Class 37: How to Find Aliens (and Factors)

CS150: Computer Science
University of Virginia
Computer Science

David Evans
http://www.cs.virginia.edu/evans

Finding Aliens

Arecibo Antenna, Puerto Rico

From http://www.adl.gatech.edu/research/tff/ph1final_radio.html

Finding the Aliens

```
for signal in signals:
   power = findPowerSpectrum (signal)
   if (isAlien (power)):
      print "Found an alien!" + signal
```


Processing Signals

- Power spectrum
- Find patterns in signal
- Eliminate natural and human-made signals
- Today:
  - BlueGene, 280Tflops/s

  ~$200 per Gigaflop
- No success finding aliens

Cray T3E-1200E (~1998)
1 Teraflop/s
= Trillion floating point operations/sec

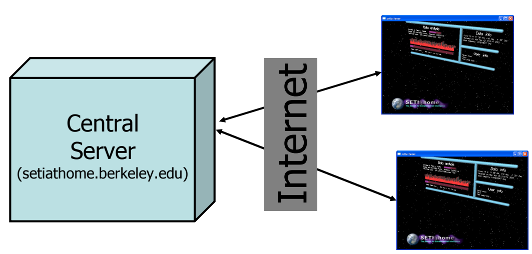
Finding the Aliens Cheaper

```
parfor signal in signals:
   power = findPowerSpectrum (signal)
   if (isAlien (power)):
      print "Found an alien!" + signal
```

Parallel for: instead of doing each element sequentially in order, we can do each element in parallel on a different machine.

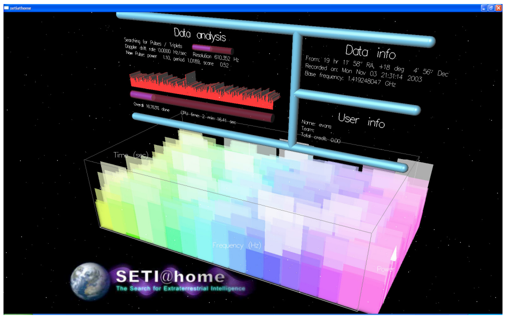
Note: python does not actually have parfor, but other languages do.

Public Distributed Computing

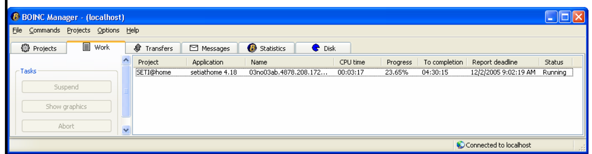


Millions of home PCs

BOINC (Seti@Home) Client



BOINC Manager



Incentives are Necessary


Top participants

Rank	Name	Recent average credit	Total Credit	Country	Participant since
1	NEZ	298,075.79	31,326,386.79	United States	5 Feb 2000 6:22:19 UTC
2	Carsten Giese	54,391.89	2,416,452.01	Germany	6 Mar 2003 18:18:36 UTC
3	Burlsey	24,429.98	3,325,772.60	United Kingdom	2 Jul 1999 21:54:32 UTC
4	TimC	23,308.50	1,273,540.60	United States	4 Mar 2002 21:33:48 UTC
5	University of Oulu - IT Administration	22,957.94	5,030,529.58	Finland	4 Nov 2004 14:26:55 UTC
6	finrod	21,034.23	1,651,804.70	United States	15 Nov 1999 16:28:00 UTC
7	modboy	17,972.80	2,811,157.32	United Kingdom	1 Jan 2004 14:34:26 UTC
8	Informatics at Edinburgh	17,649.49	3,027,702.18	United Kingdom	15 May 1999 15:37:27 UTC

Incentives are Dangerous

- People will cheat*
- How to cheat?
  - Respond with the "there are no aliens" message without actually doing all the work
- Chances of getting caught?
  - 0 (Assumes all jobs have no aliens. So far this is true.)

* Only applies in real world, not at UVa.

Preventing Cheaters

- Send the same job to multiple workers
  - Wastes computing
  - What if the cheater controls many machines?
- Instead of response being "no aliens" make clients send a response that proves they did computation
- Sometimes send out fake jobs that have aliens in them
  - Clients must find the fake aliens
  - Need to make sure the fake jobs look just like real jobs
  - (Airport security scanners work like this also)

Doug Szajda and colleagues at University of Richmond work on these problems (see link on notes)

Why is finding aliens so "easy"?

