PlanetLab: A Blueprint for Introducing Disruptive Technology into the Internet

Scott Karlin
Princeton University

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

PLANETLAB

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)

PLANETLAB
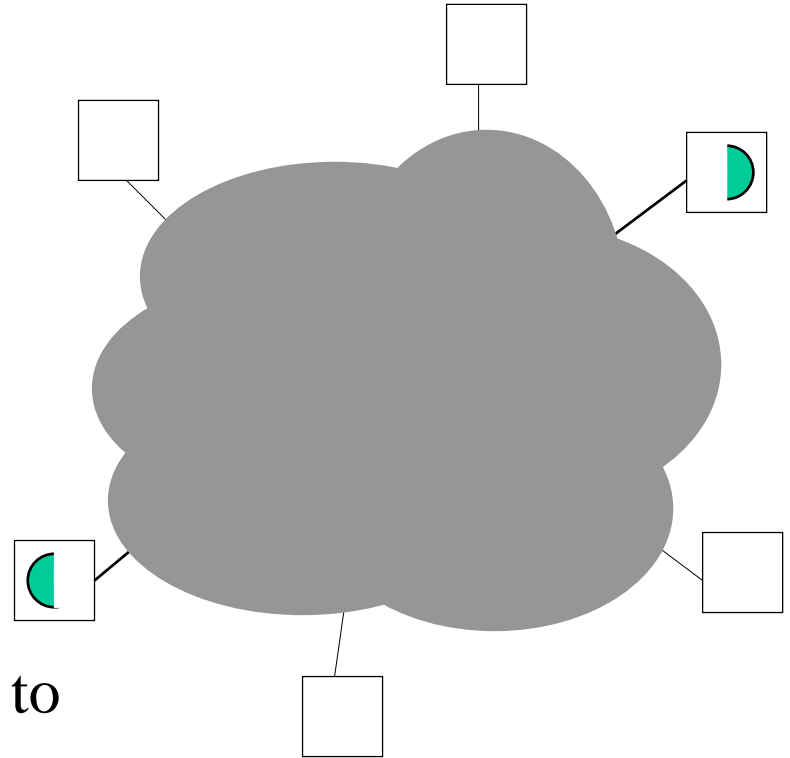
Tomorrow's Internet

Collection of
Planetary-Scale
Services

<u>Opportunities</u>

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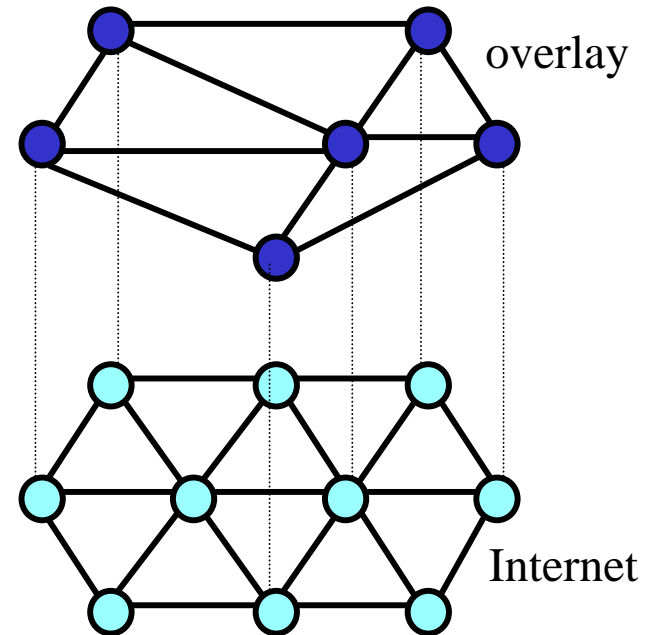 ▪ survivable storage

