

Volunteer Computing: Planting the Flag

***David P. Anderson
Space Sciences Lab
U.C. Berkeley***

March 30, 2007



Outline

- What is volunteer computing?
- History and current status
- Some research directions
- Non-technical problems



What is Volunteer computing?

Projects



Volunteers



- Helps science
- Involves public in science

Early history

- GIMPS (1996)
 - George Woltman, Scott Kurowski
- distributed.net (1997)
 - Adam Beberg, Jim Lawson, ...
- SETI@home (1999)
 - me, Eric Korpela
- Folding@Home (2000)
 - Vijay Pande



Academic projects, 1996-99

- Bayanihan
 - L. Sarmenta, MIT
 - fault-tolerance in presence of hackers
 - “volunteer computing”
- Popcorn
 - Noam Nisan, Hebrew Univ.
 - Market-based approach
- Superweb
- Charlotte
- Jet

The problem:
Java

Business attempts

- Entropia (1998)
- Popular Power
- United Devices (2000)
- Parabon
- Data Synapse

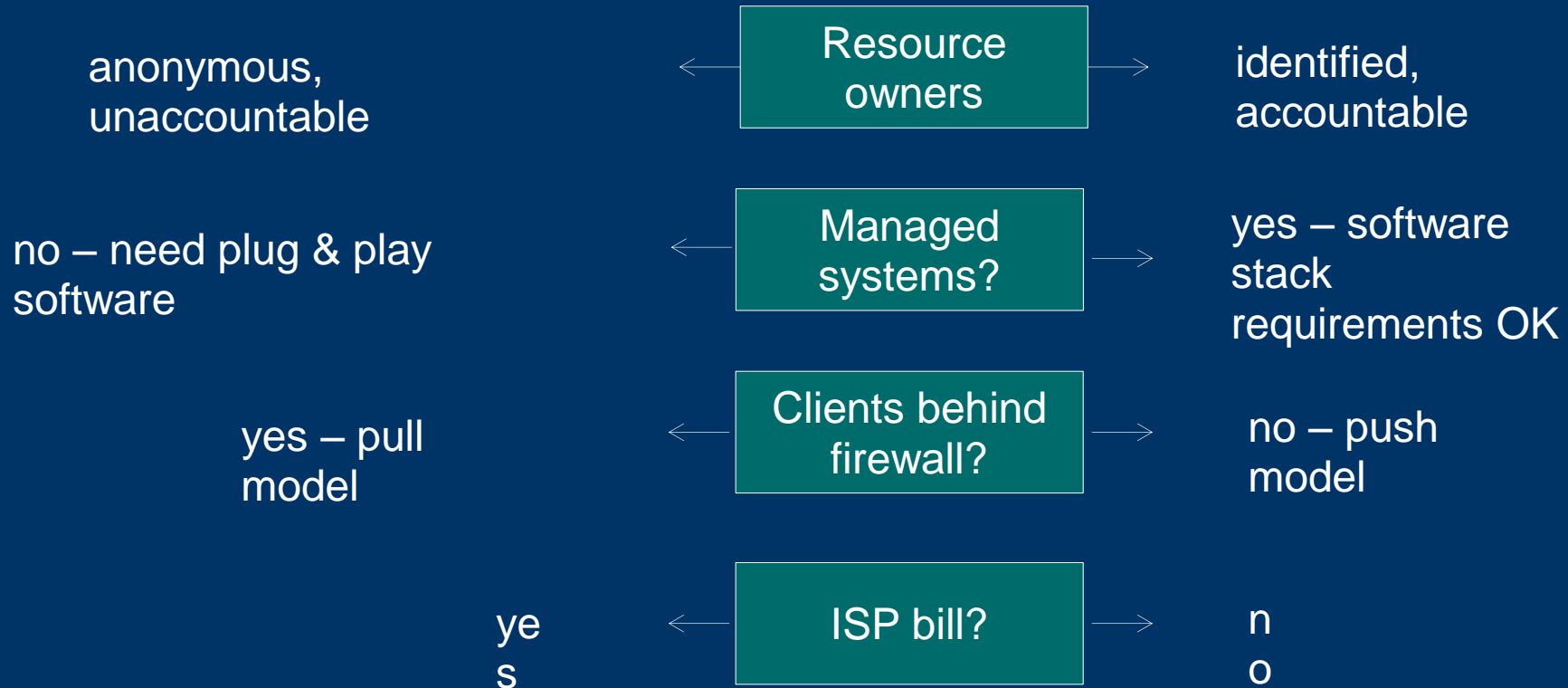
The problem: no customers

BOINC (2002)

