PATHLET ROUTING

P. Brighten Godfrey
pbg@illinois.edu
Igor Ganichev, Scott Shenker, and Ion Stoica
{igor,shenker,istoica}@cs.berkeley.edu

SIGCOMM 2009

Design for variation

"*Design for variation in outcome, so that the outcome can be different in different places, and the tussle takes place within the design, not by distorting or violating it.*"

— Clark, Wroclawski, Sollins & Braden, 2002
"Tussle in Cyberspace"

High level goals

- Goal: flexibility in network services
  - "Route to this destination", route along a specified path, VPNs, quality of service, ...

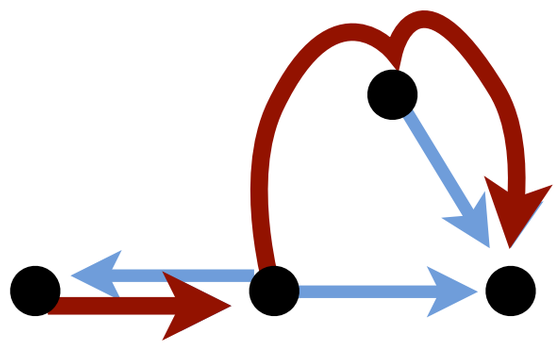
- Goal: user choice
  - Reliability, path quality, throughput, promote competition, ...

Pathlet routing's solution

- Goal: flexibility in network services
  - Represent network as a virtual topology

vnode virtual node

pathlet fragment of a path:
a sequence of vnodes



- Goal: User choice
  - Source routing within virtual topology

Outline

- The protocol
- Uses
- Experimental results
- Comparing routing protocols

Pathlet routing

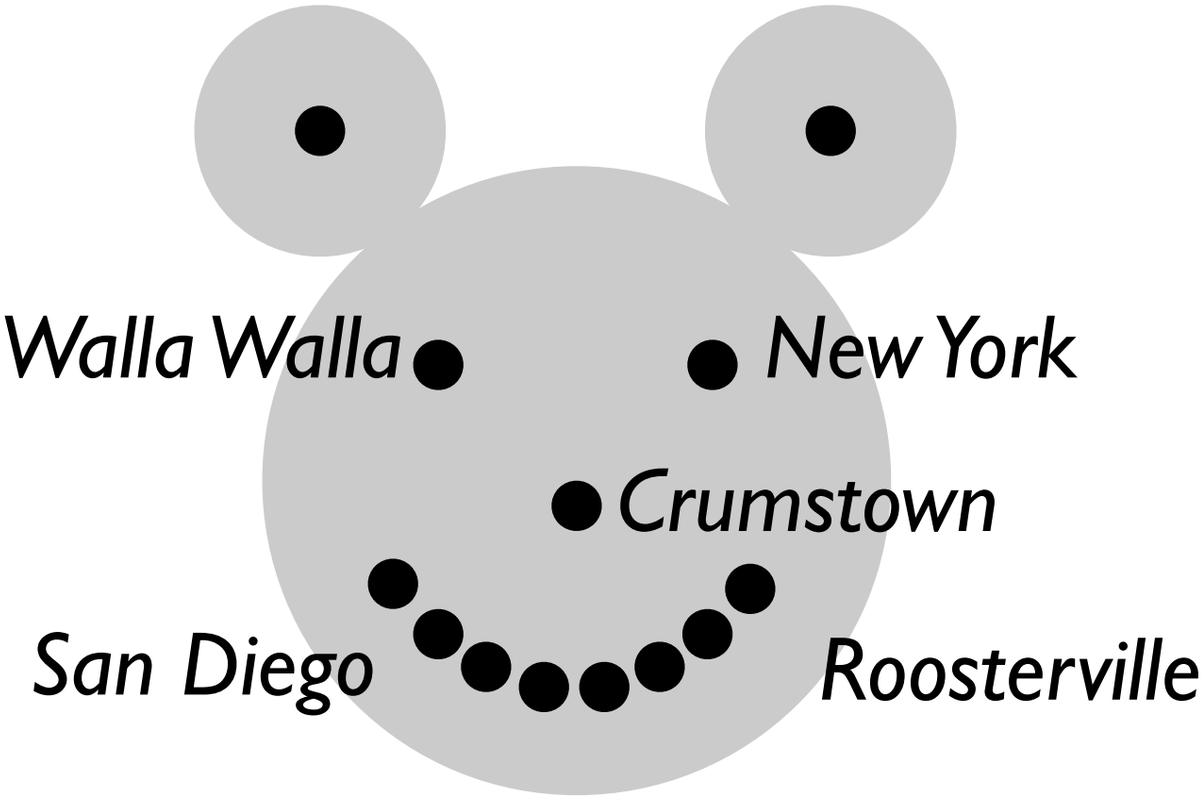
vnode virtual node

pathlet fragment of a path:
a sequence of vnodes

Source routing over pathlets.