PLANETLAB
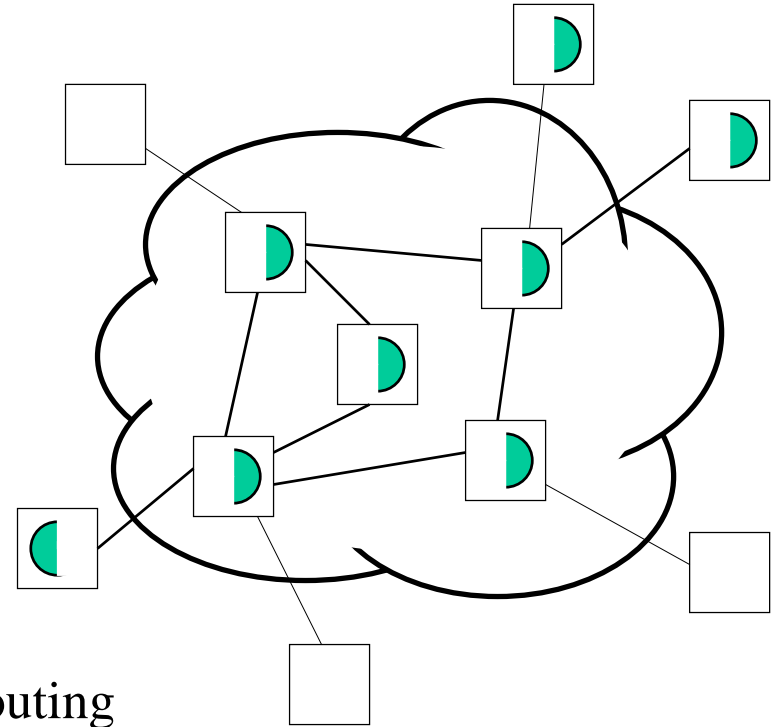
Evolving the Internet

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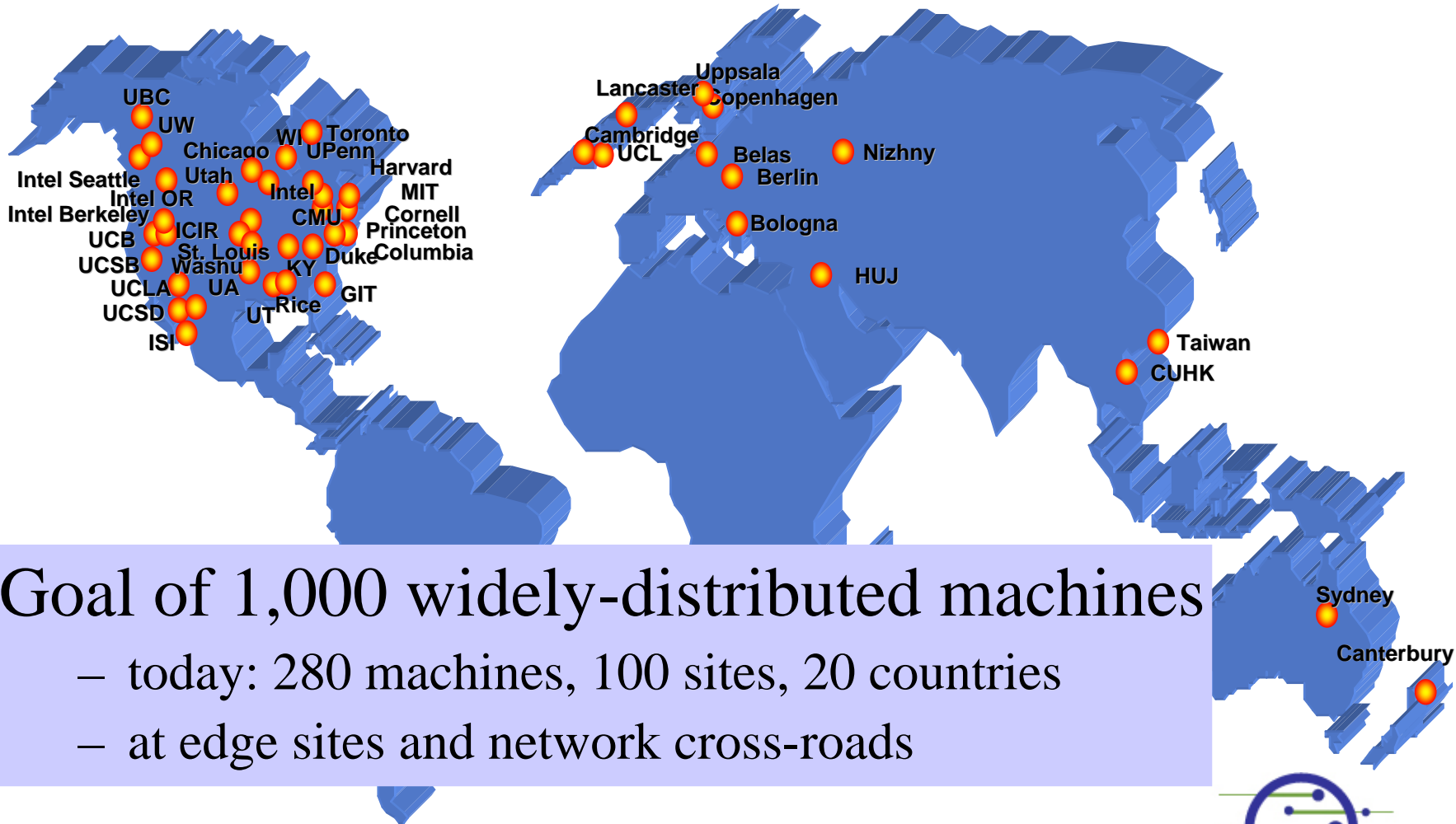   - the Internet was once deployed as an overlay on top of the telephony network

overlay

Internet

- Challenge
  - how to innovate & deploy at scale

PLANETLAB

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 (more later)
  - 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

PlanetLab is…

Goal of 1,000 widely-distributed machines
