The Case for Resilient Overlay Networks

David G. Andersen

Hari Balakrishnan, M. Frans Kaashoek, Robert Morris

MIT Laboratory for Computer Science

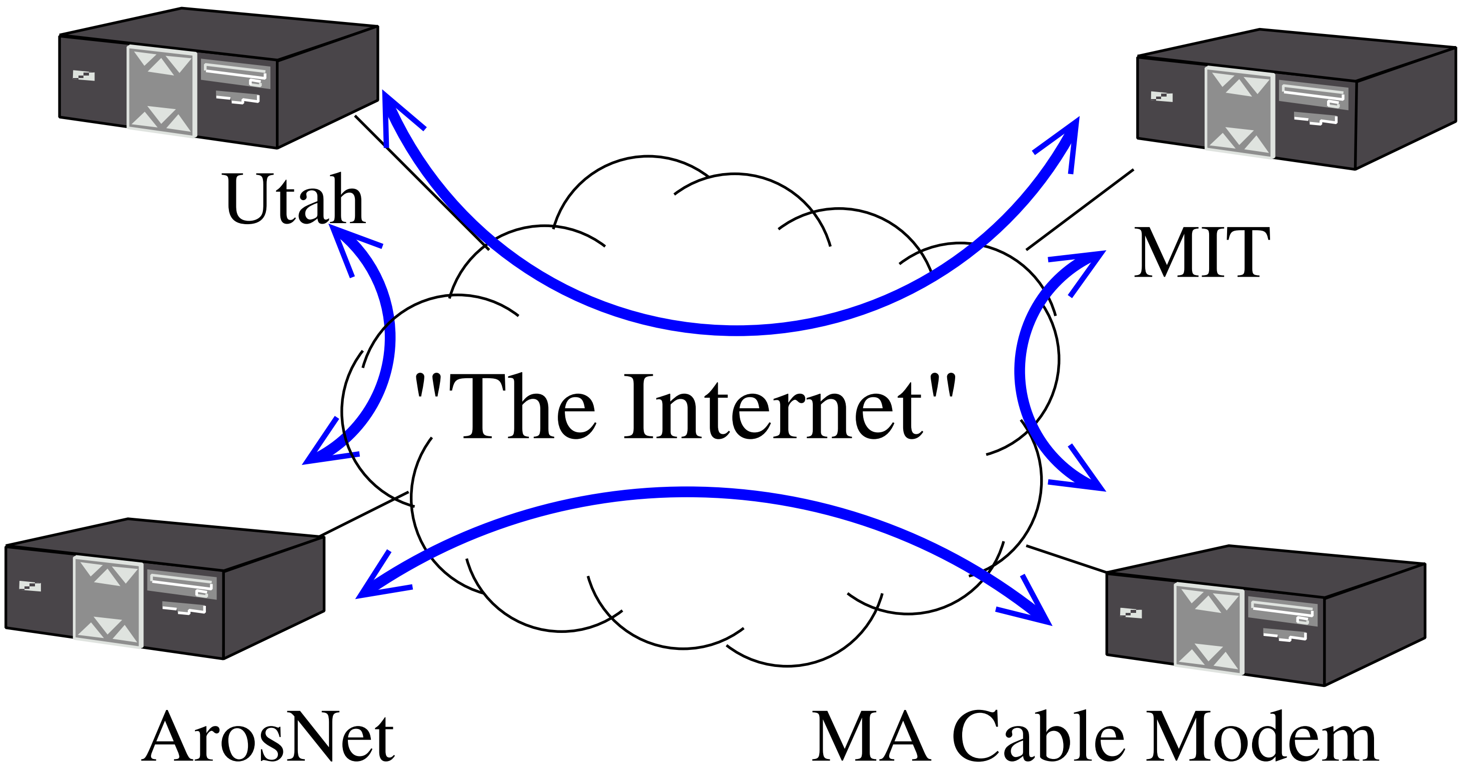
May 2001

`http://nms.lcs.mit.edu/ron/`

Idea

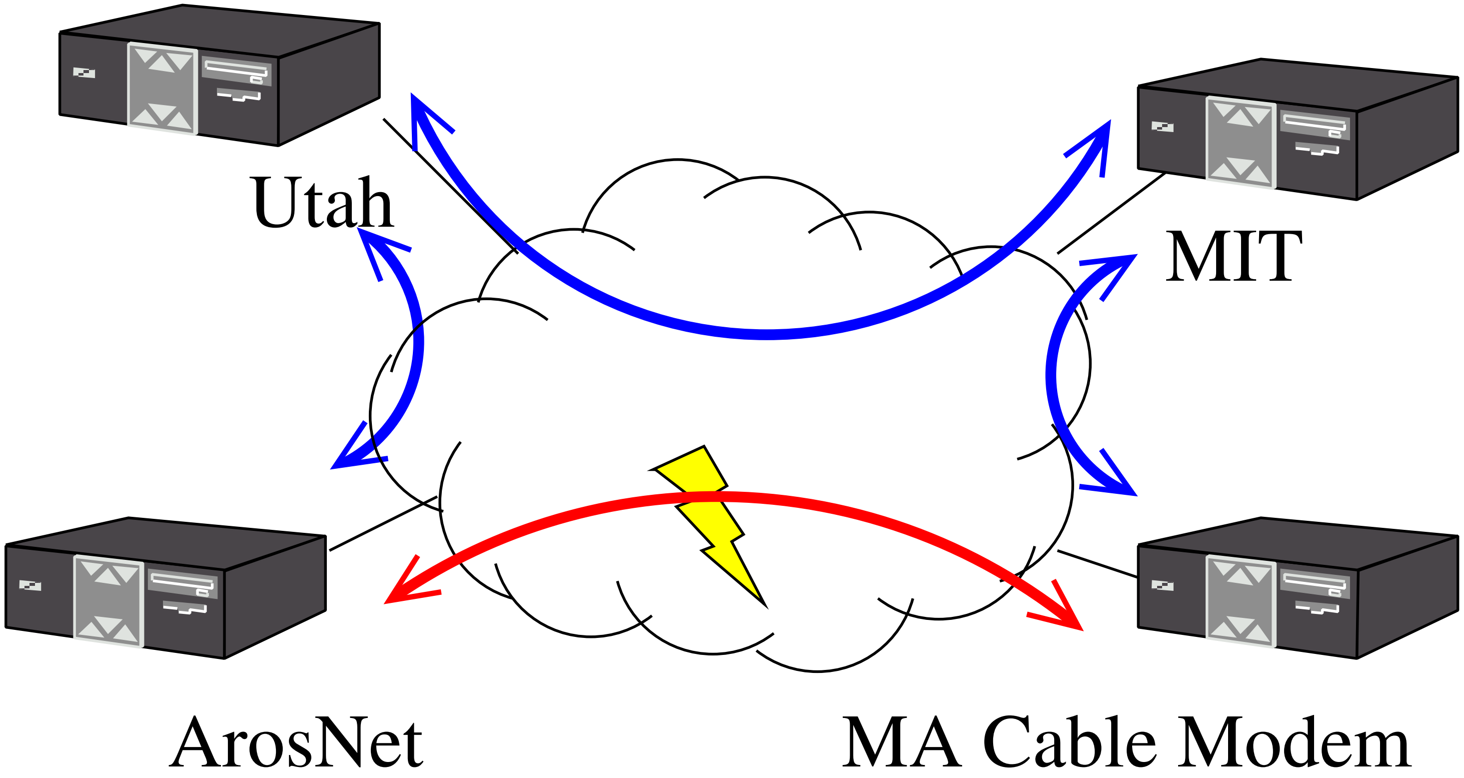
- Scalability may get in the way of deploying services and protocols that may not scale
- So do cool things in small overlays
  - ➔ More aggressive
  - ➔ Things that're less efficient

Routing around Internet Failures



People expect all-to-all communication...

Routing around Internet Failures

