

Volunteer Computing

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

May 2, 2007



Outline

- Goals of volunteer computing
 - How BOINC works
 - Some volunteer computing projects
 - Some research directions
 - Non-technical problems
-
-

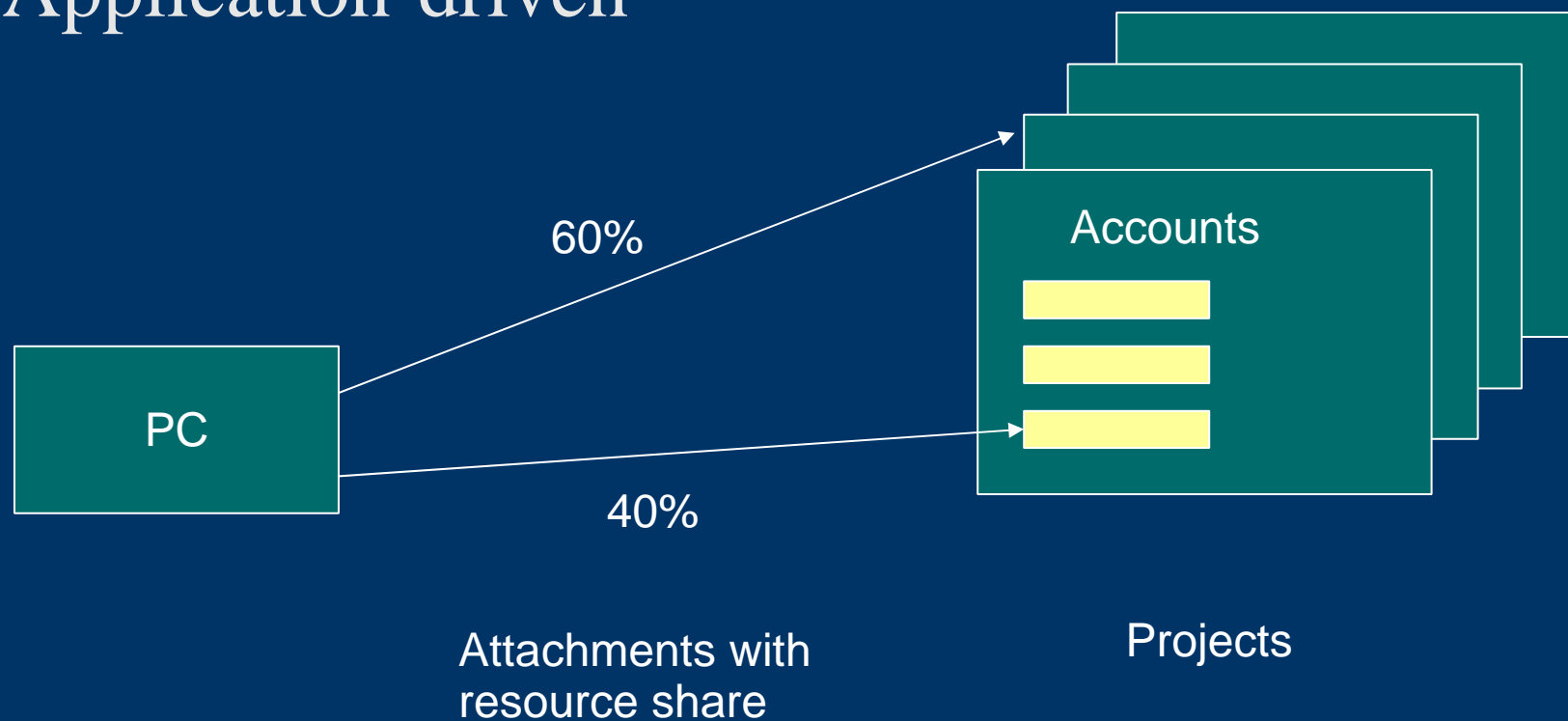
Goal:

Use all the computers in the world, all the time, to do worthwhile things

- What do we mean by “computers”?
 - Who owns the computers?
 - Individuals (60% and rising)
 - Organizations
 - What does “worthwhile” mean?
-
-

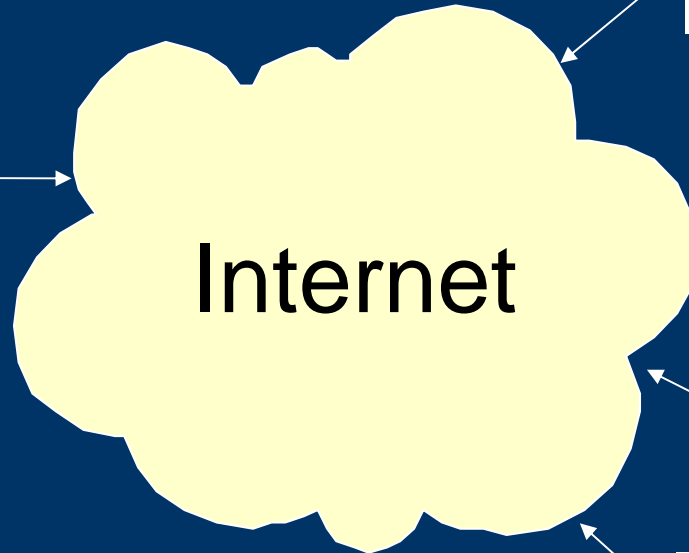
BOINC (Berkeley Open Infrastructure for Network Computing)