– today: 280 machines, 100 sites, 20 countries
– at edge sites and network cross-roads

PLANETLAB

PlanetLab's Beginnings

- Started as a grass-roots effort
  - 35 researchers gathered in March 2002
  - Academic and corporate research groups
- Research Approach for Internet-Scale Services has Significant Gap:
  - Simulation
  - Lab-Scale Emulation
  - Ask "family and friends" for accounts elsewhere
  - …
  - Deploy on the Internet (how?)
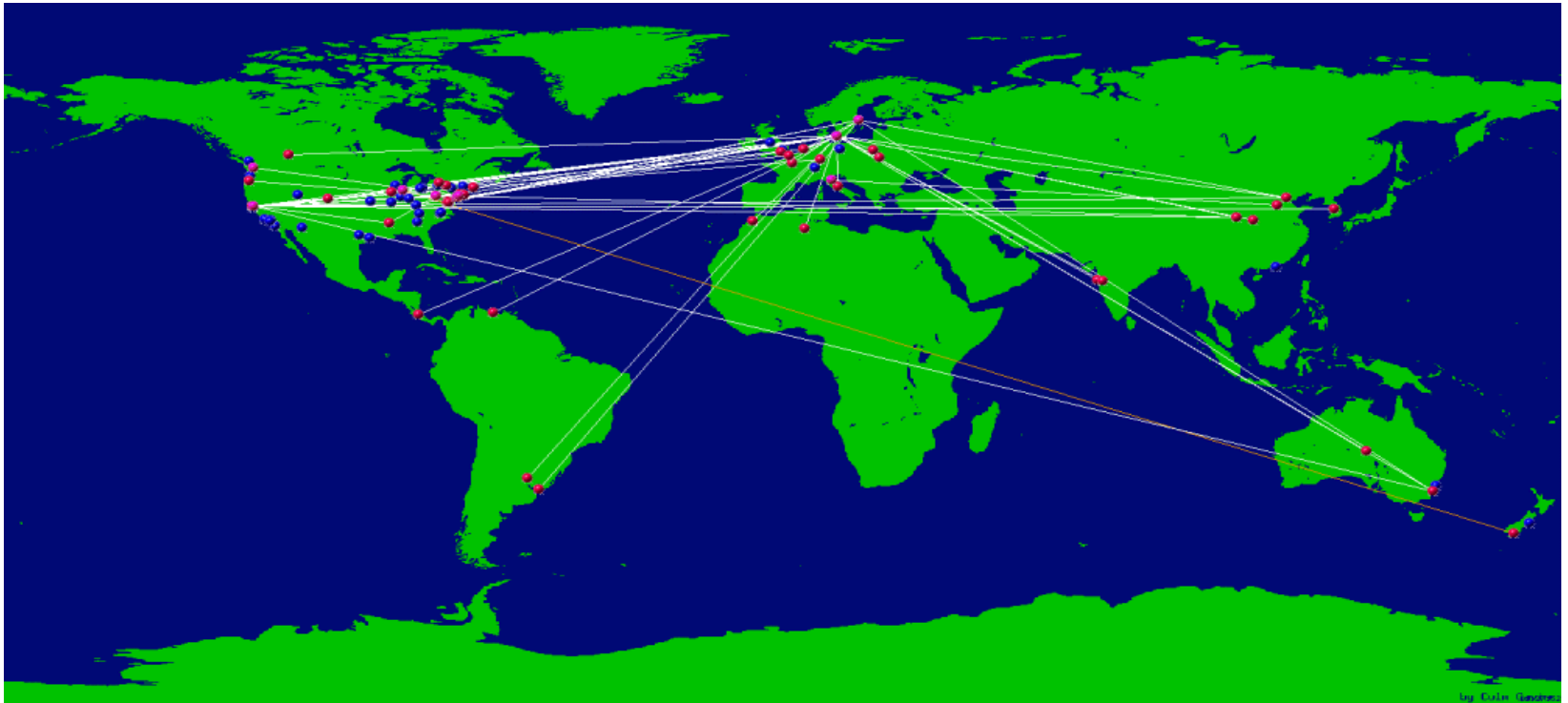- PlanetLab fills the gap