vnodes

vnode: virtual node within an AS

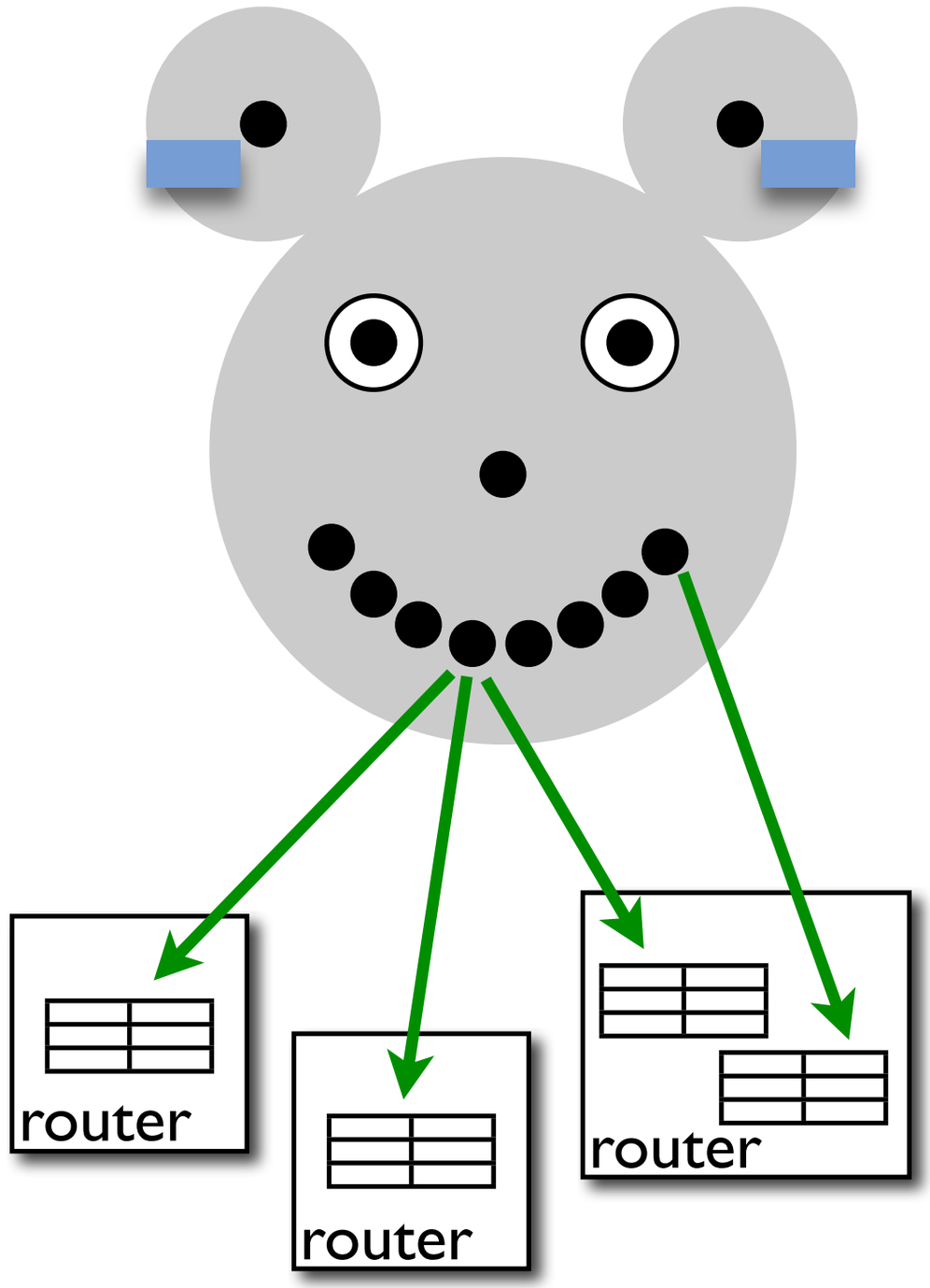


vnodes

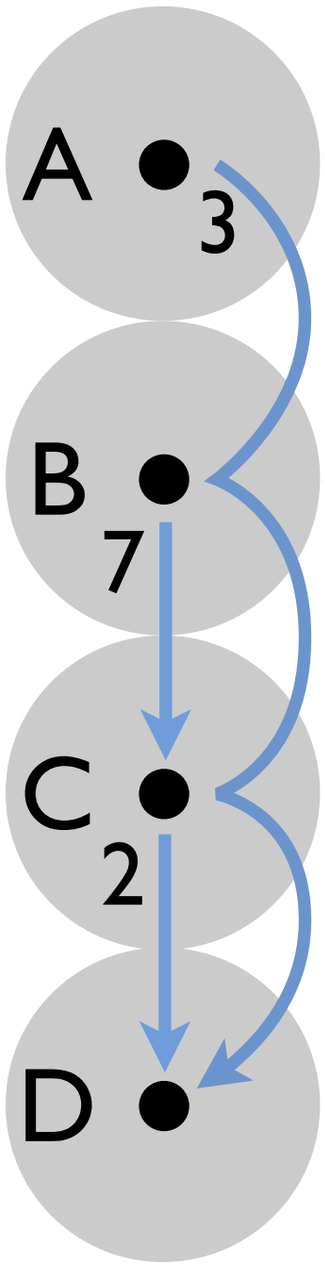
vnode: virtual node within an AS

designated ingress vnode for each neighbor

Internally: a forwarding table at one or more routers



Pathlets



Packet route field

Forwarding table

...	...
3	push 7,2; fwd to B

...	...
7	fwd to C

...	...
2	fwd to D

delivered!