Note: we haven't yet found any aliens, but its easy to set up the computation…

- Can be broken into many tasks
- Each task can be done completely independently
  - No shared memory between the tasks
- The data to describe a task and response is small compared to the computing
  - SETI@home jobs are 350KB data download, 1KB upload, 3.9 Trillion operations (several hours on PC)

"Harder" Task

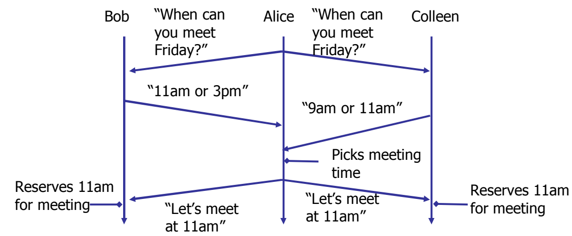
- Dividing your PS8 project work among your team

PS8 Update: you have 1 week left!

If you do not have basic functionality working by tomorrow, arrange to meet with me

Scheduling Meetings

Alice wants to schedule a meeting with Bob and Colleen

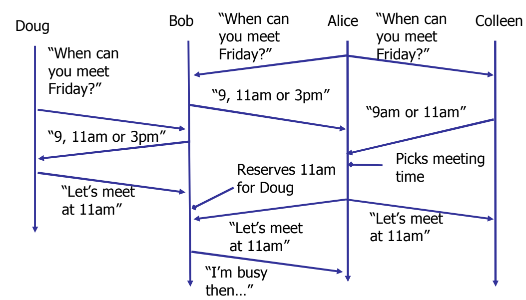


Partial Ordering of Events

- Sequential programs give use a *total ordering* of events: everything happens in a determined order
- Concurrency gives us a *partial ordering* of events: we know some things happen before other things, but not total order

Alice asks to schedule meeting before Bob replies
Alice asks to schedule meeting before Colleen replies
Bob and Colleen both reply before Alice picks meeting time
Alice picks meeting time before Bob reserves time on calendar

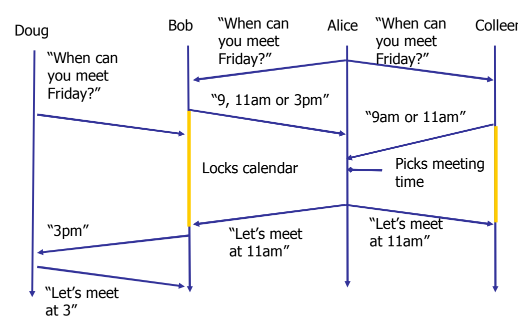
Race Condition



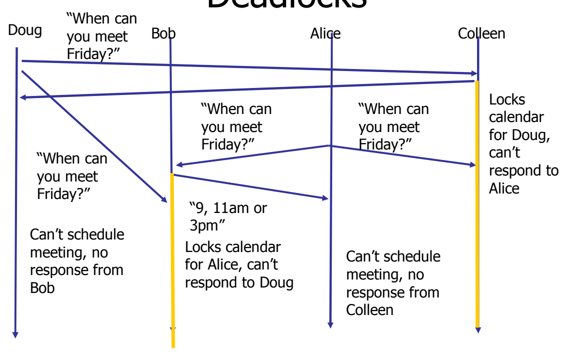
Preventing Race Conditions

- Use locks to impose ordering constraints
- After responding to Alice, Bob reserves all the times in his response until he hears back (and then frees the other times)

Locking



Deadlocks



Why multiple processors is hard?

- Too few ordering constraints: race conditions
- Too many ordering constraints: deadlocks
- Hard/impossible to reason modularly
  - If an object is accessible to multiple threads, need to think about what any of those threads could do at any time!

Worry: nearly all standard desktop computers will be multi-processors soon!

Charge

- The easiest way to solve distributed scheduling problems is to "undistribute" them:
  - Find your teammates now and make sure you know what you are doing next week
- Wednesday: Google
  - Read the paper distributed today
- Friday: Review
  - Send me your questions and topic requests
- Monday: PS8 presentations/demos