- Middleware for volunteer computing
- Open-source (LGPL)
- Application-driven
- Goals
 - lots of independent projects
 - support for diverse applications
 - client participation in multiple projects



Volunteer computing \neq Grid computing



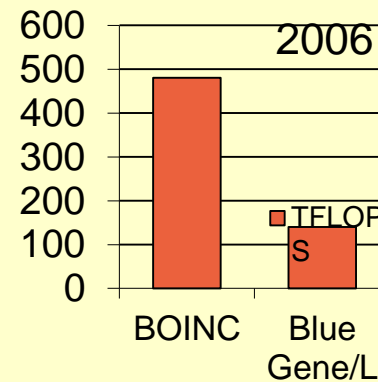
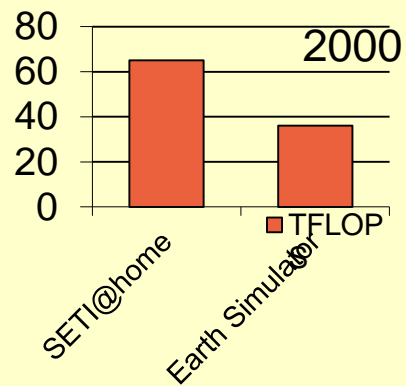
... nor is it “peer-to-peer computing”

Some BOINC-based projects

- Climateprediction.net
 - Oxford; climate change study
- Einstein@home
 - LIGO; gravitational wave astronomy
- Rosetta@home
 - U. Washington; protein study
- SETI@home
 - U.C. Berkeley; SETI
- LHC@home
 - CERN; accelerator simulation
- Africa@home
 - STI, U. of Geneva; malaria epidemiology
- IBM World Community Grid
 - several biomedical applications
- ...and about 30 others

Computing power

- Folding@home:
 - 650 TeraFLOPS
 - 200 from PCs; 50 from GPUs; 400 from PS3!
- BOINC-based projects:



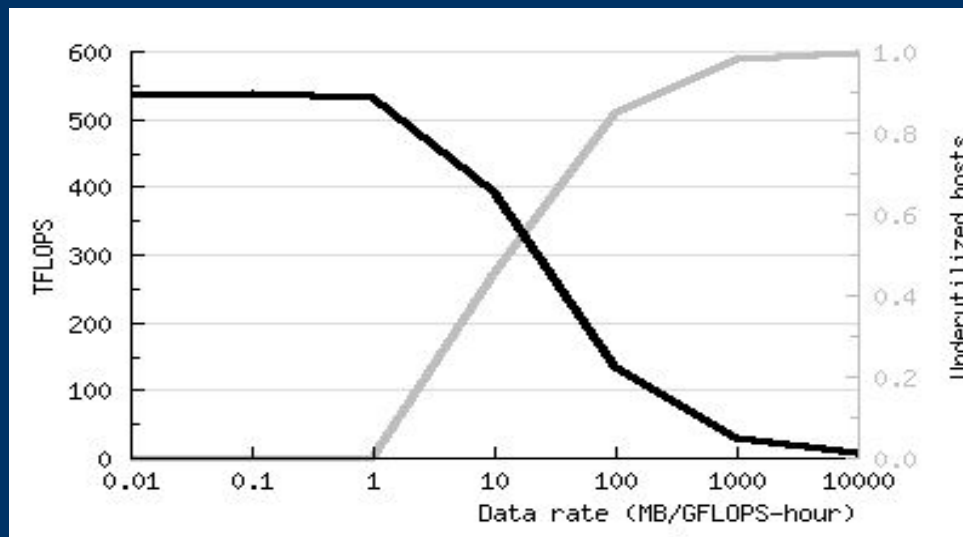
A sampling of research problems

- Data-intensive computing
- Low-latency computing
- Background utility compatibility
- Credit mechanism
- Efficient validation
- Game consoles and graphics chips
- Simulation