Dissemination

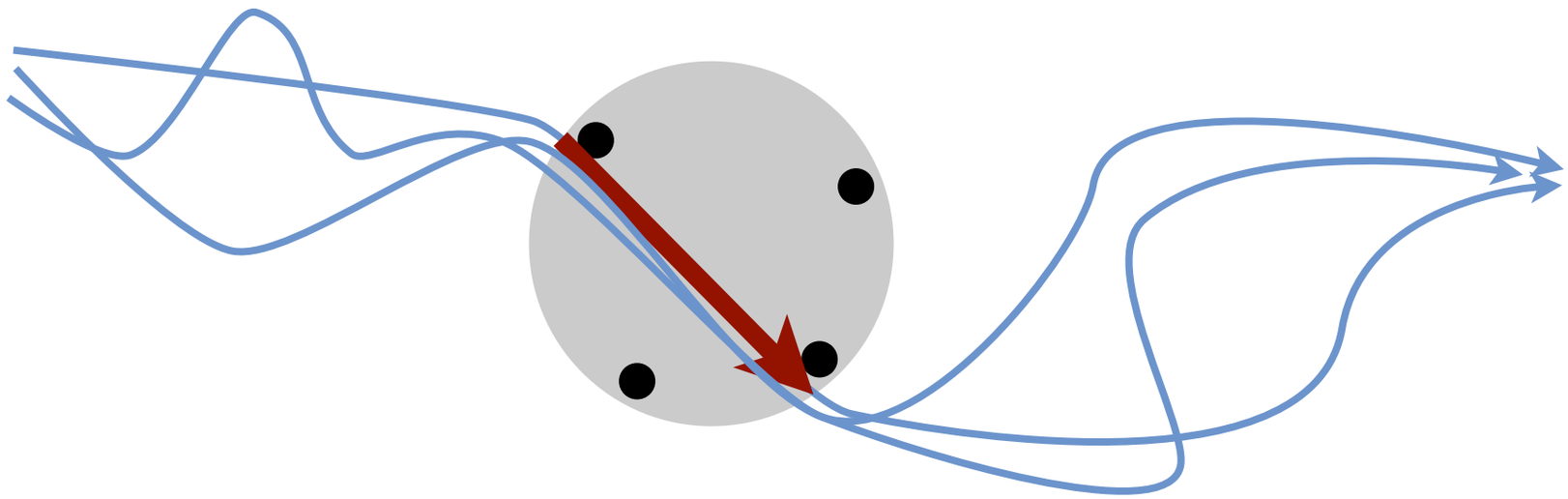
- Global gossip fine, except for scalability
- So, let routers choose not to disseminate some pathlets
- Leads to (ironic) use of path vector — only for pathlet dissemination, not route selection

Outline

- The protocol
- Uses
- Experimental results
- Comparing routing protocols

Local transit policies

Each ingress → egress pair
is either allowed or disallowed.

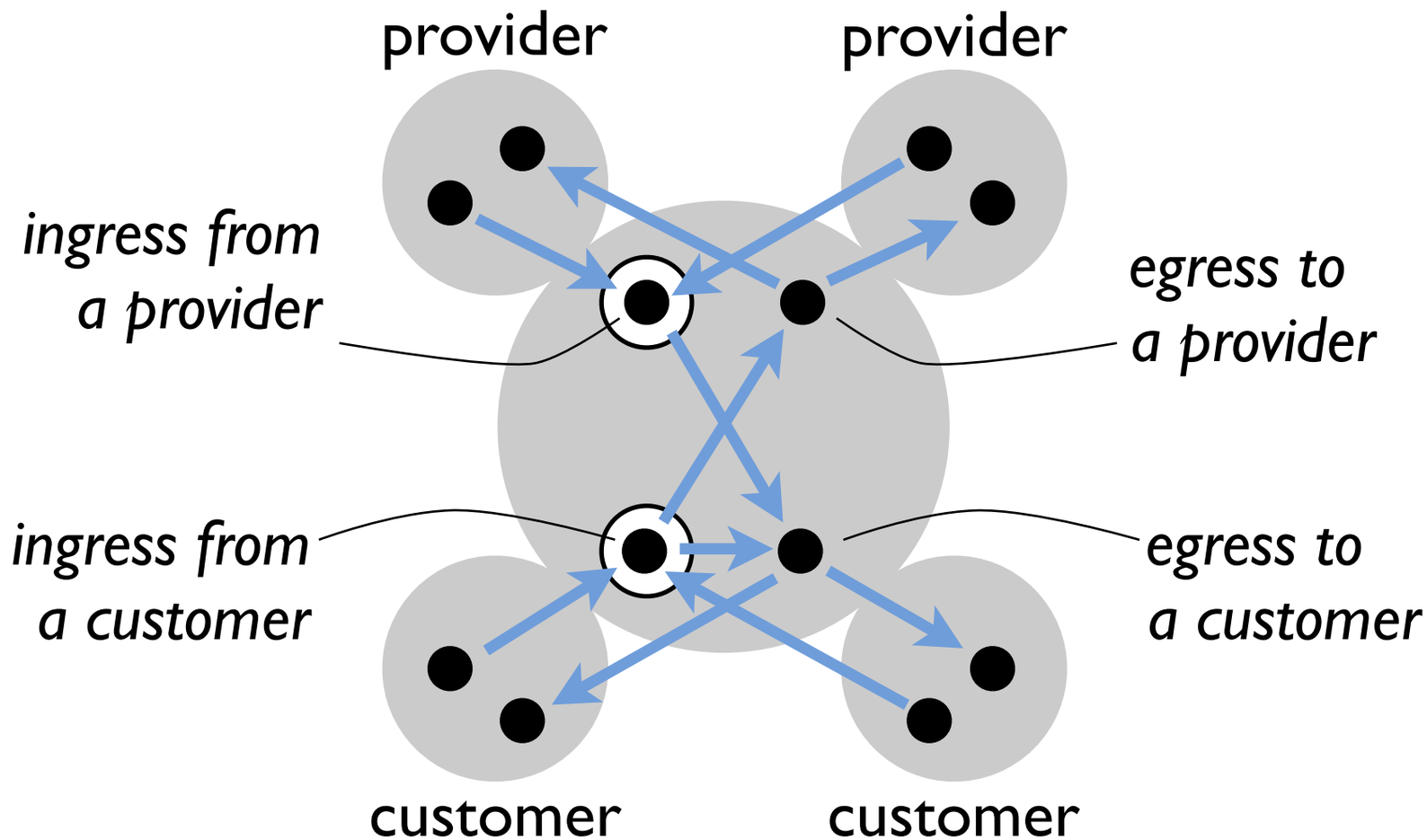


Subject to this, any path allowed!