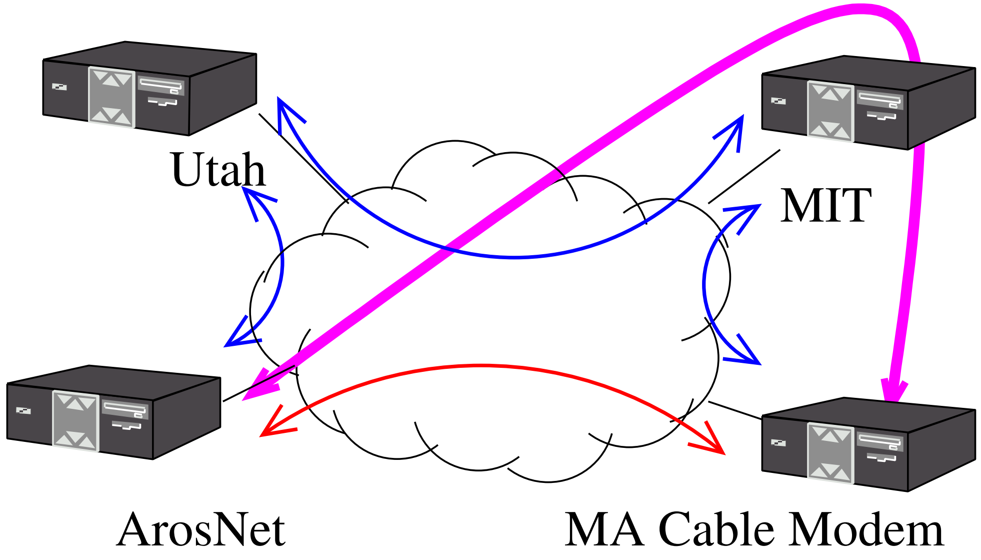


Which the Internet can't always provide.

Internet Failures

- Physical link failures (backhoes)
- Excess Traffic (14-year-olds...)
- Router misconfiguration
- The list goes on...

Routing around Internet Failures



But we think cooperating hosts can do better...

Multiparty videoconferences

Overlay Internet Service, Companies with VPNs, etc.

➜Do we need this?

The Internet Recovers Slowly

[Labovitz 00]:
"Internet ... routing convergence is an order of magnitude slower than previously thought."

