DR&E LLGrid Portal
Interactive Supercomputing for DoD

Albert Reuther, William Arcand, Chansup Byun, Bill Bergeron, Matthew Hubbell, Jeremy Kepner, Andrew McCabe, Peter Michaleas, Julie Mullen & Andrew Prout
MIT Lincoln Laboratory

HPEC Workshop
September 15, 2010

This work is sponsored by the Department of the Air Force under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.
DR&E Portal Prototype

Best of desktop + Best of supercomputing

Interactive … “what if scenarios”

Good for experts, great for novices

• HPCMP selected LLGrid for DoD wide prototype DR&E Portal
• Prototype goal: interactive pMatlab on a modest cluster (TX-DoD) over DREN alpha users with CAC authentication
Outline

• Introduction
  • Design Overview
  • Technologies
  • Summary

• LLGrid
• Interactive Supercomputing
• Parallel Matlab
What is LLGrid?

Best of desktop + Best of supercomputing

Interactive … “what if scenarios”

Good for experts, great for novices

- LLGrid is a ~400 user ~2000 processor system
- World's only desktop interactive supercomputer
  - Dramatically easier to use than any other supercomputer
  - Highest fraction of staff using (20%) supercomputing of any organization on the planet
LLGrid Interactive Supercomputing

- **Classic supercomputing**: Jobs take hours/days to run but jobs tolerate waiting in a queue.

- **Interactive supercomputing**: Jobs are large requiring answers in minutes/hours but cannot tolerate waiting in a queue.

- **Desktop computing**: Jobs take minutes on a desktop (e.g., algorithm proof-of-principles).
Why is LLGrid easier to use?

Universal Parallel Matlab programming

```matlab
Amap = map([Np 1],{},0:Np-1);
Bmap = map([1 Np],{},0:Np-1);
A = rand(M,N,Amap);
B = zeros(M,N,Bmap);
B(:,:, :) = fft(A);
```

- pMatlab runs in all parallel Matlab environments
- Only a few functions are needed
  - Np
  - Pid
  - map
  - local
  - put_local
  - global_index
  - agg
  - SendMsg/RecvMsg

- Distributed arrays have been recognized as the easiest way to program a parallel computers since the 1970s
  - Only a small number of distributed array functions are necessary to write nearly all parallel programs
- LLGrid is the first system to deploy interactive distributed arrays
Outline

• Introduction
• Design Overview
• Technologies
• Summary

• Requirements
• Phases
• Architecture
Requirements for DR&E Portal

- Cannot utilize any new networking ports
  - Hypertext Transport Protocol (http) – port 80
  - Secure Sockets Layer (ssl) – port 443
- Cannot install new software on desktop computers
- Dual-layer authentication
  - CAC Card with SSL certificates
  - PIN authentication
- Traverse multiple organizations over DREN
- Isolate users accounts from each other
- Intuitive to go from serial to parallel coding
- Desktop computer is one of computational workers
Prototype Components:
Pre-alpha cluster (TX-DoD)

• Provided an icon on scientists' and engineers’ desktops that provides them tools to do their jobs faster
  – pMatlab is first tool in the suite (extensible over time)

• Dedicated cluster at LL on DREN
  – 40 node blade system along with 8 TB of parallel storage

• Used for initial development
  – LLGrid software stack deployed and modified to work in HPCMP environment based on requirements

• Software stack copied to alpha cluster

• Maintained as a mirror system for development purposes
Prototype Components: alpha cluster testbed

- Experimental testbed on DREN
- Used for trials with alpha users
- Software stack was copied from pre-alpha and modified based on trials; changes folded back to alpha mirror (TX-DoD)
- Software stack copied to beta system
User Desktops
- Windows 7 and Mac OS X supported
- Portal connection options: WebDAV over https (port 443)

Cluster Management: LL-modified Rocks 5.2
Cluster Parallel File System: Lustre 1.8.1
Scheduler: Sun Grid Engine (SGE)

Login and Compute Nodes (15 GB image size)
- last 5 versions of Matlab, Octave, pMatlab, GridMatlab,
- lammpi, mpich, mpich2, mvapich, openmpi
1. Access Secure Portal
2. CAC Authentication Requested
3. Provide CAC with PIN
4. Credential Approved
5. Map User’s Home
6. Submit a job with a protocol file
7. Portal Watcher gets notified

8. Read & parse job description in XML
9. Send the job to scheduler via DRMAA
10. Job scheduled and dispatched
11. Job ID returned in a protocol file
12. Job ID displayed on the client system
13. Output generated and stored
Outline