Represented with few pathlets: small FIB

“All valley-free” is local

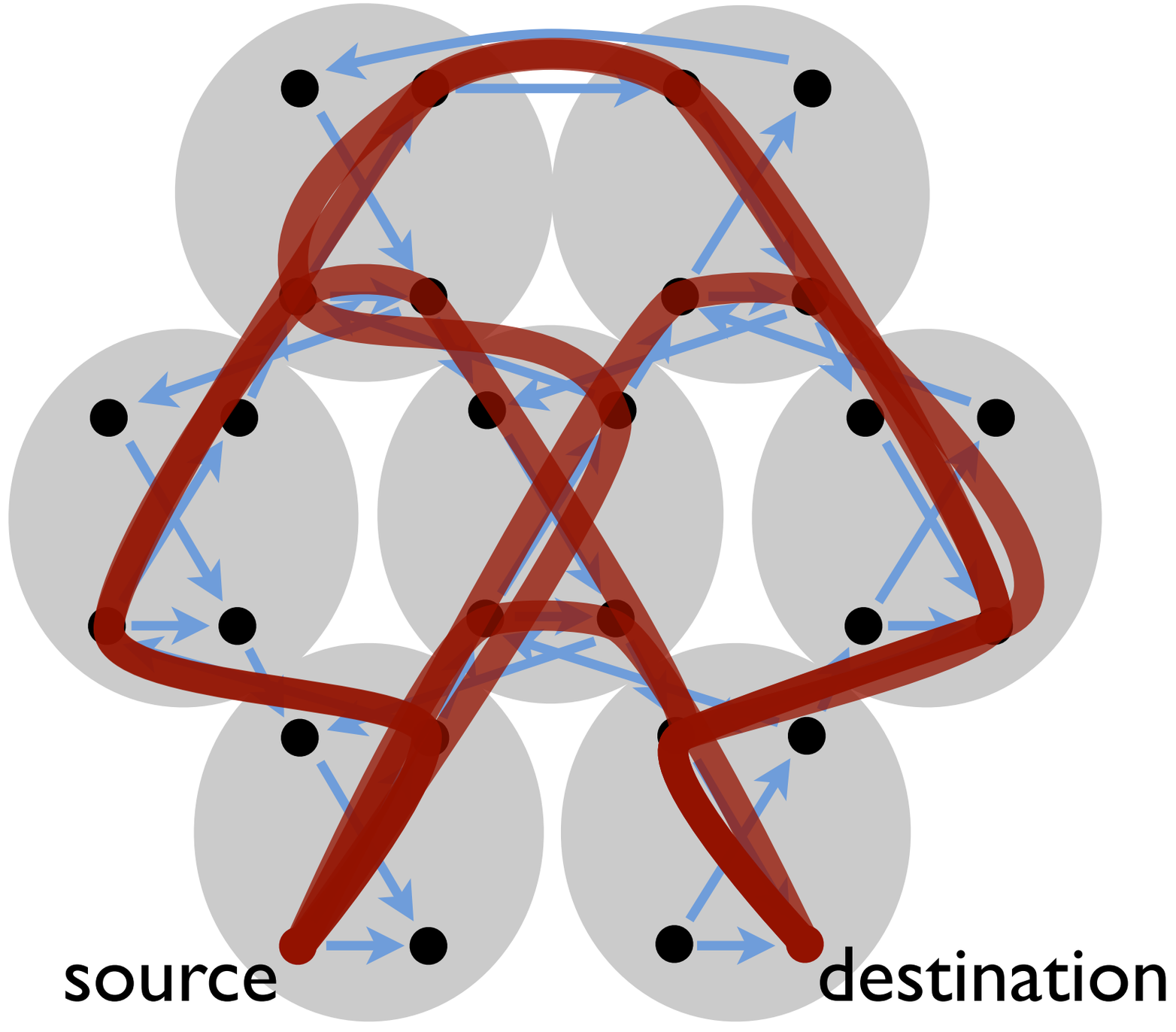
“customers can route to anyone; anyone can route to customers”