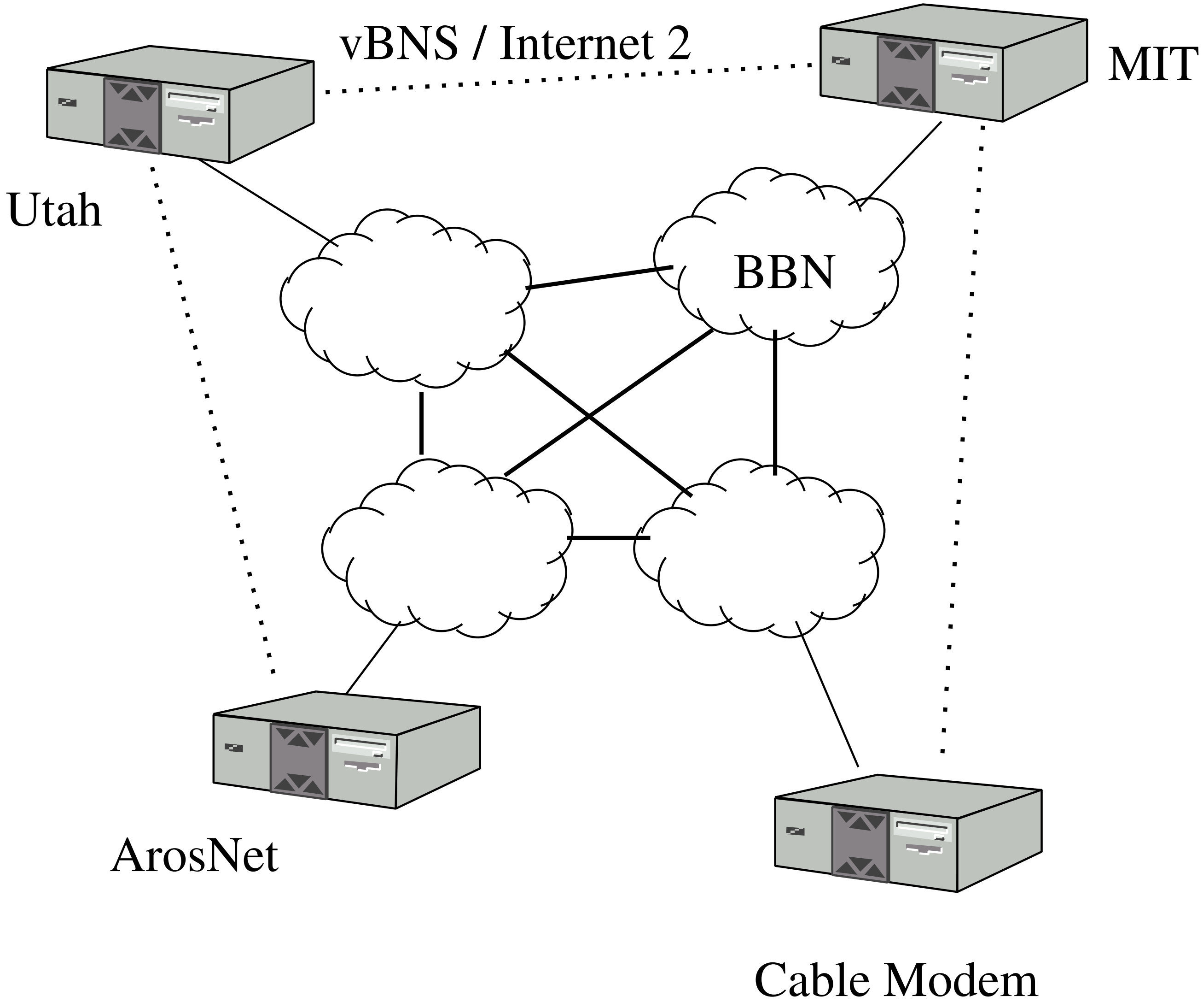
- 3 minute average recovery time
  15 minute max for *simple* failures

- Our tests: Indirect routing had 5x-10x fewer outages [Sneak preview]

