PATHLET ROUTING

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

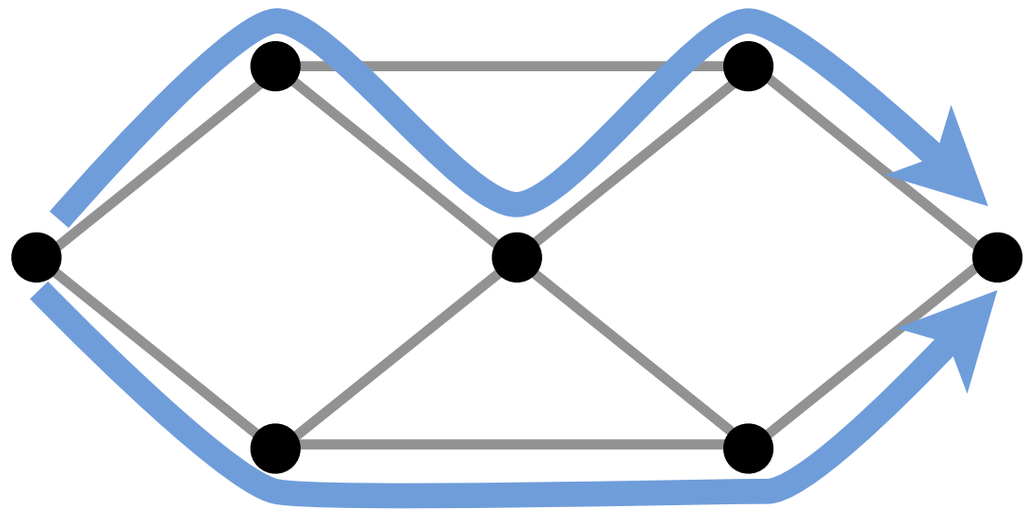
SIGCOMM 2009

Internet routing challenges

Multipath

reliability

path quality



Scalability

Policy

Internet routing challenges

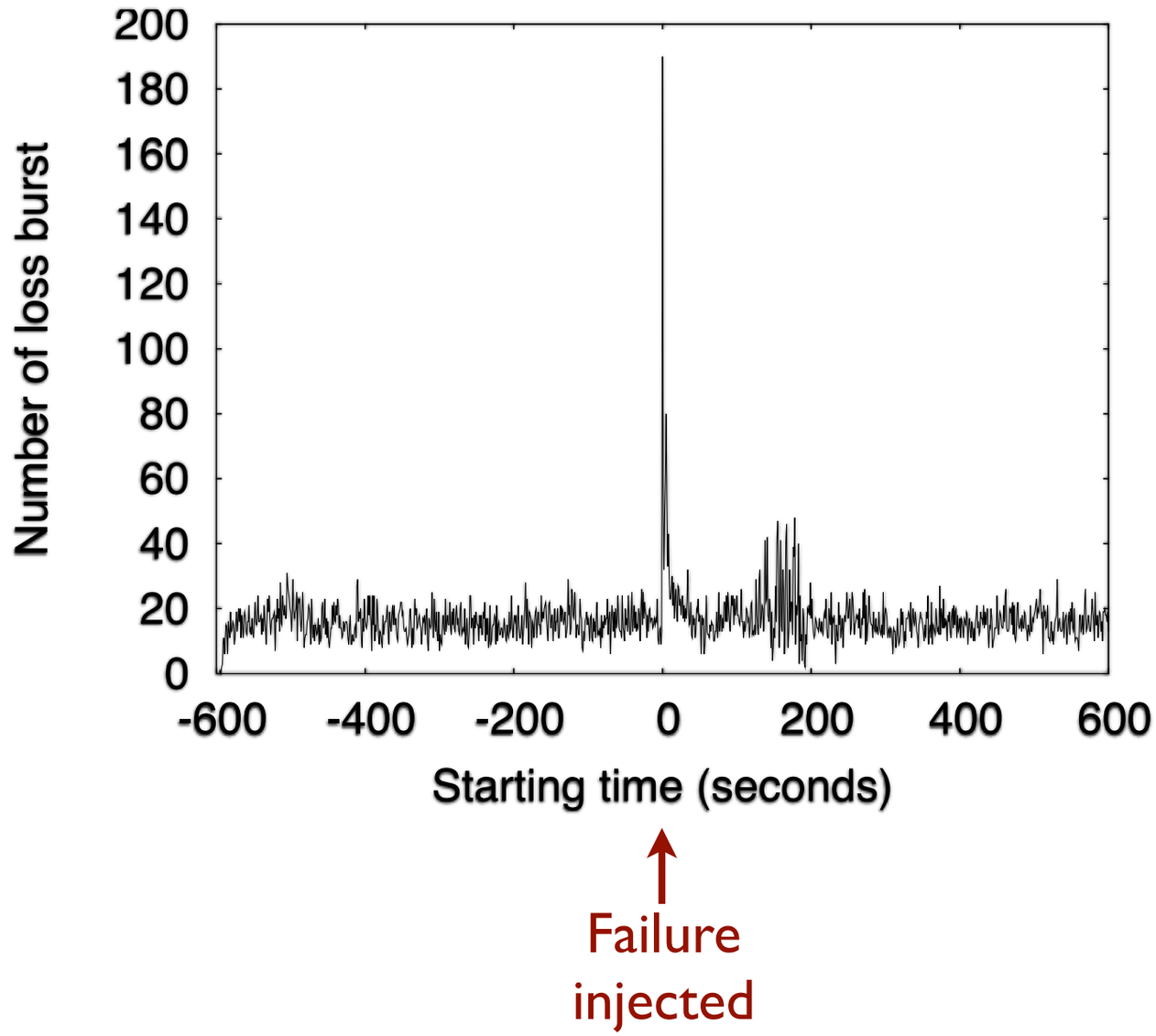
Multipath

reliability

path quality

Scalability

Policy



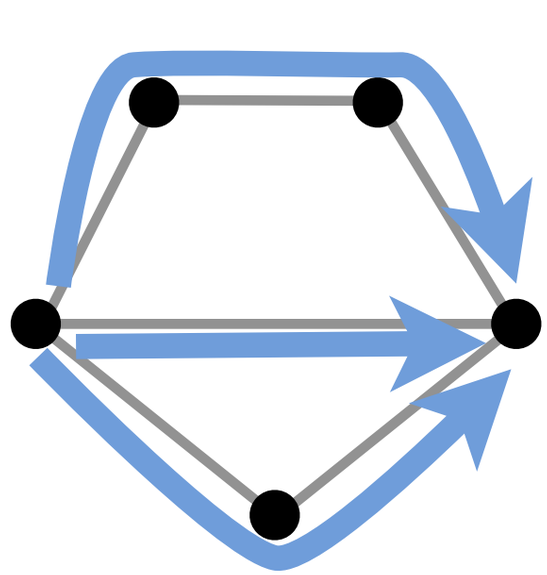
[F. Wang, Z. M. Mao, J. Wang, L. Gao, R. Bush '06]