- Middleware for volunteer computing
- Open-source (LGPL)
- Application-driven



The volunteer computing game

Projects

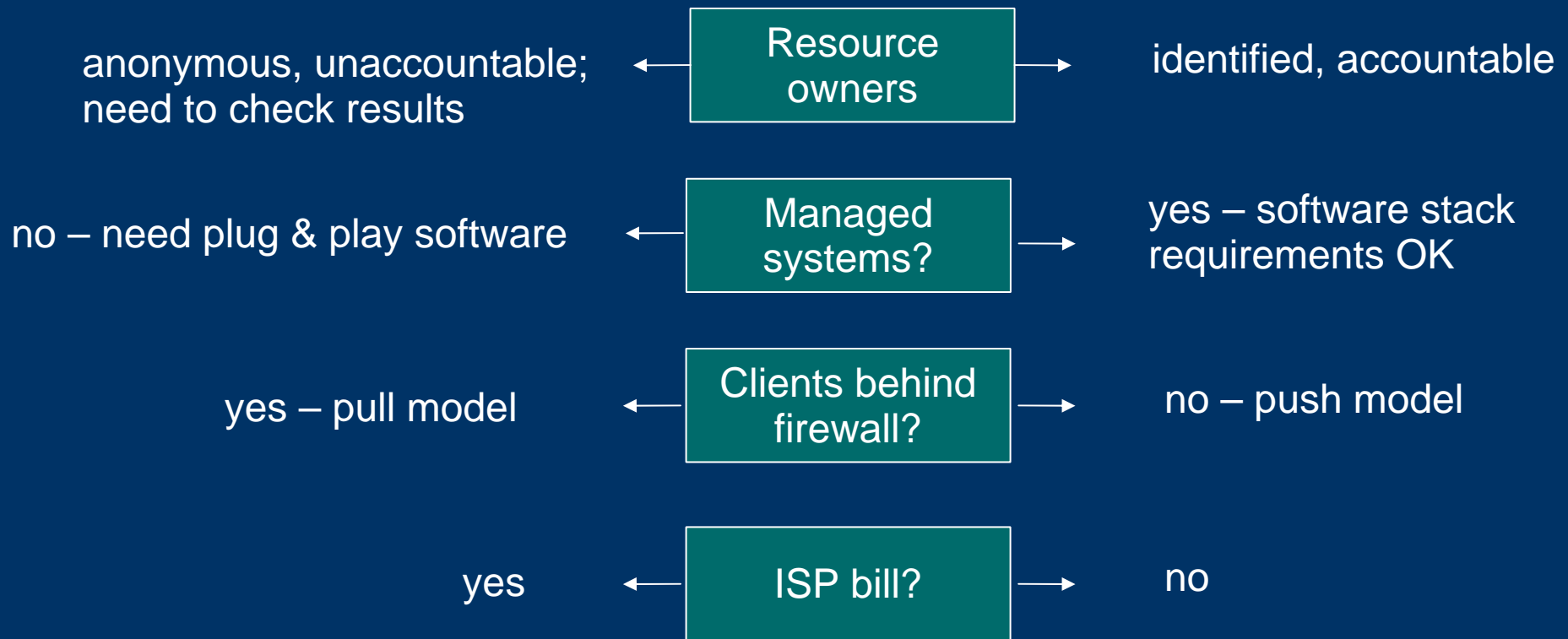


Volunteers



- Do more science
- Involve public in science

Volunteer computing \neq Grid computing



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

Types of utility

- Compute cycles (usually FLOPS)
 - with latency bound
 - with RAM
 - with storage
 - guaranteed availability
 - Storage
 - network bandwidth
 - availability
 - Network bandwidth at specific time
 - Wide/diverse deployment
 - Human steering of computation
-
-

What is resource share?

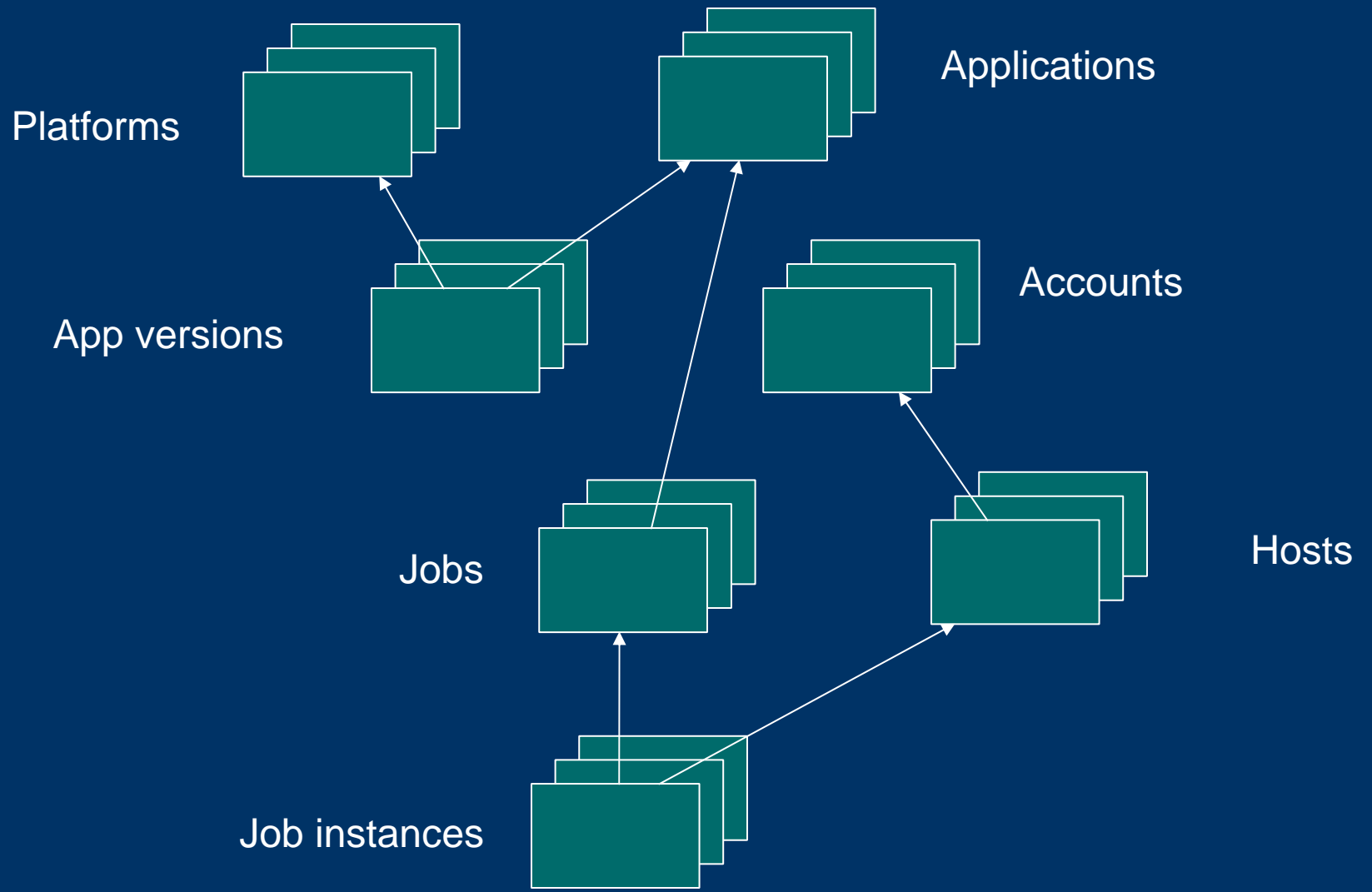
- Given: set of hosts, attachments
- “Schedule”: what projects use resources when
- Ideally:
 - Weighted Pareto optimality of utility?
 - Across multiple hosts
- In BOINC
 - determines usage of bottleneck resources on a single host, e.g.:

	Resource share	Demand
A	40%	15%
	15%	
B	40%	large
	50%	
C	20%	large
	25%	

Credit: approximation of utility

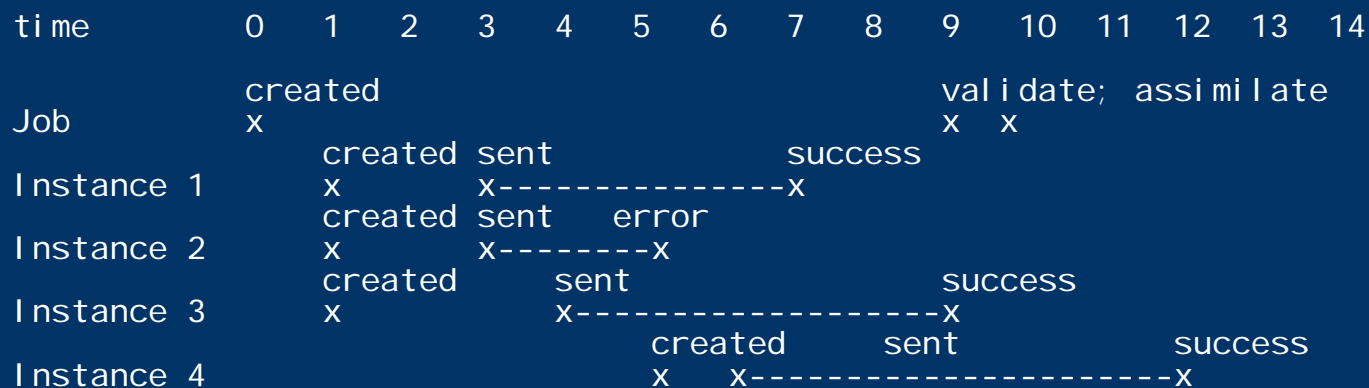
- Currently
 - projects can grant however they want
 - default: FLOPS
 - benchmarks * CPU time
 - application provides FLOP count
 - application does FP benchmark
 - cheat-proofing
 - To do:
 - cheat-proof measurements of other resources
 - projects publish “credit per day” functions
 - Normalization rule
 - Accounting rule
-
-

