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

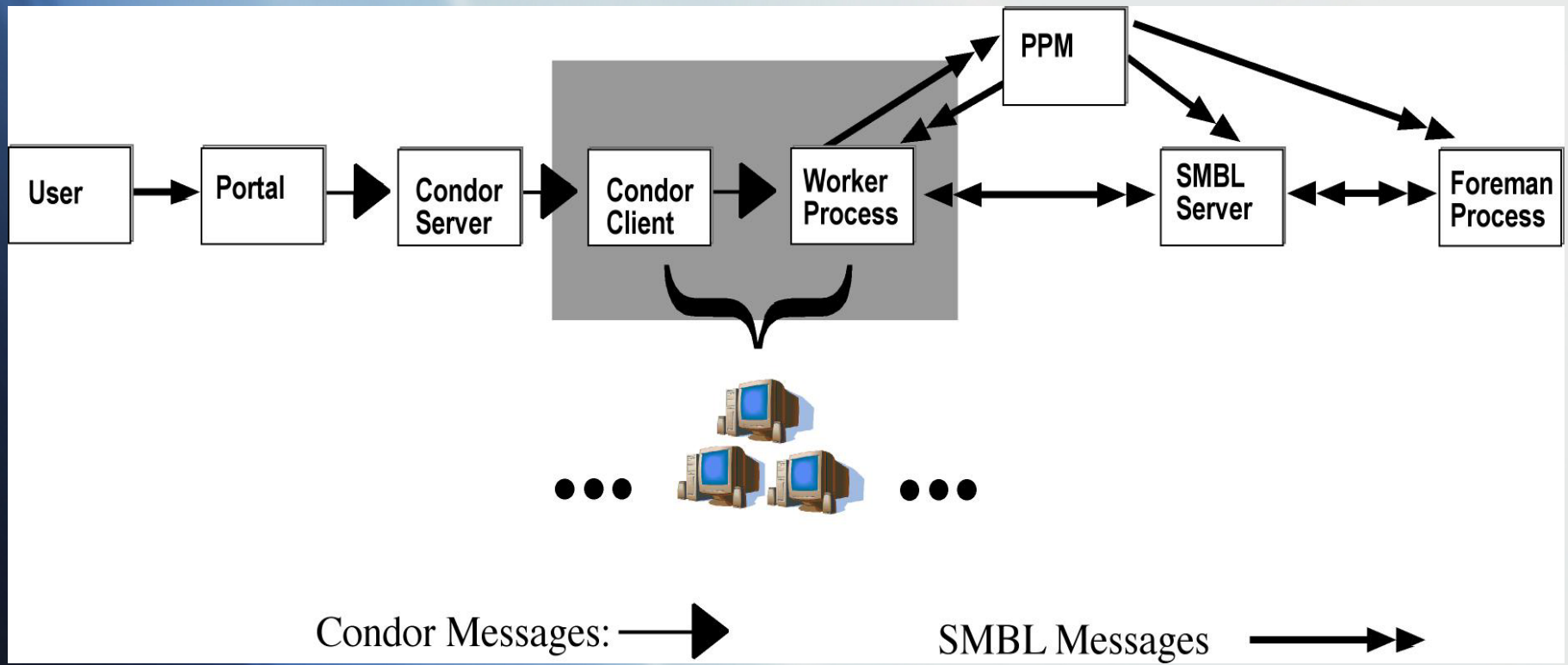
- Simple Message Brokering Library (SMBL)
  - Limited replacement for MPI
- Process and Port Manager (PPM)