Data-intensive computing – client limits

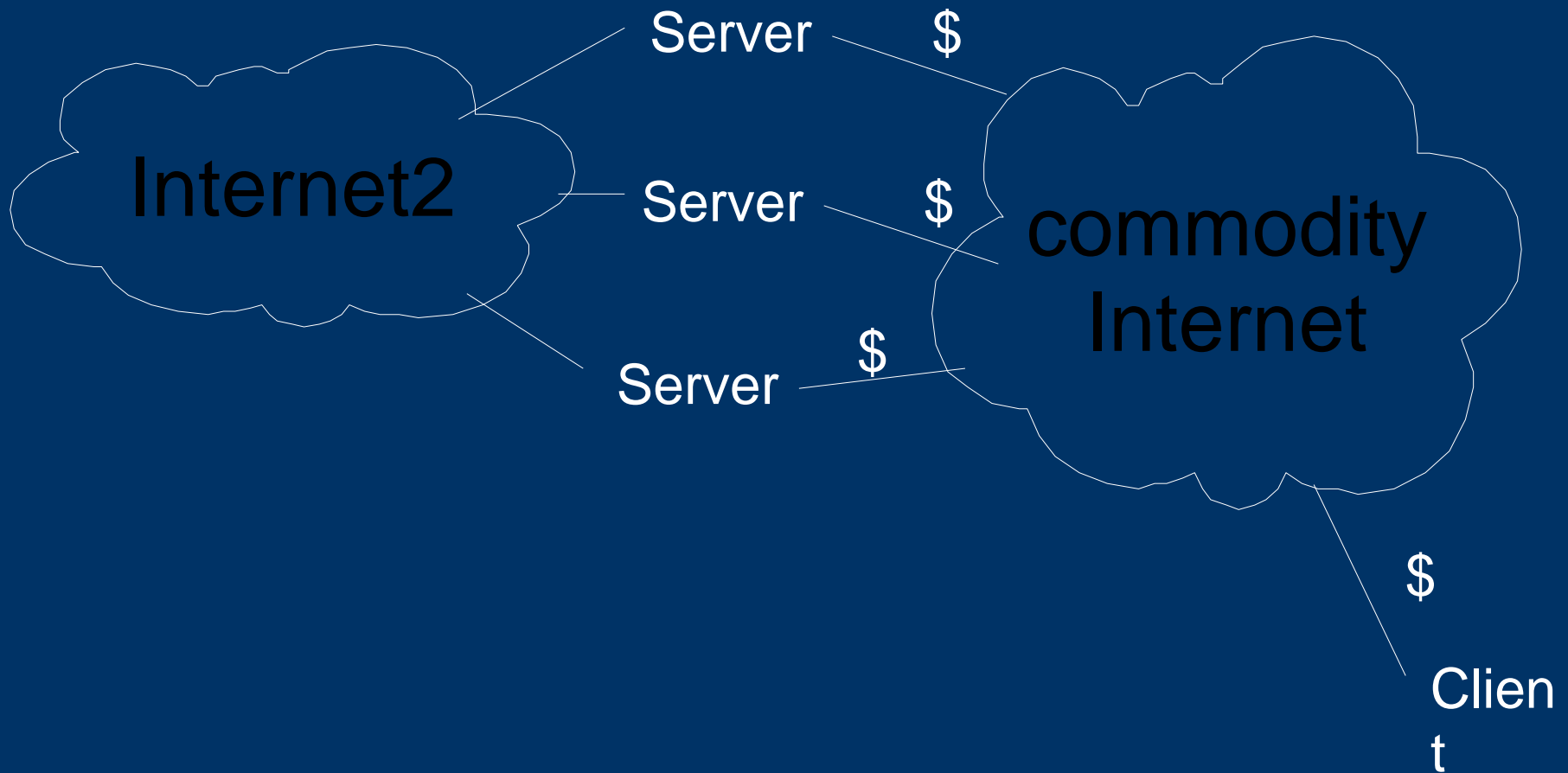
- Q = network transfer per GFLOPS/hr
- SETI@home: $Q = 0.1$ MB
- but wider range is OK:



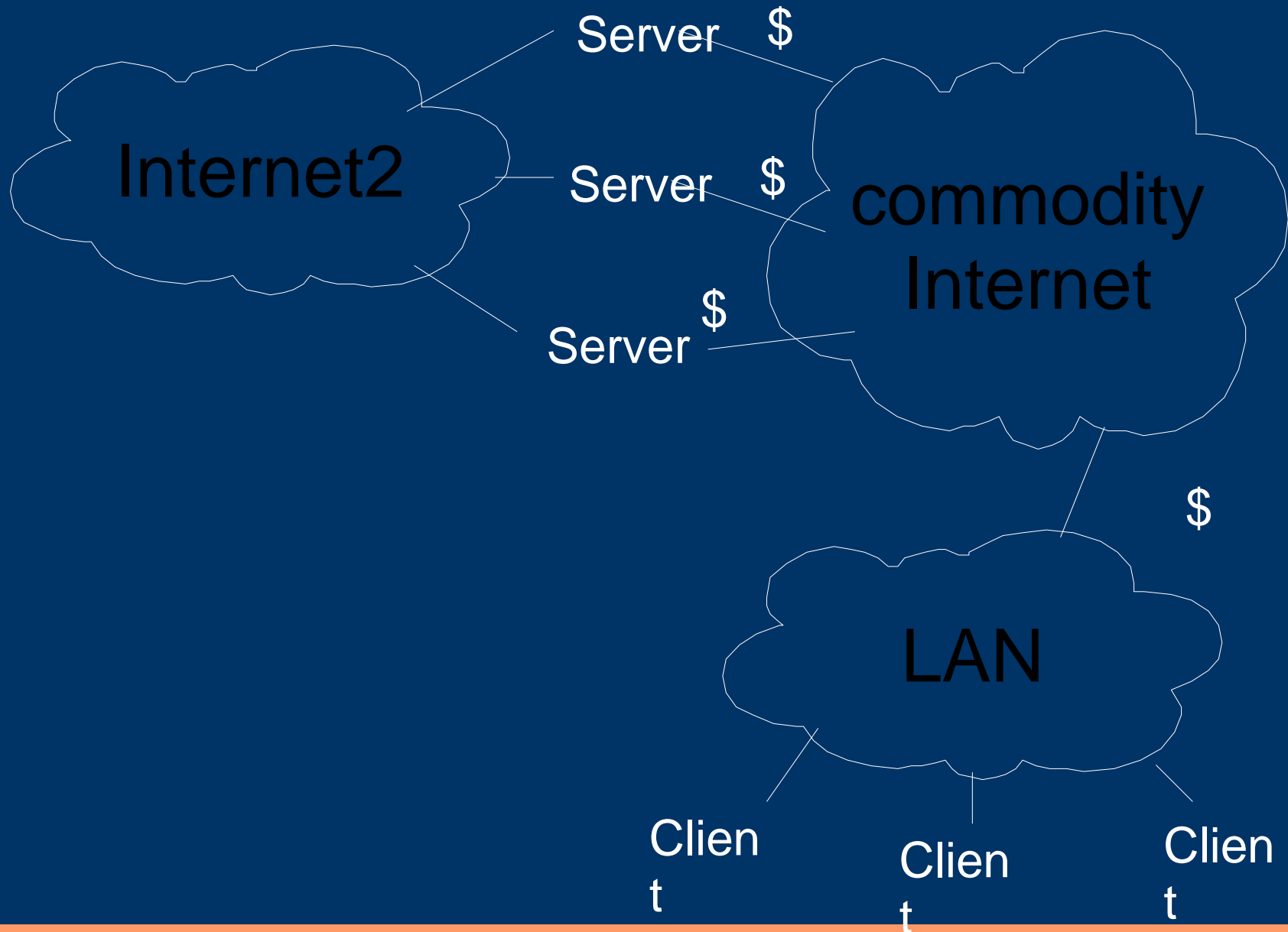
Server-side limits



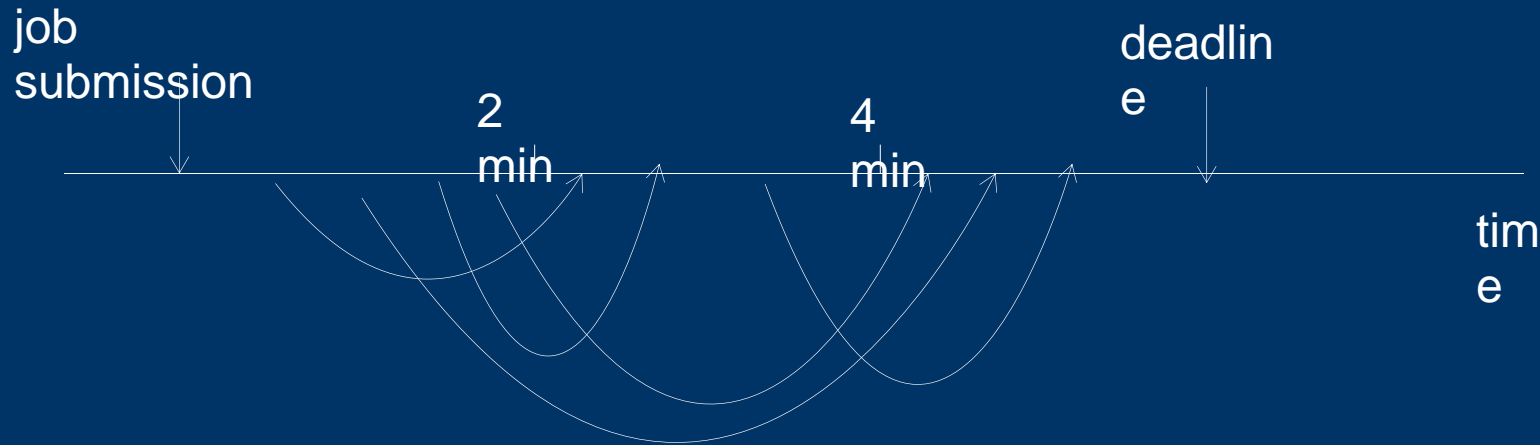
Using free networks



Using more free networks



Low-latency computing



- VC usually minimizes connection frequency
- What if you want to do 10,000 1-minute jobs in 6 minutes of wall time?

Background utility compatibility

- Background utilities
 - disk defrag
 - disk indexing
 - virus scanning
 - web pre-fetch
 - disk backup
 - Most run only when computer is idle
 - volunteer computing ==> they never run
 - A) ignore zero-priority CPU activity
 - B) Background manager
 - intelligent decision about when to run various activities
-
-

Credit mechanism

- Credit is important
- Currently based on FLOPS