How BOINC works: server DB



Job replication

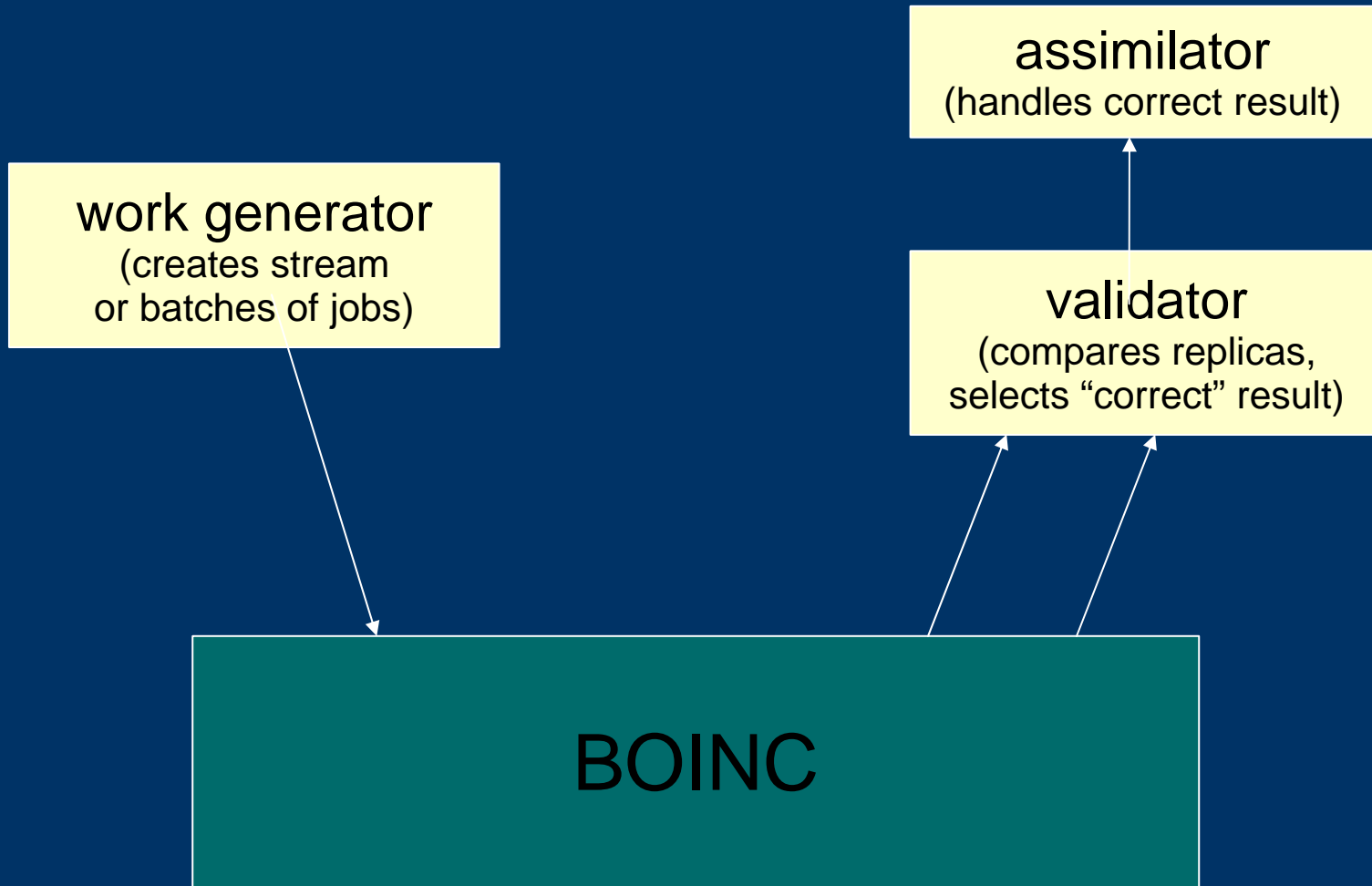
- Problem: can't trust volunteers
 - computational result
 - claimed credit
- Application-specific checks, no replication
- Replicated computing
 - do N copies, require that M of them agree
 - not bulletproof (collusion)



How to compare results?

- Problem: numerical discrepancies
- Stable problems: fuzzy comparison
- Unstable problems
 - Eliminate discrepancies
 - compiler/flags/libraries
 - Homogeneous replication
 - send instances only to numerically equivalent hosts
(equivalence may depend on app)

Work flow



Volunteer's view

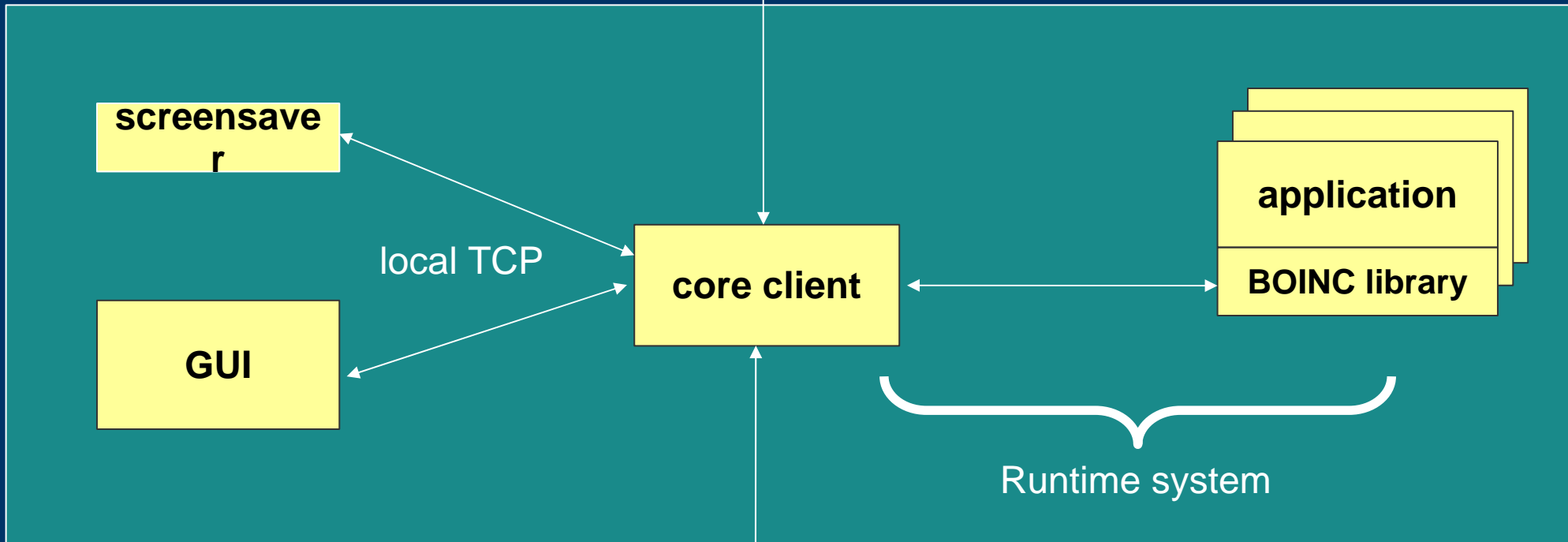
- 1-click install, zero configuration
- All platforms
- Invisible, autonomic