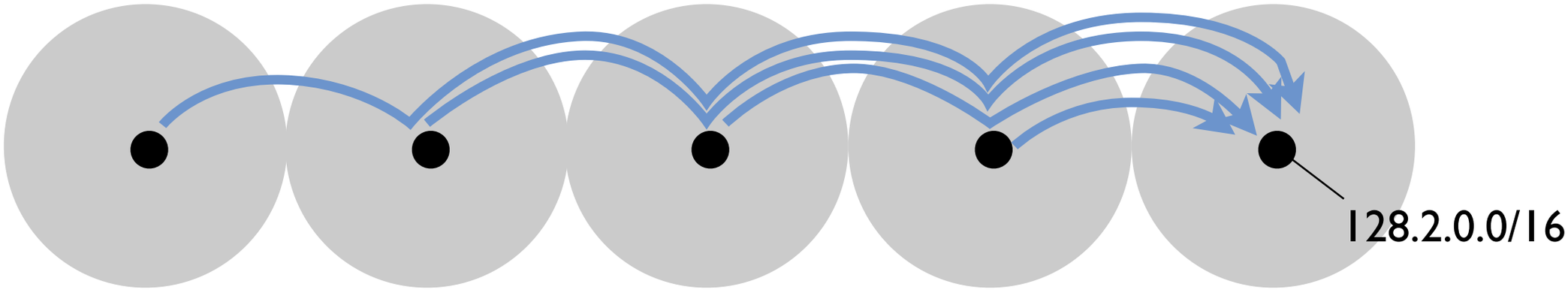
Forwarding table size: 3 + #neighbors

Choice for senders

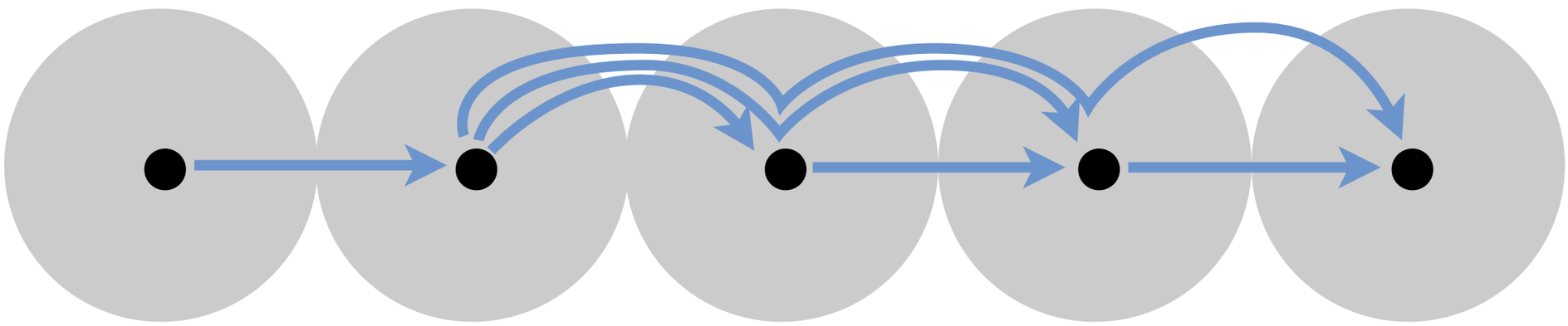
Local transit policies provide some policy control for networks, while enabling a large number of paths for senders.