… Plus …

- Condor NT
  - Job management
- Web portal
  - Job submission

The Big Picture

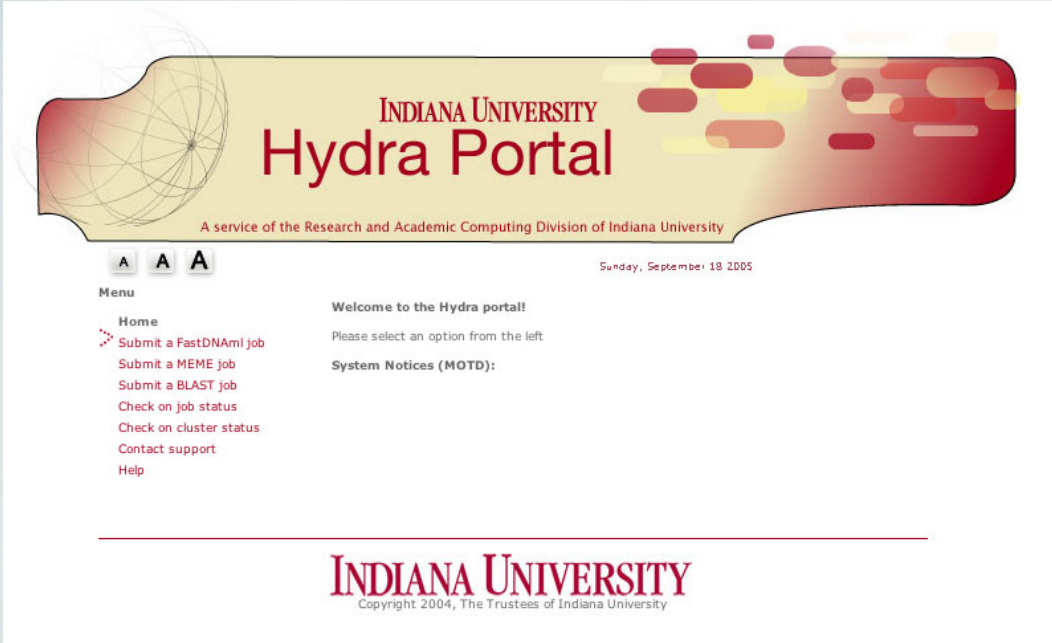
We'll discuss each part in more detail next…



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 check-pointing 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

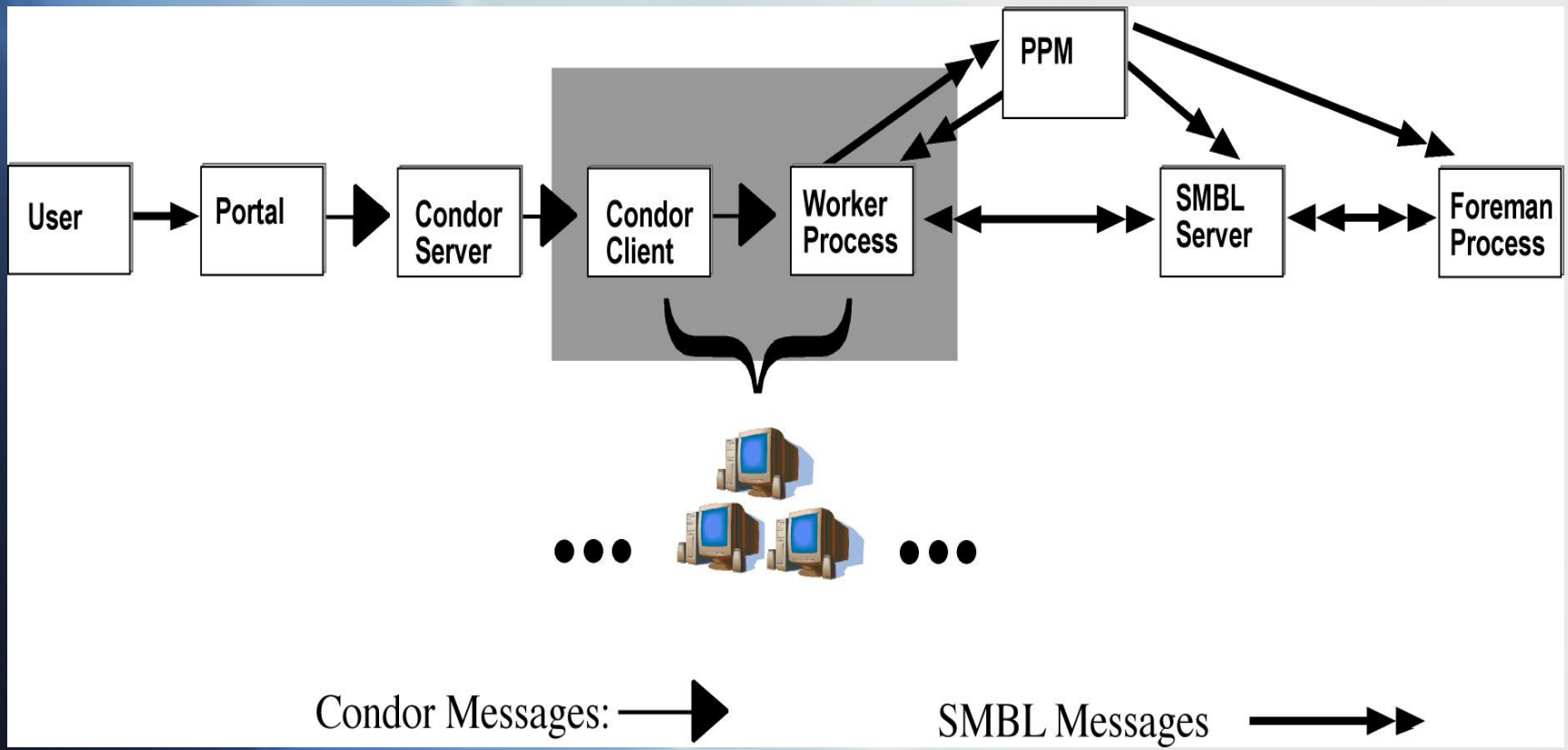
SMBL (Contd …) Managing Temporary Workers