Internet routing challenges

Multipath

reliability

path quality



Lowest latency path

Highest bandwidth path

Scalability

Path the network picked for you

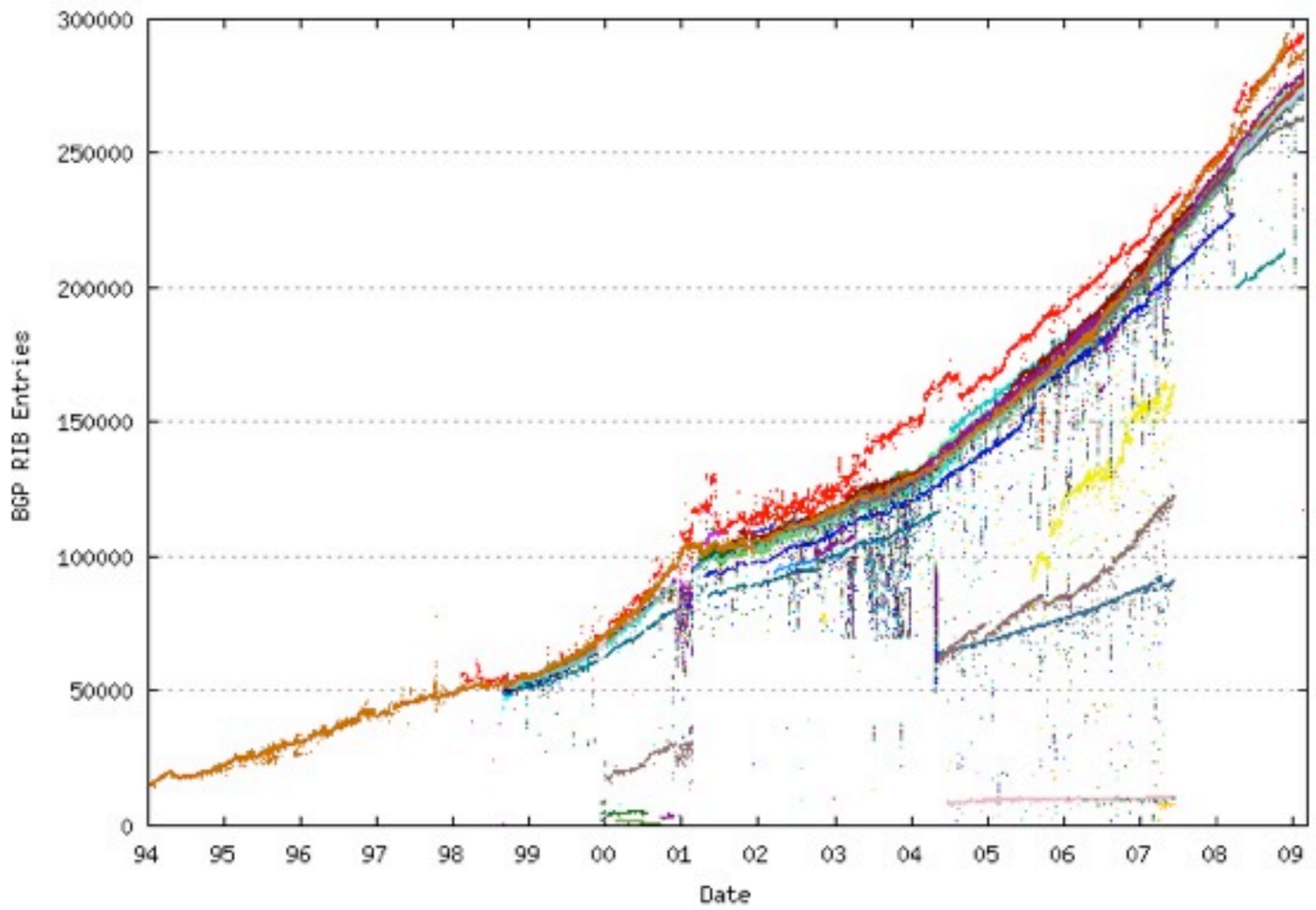
Policy

Internet routing challenges

Multipath
reliability
path quality

Scalability

Policy



Internet forwarding table size [Huston '09]