• Introduction

• Design Overview

• Technologies
  • Key Components
  • Component Descriptions

• Summary
Prototype Architecture
Key Components

1. CAC-Enabled Apache WebDAV Server
2. Linux File System Watcher
3. gridMatlab for Portal
4. grsecurity Kernel Patches
Prototype Architecture
CAC-Enabled Apache WebDAV Server

- WebDAV provides file system services across HTTP (80)
- Apache server authenticates via CAC
- Required significant modification to Apache Web Server
- File Access Monitor in Linux kernel (2.6.25+)
- Receive event notification when file events occur
- Configure actions based on file name, directory, etc.
- Enables activities to launch jobs, abort jobs, etc.
Prototype Architecture
gridMatlab for Portal

Secure Portal Technology

- Launch jobs, abort jobs, etc. by writing files to WebDAV file system
- Defined rich XML file formats for each action

Client Systems

Web Server

Scheduler

Portal Watcher

Grid

Authentication

Storage
- Role-Based Access Control (RBAC) system
- Users can only view own processes, files, etc.
- Extensive auditing and logging
- Randomization of the stack, library, heap and kernel bases
- Prevention of arbitrary code execution
Speed-up for Example Code 1 running on Lincoln Laboratory Grid (LLGrid) (Matlab/pMatlab)

<table>
<thead>
<tr>
<th>Nprocs</th>
<th>Max Time (secs)</th>
<th>Average Time (secs)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>178972.64</td>
<td>17897.64</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>25247.58</td>
<td>17448.12</td>
<td>7.08</td>
</tr>
<tr>
<td>20</td>
<td>14825.30</td>
<td>8767.93</td>
<td>12.1</td>
</tr>
<tr>
<td>36</td>
<td>7589.20</td>
<td>4832.97</td>
<td>23.6</td>
</tr>
</tbody>
</table>
Speed-up for Example Code 2 running on Lincoln Laboratory Grid (LLGrid) (Matlab/pMatlab)

Code with MATLAB Speedup on LLGrid, nreps = 1000

<table>
<thead>
<tr>
<th>Nprocs</th>
<th>Max Time (secs)</th>
<th>Average Time (secs)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>176344.0309</td>
<td>176344.0309</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>6088.6758</td>
<td>5471.0087</td>
<td>29</td>
</tr>
<tr>
<td>50</td>
<td>3245.6632</td>
<td>2699.6459</td>
<td>54</td>
</tr>
<tr>
<td>100</td>
<td>1673.5188</td>
<td>1341.6794</td>
<td>105</td>
</tr>
<tr>
<td>200</td>
<td>803.9898</td>
<td>658.7564</td>
<td>219</td>
</tr>
</tbody>
</table>
Outline

• Introduction

• Design Overview

• Technologies

• Summary
Summary

• DR&E Portal technology enables interactive, on-demand parallel Matlab from DoD desktops
  – Required Zero Footprint LLGrid
  – Several phase rollout
• Four key technologies
  – CAC-enabled Apache WebDAV Server
  – Linux File System Watcher
  – gridMatlab for Portal
  – grsecurity Kernel Patches
• Performance does not impede user experience
Backups
HPCMP DR&E Portal Prototype Demo Using LLGridZF (zero footprint)

Steps:

Account use
Please insert your smart card. 

New Users

- If you would like to request an account simply click the request button.

Request Account
Menu: Login | Logout | Status

Cookie set to: deb355cf48f8dd4af345accdea22fb06036f1b91
value=
  UserName=/C#US/O#U.S.
  Government/OU#DoD/OU#PKI/OU#CONTRACTOR/CN#HUBBELL.MATTHEW.LANGAN.1391537654
  Groups=
  RemoteIP=140.31.214.115
  LoginTime=05/14/2010 18:07:46
  ClientCertDN=/C#US/O#U.S.
  Government/OU#DoD/OU#PKI/OU#CONTRACTOR/CN#HUBBELL.MATTHEW.LANGAN.1391537654
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now customized. In addition, many keyboard shortcuts have changed for improved consistency across the desktop.

To customize keyboard shortcuts, use Preferences. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, click here if you do not want to see this message again.
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now consistent across the desktop.

In addition, many keyboard shortcuts have changed for improved usability.

To customize keyboard shortcuts, use Preferences. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, click here if you do not want to see this message again.

>> cd x:
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now custom.

In addition, many keyboard shortcuts have changed for improved
across the desktop.

To customize keyboard shortcuts, use Preferences. From there,
restore previous default settings by selecting "R2009a Windows
from the active settings drop-down list. For more information,

Click here if you do not want to see this message again.

