a gentler decrease.]
Other Applications – Parallel MEME
Other Applications – BLAST
Utilization of Idle Cycles

Red: total owner  Blue: total idle  Green: total Condor
Work in Progress/Future Direction

- Teragrid’ize Hydra cluster – allow TG users to access resource
- New Portal – JSR 168 compliant, certificate based authentication capability
- Range of applications – Virtual machines, so forth
Summary

- Large parallel computing facility created at very low cost
  - SMBL parallel message passing library that can deal with ephemeral resources
  - PPM port broker that can handle multiple parallel sessions
Links and References

- Hydra Portal – http://hydra.indiana.edu
- Condor home page -- http://www.cs.wisc.edu/condor/
- IU Teragrid home page – http://iu.teragrid.org
- Parallel FastDNAml – http://www.indiana.edu/~rac/hpc/fastDNAml
- Meme -- http://meme.sdsc.edu/meme/intro.html