Internet routing challenges

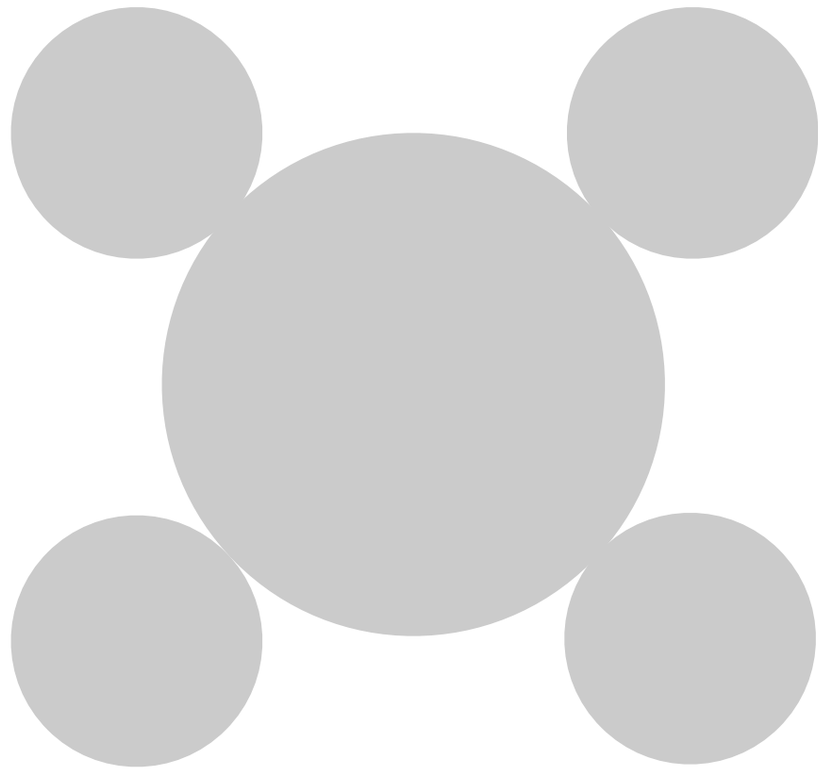
Multipath

reliability

path quality

Scalability

Policy



Internet routing challenges

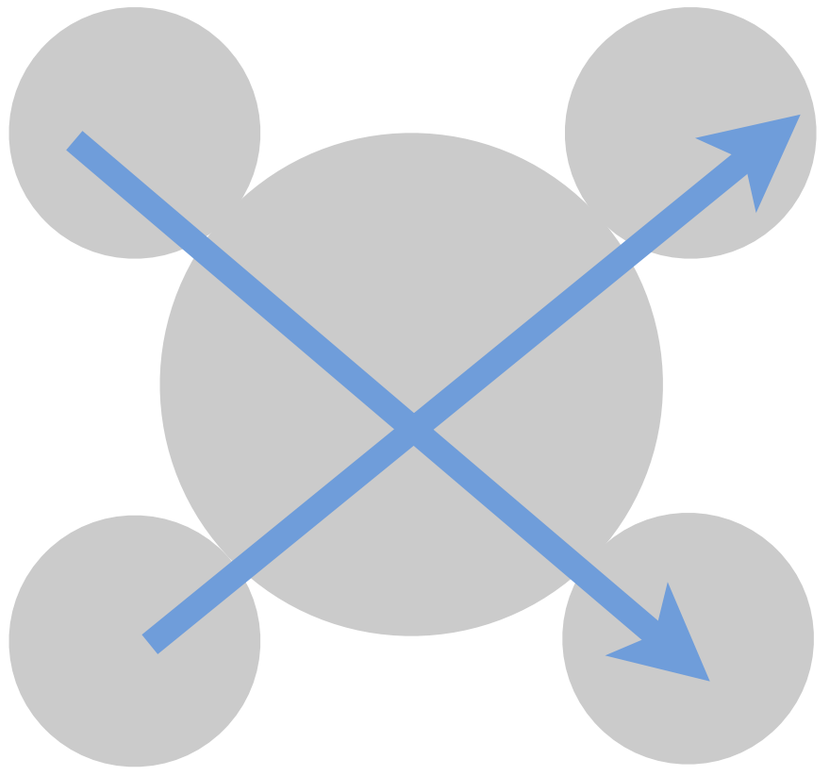
Multipath

reliability

path quality

Scalability

Policy



Internet routing challenges

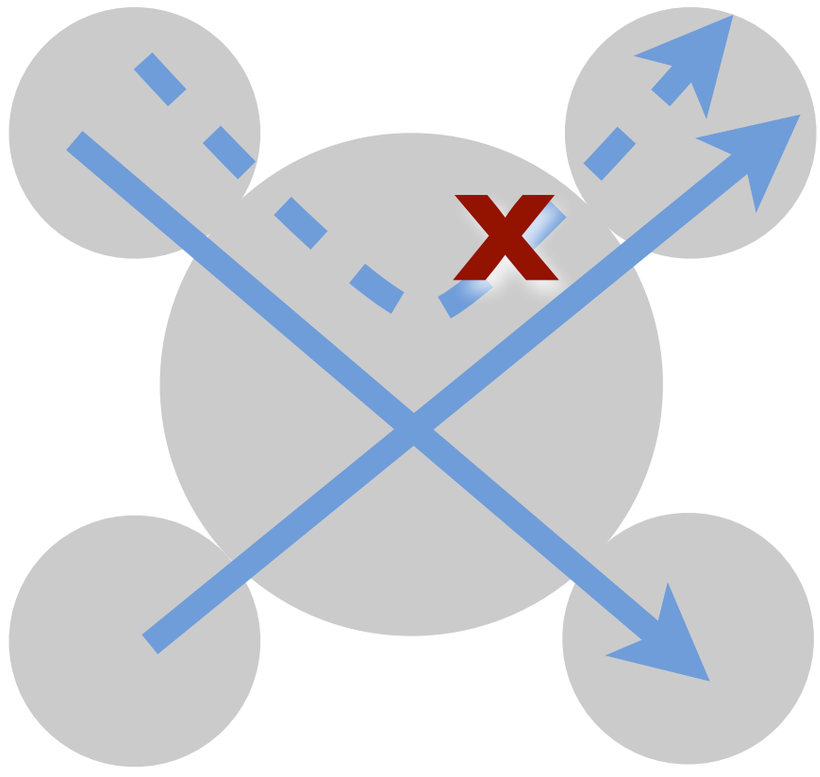
Multipath

reliability

path quality

Scalability

Policy



Pathlet routing

vnode virtual node

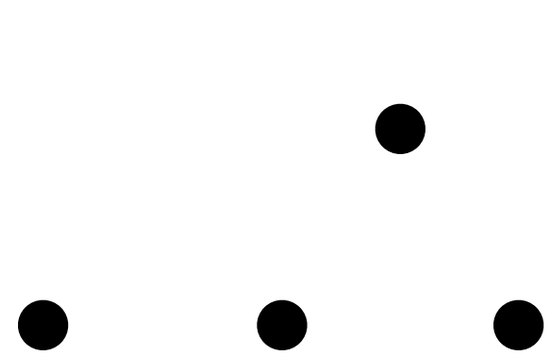
pathlet fragment of a path:
a sequence of vnodes

Source routing over pathlets.

Pathlet routing

vnode virtual node

pathlet fragment of a path:
a sequence of vnodes