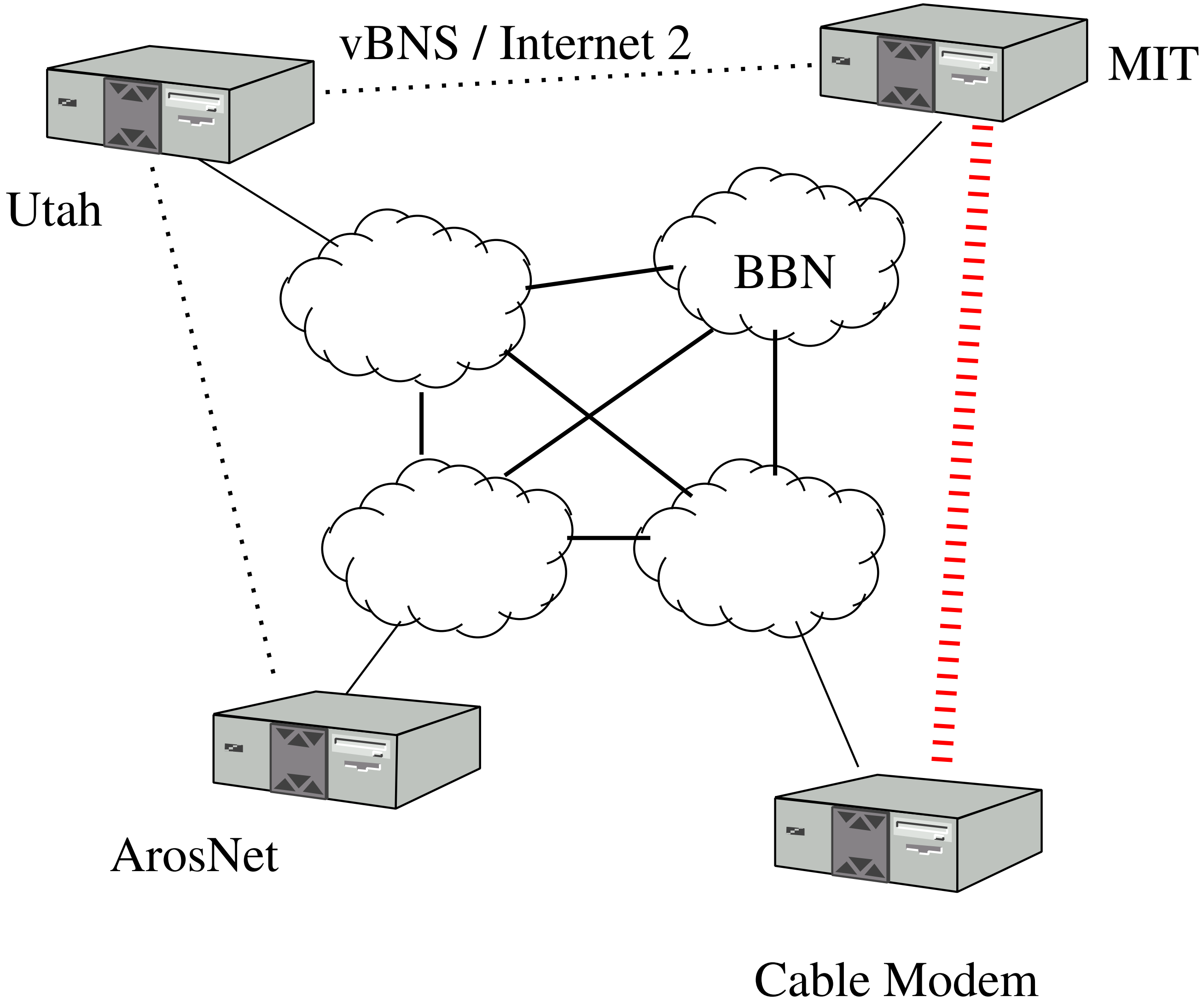
Internet Trade-offs

- Scalability and heterogeneity:
  ➔Slow Recovery
  (Is this a fundamental trade-off?)
- RON takes a different approach:
  Fast recovery for small groups in an overlay
- Exploit redundancy in the Internet

A More Realistic Picture



Hidden Links



Policy and AUPs

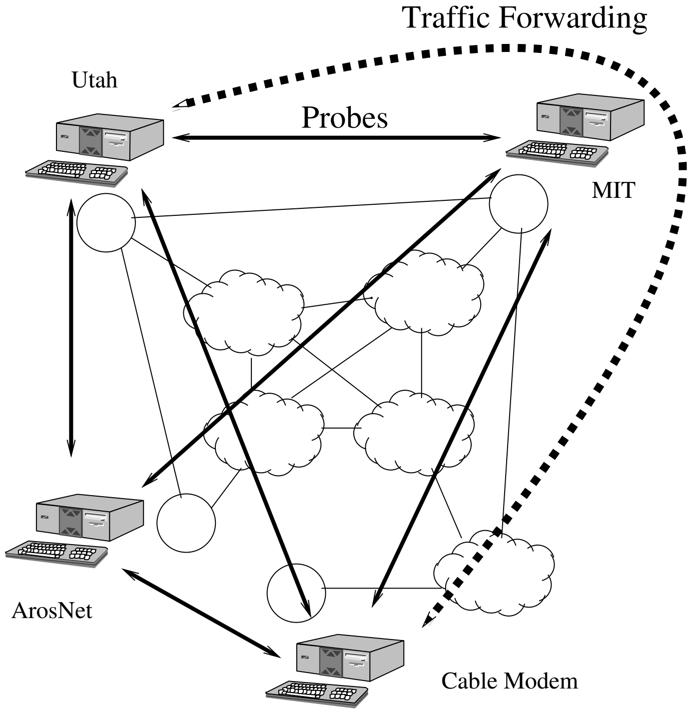
- WAN routing policy expression is a sledgehammer
- But we need policy control (Internet2, etc)

✘ RON could violate AUPs

✔ But RON can provide flexible policies

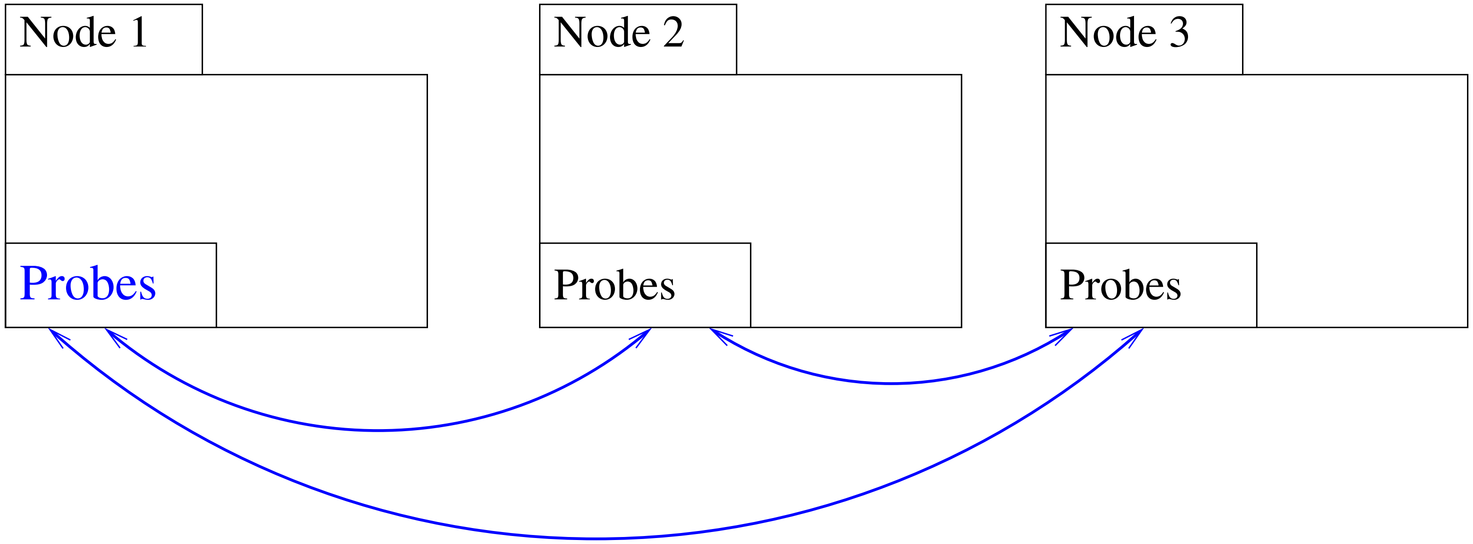
  – More complex routing decisions
  – Multiple routing tables
  – Deeper packet inspection