The screenshot displays the BOINC Manager application window. The main window title is "BOINC Manager - (localhost)". The interface includes a menu bar (File, View, Tools, Activity, Advanced, Help) and a toolbar with icons for Projects, Tasks, Transfers, Messages, Statistics, and Disk. A left sidebar contains "Commands" (Show graphics, Suspend, Abort) and "Web sites" (SETI@home, Message boards, Help). The central area features a table of tasks with columns for Project, Application, Name, CPU time, Progress, To compl..., and Report deadline. The status bar at the bottom indicates "Connected".

Project	Application	Name	CPU time	Progress	To compl...	Report deadline
ABC@home	ABC finder 1.01	abc_wu_7024860000_5000...	---	0.000 %	02:43:34	2/5/2007 2:30:12 PM
rosetta@home	Rosetta 5.45	2tif_TREEJUMP_ABRELAX_...	00:00:15	1.000 %	21:13:57	2/19/2007 12:58:15
climateprediction.net	HadCM3 Coupled Model...	hadcm3pbb_ck11_0585990...	00:52:06	0.038 %	7628:22...	1/11/2008 6:15:07 F
climateprediction.net	HadCM3 Coupled Model...	hadcm3pbb_b8va_0579878...	00:18:52	0.017 %	7631:06...	1/11/2008 6:15:28 F
BOINC alpha test	Upper Case 5.14	wu_1169574347_1202_0	00:00:07	99.659 %	00:00:06	2/11/2007 12:07:45
Project TANPAKU	BD 5.08	pep3_259_69_2_1_0	00:23:09	11.819 %	01:59:52	1/28/2007 9:00:43 A
Project TANPAKU	BD 5.09	pep3_223_129_10_2_1	00:16:32	4.508 %	02:12:25	2/5/2007 9:18:36 A
SETI@home	SETI@home Enhanced ...	20se03aa.22508.4609.1015...	00:01:22	0.680 %	20:48:07	2/23/2007 1:45:17 F
SETI@home	SETI@home Enhanced ...	07mr00aa.3854.2978.8111...	00:00:37	0.133 %	34:58:27	3/9/2007 7:39:45 A