Emulating BGP



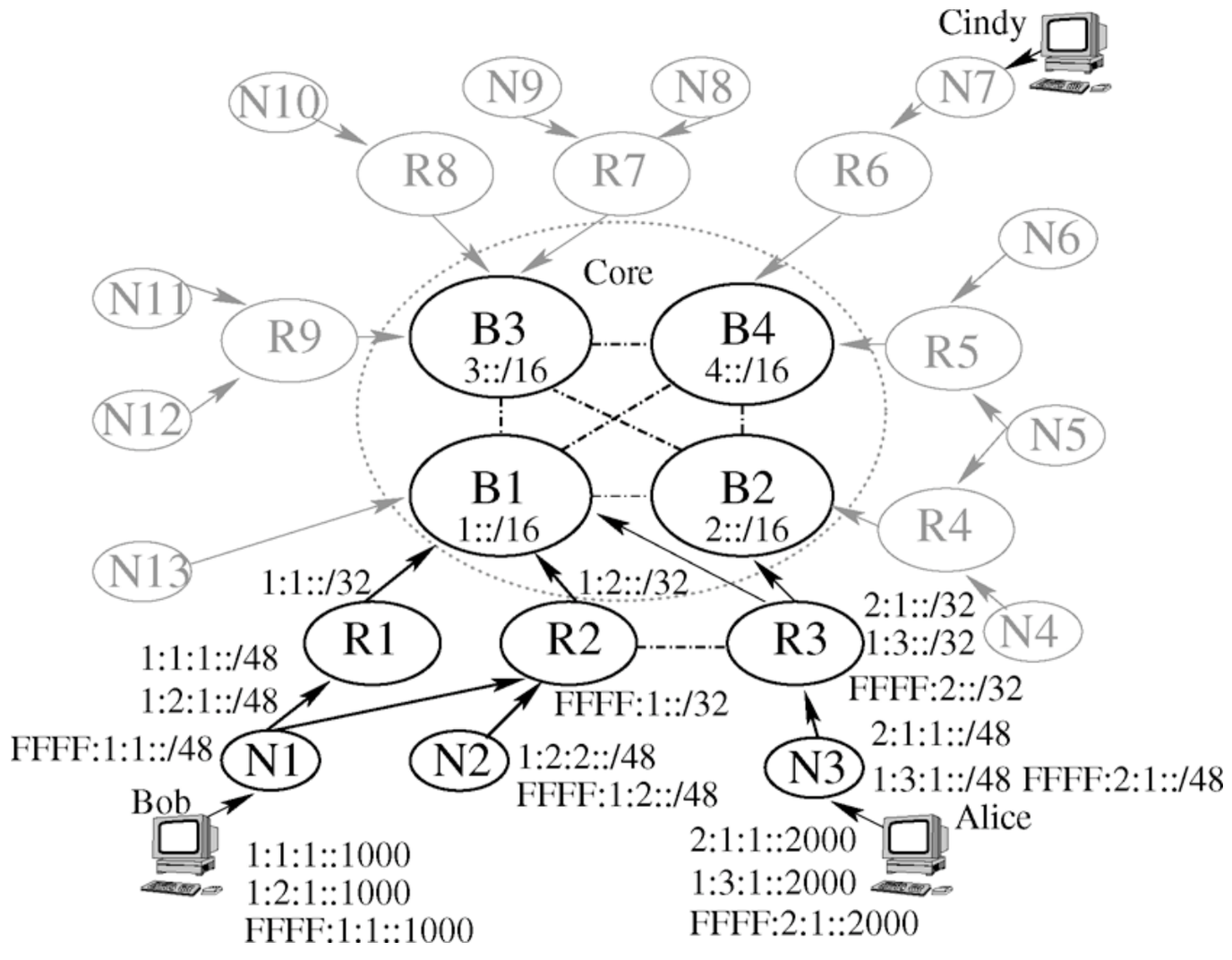
Mixed policies



local BGP-like local local local

Emulating NIRA

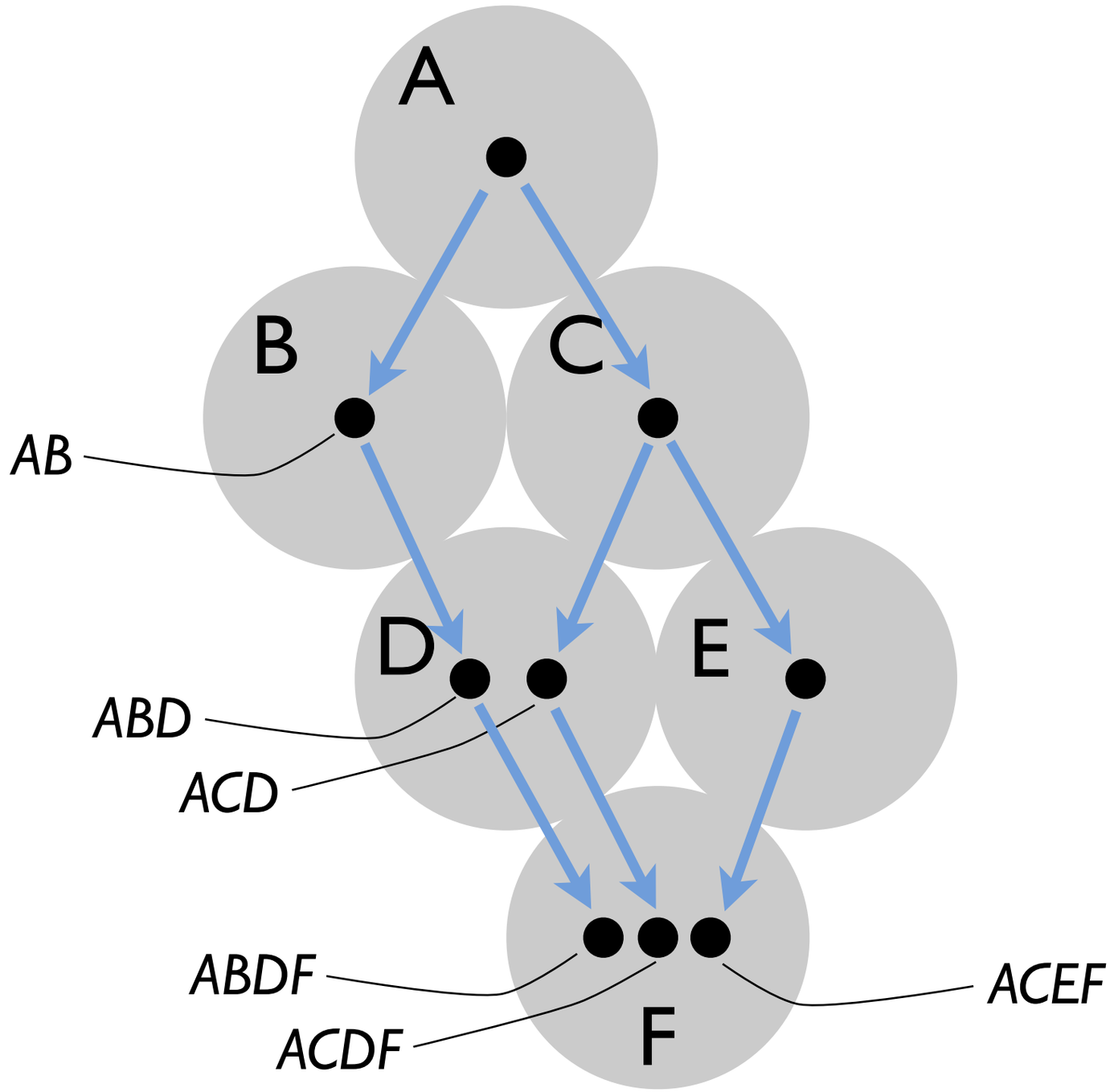
Tricky bit: policy can depend on previous hops!



Emulating NIRA

NIRA: carry state about previous hops in destination IP address.

