Using Grid Computing Within the Department of Defense

June 24, 2008

LtCol Peter Garretson

Chief, Future Technology Branch

HQ USAF*

Mr. Richard Kundiger

Cyberspace Architect

Defense Contract Management Agency*

*** Information in this brief does not represent any official position of the U.S. Air Force or the Defense Contract Management Agency. It is solely a thought piece from the authors.**

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
24 JUN 2008	**Technical Briefing**	**01-06-2008 to 30-06-2008**

4. TITLE AND SUBTITLE
Using Grid Computing Within the Department of Defense

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Richard Kundiger; Peter Garretson

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Defense Contract Management Agency,6350 Walker Ln,Alexandria,VA,22310

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Future Technology Branch, USAF, HAF/A8XC Air Force, Pentagon, Washington, DC

10. SPONSOR/MONITOR'S ACRONYM(S)
HAF A8XC DCMA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
*** Information in this brief does not represent any official position of the U.S. Air Force or the Defense Contract Management Agency. It is solely a thought piece from the authors.**

14. ABSTRACT
The Department of Defense has massive amounts of data that needs analysis. Purchase and maintenance of supercomputers to process this information, or leasing of supercomputer resources, is very expensive. DoD already has a supercomputer available to it via distributed or grid computing by making use of the millions of PCs within the department. Using CPU and memory scavenging agents DoD has to potential to harness over 70 Petaflops of compute capacity using existing equipment with minimal additional investment. This platform could then be used to perform advanced modeling, advanced simulation, deep data mining and evolutionary analysis.

15. SUBJECT TERMS
grid cloud computing advanced modeling simulation deep data mining evolutionary iterative analysis large data sets compute node boinc hpc high process computing supercomputer flops peta giga parallel parametric predictive tera roadrunner los alamos hadoop google ibm mapreduce BOINC LHC SETI distributed

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**3**	**8**	

Grid Computing Within DoD

What is Grid Computing?

- Grid Computing is a form of computing that makes use of computer resources across many disparate platforms and devices to perform work

- Allows massive parallelization of tasks and entirely new kinds of processing by pooling the total resources from many machines to create a supercomputing environment from common, inexpensive, hardware

- It can be configured to scavenge “spare” CPU cycles, unused cores, GPU’s, Memory, and disk storage to leverage DoD’s existing, under-utilized, infrastructure

Grid Computing Within DoD

Background

- The Department of Defense has massive amounts of data that needs analysis
  - Existing methods of analyzing this data are slow and ineffective
  - Many disparate systems are purchased or leased to perform analysis
- Modeling and simulation is typically outsourced
  - Computers capable of doing advanced modeling and simulation are very expensive to purchase, manage, house, cool and power
- DoD's problems are not unique. Research, Industry, and Academia are increasingly turning to Grid Computing

Grid Computing Within DoD

Premise

- Provide a platform that allows DoD to self perform:
  - Advanced modeling
  - Advanced simulation
  - Deep data mining and analysis of extremely large data sets.
  - Iterative or “evolutionary” analysis