BOINC client structure

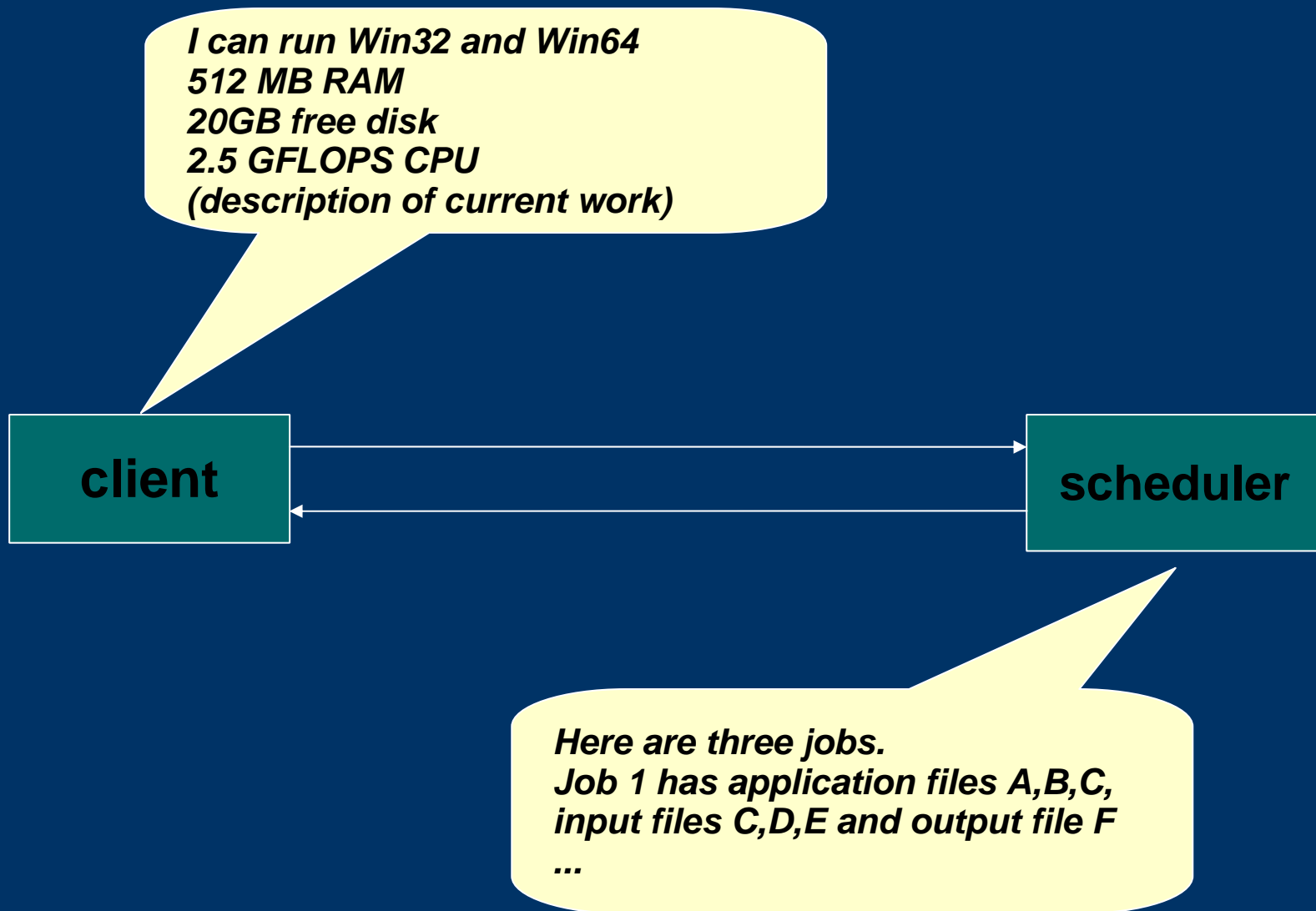
schedulers, data servers



user preferences, control

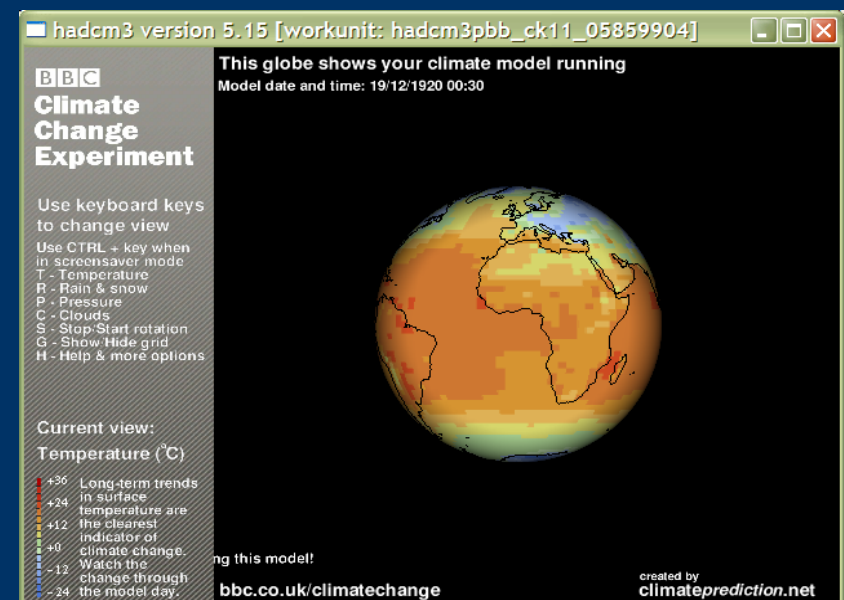


Communication: “Pull” model



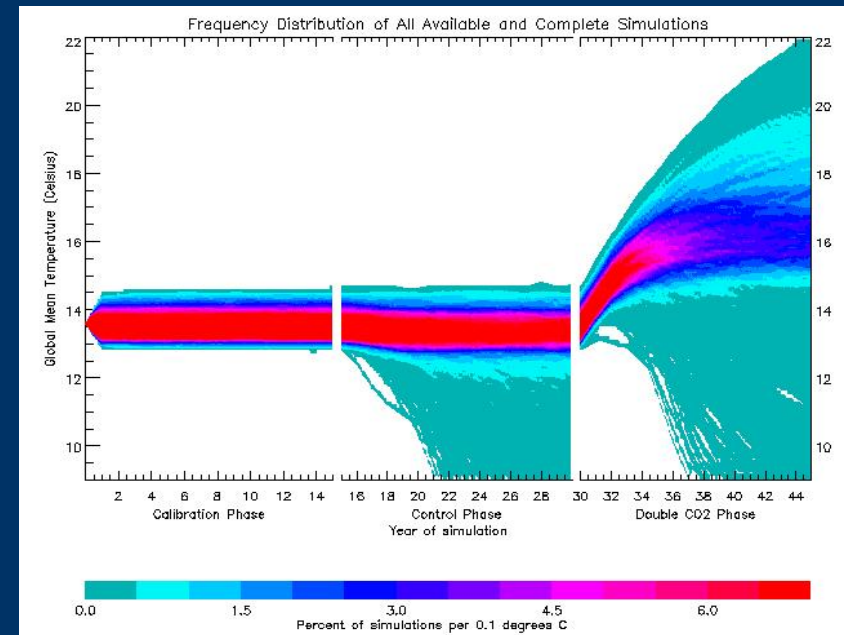
Example: ClimatePrediction.net

- Application: UK Met Office Unified Model
- State-of-the-art global climate model
 - 1 million lines of FORTRAN
- High-dimensional search space
 - model parameters
 - boundary conditions
 - perturbed initial conditions



ClimatePrediction.net

- Using supercomputers:
 - 1 day per run
 - 10-20 total runs
- Using BOINC:
 - 6 months per run
 - 50,000 active hosts
 - 171,343 runs completed
 - Nature papers
 - 60-fold savings

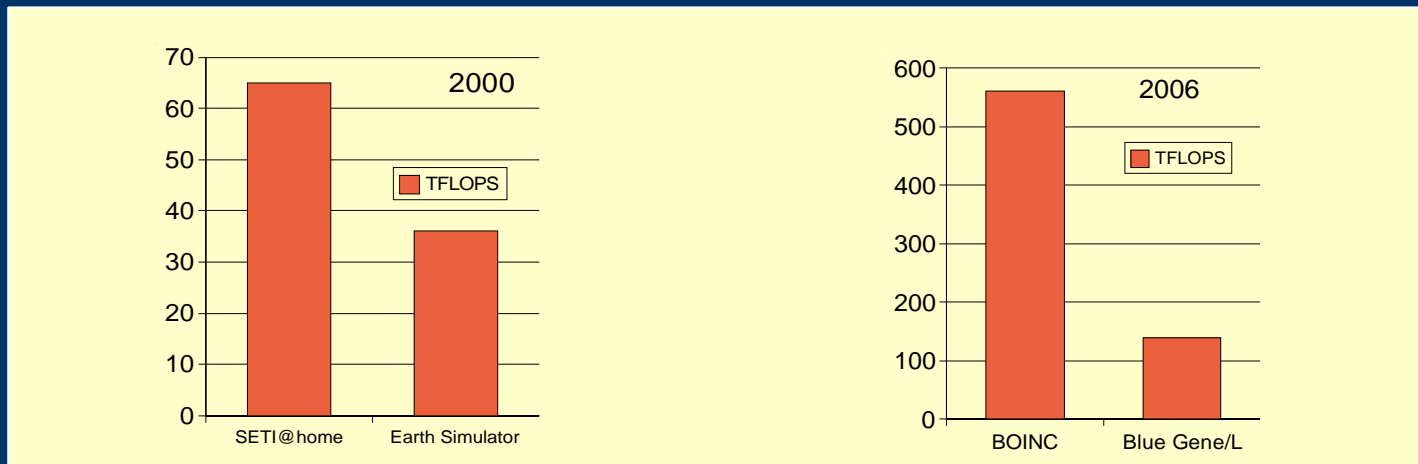


Some other BOINC-based projects

- 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:

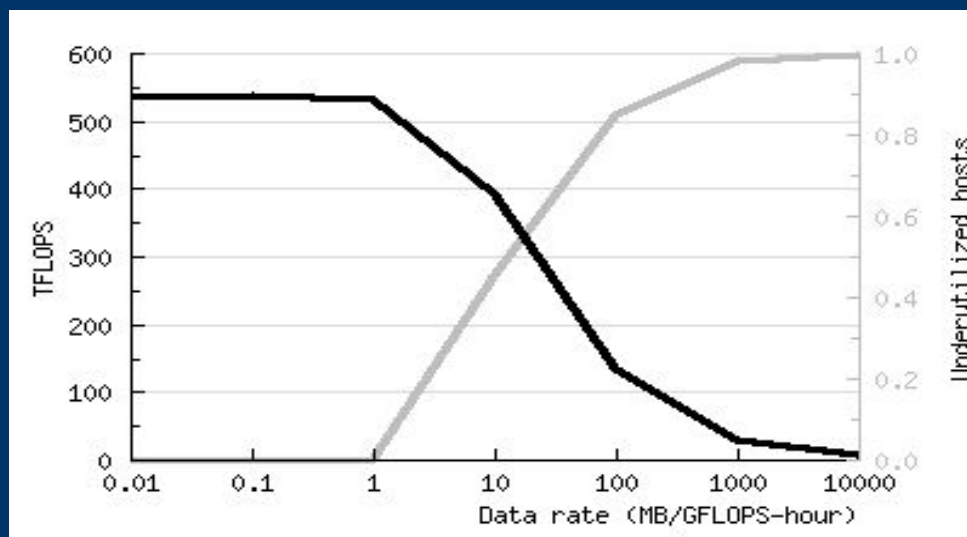


A sampling of research problems

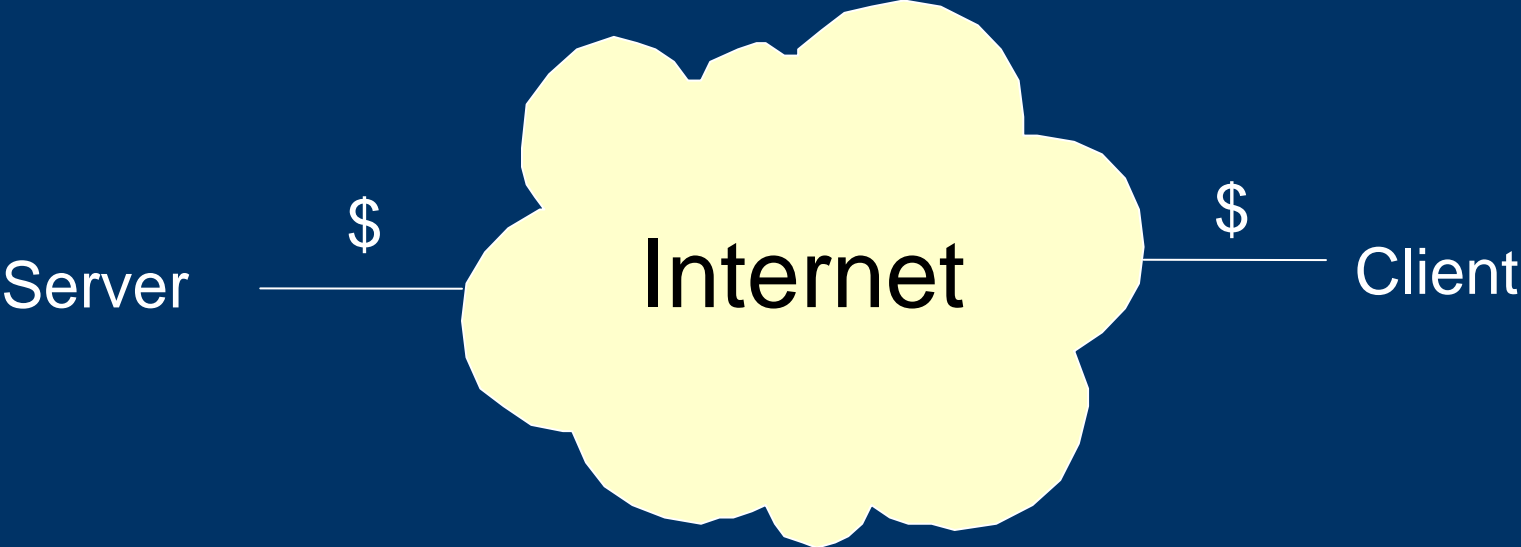
- Data-intensive computing
 - Low-latency computing
 - Background utility compatibility
 - Credit mechanism
 - Efficient validation
 - Game consoles and graphics chips
 - Simulation
-
-

Data-intensive computing – client limits

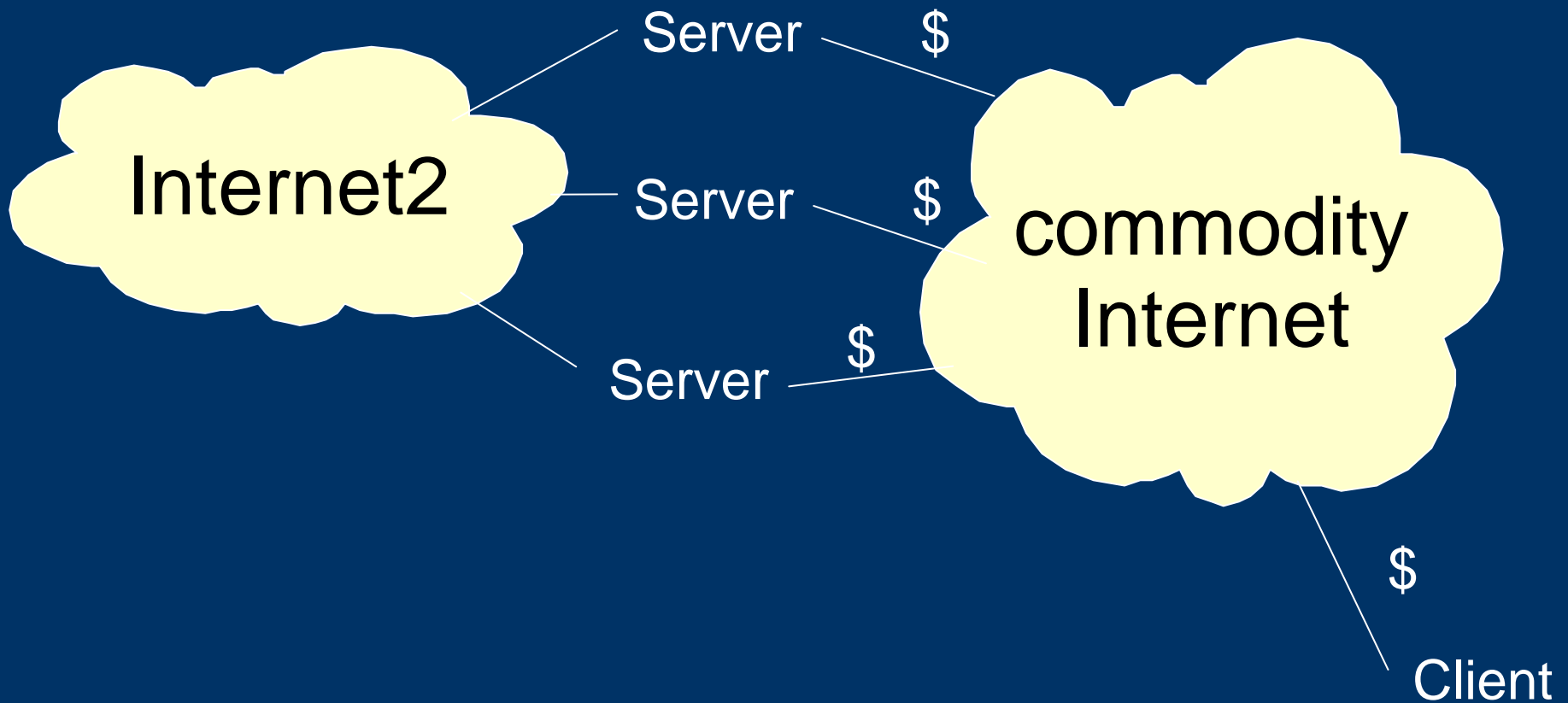
- Q = network transfer per GFLOPS/hr
- SETI@home: $Q = 0.1$ MB
- but wider range is OK:



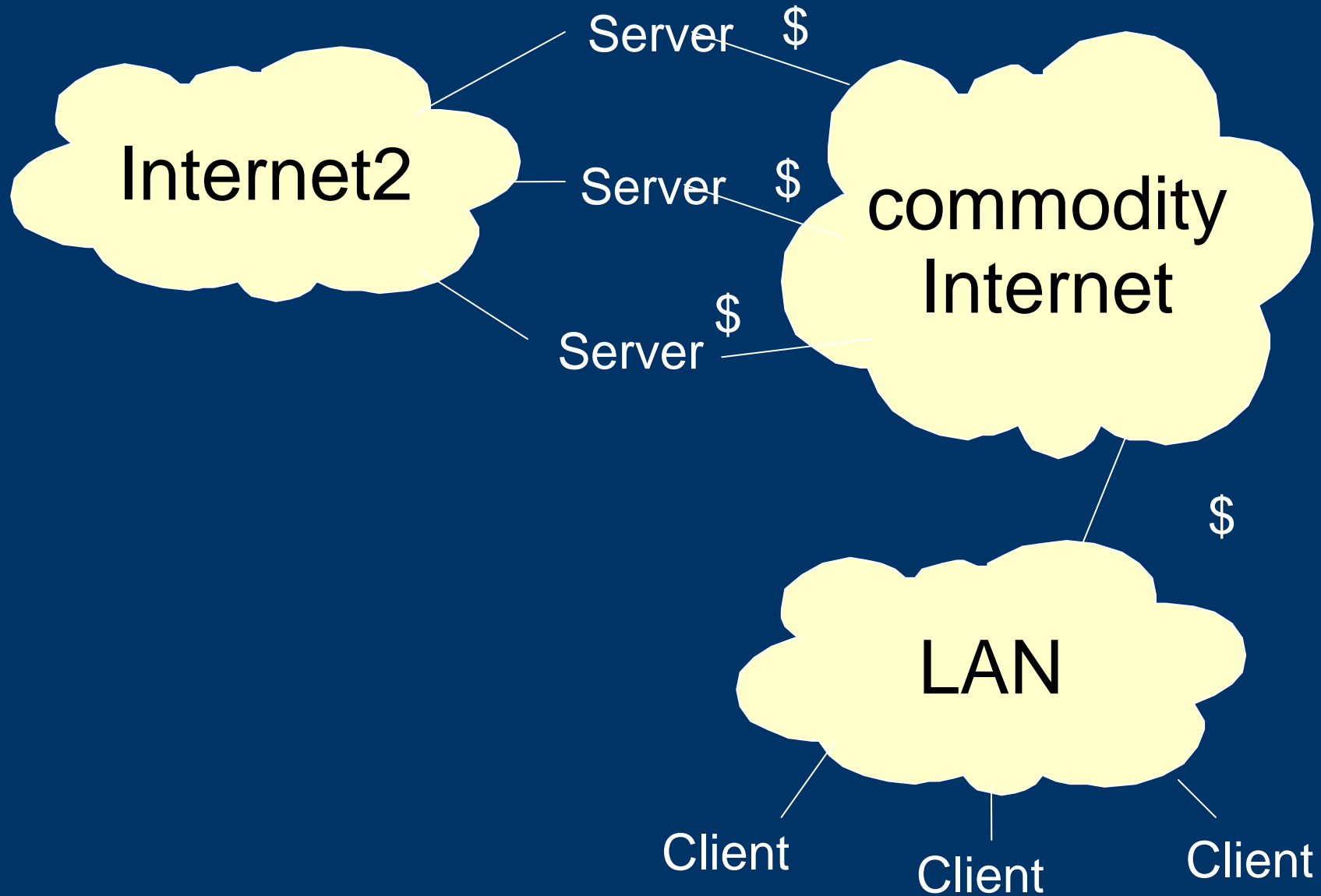
Server-side limits



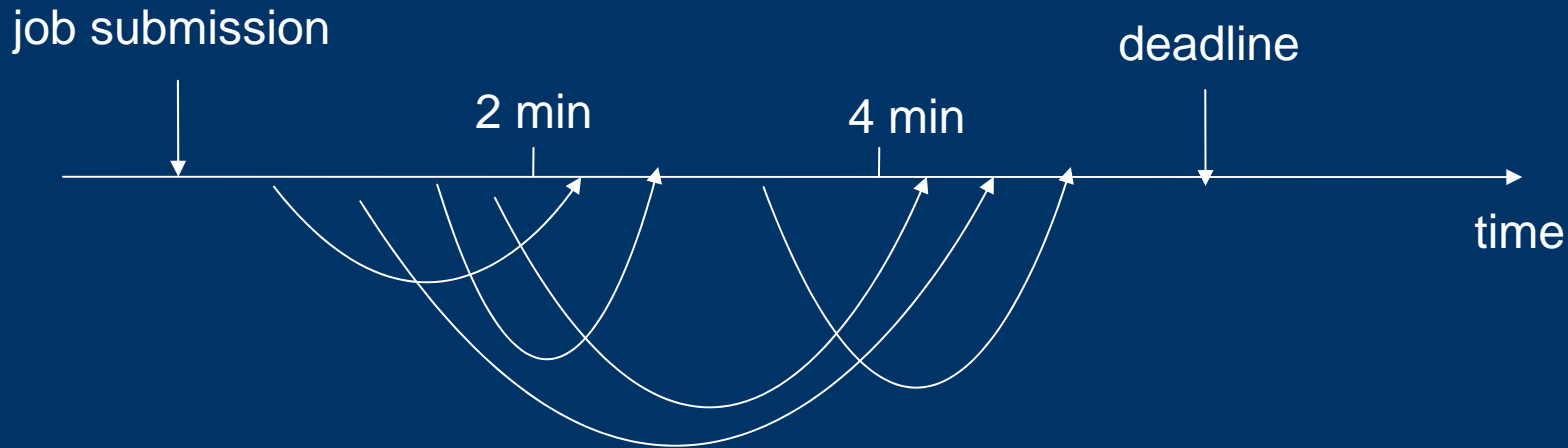
Using free networks



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

- Already described



Efficient validation

- How to validate provably and efficiently with replication factor approaching 1?