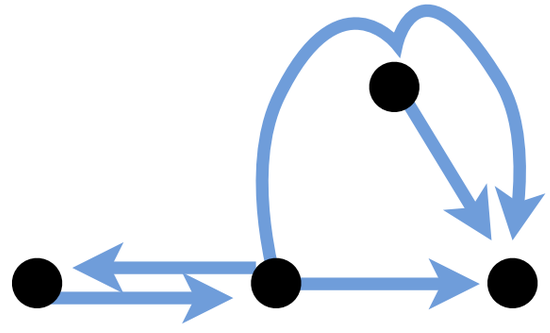


Source routing over pathlets.

Pathlet routing

vnode virtual node

pathlet fragment of a path:
a sequence of vnodes

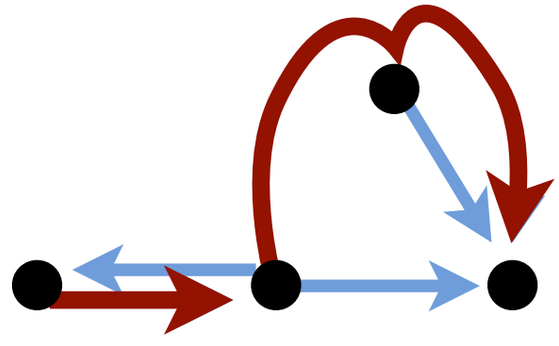


Source routing over pathlets.

Pathlet routing

vnode virtual node

pathlet fragment of a path:
a sequence of vnodes



Source routing over pathlets.

Pathlet routing

vnode virtual node

pathlet fragment of a path:
a sequence of vnodes

virtual graph:
flexible way to define
policy constraints

Source routing over pathlets.

Pathlet routing

vnode virtual node

pathlet fragment of a path:
a sequence of vnodes

Source routing over pathlets.

