PlanetLab: A Blueprint for Introducing Disruptive Technology into the Internet

Larry Peterson
Princeton University / Intel Research
PlanetLab vs Grid

• The Grid is the next generation supercomputer

• PlanetLab is the next generation Internet
Innovator’s Dilemma

• The Internet is an enormous success story
  – commercially
  – impact on our daily lives
  – global reach
• Success has an unexpected cost: *ossification*
  – difficult to deploy disruptive technologies
    ▪ correct vulnerabilities
    ▪ introduce new capabilities
Today’s Internet

Best-Effort Packet Delivery Service

Limitations
- the Internet is “opaque” making it difficult to adapt to current network conditions
- applications cannot be widely distributed (typically split into two pieces: client and server)
Tomorrow’s Internet

Collection of Planetary-Scale Services

Opportunities

– multiple vantage points
  - anomaly detection, robust routing
– proximity to data sources/sinks
  - content distribution, data fusion
– multiple, independent domains
  - survivable storage
Berkeley: OceanStore

RAID distributed over the whole Internet
Intel: Netbait

Detect and track Internet worms globally
Washington: ScriptRoute

Internet Measurement Tool
Princeton: CoDeeN

Open Content Distribution Network
Evolving the Internet

• Add a new layer to the network architecture
  – overlay networks
    ▪ purpose-built virtual networks that use the existing Internet for transmission
    ▪ the Internet was once deployed as an overlay on top of the telephony network

• Challenge
  – how to innovate & deploy at scale
The Story so Far

• The Internet is a tremendous success, but…
  – the architecture has fundamental limits
  – its very success makes it hard to change

• The research community is teeming with innovative planetary-scale services
  – exploit multiple points-of-presence throughout the net

• Overlays offer an attractive way to introduce disruptive technology into the Internet, but…
  – there is a high barrier-to-entry
PlanetLab is...

Goal of 1,000 widely-distributed machines
- today: 185 machines, 75 sites, 16 countries
- at edge sites and network cross-roads
PlanetLab is…

A common software package

• Main components
  – Linux kernel w/ extensions to support isolation
  – bootstrapping and software distribution mechanisms
  – collection of *unbundled management* services

• Collectively support distributed virtualization
  – run many overlay networks simultaneously
  – each service (overlay) runs in a *slice* of PlanetLab’s global resources
Slices
PlanetLab is...

A test-bed for experimenting with network services

- 120+ active research projects
- Advantages
  - experiment at scale
  - experiment under real-world conditions
  - potential for real workloads and users
PlanetLab is…

A deployment platform

• Continuously-running services
  – CoDeeN content distribution network (Princeton)
  – Sophia distributed query processing engine (Princeton)
  – PIER distributed query processing engine (Berkeley)
  – ScriptRoute network measurement tool (Washington)
  – NetBait worm detection service (Intel)
  – Chord scalable object location service (MIT, Berkeley)
  – OceanStore storage system (Berkeley)
PlanetLab is…

A microcosm of the next Internet

• Fold services back into PlanetLab
  – evolve core technologies to support overlays and slices

• Examples
  – Sophia used to monitor health of PlanetLab nodes
  – Chord provides scalable object location

• Long-term goals
  – develop open protocols and standards
    ▪ allow federation of public & private “PlanetLabs” to co-exist
  – discover common sub-services
PlanetLab is…

A research community

• Started as a grass-roots effort
  – 35 researchers gathered in March of 2002
  – Intel provided seed funding
  – self-organized into five working groups

• Next Phase: Academic/Industrial Consortium
  – hosted by Princeton (w/ Berkeley and Washington)
  – build-out and operate the infrastructure
  – lower the barrier to entry for research and teaching
Software Architecture

- Support distributed virtualization
  - *slice*: a network of virtual machines
  - multiple services run concurrently (some long-lived)
  - deploy version $i$ of PlanetLab on version $i+1$

- Per-Node Components
  - create and isolate virtual machines

- Global Components (Services)
  - create slice across a set of nodes
  - monitor node health
  - routing underlay
Per-Node Components