- Do it by using untapped, existing computational potential

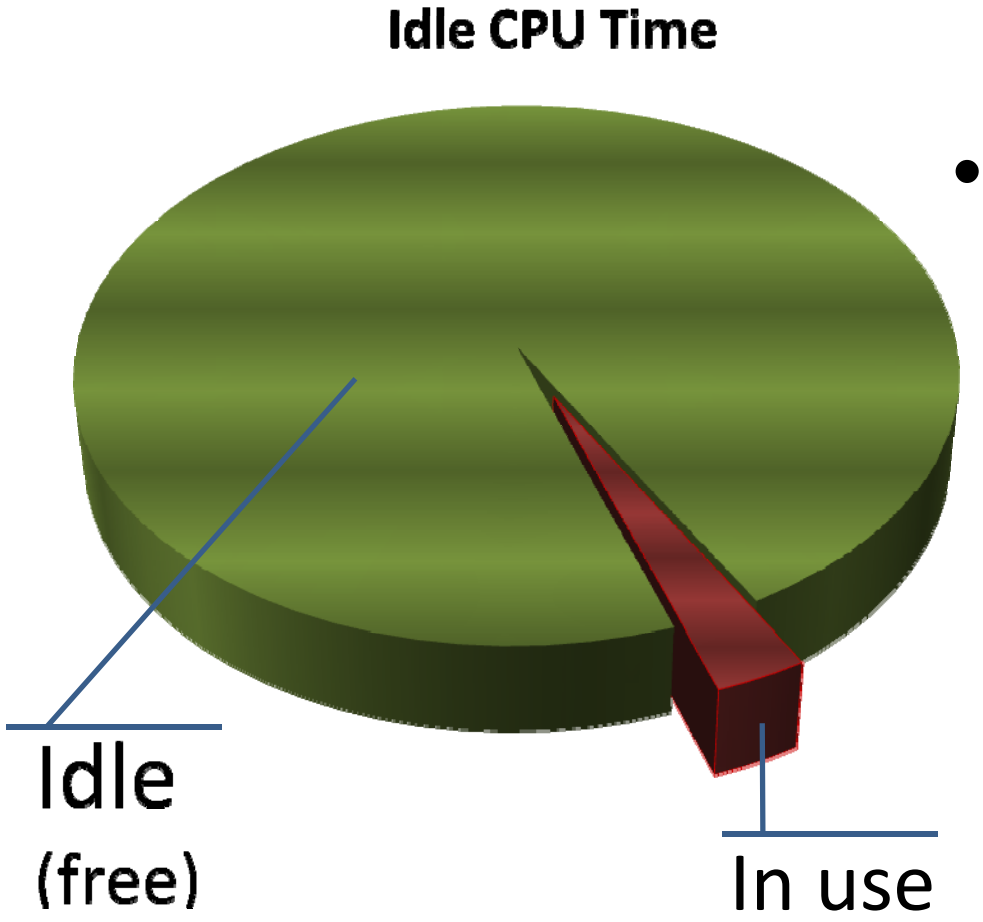
Grid Computing Within DoD

Premise

- There is a huge untapped potential
- CPU time is the amount of time the CPU is actually executing instructions
- The rest of the time the CPU sits idle, awaiting system or user instructions

Grid Computing Within DoD

Premise



- CPU time at idle is typically between 3%-7% leaving 93-97% free

Grid Computing Within DoD

Premise

- Many DoD PC hard drives go unused, or very minimally used, due to availability of networked storage
  - This may lead to 10's or 100's of free Gigabytes of space
- Computer memory (RAM) is increasing leaving hundreds of Megs, if not Gigs, of unused memory at idle
- Gartner identifies Grid computing as one of the top ten most disruptive technologies to shape information technology over the next five years.

Gartner

Grid Computing Within DoD

Premise

- Modern computers can provide a conservative estimate of 3 GigaFLOPS each of computing potential.
  - A GigaFLOP is a billion floating point operations per second
- DoD has approximately 5 million computers
- A PetaFLOP is One Quadrillion Floating Point Operations / Second
- Until June of 2009 no computer had ever exceeded the PetaFLOP level.