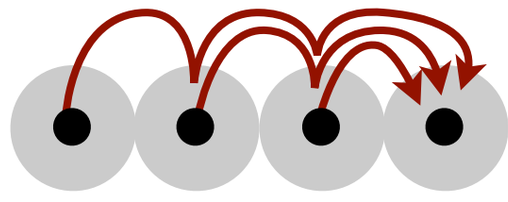
virtual graph:
flexible way to define
policy constraints

provides many path
choices for senders

Flexibility

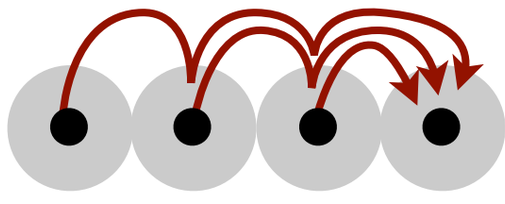
Flexibility

- can emulate BGP, source routing, MIRO, LISP, NIRA

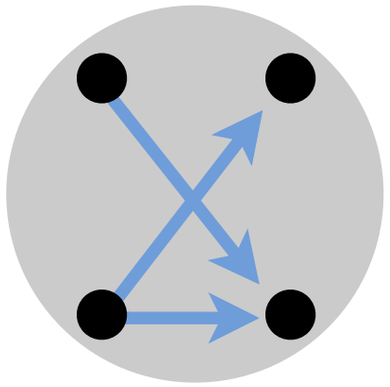


Flexibility

- can emulate BGP, source routing, MIRO, LISP, NIRA

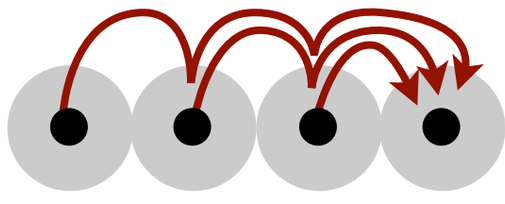


- local transit policies provide multipath and small forwarding tables

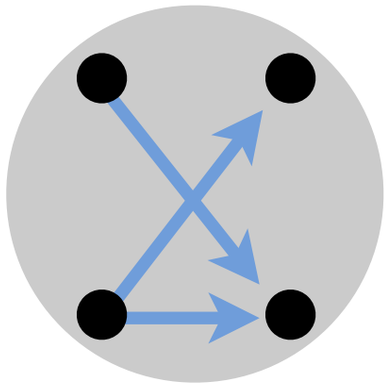


Flexibility

- can emulate BGP, source routing, MIRO, LISP, NIRA



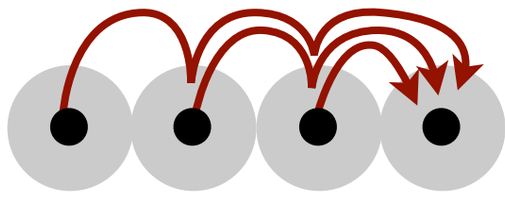
- local transit policies provide multipath and small forwarding tables



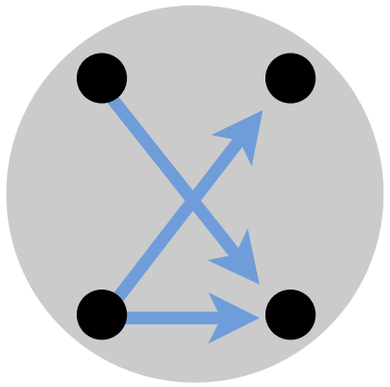
- coexistence of different styles of routing policy

Flexibility

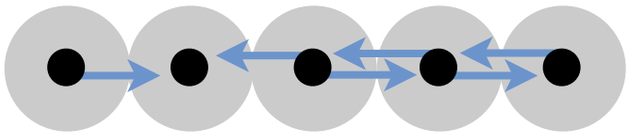
- can emulate BGP, source routing, MIRO, LISP, NIRA



- local transit policies provide multipath and small forwarding tables

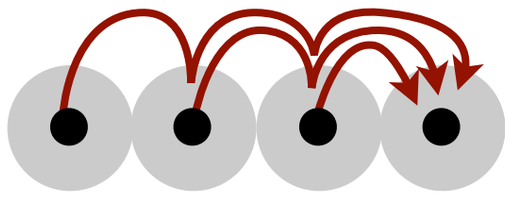


- coexistence of different styles of routing policy

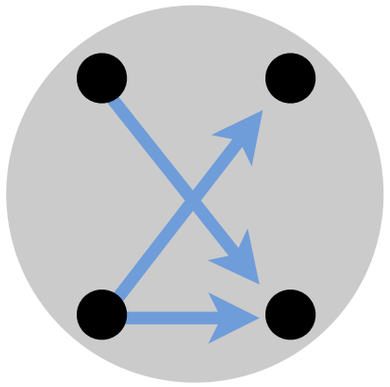


Flexibility

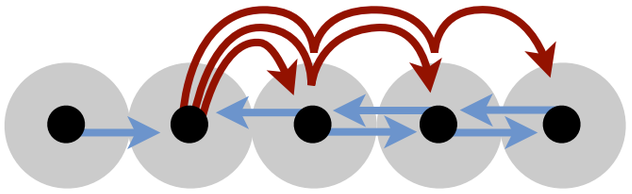
- can emulate BGP, source routing, MIRO, LISP, NIRA