Berkeley: OceanStore

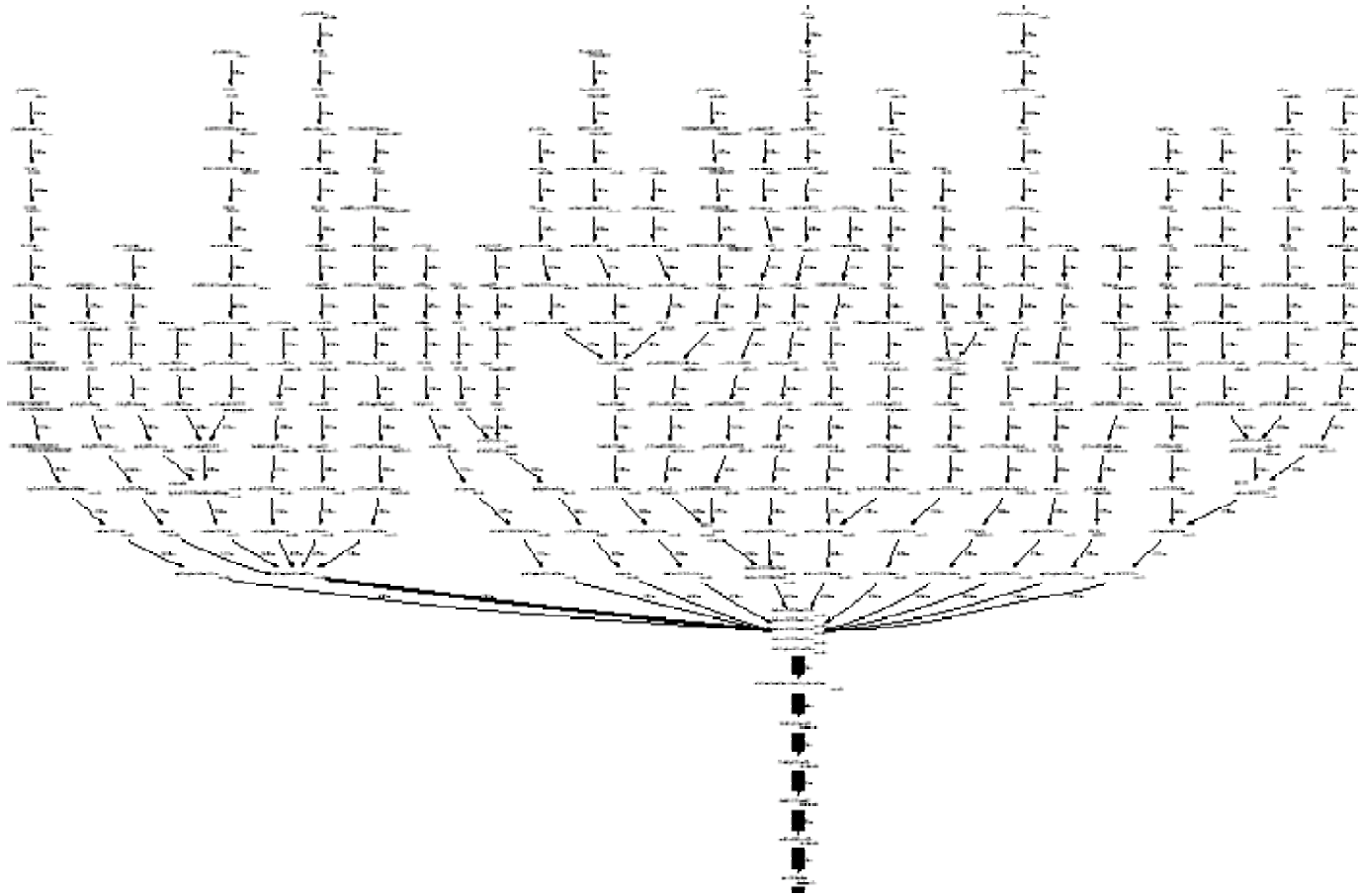RAID distributed over the whole Internet

9
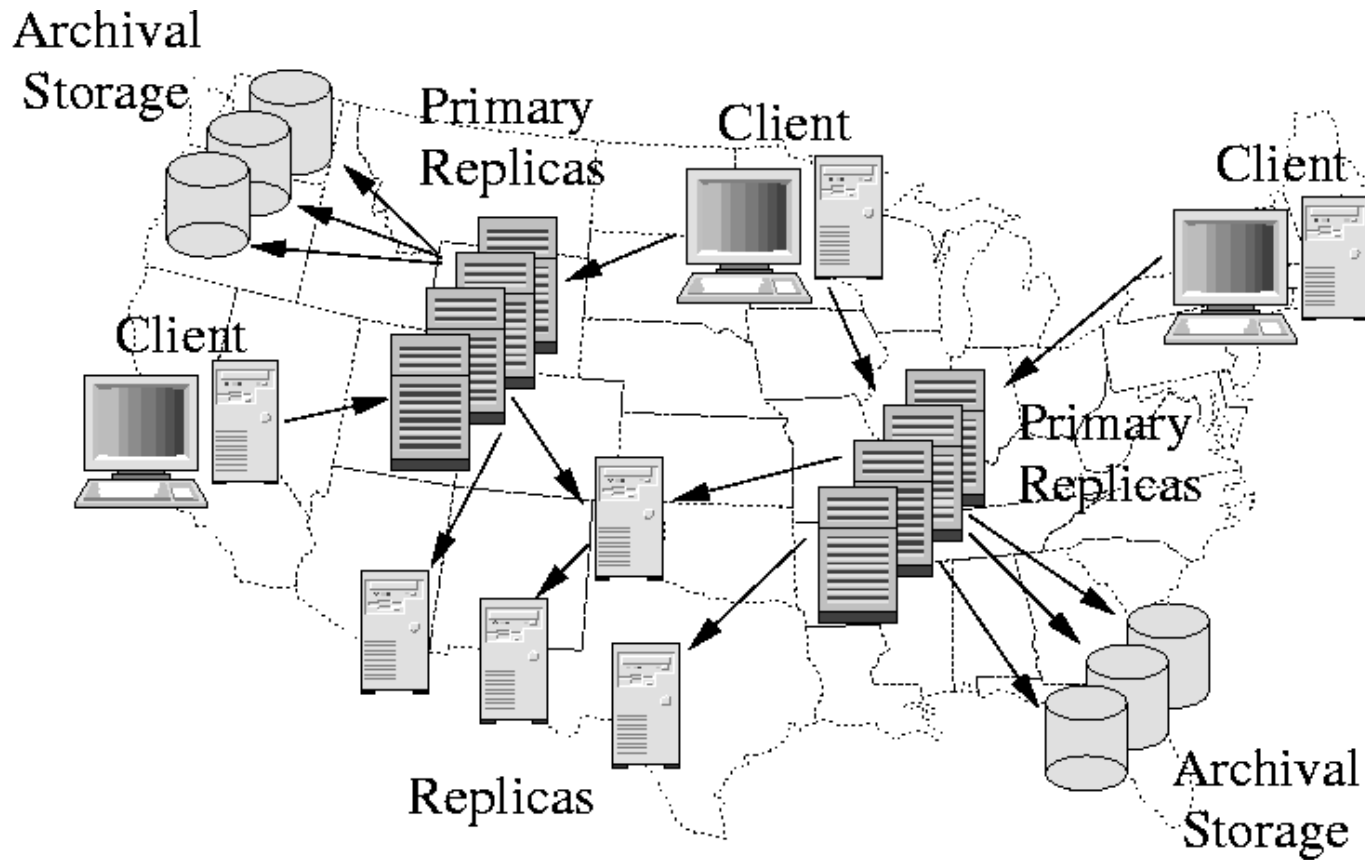
Intel: Netbait

Detect and track Internet worms globally

PLANETLAB

Washington: ScriptRoute

Internet Measurement Tool

PLANETLAB

Princeton: CoDeeN

Open Content Distribution Network

PLANETLAB

PlanetLab is…

A common software architecture

- Distributed virtualization
  - *slice* → a network of virtual machines
  - isolation
    - isolate services from each other
    - protect the Internet from PlanetLab

- Unbundled Management
  - OS defines only local (per-node) behavior
    - global (network-wide) behavior implemented by services
  - multiple competing services (overlays) running in parallel
    - shared, unprivileged interfaces