The Approach



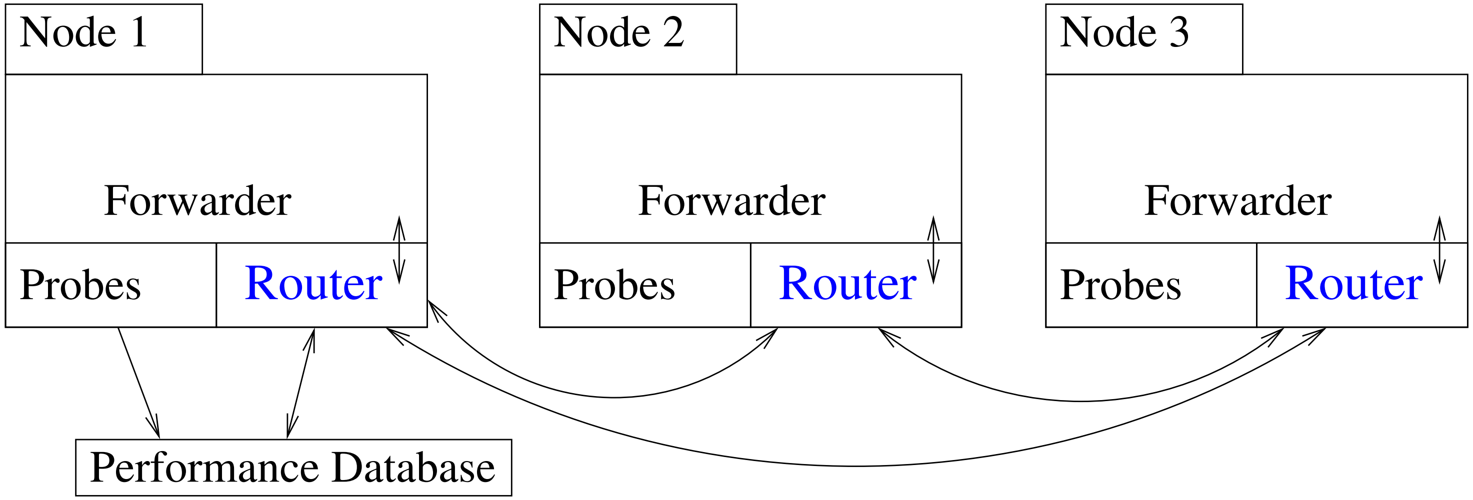
- Measure *all* inter-node paths
- Exchange routing information
- Route along best path

Architecture



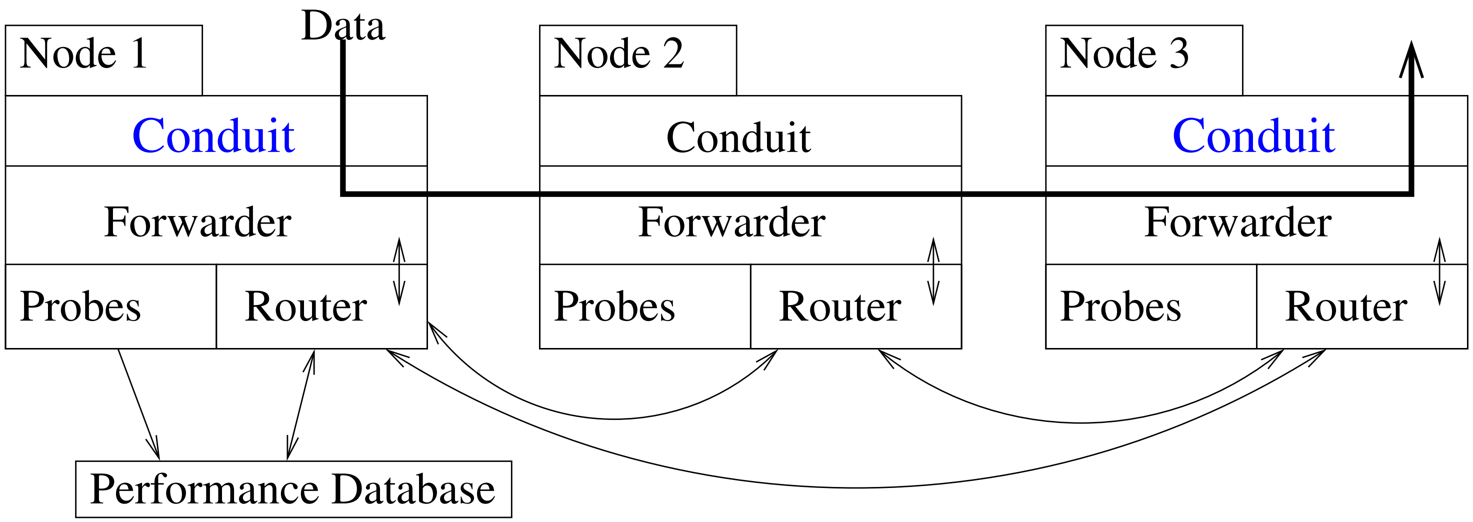
- Probe between nodes to establish best route
  - Active, application probing of N^2 paths
  - Passive measurements

