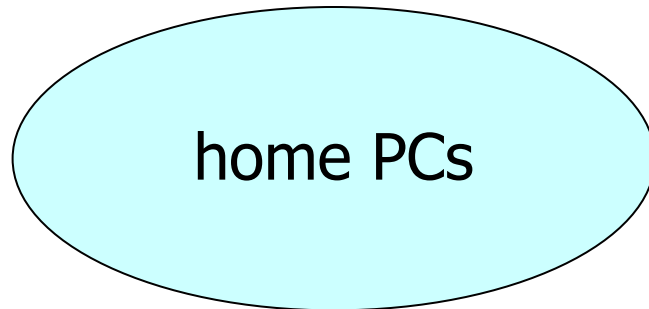


Climate Prediction and BOINC

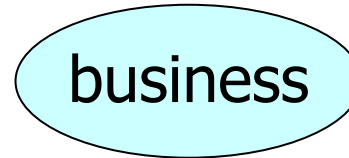
A photograph of a desert landscape. In the foreground, there is a sandy path leading through sparse, dry vegetation. A large, prominent Joshua tree stands on the right side. In the middle ground, a large, light-colored rock formation with several large boulders is visible. The background shows rolling hills under a clear blue sky.

David P. Anderson
Space Sciences Laboratory
University of California – Berkeley
davea@ssl.berkeley.edu

Public-resource computing



your computers



95 96 97 98 99 00 01 02 03 04

GIMPS, distributed.net

SETI@home, folding@home

fight*@home

climateprediction.net

names:

public-resource computing

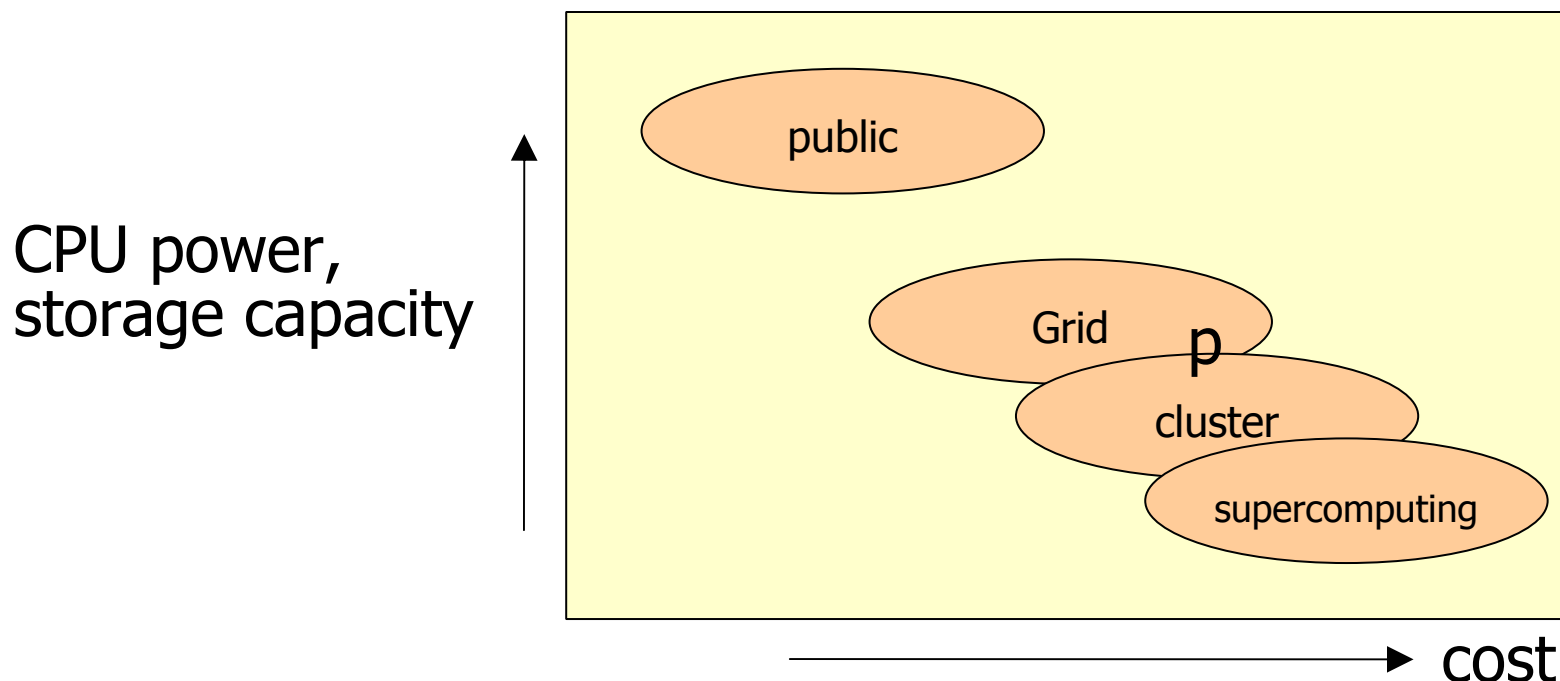
peer-to-peer computing

Internet distributed computing

"@home" computing

The potential of public computing

- SETI@home: 500,000 CPUs, 70 TeraFLOPs
- 1 billion Internet-connected PCs in 2010, 50% privately owned
- If 100M participate:
 - > 100 PetaFLOPs
 - > 1 Exabyte (10^{18}) storage



Public/Grid differences

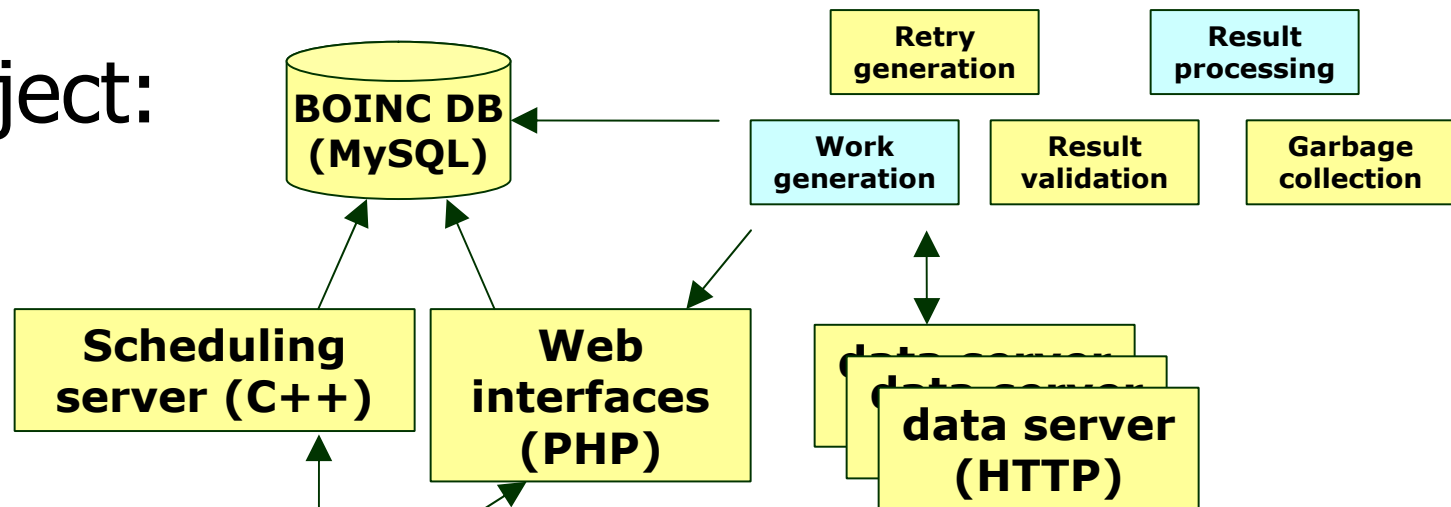
	Public	Grid
Managed resources?	no	yes
Secure resources?	no	yes
Always on?	no	yes
Always connected?	no	yes
Network bandwidth	Expensive, scarce	abundant
Network connection	1 way (pull)	2 way (pull or push)
Must be unobtrusive	yes	no
Credit system	yes	maybe
How to get resources?	complex	complex
public education/outreach	yes	no
self-upgrading?	yes	no