PLANETLAB

Slices
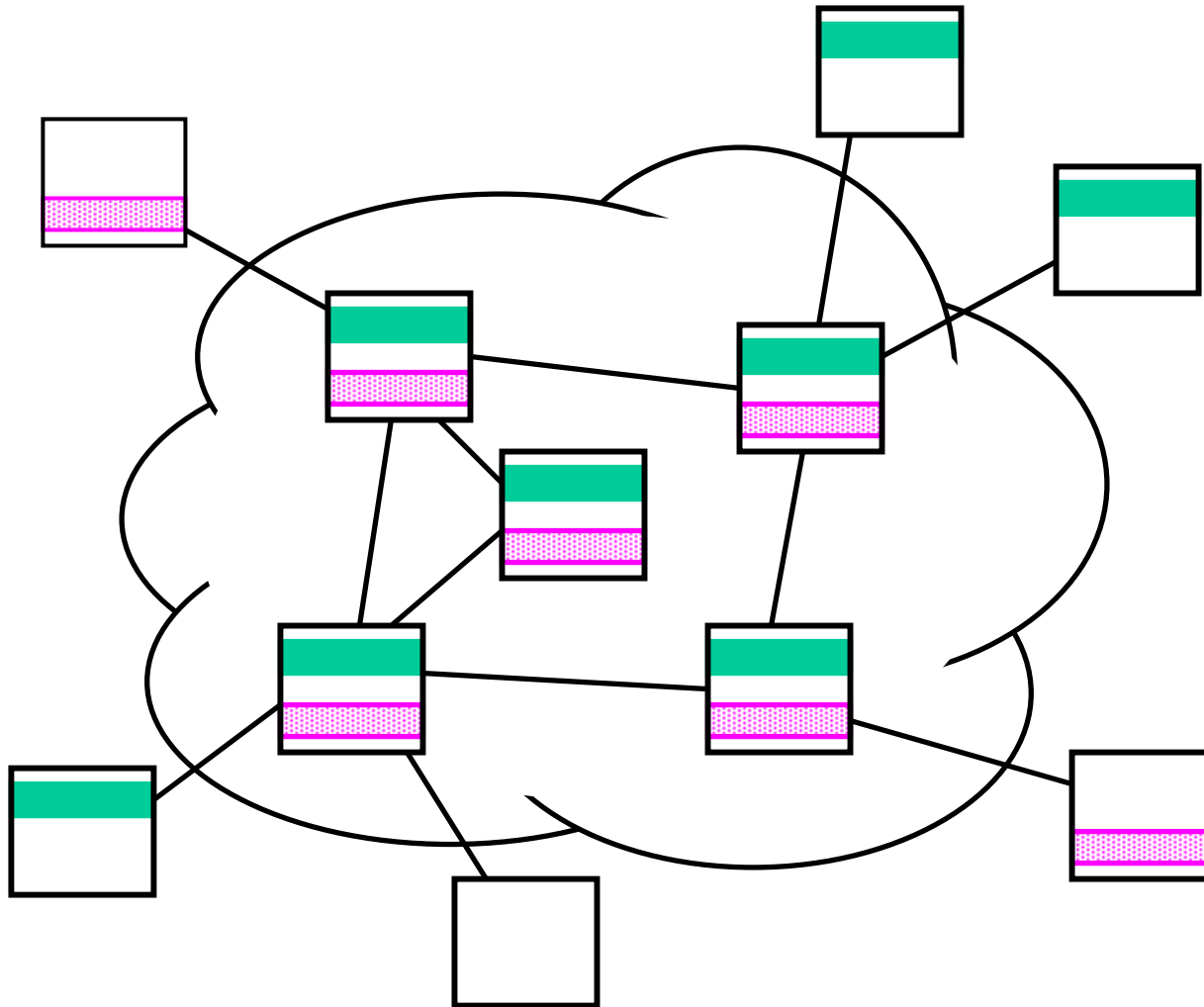
PLANETLAB

PlanetLab is…

A test-bed for experimenting with network services

- 450 active research projects
- Advantages
  - experiment at scale
  - experiment under real-world conditions
  - potential for real workloads and users

PLANETLAB

PlanetLab is…

A deployment platform

- Continuously-running services
  - CoDeeN content distribution network (Princeton)
  - Sophia distributed query processing engine (Princeton)
  - ScriptRoute network measurement tool (Washington)
  - Chord scalable object location service (MIT, Berkeley)
  - …

design ⟶ deploy ⟶ measure ⟶ (design)

PLANETLAB

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

Roadmap

- Introduction
- PlanetLab Overview
- **Virtualization**
- **Infrastructure Services**
- Growth Strategy
- Project Status
- Summary

PLANETLAB

Virtualization Levels

- Hypervisors (e.g., VMWare)
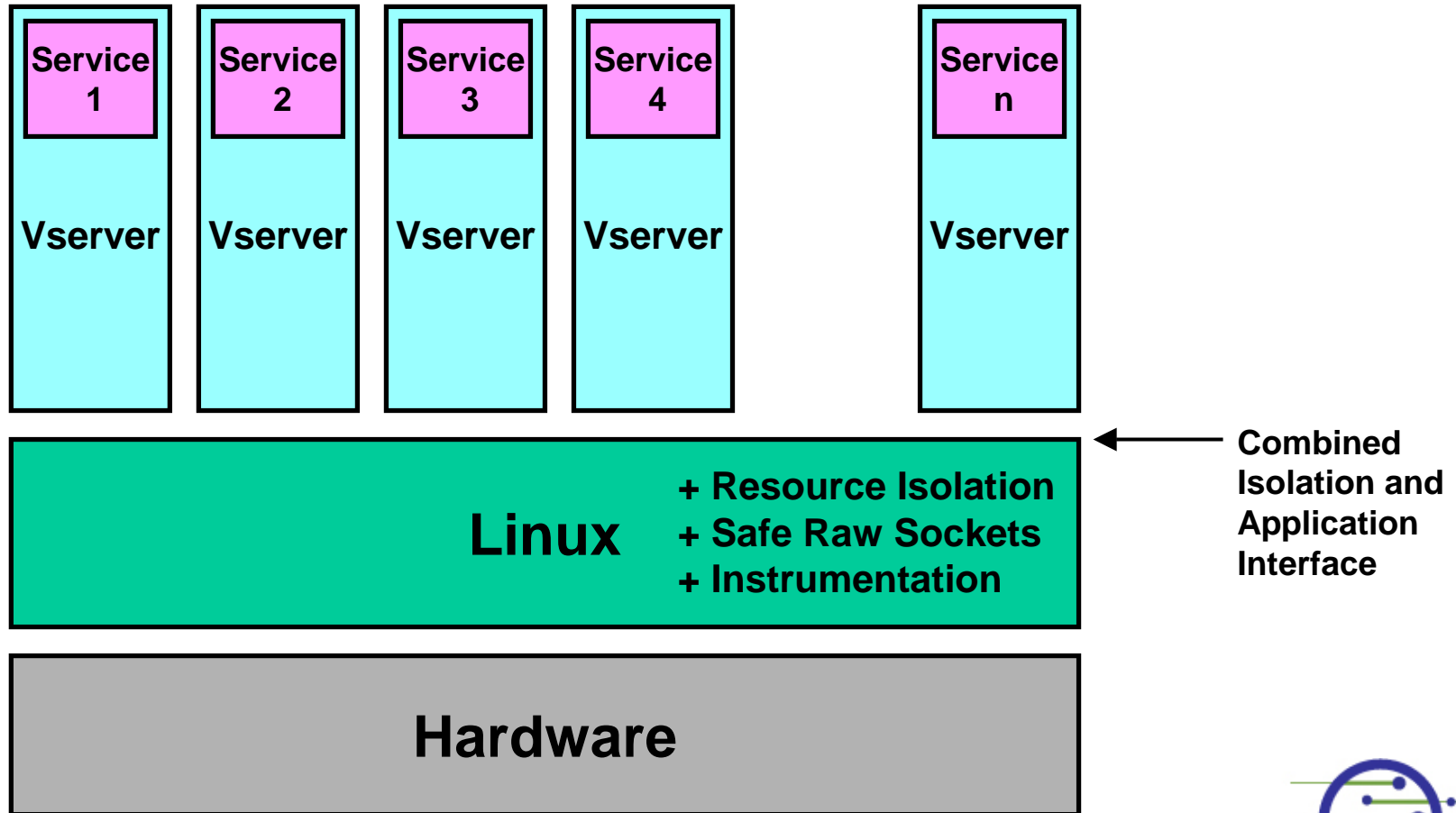 - don't scale well
  - don't need multi-OS functionality