Architecture



- Probe between nodes to establish best route
- Link-state routing protocol between nodes

Architecture



- Probe between nodes to establish best route
- Link-state routing protocol between nodes
- Data handled by application-specific conduit

  Forwarded in UDP

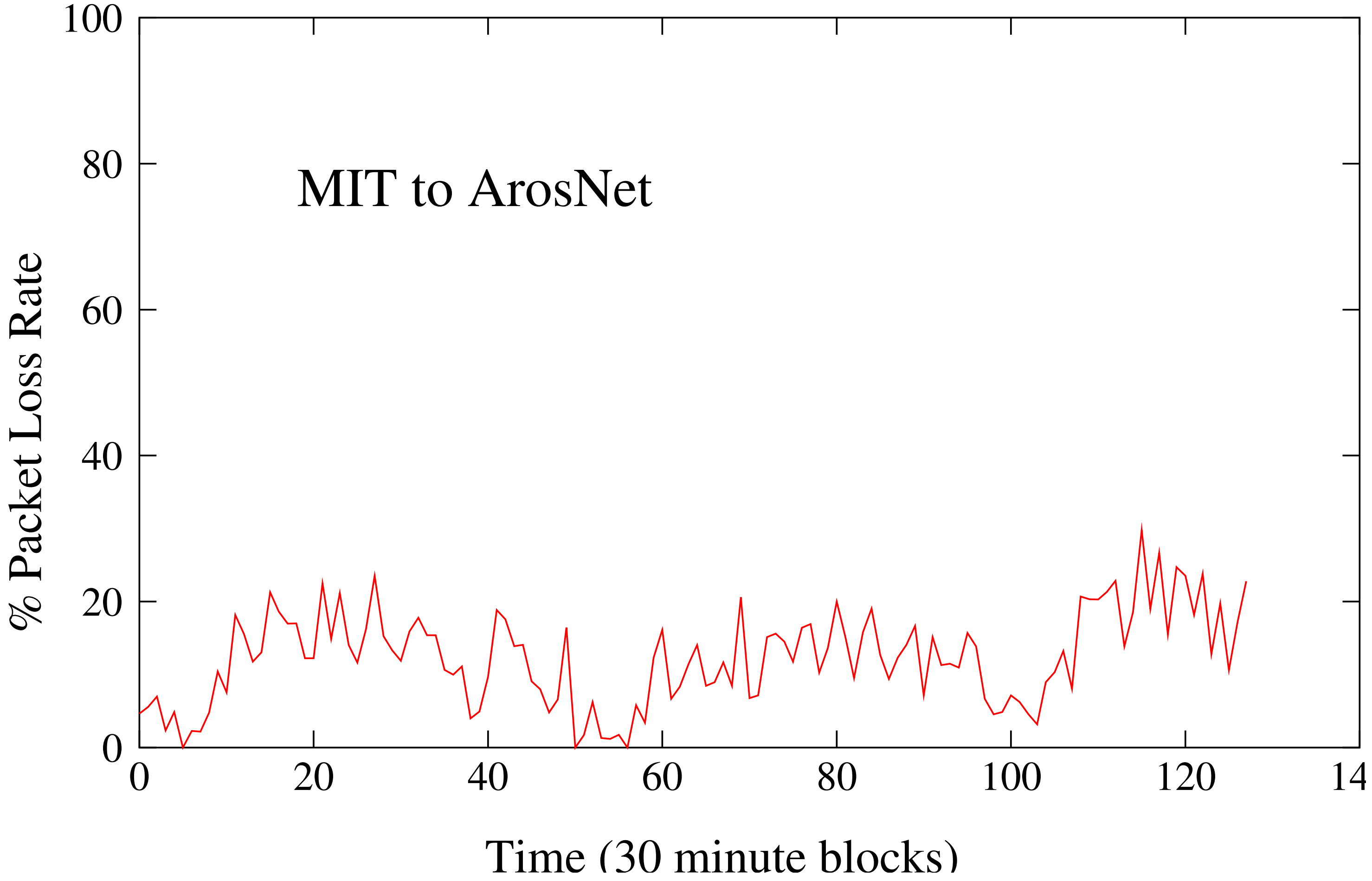
Conduits: Gateways into the RON

- IP off the wire conduit
  (Used for evaluation)
- Emulates `sendto` and `recvfrom`
- The application itself
- Interface: `send`, `register`, `callback`