Grid Computing Within DoD

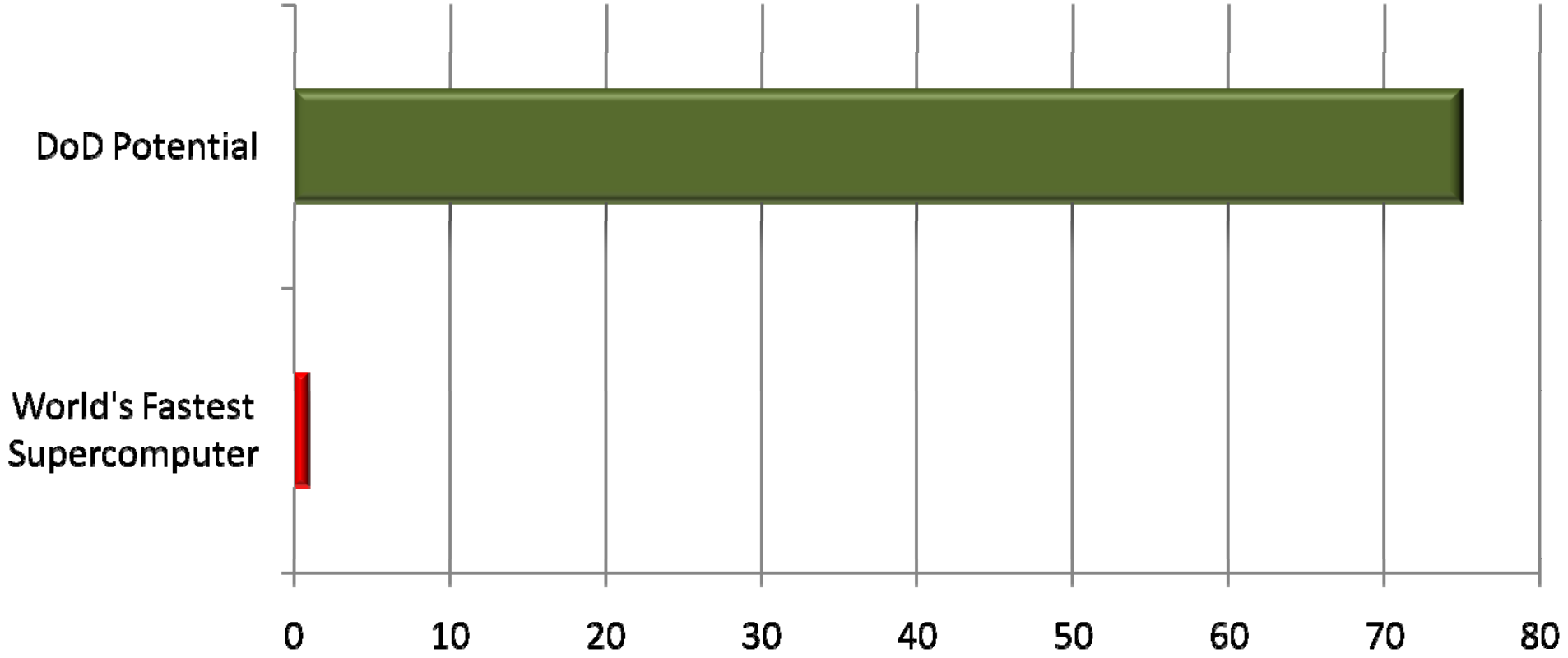
Premise

- The largest supercomputer in the world today is 1 PetaFLOP
- DoD's theoretical computing potential is 150 PetaFLOPS
- 75 PetaFLOPS if only used half the day as compute nodes in a Grid during off hours
  - (3GigaFLOPS * 5,000,000 PCs) / 2 (12 hours for work, 12 for Grid computing) = 75 PetaFLOPS

Grid Computing Within DoD

Premise

- DoD already owns the largest supercomputer in the world 75 fold



Grid Computing Within DoD

Examples in Research

- Search for Extraterrestrial Intelligence (SETI@HOME)
  - Probably one of the most well known
  - Uses Berkeley Open Infrastructure for Network Computing (BOINC)
  - Over 3 million computers
- Large Hadron Collider by CERN Labs (LHC@HOME)
  - Also uses BOINC
  - Used to simulate particles travelling around the collider.
- Folding at home by Stanford University (Folding@HOME)
  - Studies protein folding, misfolding, aggregation and related diseases. (Alzheimers, Cancer, HIV)
  - Helps research protein based nanomachines
- And growing…

Grid Computing Within DoD

Examples in Industry

PARABON COMPUTATION

CLUSTER RESOURCES

CITRIX

Mellanox TECHNOLOGIES

Sun microsystems

NEC

intel

Microsoft

CRAY
THE SUPERCOMPUTER COMPANY

PARTEC CLUSTER COMPETENCE CENTER

IBM

hp
invent

QLOGIC

FUJITSU

DELL

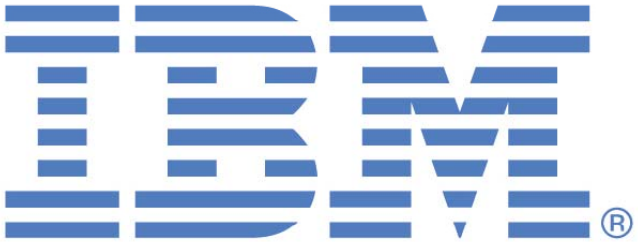
SUPERMICRO
A Server Solutions Manufacturer

nVIDIA

Grid Computing Within DoD

Examples in Education

- October 2007 - IBM and Google unite to provide Grid computing data centers for university education and research
- Universities involved are: Stanford University, MIT, University of California, Berkeley, the University of Maryland and the University of Washington
- Utilizes a public version of Google's MapReduce and Google File System technologies known as Hadoop