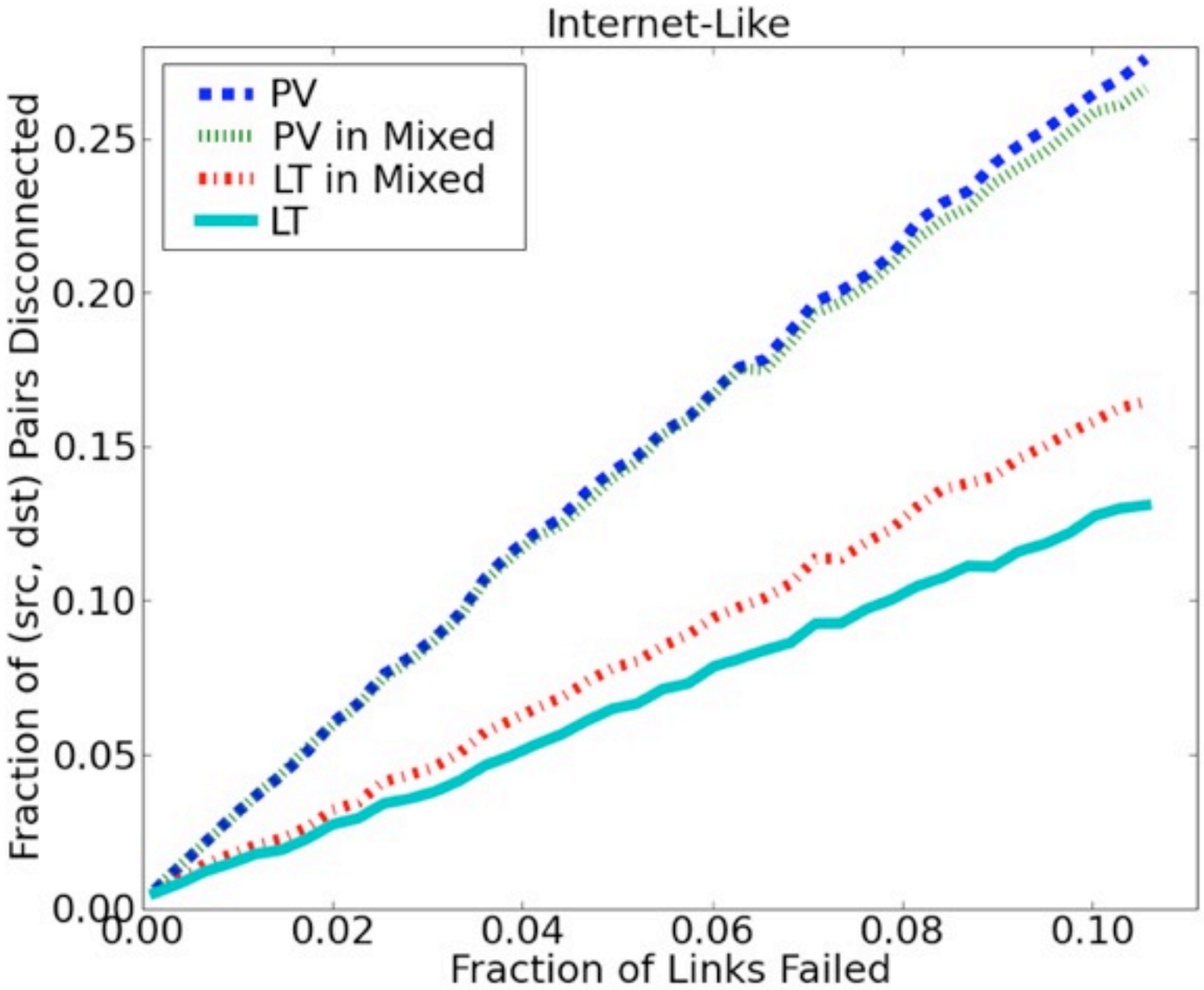
Pathlets: carry state about previous hops in vnode.