- Paravirtualization (e.g., Xen, Denali)
  - not yet mature
  - requires OS tweaks

- Virtualize at system call interface (e.g., Jail, Vservers)
  - reasonable compromise
  - doesn't provide the isolation that hypervisors do

- Unix processes
  - isolation is problematic

- Java Virtual Machine
  - too high-level

PLANETLAB

Vservers

- 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 1000 of vservers per node (29MB each)
- Isolation
  - kernel schedulers (processor and link bandwidth)
  - address spaces
- Node Manager
  - privileged security context
  - interface for creating virtual machines

PLANETLAB

Virtual Machines

Service 1	Service 2	Service 3	Service 4		Service n
Vserver	Vserver	Vserver	Vserver		Vserver

Linux + Resource Isolation
+ Safe Raw Sockets
+ Instrumentation

← **Combined Isolation and Application Interface**
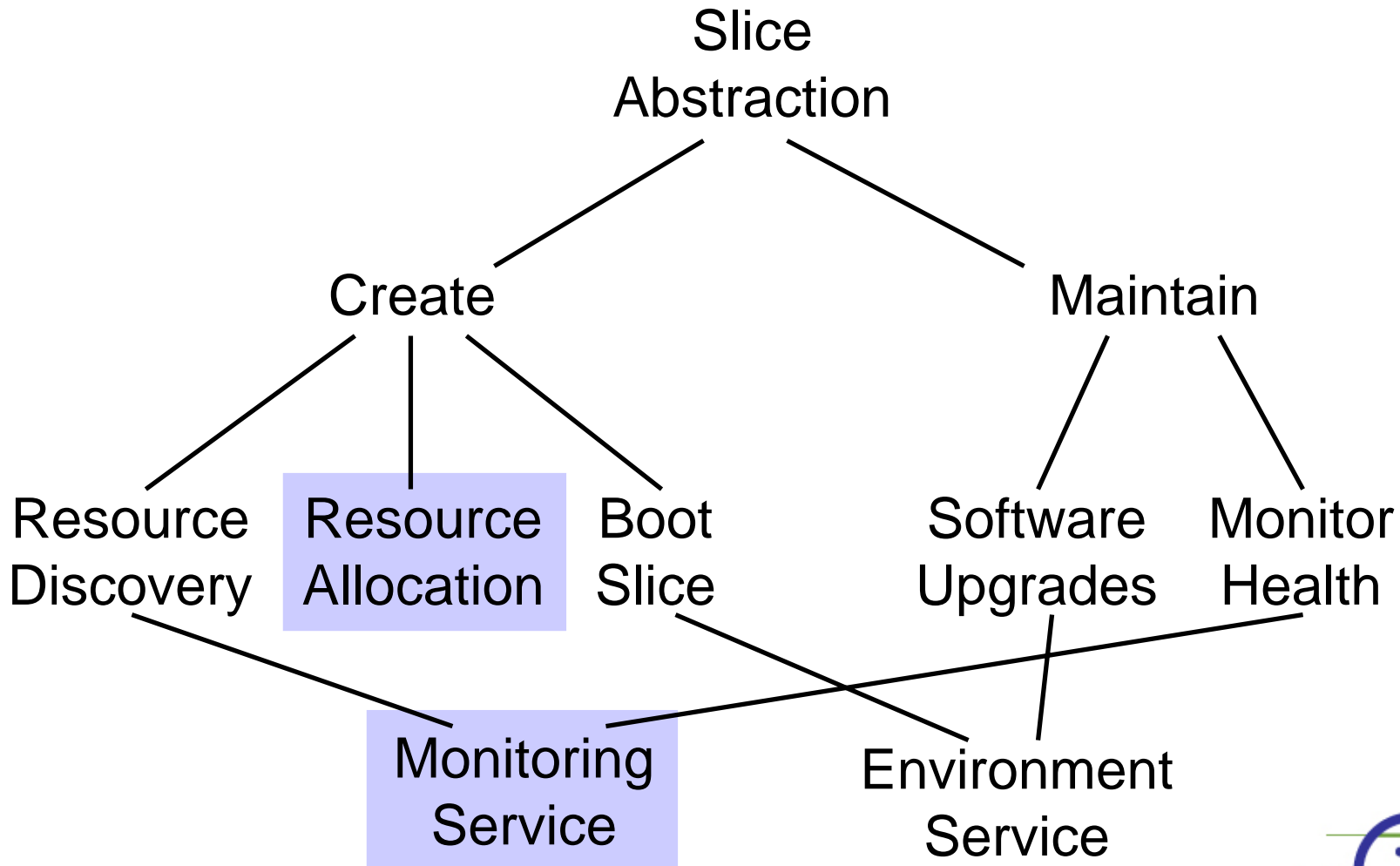
Hardware

PLANETLAB

Network Virtualization

- Standard raw sockets
  - privileged operation
  - access to all packets to/from host

- Safe raw sockets
  - bound to a specific UDP/TCP port (+ related ICMP)
  - ensure that outgoing packets do not spoof

- Other issues
  - rate limiting exceptional packets
  - allowing multiple virtual networks to co-exist

PLANETLAB

Infrastructure Services

Slice
Abstraction

Create   Maintain

Resource   Resource   Boot   Software   Monitor
Discovery   Allocation Slice   Upgrades   Health

Monitoring   Environment
Service   Service

PLANETLAB

Monitoring Services

- Serve several purposes
  - discover/select resources for a slice
  - monitor node/network health (manage PlanetLab)
  - measure/monitor Internet activity (application of PL)
- Exploit sensors on each node
  - local state (/proc) + local view of the network (ping)
  - http://localhost:33080/nodes/ip/name
- Multiple services being built
  - Sophia: distributed Prolog engine
  - PIER: distributed SQL query processor
  - IrisNet: XML-based queries