BOINC: Berkeley Open Infrastructure for Network Computing

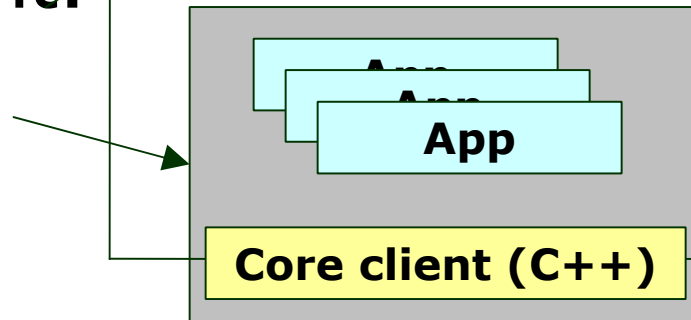
- Goals for computing projects
 - easy/cheap to create and operate projects
 - support a wide range of applications
 - no central authority
- Goals for participants
 - easy to participate in multiple projects
 - resource allocation among projects
 - invisible use of disk, CPU, network
- <http://boinc.berkeley.edu>

Structure of BOINC

• Project:



• Participant:



Data architecture

- Files
 - immutable, replicated
 - may originate on client or project
 - may remain resident on client
 - priority, expiration date
- Executables are digitally signed
- Upload certificates: prevent DOS

```
<file_info>
  <name>arecibo_3392474_jun_23_01</name>
  <url>http://ds.ssl.berkeley.edu/a3392474</url>
  <url>http://dt.ssl.berkeley.edu/a3392474</url>
  <md5_cksum>uwi7eyufiw8e972h8f9w7</md5_cksum>
  <nbytes>10000000</nbytes>
</file_info>
```

Computational abstractions

- Applications
- Platforms
- Application versions
 - may involve many files
- Work units: inputs to a computation
 - soft deadline; CPU/disk/mem estimates
- Results: outputs of a computation
- Redundant computing
- Benchmarking and credit

Project web site features

- Download core client
- Create account
- Edit preferences
 - General: disk usage, work limits, buffering
 - Project-specific: allocation, graphics
 - venues (home/school/work)
- Profiles
- Teams
- Message boards, adaptive FAQs

Project management tools

- Python scripts for project creation/start/stop
- Remote debugging
 - collect/store crash info (stack trace)
 - web-based browsing interface
- Strip charts
 - record, graph system performance metrics
- Watchdogs
 - detect system failures; dial pager

Anonymous platform mechanism

- User compiles applications from source, registers them with core client
- Report platform as “anonymous” to scheduler
- Purposes:
 - obscure platforms
 - security-conscious participants
 - performance tuning of applications

BOINC projects

- SETI@home
- Predictor@home (Scripps Institute)
- Climateprediction.net
- Einstein@home (gravity waves, LIGO)
- CERN
- UCB/Intel network performance study

Requirements of CPDN

- Long work units (CPU months)
 - preemptive scheduling
 - “trickle” messages
- Multi-phase application
 - process control
 - CPU time, fraction done reporting
- Data storage
 - disk usage preferences
 - query/upload/delete of large output files

Conclusion

- Public-resource computing
 - the ultimate paradigm for scientific computing
 - connects public to science
- BOINC: enabling technology for PRC
- CPDN: an ideal application for PRC (and a good complement to SETI)