Preliminary Investigation

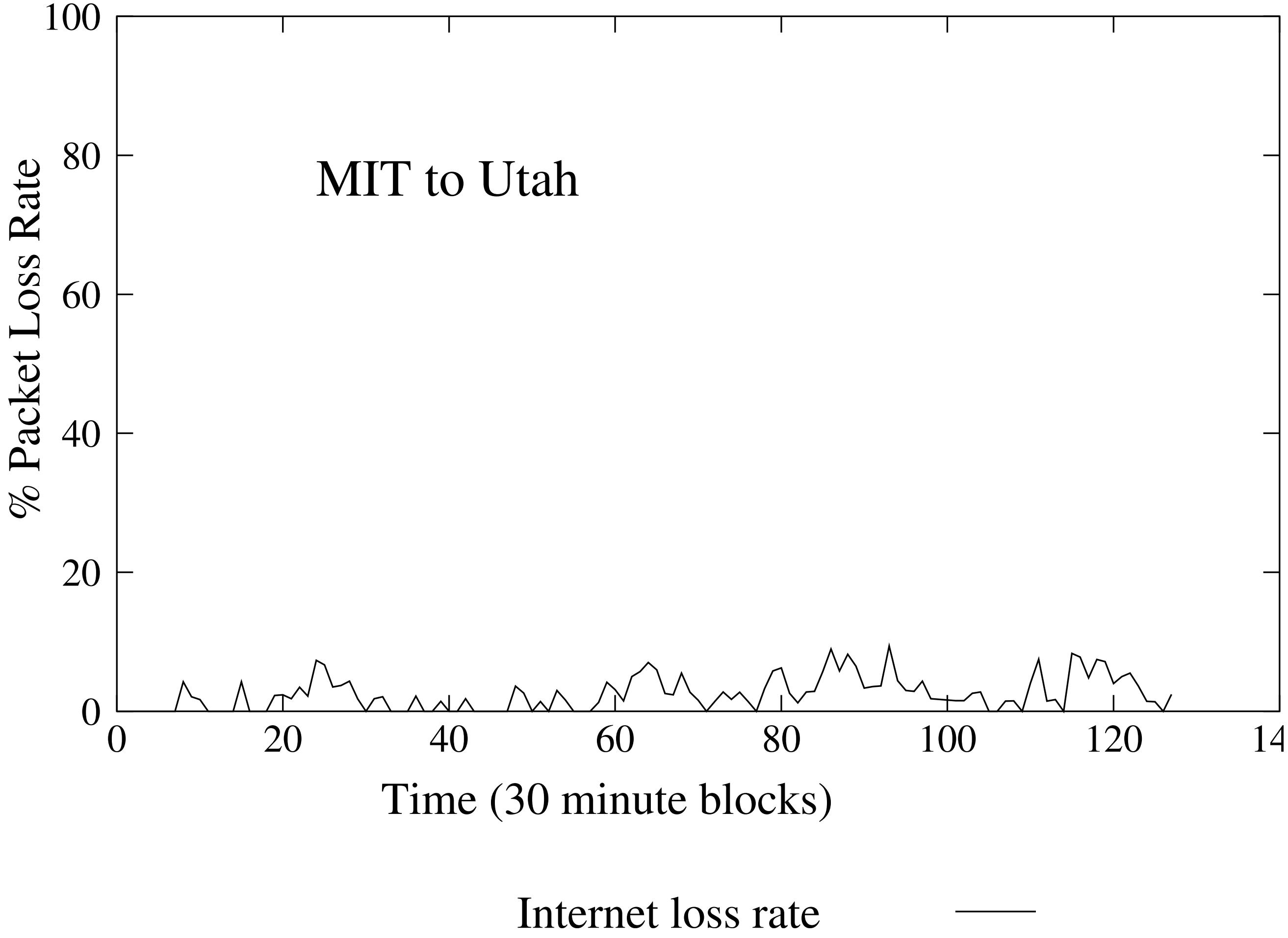
- Tested between 4 hosts
- 70 hours of ping-style measurements
- Looks promising, but we suspected it would.

From MIT to ArosNet on the Internet



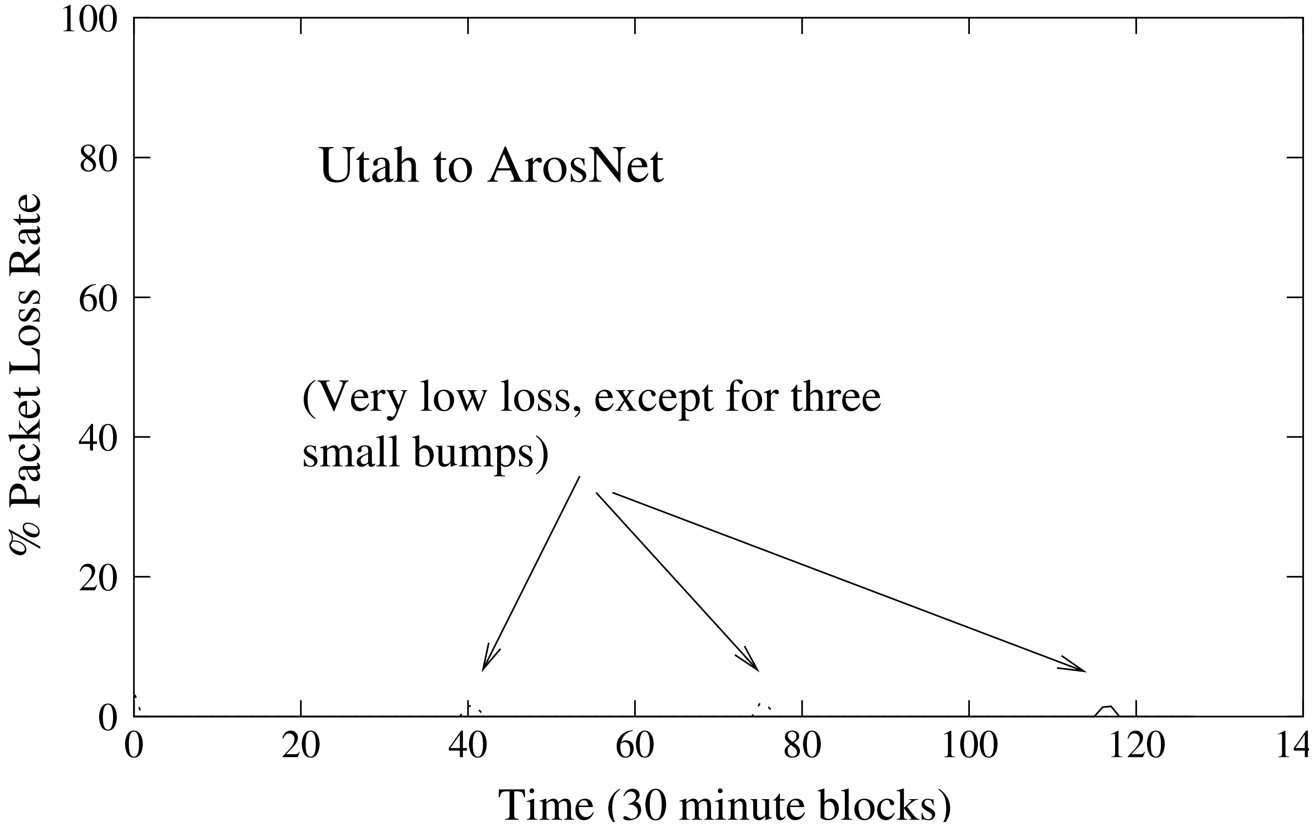
The direct Internet path is quite bad.