- 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 straight forward

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

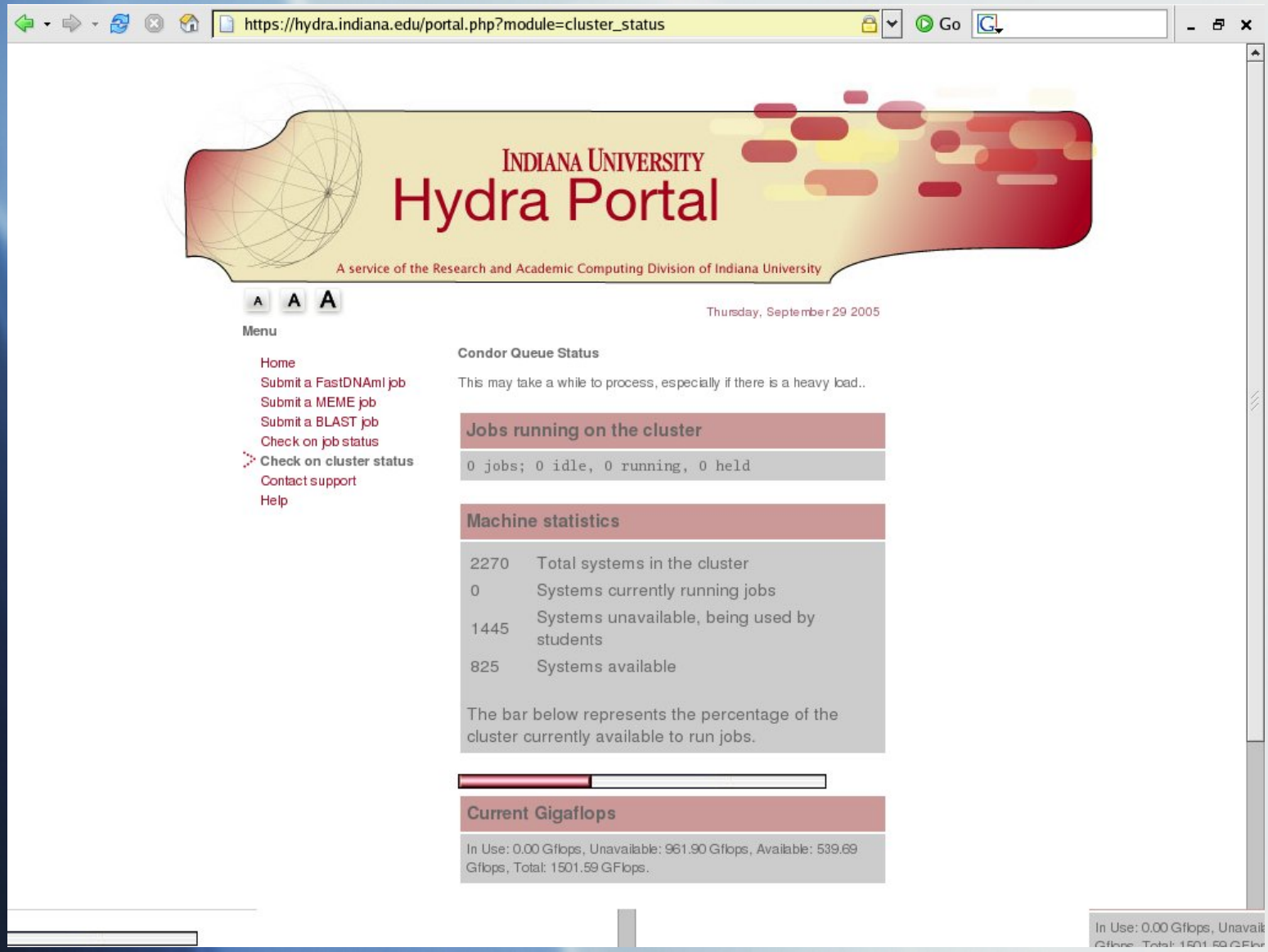
System Layout

- 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)