- Should reflect, e.g. RAM size?
- Goal: maximize utility to projects
- Project-specific pricing schedules?
- Credit for other resources
 - disk usage
 - network transfers

Efficient validation

- Job output
 - results
 - claimed credit
 - Replicated computing
 - do job N times, require quorum of M
 - Dealing with numerical discrepancies
 - homogeneous replication
 - eliminate discrepancies
 - no replication
 - How to validate provably and efficiently?
-
-

Game consoles and graphics chips

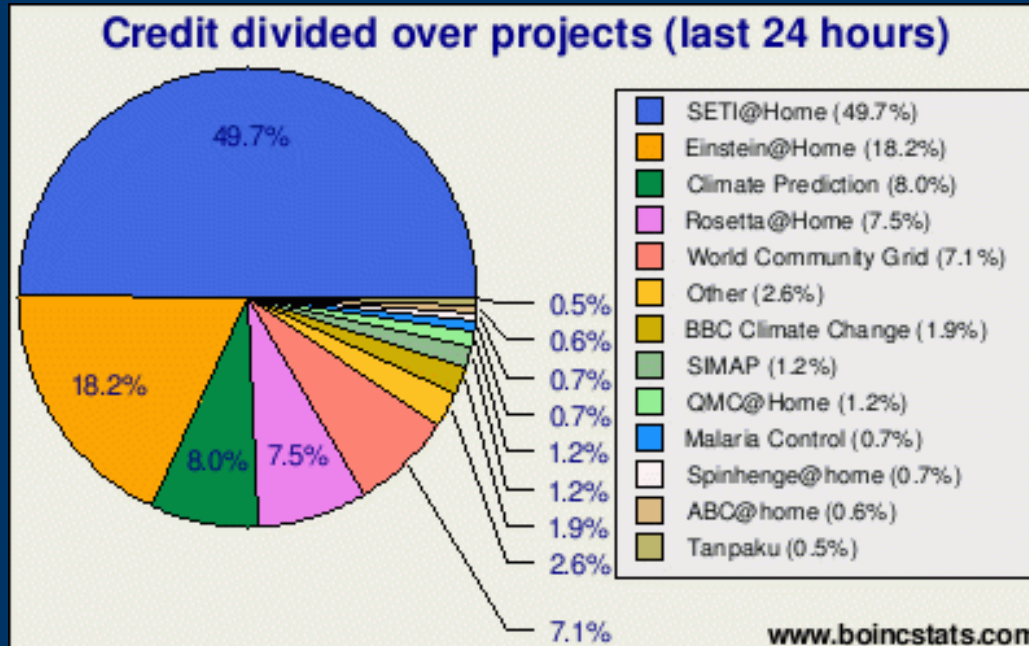
- NVIDIA, ATI, Cell
 - 10X CPU and gaining?
 - Folding@home:
 - ATI version
 - Sony PS3 version
 - BOINC and Einstein@home on PS3
 - How to make this available to other projects?
-
-