Outline

- The protocol
- Uses
- Experimental results
- Comparing routing protocols

Improved connectivity



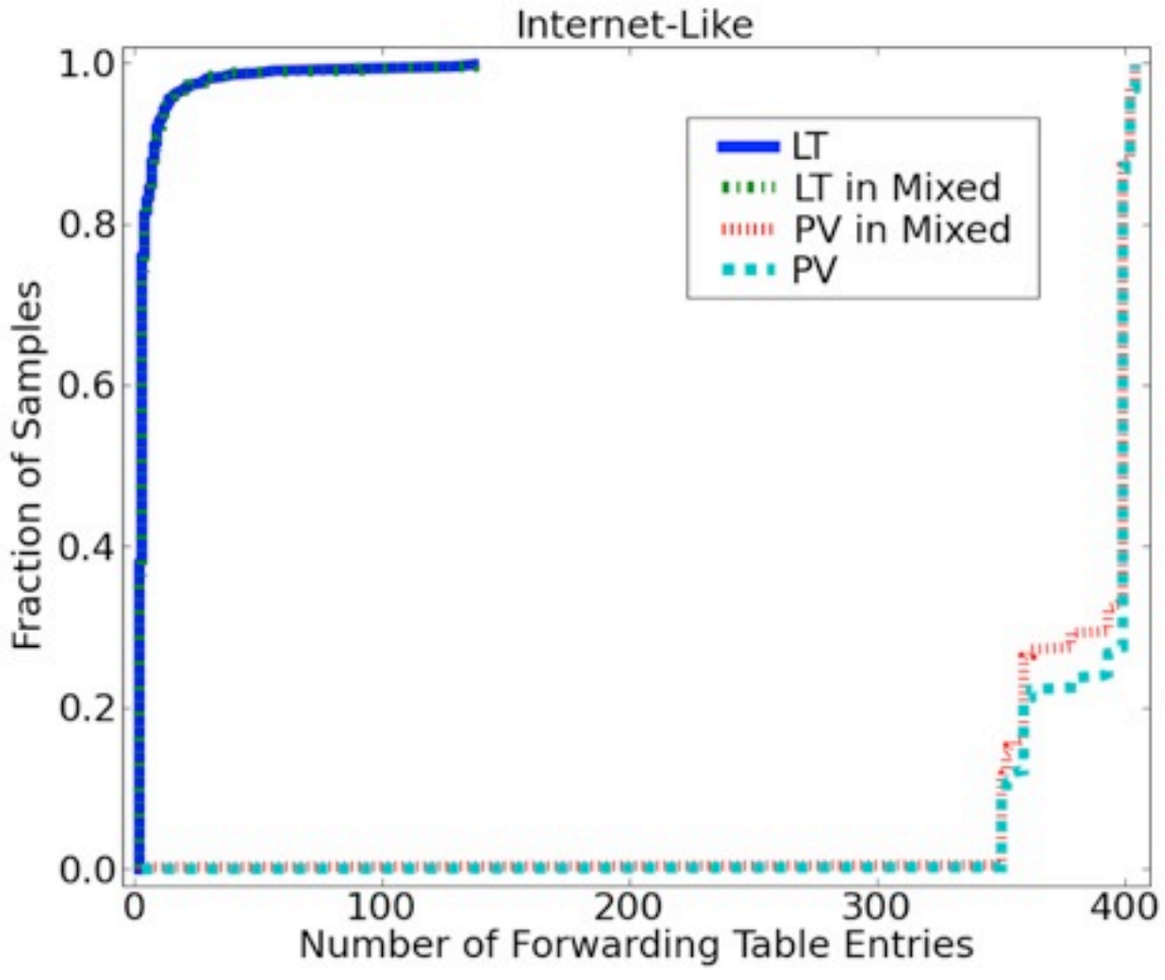
BGP-style

Mixed

LT policies

Tiny forwarding tables

Forwarding table size CDF

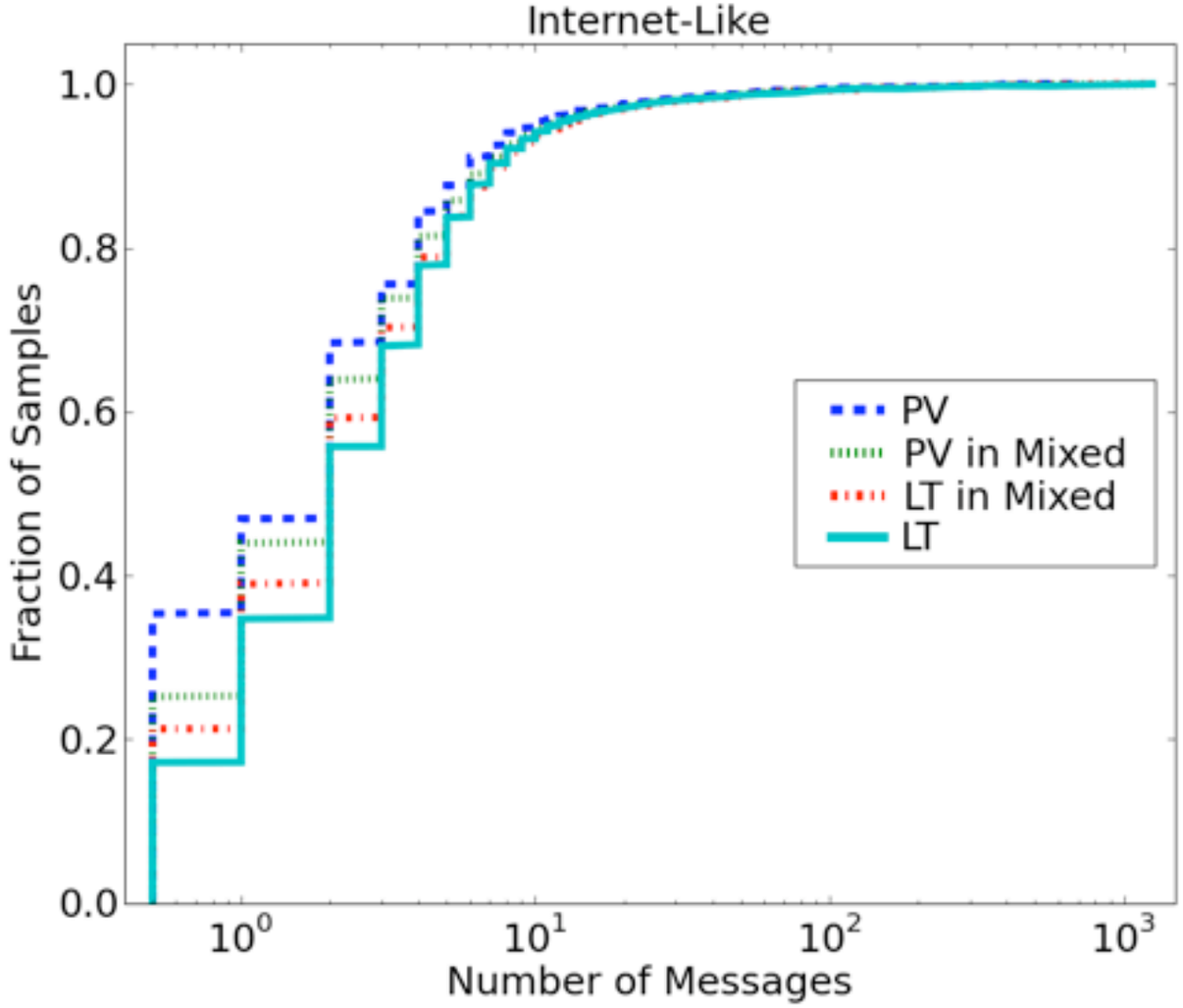


current Internet (CAIDA/APNIC):

BGP — 132,158+ entries: one per IP prefix

pathlet routing, valley-free LT policies — 2,264 entries, max 8.48 entries, mean

Control overhead



2.23x more messages,
1.61x more memory
in LT than PV

This can likely be
improved.

Questions

- Are either of these protocols viable?
- Would ASes actually use "local" policies (permitting many routes) or would they stick with BGP-style?
- Are there security vulnerabilities in NIRA or PR that are not in the current Internet?