• Node Manager
  – responds to requests to create a virtual machine
  – defines spec for VM
    ▪ resources consumed
    ▪ network name space consumed
  – performs admission control
• Vserver: virtualizes at system call interface
  – each vserver runs in its own security context
    ▪ private UID/GID name space
    ▪ limited superuser capabilities (e.g., no CAP_NET_RAW)
  – uses `chroot` for file system isolation
  – scales to hundreds of vservers per node
Per-Node (cont)

- **plkmod**: kernel module that enforces VM isolation
  - processor and link scheduling
  - virtualizes the network
    - safe raw sockets
    - port-space isolation
    - address space sandboxing

- **Sensors**: uniform interface to node status info
  - HTTP-based
  - core set + user-defined

- **Admin Slice**: local admin control
  - set bw limits
  - run tcpdump
Creating Slices

• Two-stage process
  – discover available resources
    ▪ use monitoring service
  – create virtual machine on each selected node
    ▪ contact broker for rights to resources (receive tickets)
    ▪ contact node manager to redeem tickets
    ▪ node manager implements admission control

• Status
  – prototypes of mechanisms
  – simple policies in the near-term
  – create a market for resources in the long-term
Monitoring Services

• Serve several purposes
  – discover/select resources for a slice
  – monitor node/network health
  – measure/monitor Internet activity

• Exploit sensors
  – local state + local view of the network

• Multiple services being built
  – Sophia: distributed Prolog engine
  – PIER: distributed SQL query processor
  – IrisNet: XML-based queries
  – service-specific mechanisms (e.g., CoDeeN)
Routing Underlay

Overlay Services

Library of Routing Services
  • $k$-disjoint paths
  • $k$-nearest neighbors

Topology Probing Layer
  • peering graph
  • path from $x$ to $y$
  • latency from $x$ to $y$

Raw Topology Information
  • local BGP feed
## Current Institutions

<table>
<thead>
<tr>
<th>Institution</th>
<th>Institution</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academia Sinica, Taiwan</td>
<td>Princeton University</td>
<td>University of Kansas</td>
</tr>
<tr>
<td>Boston University</td>
<td>Purdue University</td>
<td>University of Kentucky</td>
</tr>
<tr>
<td>Caltech</td>
<td>Rensselaer Polytechnic Inst.</td>
<td>University of Maryland</td>
</tr>
<tr>
<td>Carnegie Mellon University</td>
<td>Rice University</td>
<td>University of Maryland</td>
</tr>
<tr>
<td>Chinese Univ of Hong Kong</td>
<td>Rutgers University</td>
<td>University of Massachusetts</td>
</tr>
<tr>
<td>Columbia University</td>
<td>Stanford University</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Cornell University</td>
<td>Technische Universitat Berlin</td>
<td>University of North Carolina</td>
</tr>
<tr>
<td>Datalogisk Institut Copenhagen</td>
<td>The Hebrew Univ of Jerusalem</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Duke University</td>
<td>University College London</td>
<td>University of Rochester</td>
</tr>
<tr>
<td>Georgia Tech</td>
<td>University of Arizona</td>
<td>USC / ISI</td>
</tr>
<tr>
<td>Harvard University</td>
<td>University of Basel</td>
<td>University of Technology Sydney</td>
</tr>
<tr>
<td>HP Labs</td>
<td>University of Bologna</td>
<td>University of Tennessee</td>
</tr>
<tr>
<td>Intel Research</td>
<td>University of British Columbia</td>
<td>University of Texas</td>
</tr>
<tr>
<td>Johns Hopkins</td>
<td>UC Berkeley</td>
<td>University of Utah</td>
</tr>
<tr>
<td>Lancaster University</td>
<td>UCLA</td>
<td>University of Virginia</td>
</tr>
<tr>
<td>Lawrence Berkeley Laboratory</td>
<td>UC San Diego</td>
<td>University of Washington</td>
</tr>
<tr>
<td>MIT</td>
<td>UC Santa Barbara</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>University of Cambridge</td>
<td>Uppsala University, Sweden</td>
</tr>
<tr>
<td>National Tsing Hua Univ.</td>
<td>University of Canterbury</td>
<td>Washington University in St Louis</td>
</tr>
<tr>
<td>New York University</td>
<td>University of Chicago</td>
<td>Wayne State University</td>
</tr>
<tr>
<td>Northwestern University</td>
<td>University of Illinois</td>
<td></td>
</tr>
</tbody>
</table>

September 15, 2003
More Information

www.planet-lab.org