- local transit policies provide multipath and small forwarding tables



- coexistence of different styles of routing policy



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"

Outline

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

Pathlet routing

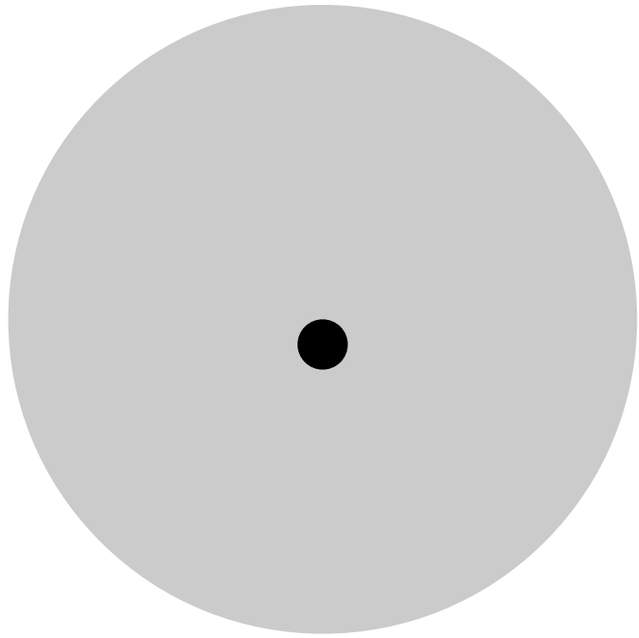
vnode virtual node

pathlet fragment of a path:
a sequence of vnodes

Source routing over pathlets.

