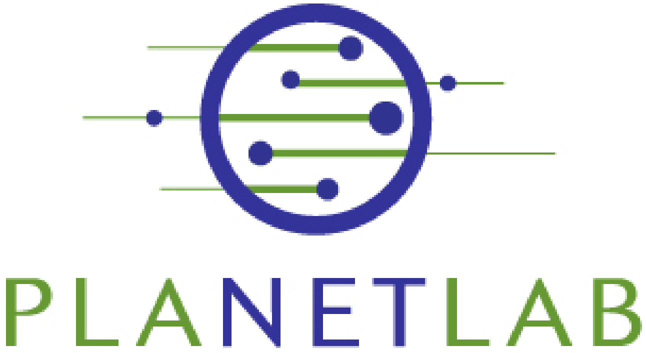
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chapter 8
PLANETLAB

Computer Networks
Timothy Roscoe
Summer 2007

Overview

- PlanetLab
  - *Systems* research in networking
    - Many other approaches…
  - The state of Internet research today
  - The possible future of the Internet
- Other cool stuff from Prof. Wattenhofer
- Questions about the course, exam, life, etc.



PlanetLab is…

- Large collection of machines spread around the world for distributed systems research
  - I can deploy and run my code in Seoul, San Francisco, Rio de Janeiro, Moscow, Mumbai, …
- Focus/catalyst for networking and systems community
  - Most major Universities now host sites
- Intel project ⇒ consortium of companies and universities

The value proposition

- Institutions join, provide nodes
  - ≥ IA32 architecture servers
  - Hosted outside the firewall
  - Provide power, cooling, & bandwidth
- In exchange, researchers get to use a small “slice” of many machines worldwide.

What is PlanetLab good for?

- *Planetary-Scale* networked applications:
  - **Low latency** to widely spread users
  - **Span boundaries:** jurisdictional and administrative
  - **Simultaneous viewpoints:** on the network or sensors
  - **Hardware deployment** is undesirable
- Long-running services, not just experiments
- Overlay networks

PlanetLab is not…

- A distributed supercomputer
- A simulation platform
- An Internet emulator
- An arena for repeatable experiments
- Completely representative of the current Internet

PlanetLab is…

- An opportunity to qualitatively validate distributed systems research in a real deployment environment
- An opportunity to gain valuable experience about what works and what doesn’t in the wide area at scale

Why is it successful?

- Community “P2P”-like model
  - “network effects”
  - Lots of benefit from small entry fee
- *Sliceability*
  - Enables multiple approaches
  - Sharing of the platform
- Virtual machine interface
  - Emphasis on *multiplexing the machine*
  - *Isolation* left to the VMM

Motivation

- New class of services & applications emerging
  - Spread over a sizable fraction of the net
  - CDNs, P2P as the first examples
- Architectural components are beginning to emerge
  - Distributable hash tables provide scalable translation
  - Distributed storage, caching, instrumentation, mapping, ...
- The next Internet will start as an overlay on the current one
  - as did the last one…
  - it will be defined by its services, not its transport
    - translation, storage, caching, event notification, management
- There was NO vehicle to try out the next *n* great ideas in this area

Lots of work done in big distributed systems…

- Researchers had no vehicle to try out their next *n* great ideas in this space
  - Lots of architectures
  - Lots of simulations
  - Lots of emulation on large clusters
  - Lots of folks calling their 17 friends before the next deadline
- but *not* the surprises and frustrations of experience at scale to drive innovation

Origins and progress

- “Underground” meeting March 2002
- Intel seeds
  - First 100 nodes
  - Operational support
- First node up July 2002
- By SOSP (deadline March 2003) 25% of accepted papers refer to PlanetLab
- Large presence at SIGCOMM
- 11 out of 27 papers in NSDI 2004
- Now…

So what are people running?

Ping!

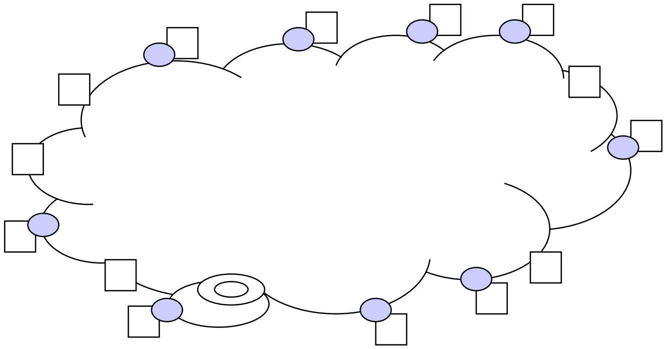
What do people use it for? Some early examples:

- Overlay Networks
- RON, ROM++, ESM, XBone, ABone, etc.
- Network measurement
  - Scriptroute, *Probe, I3, etc.
- Application-level multicast
  - ESM, Scribe, TACT, etc.
- Wide-area distributed storage
  - Oceanstore, SFS, CFS, Palimpsest, IBP
- Resource allocation
  - Sharp, Slices, XenoCorp, Automated contracts
- Distributed query processing
  - PIER, IrisLog, Sophia, etc.
- Content Dist. Networks
  - CoDeeN, ESM, UltraPeer emulation, Gnutella mapping
- Management and Monitoring
  - Ganglia, InfoSpect, Scout Monitor, BGP Sensors, etc.
- Distributed Hash Tables
  - Chord, Tapestry, Pastry, Bamboo, etc.
- Virtualization and Isolation
  - Denali, VServers, SILK, Mgmt VMs, etc.
- Router Design implications
  - NetBind, Scout, NewArch, Icarus, etc.
- Testbed Federation
  - NetBed, RON, XenoServers
- Etc., etc., etc.

What is it good for?

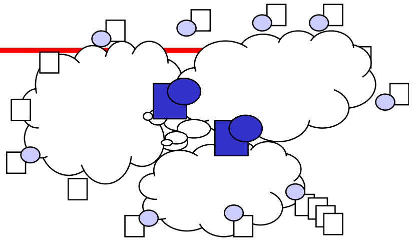
- PlanetLab addresses the related problems of:
  - *Deploying* widely-distributed services
  - *Evaluating* competing approaches in a realistic setting
  - *Evolving* the network architecture to better support such services
- Only game in town for most networking research
  - Other than building into Azureus…
- See demo…

Guidelines

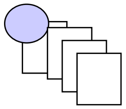


- Thousand viewpoints on “the cloud” is what matters
  - not the thousand servers
  - not the routers, per se
  - not the pipes

Guidelines

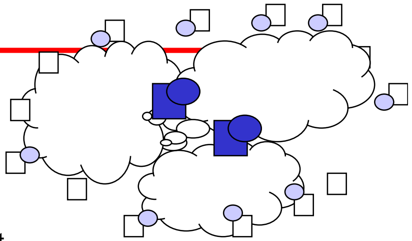


- Each service needs an overlay covering many points
  - logically isolated
- Many concurrent services and applications
  - must be able to slice nodes => VM per service
  - service has a slice across large subset
- Must be able to run each service / app over long period to build meaningful workload
  - traffic capture/generator must be part of facility
- Consensus on “a node” more important than “which node”



Guidelines

Management, Management, Management



- Platform as a whole must be up a lot
  - global remote administration and management
    - mission control
  - redundancy within
- Each service will require its own remote management capability
- Platform nodes cannot “bring down” their site
  - generally not on main forwarding path
  - proxy path
  - must be able to extend overlay out to user nodes?
- Relationship to firewalls and proxies is key

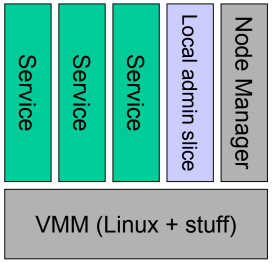
PlanetLab relationships

- PlanetLab ⇔ member institutions
  - Shared control of nodes
- PlanetLab ⇔ research users
  - *Distributed virtualization*, slices
- PlanetLab ⇔ research builders
  - Shared interfaces, *unbundled mgmt*
- PlanetLab ⇔ rest of the Internet
  - Isolation, security, packet auditing
  - See web interface for nodes…

Distributed Virtualization

- *Services* run in *slices*.
- Slice: set of virtual machines (*slivers*)
- Created by *slice creation service* acting as a broker

Node structure:



Service

Service

Service

Local admin slice

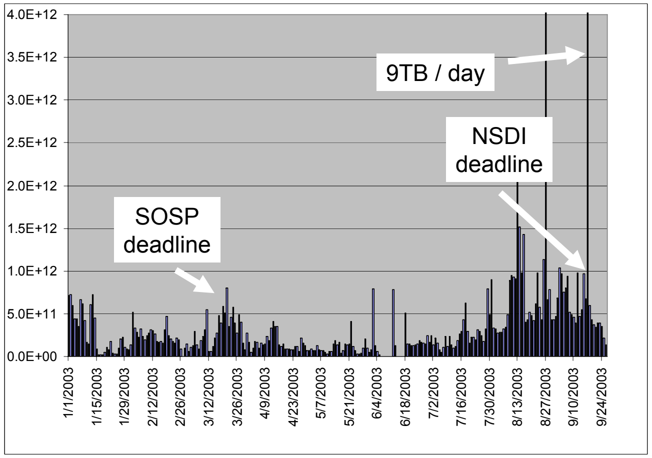
Node Manager

VMM (Linux + stuff)

Requirements for slices

- Isolation of virtual machines:
  - Allocate resources
  - Partition or contextualize namespaces
  - Provide stable programming base
- Isolation of PlanetLab:
  - Resource accounting and limits
  - Auditing of slice *actions* (eg. packets)
    - Unexpected requirement!