PLANETLAB

Resource Allocation

- Multiple allocators
  - PlanetLab Central (PLC)
  - SHARP
  - Emulab
  - …
- Interfaces
  - PLC (API + GUI)
  - node manager (initially private)
- Evolution
  - centralized → decentralized
  - policy-heavy → user-specified → market-based

PLANETLAB

PlanetLab Central API

create_slice(name, credentials)

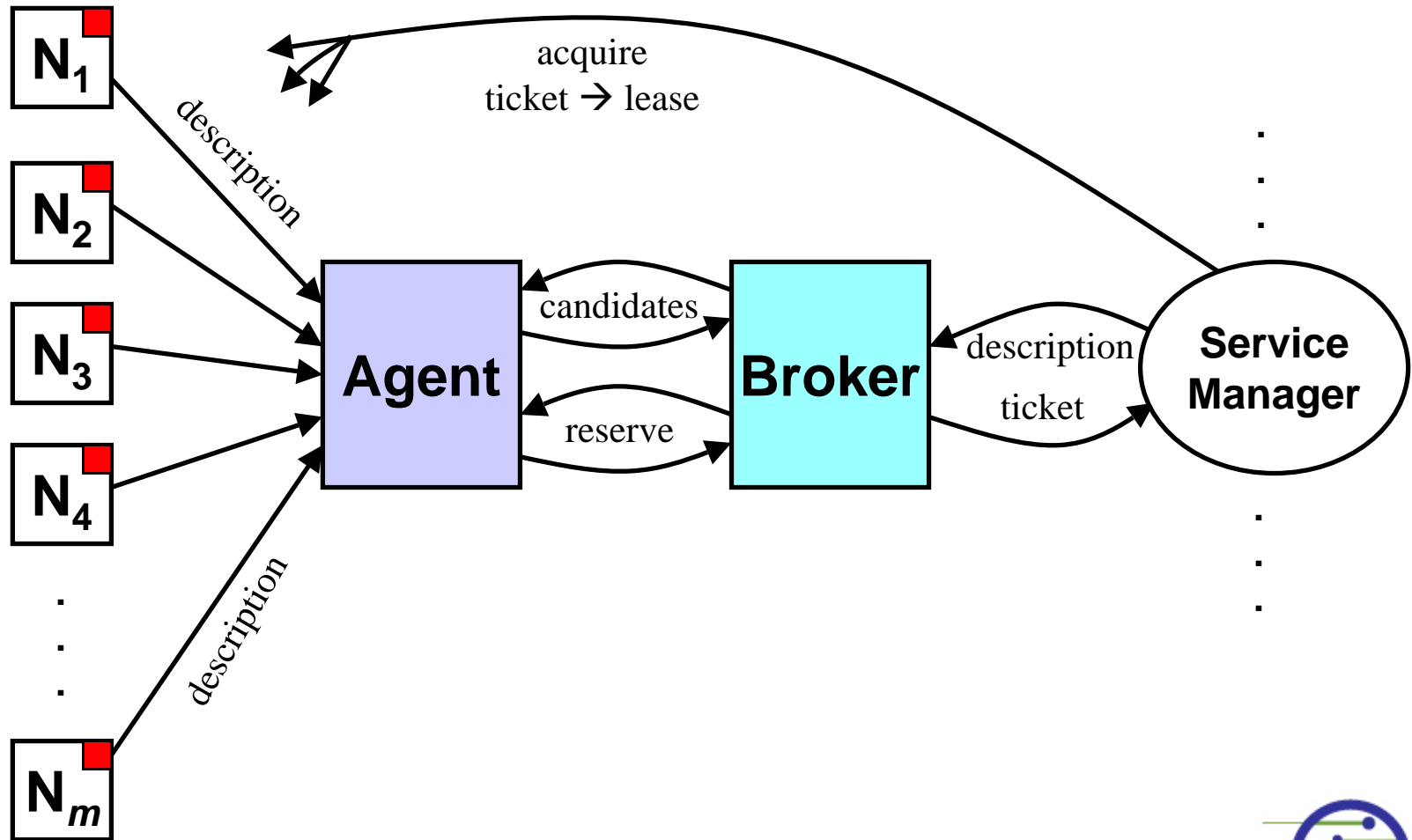delete_slice(name, credentials)

set_state(name, boot_state, credentials)

set_resources(name, resource_spec, credentials)

instantiate_slice(name, nodes[ ], credentials)