From MIT to Utah on the Internet



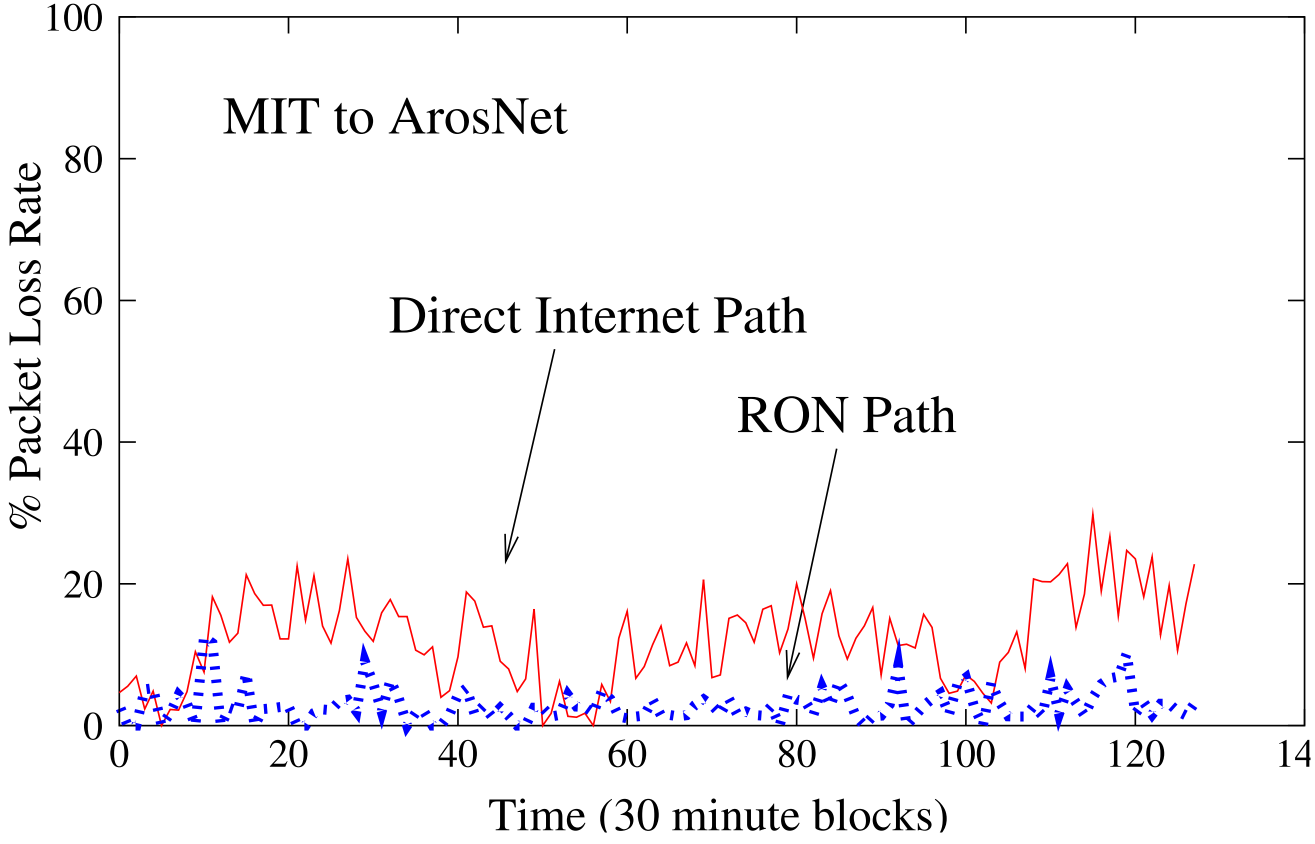
But the path from MIT to Utah looks good...

From Utah to ArosNet on the Internet



As does the path from Utah to ArosNet...

From MIT to ArosNet with RON



Putting them together...

Other results

- Big latency reduction between MIT - ArosNet
- Big latency reduction between Cable Modem - Utah
- Real results are hiding in a thesis

Some Research Questions

- Is this a stupid idea?
- How many intermediate hops?
- How do we best choose routes?
- How frequently do we probe?
- What routing policies can we express?
- **How do RONs interact?**

`http://nms.lcs.mit.edu/ron/`

Status

http://nms.lcs.mit.edu/ron/