Aggregate bandwidth usage



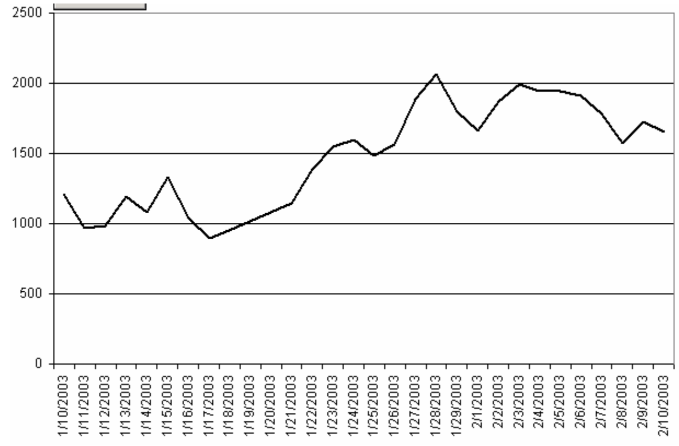
9TB / day

NSDI deadline

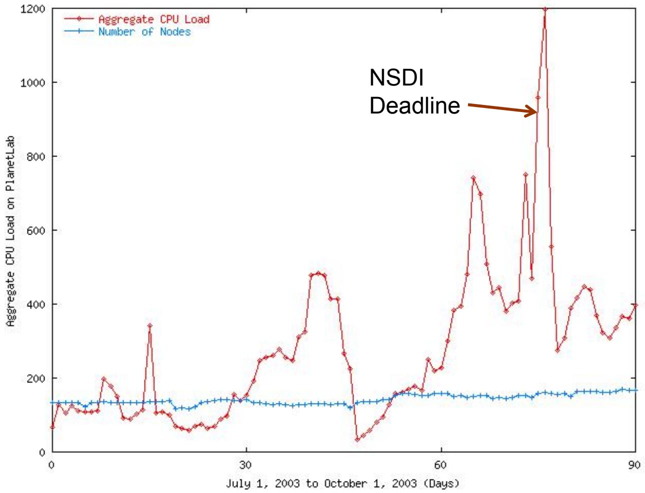
SOSP deadline

Pre-SIGCOMM 2003 deadline

(GB per day)



Aggregate CPU load



NSDI Deadline

How to use PlanetLab

- Don’t expect:
  - Repeatable experiments, other than very long-running studies
  - Large numbers of nodes (approx 500)
  - Lots of CPU (machines are loaded!)
  - High availability (machines reboot without warning)
- Do expect:
  - The unexpected!
  - Real experience running a service
  - Real users (if you want them)
  - Lots of interesting challenges
  - To find out if your idea really works

Best practice

1. Build a real system
2. Debug it in the lab on your own network
3. Try it out on PlanetLab to ensure it works
4. Experiment on EmuLab for repeatability
5. Use simulation for scalability
6. Cross-validate your results!
7. Deploy on PlanetLab to get real experience
8. Publish, graduate, get job as prestigious professor

Lessons from PlanetLab

- Nothing works as expected at scale!
  - Many unintended and unexpected consequences of algorithmic choices
  - Simulation results do not carry over well
    - Simulate, deploy, measure, edit cycle
- Evaluating competing approaches “in the wild” refines techniques
- The ability to try things out “for real” seems to stimulate ideas

What is PlanetLab doing to the Internet?

- PlanetLab functions as a disruptive technology
  - Applications use the network differently
  - The network sometimes reacts in a hostile manner
  - Leads to new requirements on infrastructure

Operational experience

- UDP replaces TCP?
  - N-to-N applications are different
  - Removes abstraction barriers
    - Aggressive application timeouts
    - Late data choice
    - Detailed information about network perf.
  - Still congestion controlled
  - DCCP not yet available

Operational experience

- The Internet is extraordinarily brittle
  - Innovation resembles Denial-of-Service
  - IDSes default to “attack warning”
  - “Common cases” burned into routers
- PlanetLab now supports *full packet auditing* to end-users

Long-term aims

- PlanetLab incubates the next Internet
  - Now: GENI (PlanetLab + pipes + $400,000,000)
- New networks deployed as overlays over existing Internet
- *Service-oriented* network becomes the norm
- Computation as a localizable network resource

Conclusion

- Think of PlanetLab as a communal **shared artefact** for researchers
- Provides many diverse, overlapping projects around the world with a stable **place to stand** to change things
- Forum for **exchange** and **composition** of services and applications
- **Selection environment** based on real deployment and use
- **Bottom-up approach** to changing the world