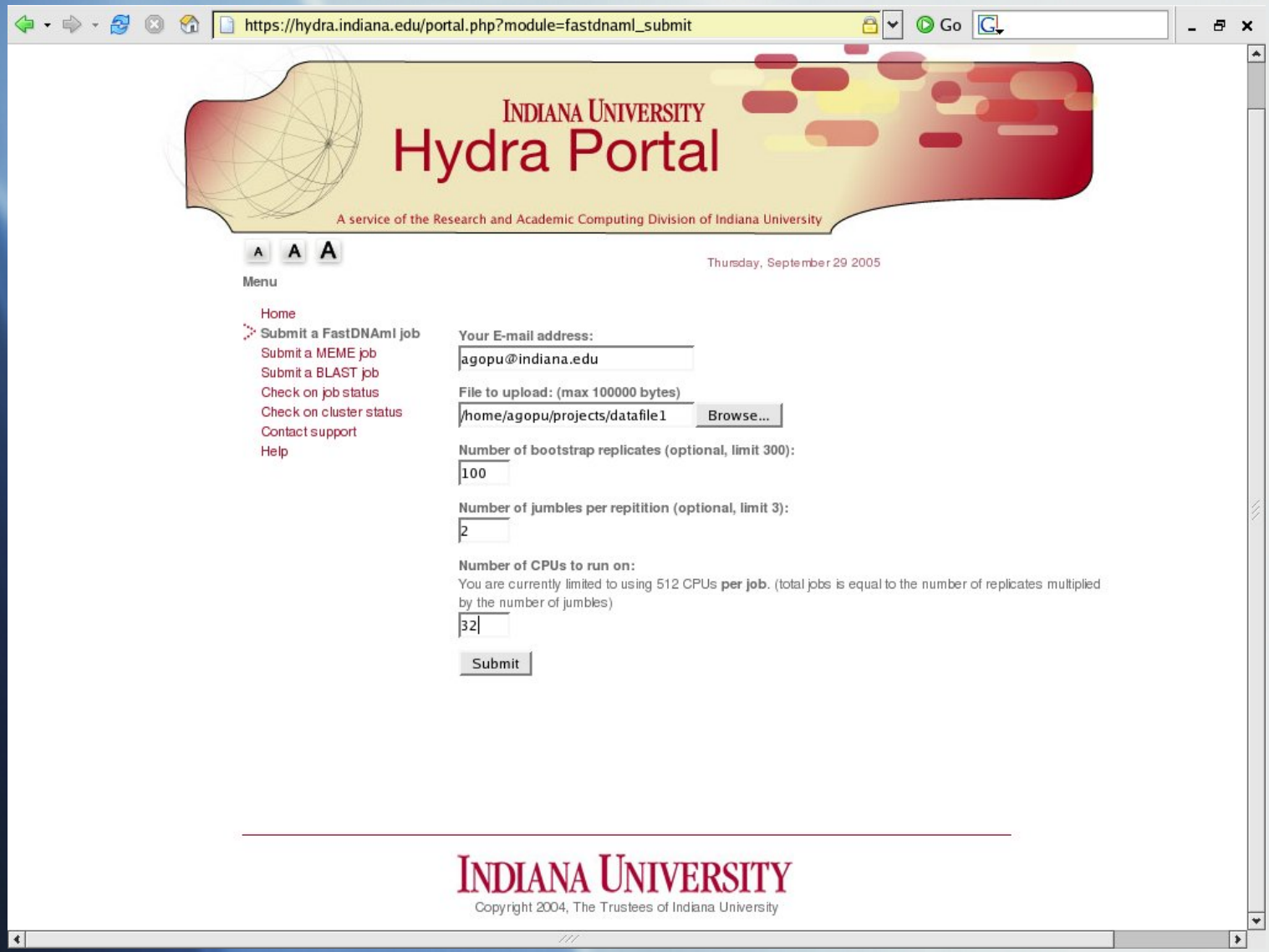
Applications

- 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

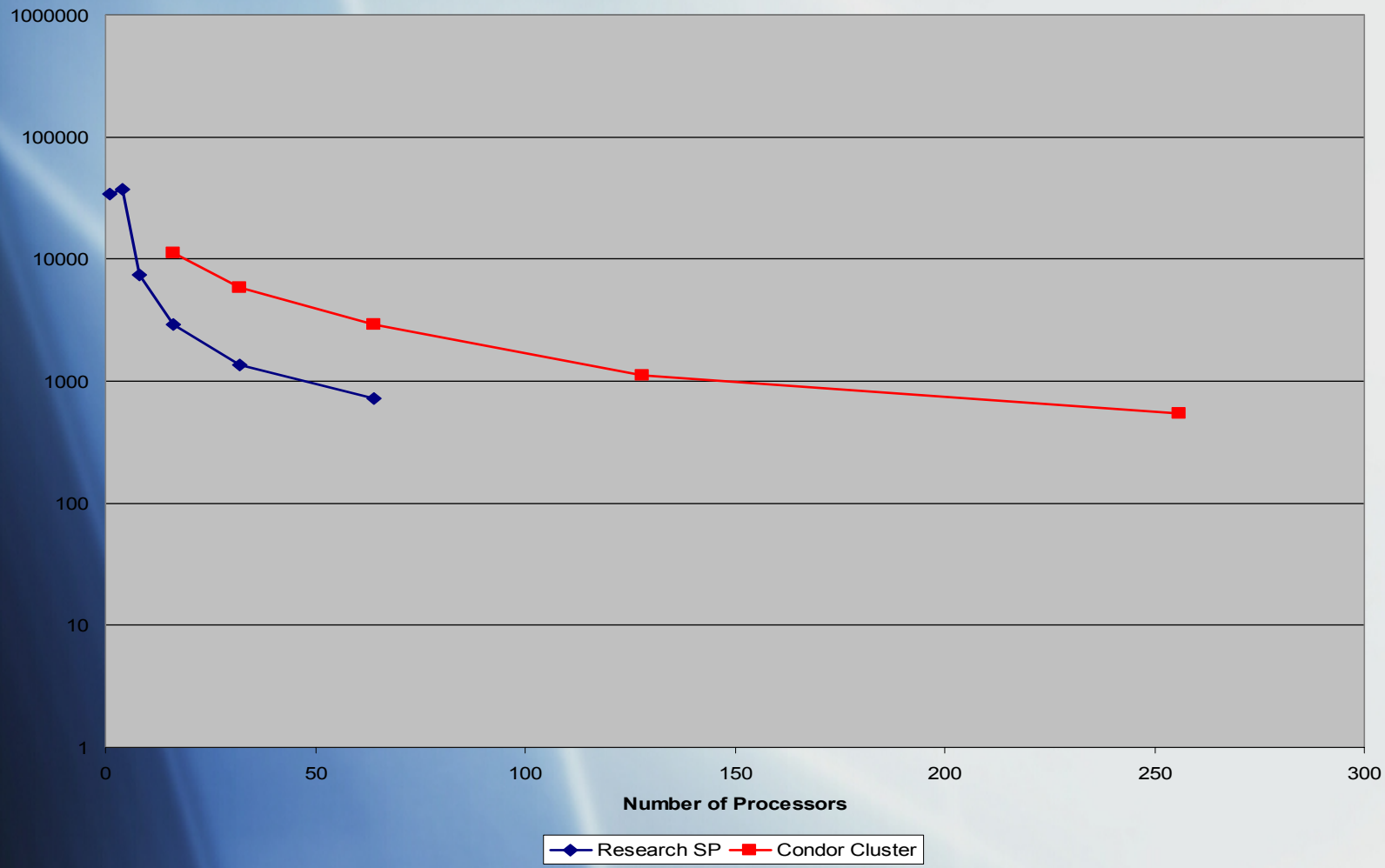
Portal


https://hydra.indiana.edu/portal.php?module=cluster_status
Go
INDIANA UNIVERSITY
Hydra Portal
A service of the Research and Academic Computing Division of Indiana University
A A A
Thursday, September 29 2005
Menu
Home
Submit a FastDNAml job
Submit a MEME job
Submit a BLAST job
Check on job status
Check on cluster status
Contact support
Help
Condor Queue Status
This may take a while to process, especially if there is a heavy load..
Jobs running on the cluster
0 jobs; 0 idle, 0 running, 0 held
Machine statistics
2270 Total systems in the cluster
0 Systems currently running jobs
1445 Systems unavailable, being used by students
825 Systems available
The bar below represents the percentage of the cluster currently available to run jobs.
Current Gigaflops
In Use: 0.00 Gflops, Unavailable: 961.90 Gflops, Available: 539.69 Gflops, Total: 1501.59 GFlops.