Grid Computing Within DoD

Examples in Industry

- June 2008 - IBM proves they have built the first Pflop supercomputer which is known as Roadrunner

- IBM held the previous record at 478 Teraflop (Tflops) with BlueGene/L

- Roadrunner uses a mix of AMD Opterons and IBMs Cell (PowerPC) processors

- Contains 294,912 Processors in 72 racks covering 6,000 square feet.
  - DoD networks contains a *minimum* 5 million CPUs, over 17x larger than roadrunner

- Made up of mostly commodity parts

- In use by the DoE in the Los Alamos National Lab

Cost over $133 Million Dollars!

Grid Computing Within DoD

Examples in Industry

- Microsoft HPC++ (Windows HPC Server 2008)
  - Scales to thousands of processing cores
  - Benchmarked at 18 TeraFLOPS on 2,048 processors spread across 256 compute nodes
  - Microsoft is adding parallel computing extensions to .NET 3.5

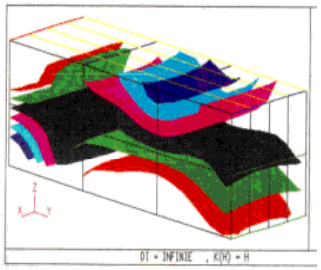
Grid Computing Within DoD

A DoD System is Deployable

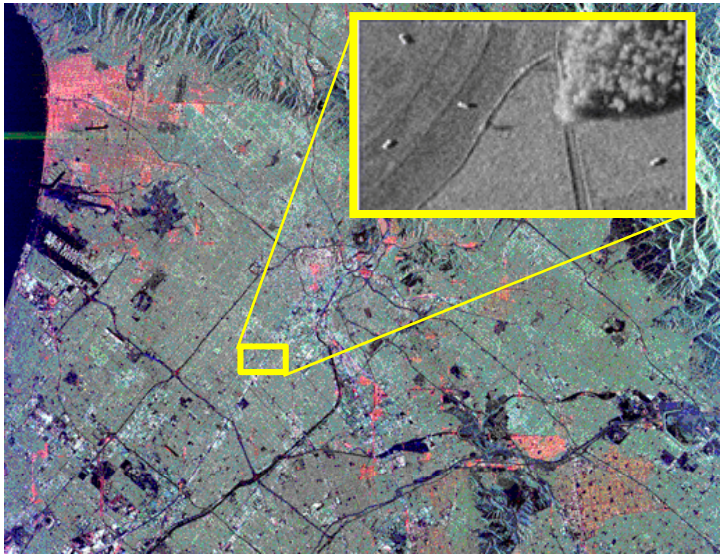
- A deployed thin client with limited permissions
  - Administrative policy regulates PC and network load
  - End user has no access to client or job data, which is encrypted
- A centrally controlled scheduler / server controls tasks
  - Jobs are signed / certified, encrypted
- Security is comparable to existing net-administered maintenance
  - Application sensitivity is addressed on a case-by-case basis
  - Solution diffuses rather than concentrates data, into small and useless snippets of information
- SIPR, being a more tightly controlled environment, is perfect for secure Grid computing.

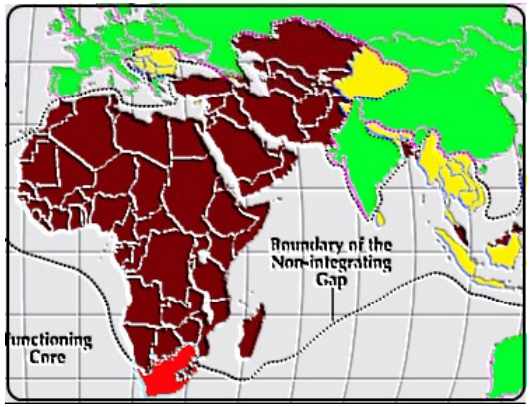
Grid Computing Within DoD

Imagine the Possibilities



- Massively parallel wide-area search, data-fusion, target recognition
- Complete sets of parametric analysis
- Large, interactive predictive models (global instability)
- Campaign analysis models





Questions?