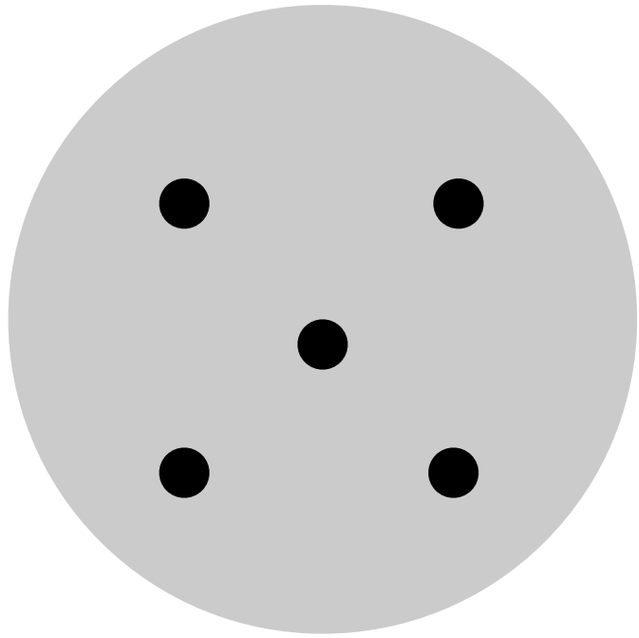
vnodes

vnode: virtual node
within an AS



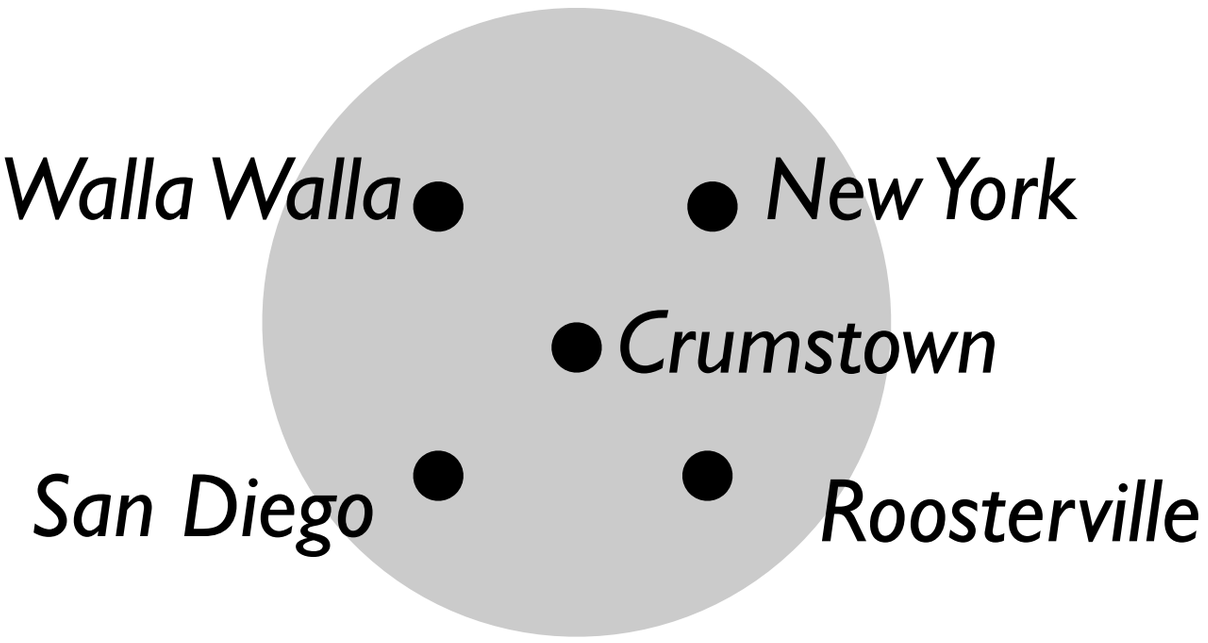
vnodes

vnode: virtual node
within an AS



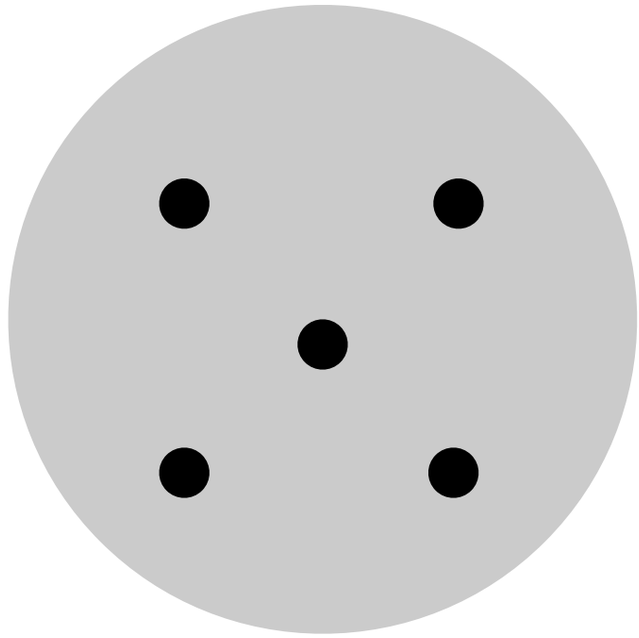
vnodes

vnode: virtual node within an AS



vnodes

vnode: virtual node
within an AS

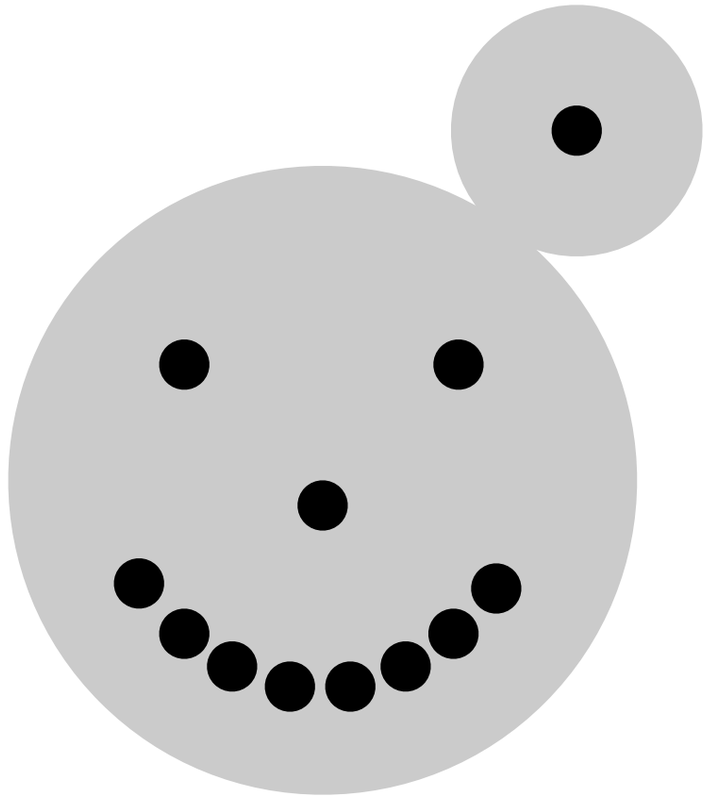


vnodes

vnode: virtual node