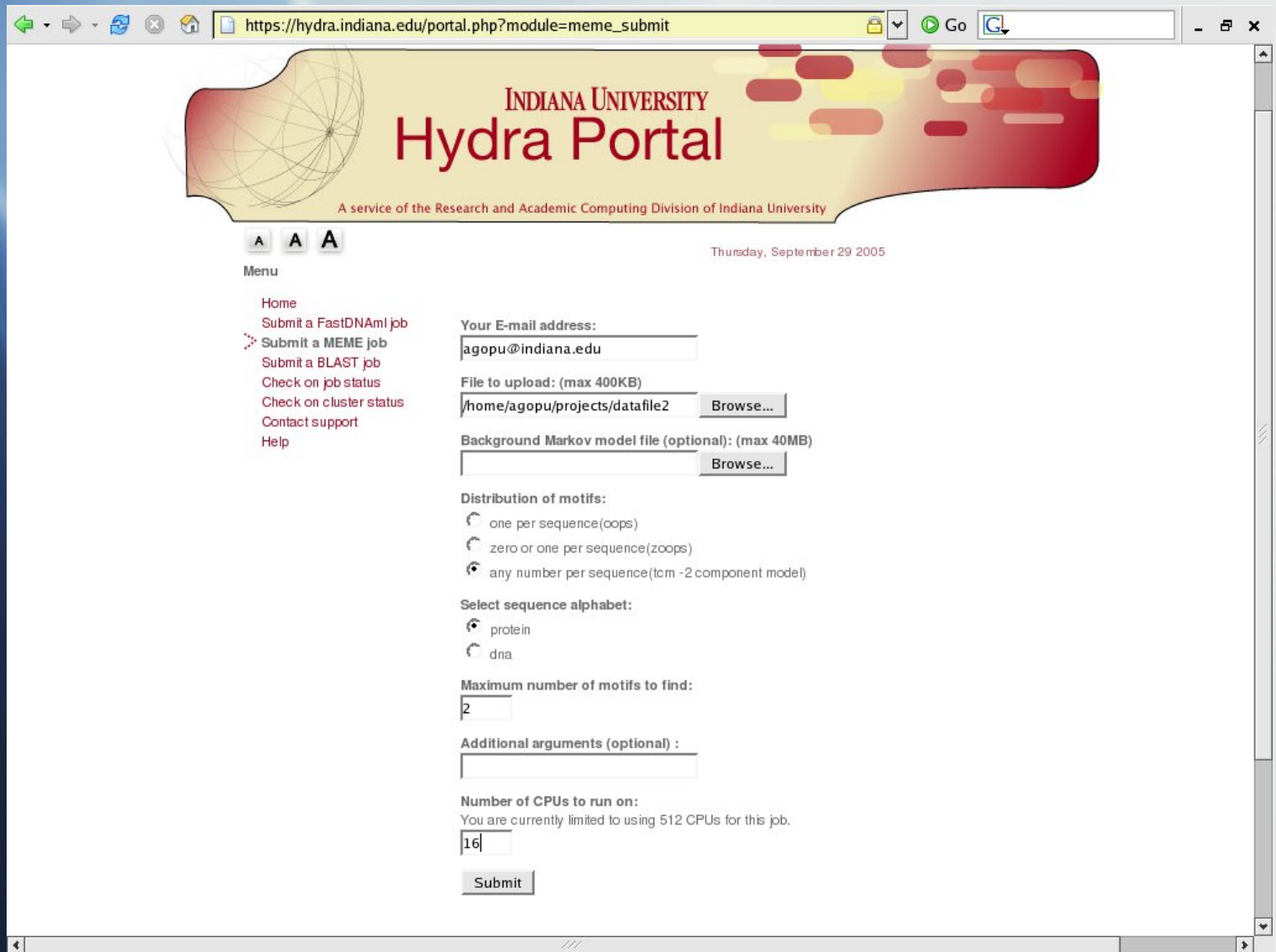
Applications – FastDNAml



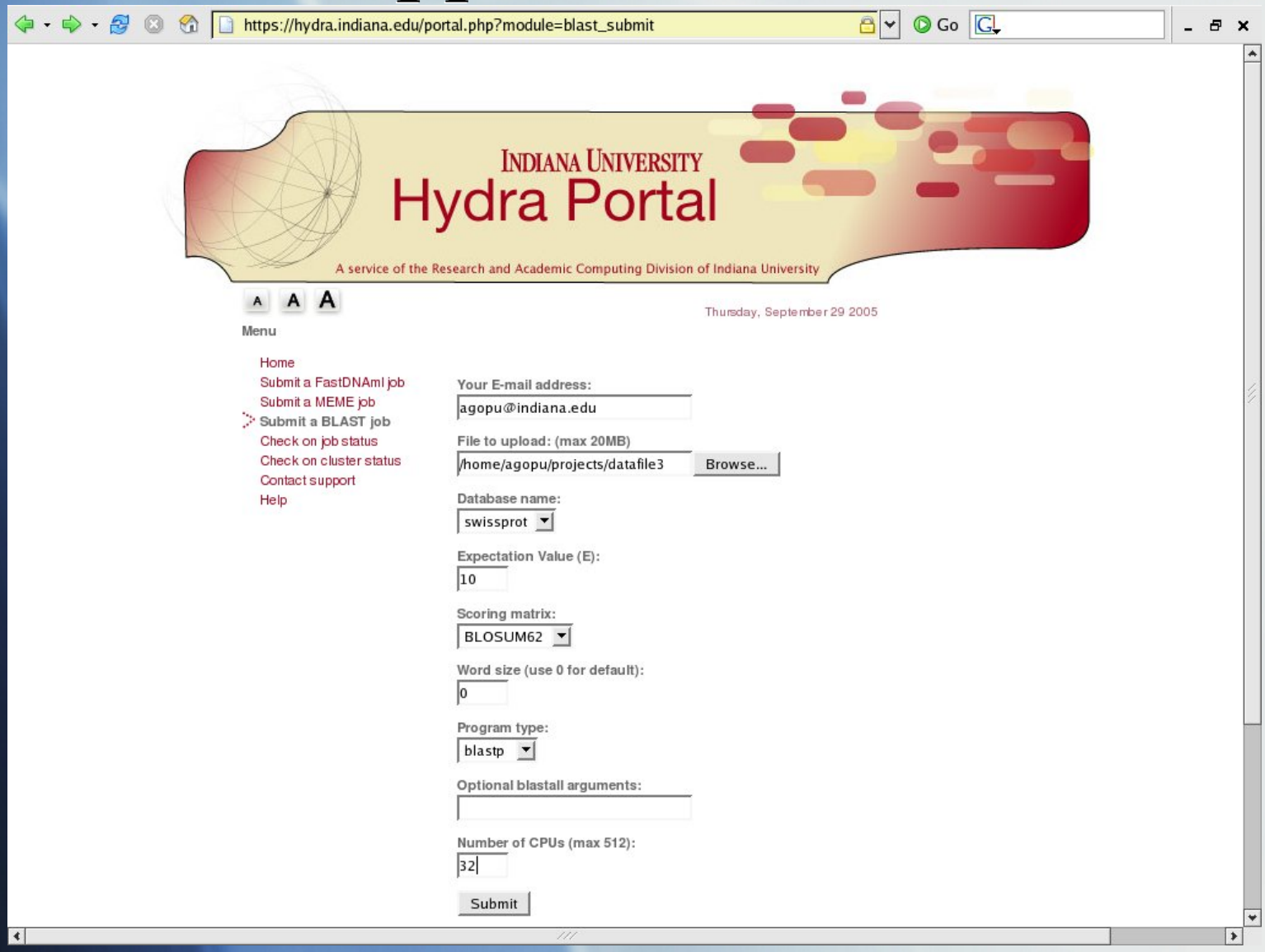
FastDNAml-p Performance



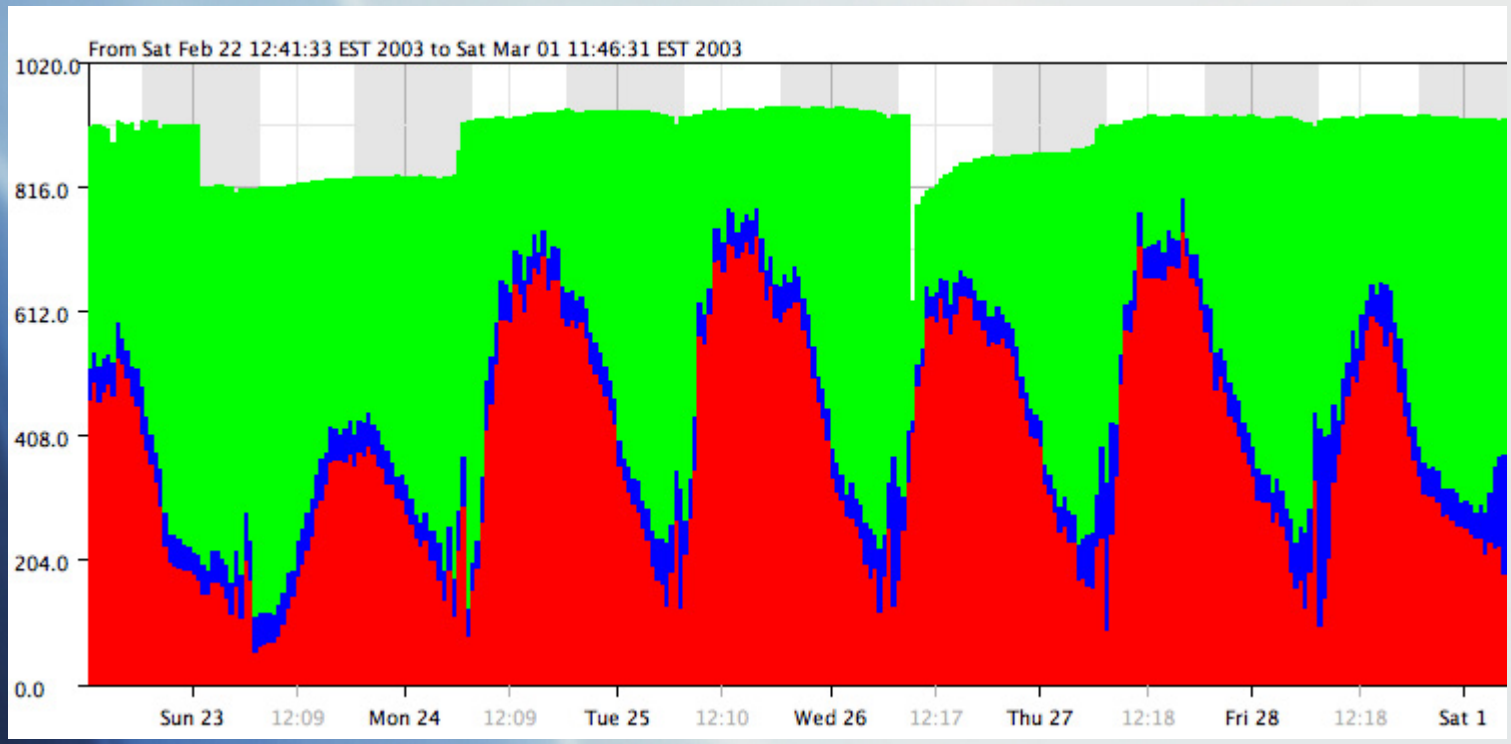
Other Applications – Parallel MEME



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
- SMBL (Open Source) Home – http://smbl.sourceforge.net

Links and References

- Hydra Portal – http://hydra.indiana.edu
- SMBL home page – http://smbl.sourceforge.net
- 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
- Blast -- http://www.ncbi.nlm.nih.gov/BLAST
- Meme -- http://meme.sdsc.edu/meme/intro.html