Game consoles and graphics chips

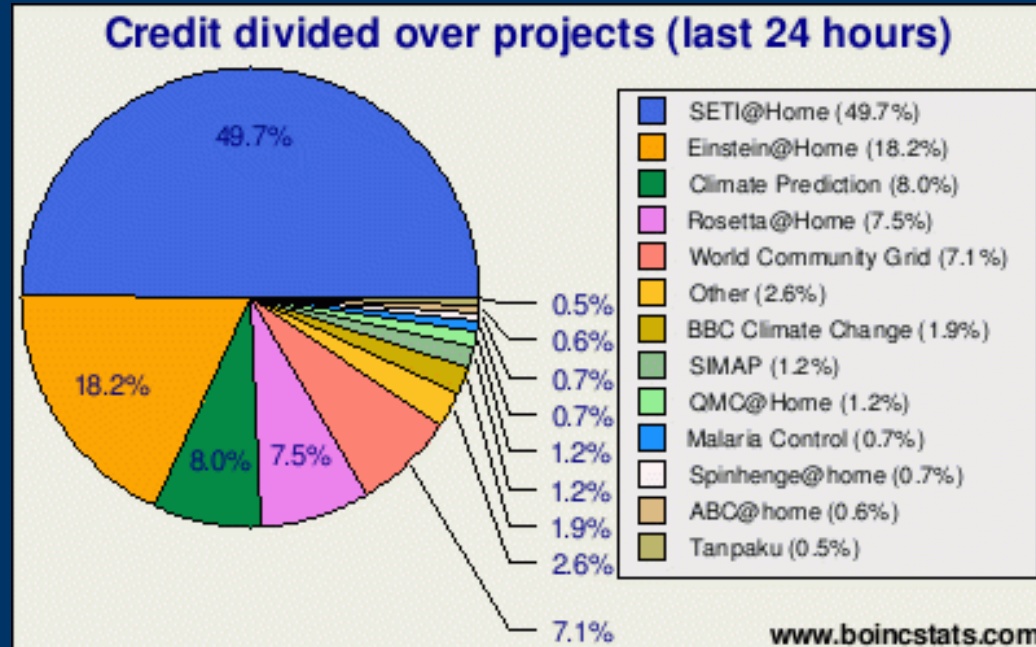
- NVIDIA, ATI, Cell
 - 10X CPU and gaining?
 - Folding@home:
 - ATI version (50 GFLOPS)
 - Sony PS3 version (100 GFLOPS)
 - BOINC and Einstein@home on PS3
 - How to make this available to other projects?
-
-

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

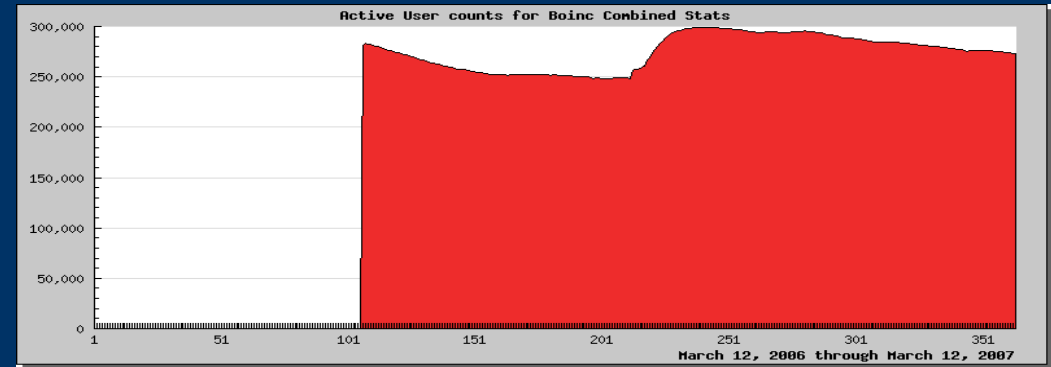
- How to increase the number of volunteers?
 - currently 1 in 1000 PC owners
- How to increase the number of projects?
 - currently stuck at about 50
- How to get volunteers to diversify?



How to attract and retain volunteers?

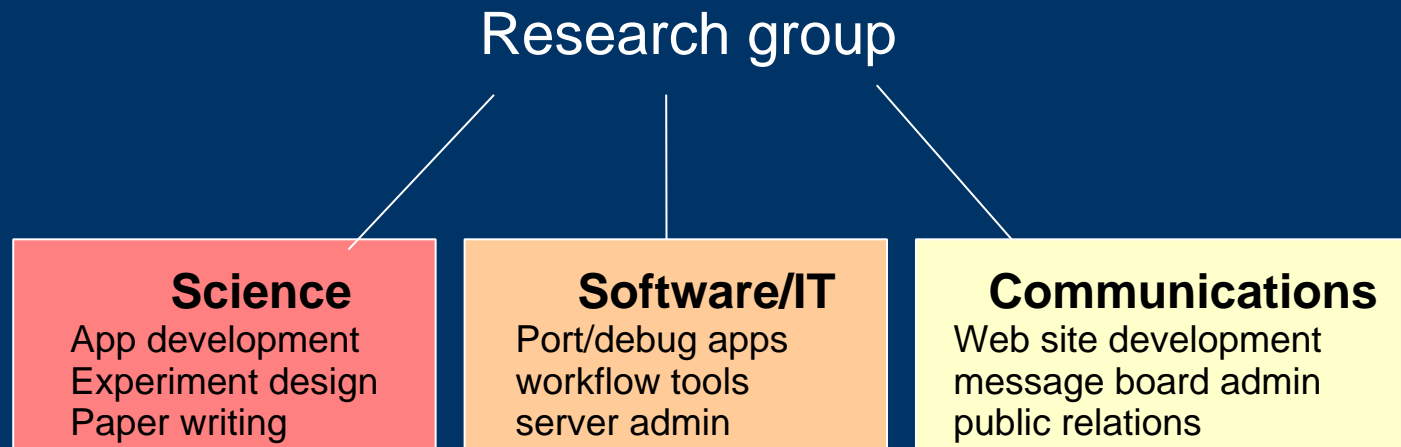
Active hosts:

- 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



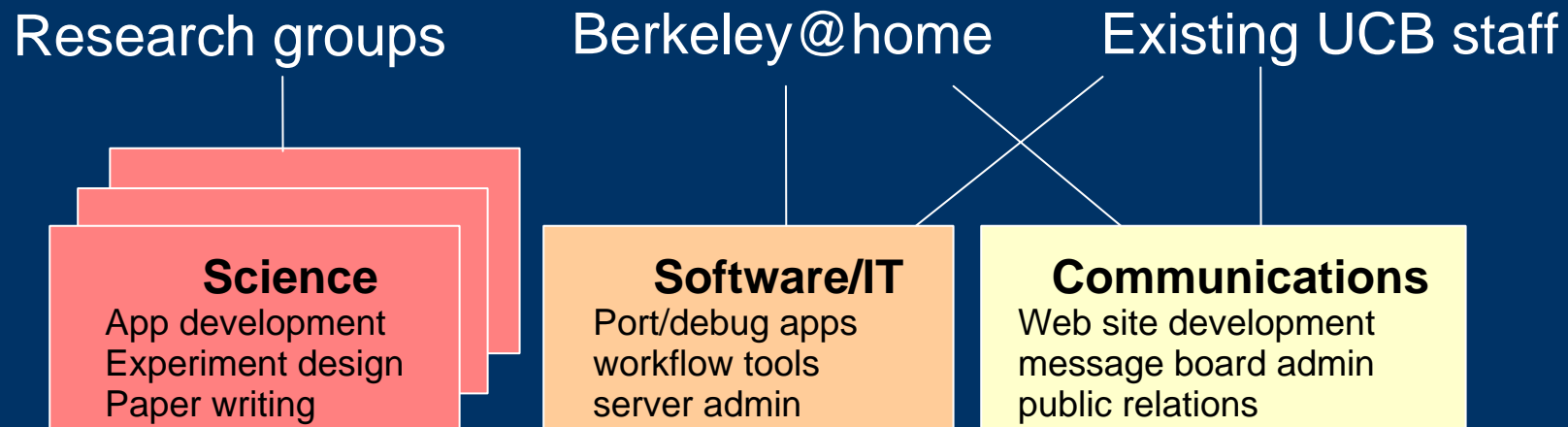
Why aren't there more projects?

- 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
- Social impacts
- Contact me about:
 - Using BOINC
 - Research based on BOINC

davea@ssl.berkeley.edu