>> cd x:

>> cd ..\Examples\Blurimage/
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now customized to your settings.

In addition, many keyboard shortcuts have changed for improved usability across the desktop.

To customize keyboard shortcuts, use Preferences. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, Click here if you do not want to see this message again.

```plaintext
>> cd x:
```

```plaintext
>> cd ../../../Examples/Blurimage/
```
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now customizable. In addition, many keyboard shortcuts have changed for improved consistency across the desktop.

To customize keyboard shortcuts, use Preferences. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, Click here if you do not want to see this message again.

```matlab
>> cd x:
>> cd ./Examples/Blurimage/
```

```matlab
>> eval(pRUN('pBlurimage',4,'portal'))
```
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now customizable. In addition, many keyboard shortcuts have changed for improved consistency across the desktop.

To customize keyboard shortcuts, use Preferences. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, click here if you do not want to see this message again.

```matlab
>> cd x;
>> cd ./Examples/Blurimage/
>> eval(pRUN('pBlurimage',4,'portal'))
```
MATLAB desktop keyboard shortcuts, such as Ctrl+S, are now customized. In addition, many keyboard shortcuts have changed for improved consistency across the desktop.

To customize keyboard shortcuts, use `Preferences`. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, click here if you do not want to see this message again.

```matlab
>> cd x:
>> cd ../Examples/Blurimage/
>> eval(pRUN('pBlurimage',4,'portal'))
```

Substituting `pBlurimage` on 4 processor(s).

`MPI_Abort_portal`: TODO: Wait until watcher confirms that all tasks have completed.

`MPI_Abort_portal`: TODO: Also need to print out the `MPI_Abort` status.

`TODO: MPI_Abort_portal.m: Generate MatMPI_log_entry call`
To customize keyboard shortcuts, use Preferences. From there, restore previous default settings by selecting "R2009a Windows" from the active settings drop-down list. For more information, click here if you do not want to see this message again.

```
>> cd x:
>> cd ./Examples/Blurimage/
>> eval(pRUN('pBlurimage',4,'portal'))
Submitting pBlurimage on 4 processor(s).
MPI_Abort_portal: TODO: Wait until watcher confirms that all
MPI_Abort_portal: TODO: Also need to print out the MPI_Abort
TODO: MPI_Abort_portal.m: Generate MatMPI_log_entry call
Warning: Directory already exists.
> In MPI_RunG at 105
    In MPI_Run at 94
    In pRUN at 44
TODO: MatMPI_grid_resource_policy.m - Need to check grid avai
TODO: MatMPI_grid_resource_policy.m - Need to return total
TODO: MatMPI_grid_resource_policy.m - Need to return unclai
Launching MPI rank: 3 on: portal_03
Launching MPI rank: 2 on: portal_02
Launching MPI rank: 1 on: portal_01
Launching MPI rank: 0 on: DREN_PC
```
MATLAB 7.10.0 (R2010a)

New to MATLAB? Watch this Video, see Demos, or read Getting Started.

```matlab
>> eval(pRUN('pBlurimage',4,'portal'))
Submitting pBlurimage on 4 processor(s).
MPI_Abort_portal: TODO: Wait until watcher confirms that all
MPI_Abort_portal: TODO: Also need to print out the MPI_Abort_portal
TODO: MPI_Abort_portal.m: Generate MatMPI_log_entry call
Warning: Directory already exists.
> In pRUN at 105
  In MPI_RunG at 94
  In pRUN at 44
TODO: MatMPI_grid_resource_policy.m - Need to check grid ava
TODO: MatMPI_grid_resource_policy.m - Need to return total
TODO: MatMPI_grid_resource_policy.m - Need to return unclai
Launching MPI rank: 3 on: portal_03
Launching MPI rank: 2 on: portal_02
Launching MPI rank: 1 on: portal_01
Launching MPI rank: 0 on: DREN_PC
Job <66> launched via SGE-HPC
Time to write XML file to notify a job submission = 0.8293
Time to confirm a job launch = 49.759
Job submitted to LLgrid via SGE-HPC resource manager
TODO: MatMPI_log_entry need to be update to log entries
TODO: MatMPI_log_entry need to be update to log entries
Launch Time (sec) = 0.7072
Compute Time (sec) = 7.6696
Performance (Gigaflops) = 4.48
```
Phases

• **Prototype**
  - On-demand interactive parallel MATLAB delivered to alpha/beta users

• **Phase I**
  - On-demand interactive parallel MATLAB delivered to DoD researchers and engineers

• **Phase II**
  - A suite of on-demand interactive applications and an easy-to-use batch environment delivered to DoD researchers and engineers