Simulating volunteer computing

- Ad-hoc development of scheduling policies
 - slow, noisy
 - jeopardizes running projects
 - Simulation-based R&D
 - client simulator
 - client scheduling policies
 - Project simulator
 - server scheduling policies
 - Global simulator
 - study data-intensive, low-latency, etc.
-
-

The hard non-technical problems

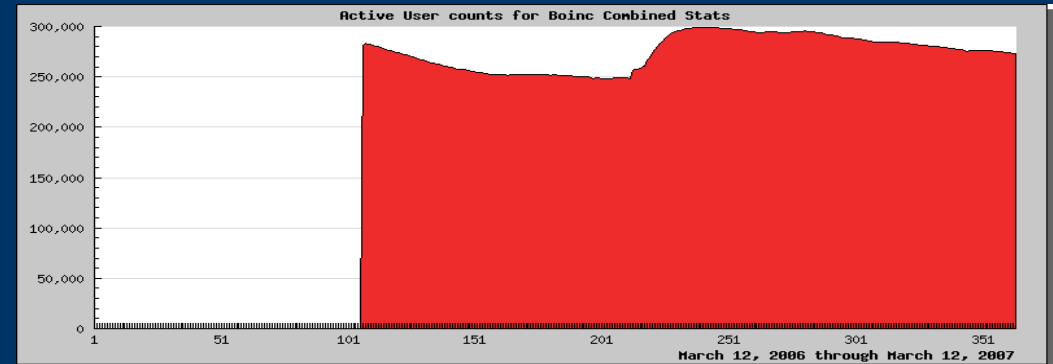
- How to increase the number of volunteers?
 - currently 1 in 1000 PC owners
- How to increase the number of projects?
 - currently stuck at about 50
- How to get volunteers to diversify?



How to attract and retain volunteers?

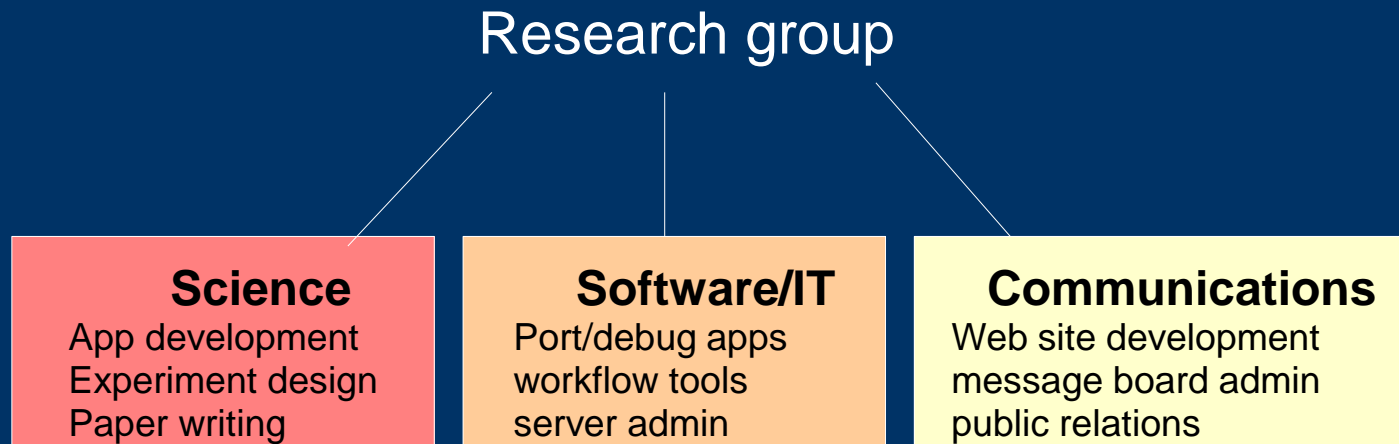
Active hosts:

- Retention
 - reminder emails
 - frequent science updates
- Recruitment
 - Viral
 - “email a friend”, referral reward
 - Organizational
 - World Community Grid: “partner” program
 - Media coverage
 - need more discoveries
 - Bundling



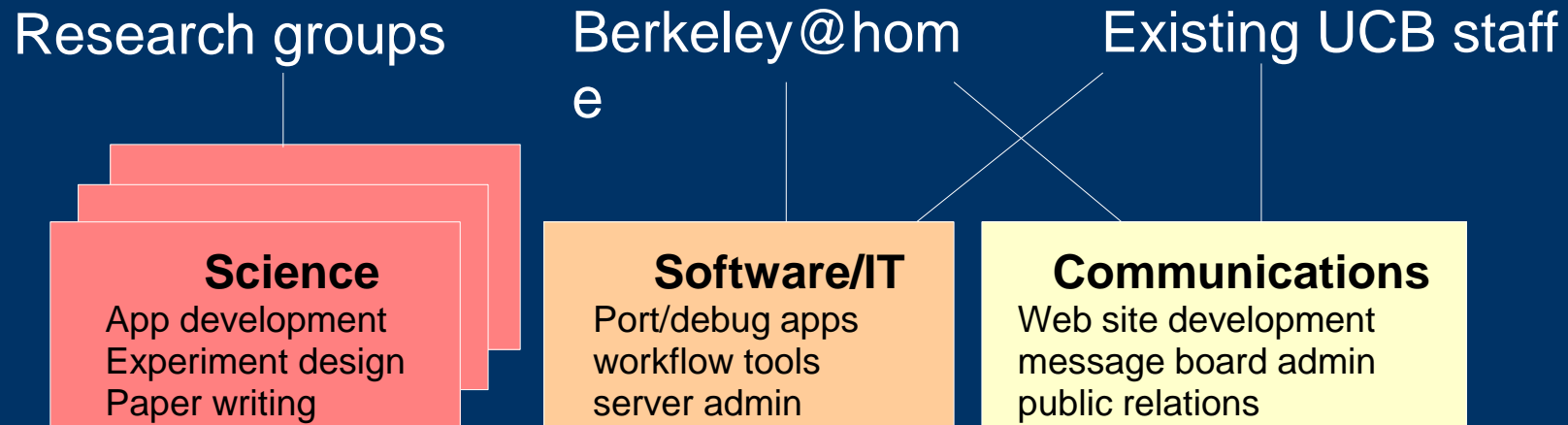
Why aren't there more projects?

- Lack of PR among scientists
- IT antipathy
- Creating a BOINC project is expensive:



Meta-projects

- Virtual Campus Supercomputing Center
 - Deployment and publicity:
 - PC labs, staff/faculty desktops
 - students
 - alumni
 - public



- IBM World Community Grid

Encouraging change

- Cross-project credit system
 - encourage competition in total credit, not per-project
 - Account Managers
 - Make it easier to discover/attach/detach projects
 - GridRepublic, BAM!
 - Science Stock Market?
 - encourage participation in new high-potential projects
 - Scientific Mutual Funds?
 - e.g. American Cancer Society BOINC “portfolio”
-
-

Conclusion

- Volunteer computing: a new paradigm
 - distinct research problems, software requirements
 - big accomplishments, potential
 - time to plant the flag!
- Social impacts
- Contact me about:
 - Using BOINC
 - Research based on BOINC
 - Organizational use of BOINC

davea@ssl.berkeley.edu