boot_state = {ssh_keys, environment_service}

resource_spec = {share, duration}

PLANETLAB

Dynamic Slice Creation
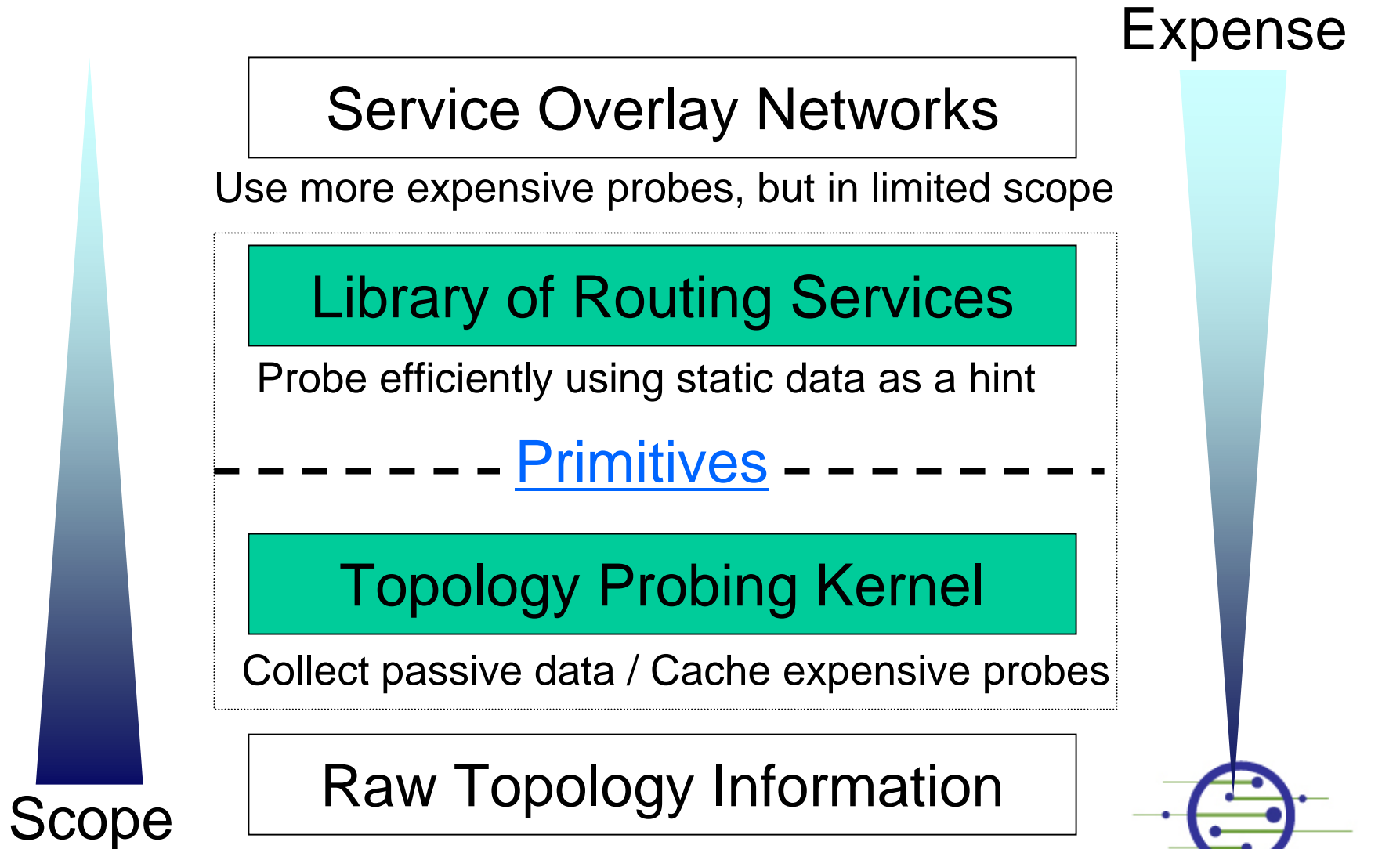
PLANETLAB

Routing Underlay Service

- Discovering efficient topology requires expensive/disturbing network probes

- Single overlay network
  - aggressive probing does not scale (RON < 50)

- Multiple overlay networks
  - Redundant probing to discover the same topological information
  - 1GB-per-day of ping traffic on PlanetLab
    - one ping-per-sec-per-node across 125 nodes

PLANETLAB

Routing Underlay

- Sits between overlays and the Internet
- Exposes topological information
  - already collected by the Internet (BGP tables)
  - caches active measurements
- Enables cost-effective network probes
  - primitives: interface to shared probes
  - layered architecture: hierarchical probes

PLANETLAB

Hierarchical Probes

Expense

Service Overlay Networks

Use more expensive probes, but in limited scope

Library of Routing Services

Probe efficiently using static data as a hint

Primitives

Topology Probing Kernel

Collect passive data / Cache expensive probes

Raw Topology Information

Scope

PLANETLAB

Roadmap

- Introduction
- PlanetLab Overview
- Virtualization
- Infrastructure Services
- **Growth Strategy**
- **Project Status**
- **Summary**

PLANETLAB

Growth Strategy

- Phase 0: Seed the testbed
  - 100 centrally managed machines
  - pure testbed (no expected client workload)
- Phase 1: Scale the testbed
  - grow to 1000 nodes with user-provided hardware
  - continuously running services (researchers as clients)
- Phase 2: Cultivate a user community
  - non-researchers as clients
  - PlanetLab spinoffs interpreted as success

PLANETLAB

Status

- Funding
  - Intel Seed Funding
  - NSF
  - PlanetLab Consortium
- Transition Phase (through mid-2004)
  - Moving "ops" from Intel to Princeton

PLANETLAB

PlanetLab Consortium

- Princeton, Berkeley, University of Washington
  - Initial Platinum Members: Intel, HP, Google

- Build out the PlanetLab infrastructure
  - operations and engineering support
  - equipment renewal
  - bandwidth at network crossroads

- Broaden and catalyze the community
  - academic and corporate researchers
  - lower the barrier to entry for research and teaching
  - drive the research agenda

Current Institutions

Academia Sinica, Taiwan

Boston University

Caltech

Carnegie Mellon University

Chinese Univ of Hong Kong

Columbia University

Cornell University

Datalogisk Institut Copenhagen

Duke University

Georgia Tech

Harvard University

HP Labs

Intel Research

Johns Hopkins

Lancaster University

Lawrence Berkeley Laboratory

MIT

Michigan State University

National Tsing Hua Univ.

New York University

Northwestern University

Princeton University

Purdue University

Rensselaer Polytechnic Inst.

Rice University

Rutgers University

Stanford University

Technische Universitat Berlin

The Hebrew Univ of Jerusalem

University College London

University of Arizona

University of Basel

University of Bologna

University of British Columbia

UC Berkeley

UCLA

UC San Diego

UC Santa Barbara

University of Cambridge

University of Canterbury

University of Chicago

University of Illinois

University of Kansas

University of Kentucky

University of Maryland

University of Massachusetts

University of Michigan

University of North Carolina

University of Pennsylvania

University of Rochester

USC / ISI

University of Technology Sydney

University of Tennessee

University of Texas

University of Toronto

University of Utah

University of Virginia

University of Washington

University of Wisconsin

Uppsala University, Sweden

Washington University in St Louis

Wayne State University

PLANETLAB

Summary

- PlanetLab: an open, global network test-bed for pioneering novel planetary-scale services.

- A model for introducing innovations into the Internet through the use of overlay networks.

- A collaborative effort involving hundreds of academic and corporate researchers from around the world.

More Information

www.planet-lab.org

PLANETLAB