within an AS

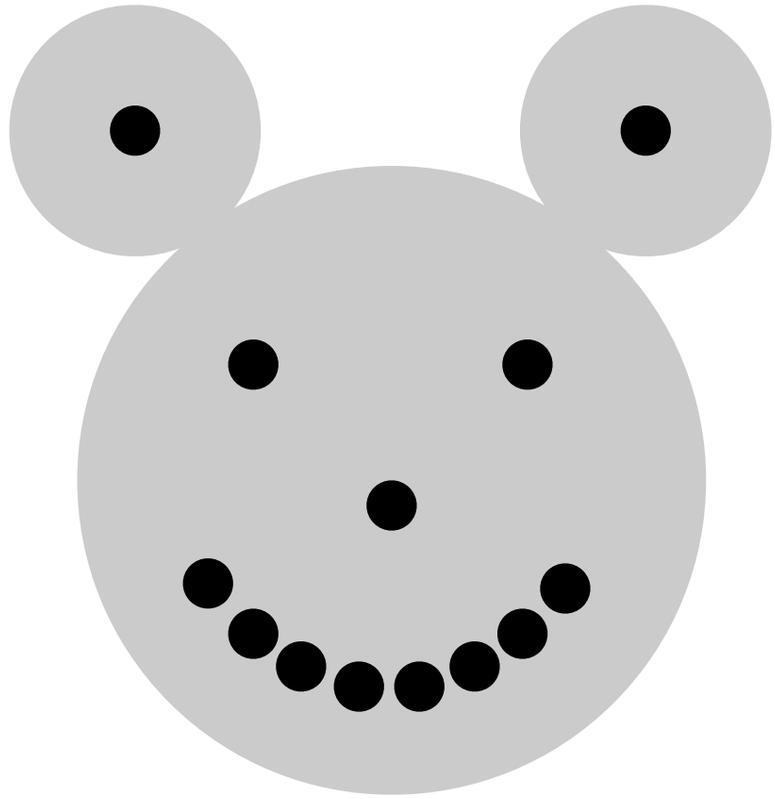
vnodes

vnode: virtual node within an AS



vnodes

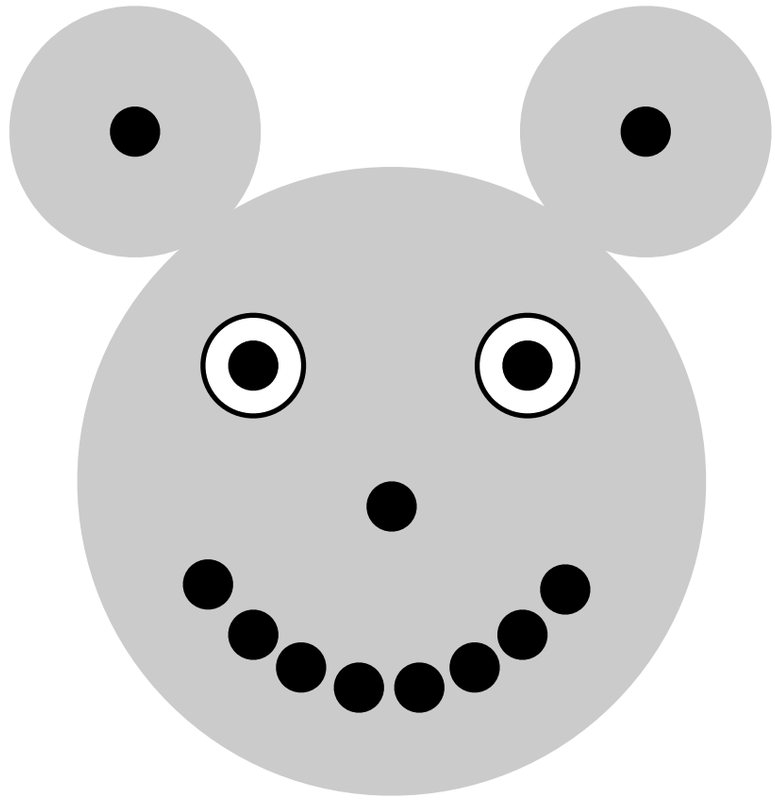
vnode: virtual node within an AS



vnodes

vnode: virtual node within an AS

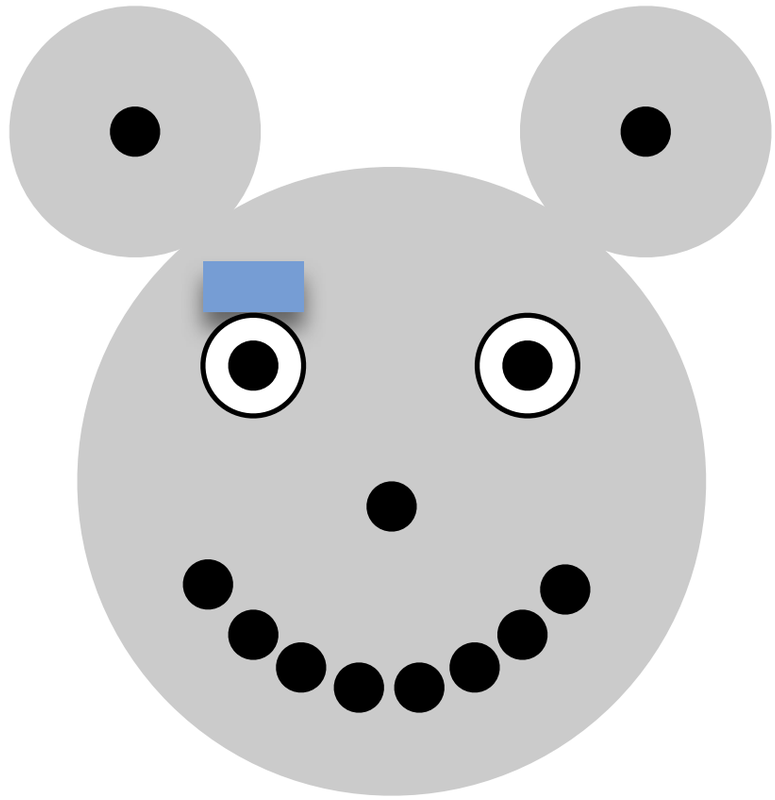
designated ingress vnode for each neighbor



vnodes

vnode: virtual node within an AS

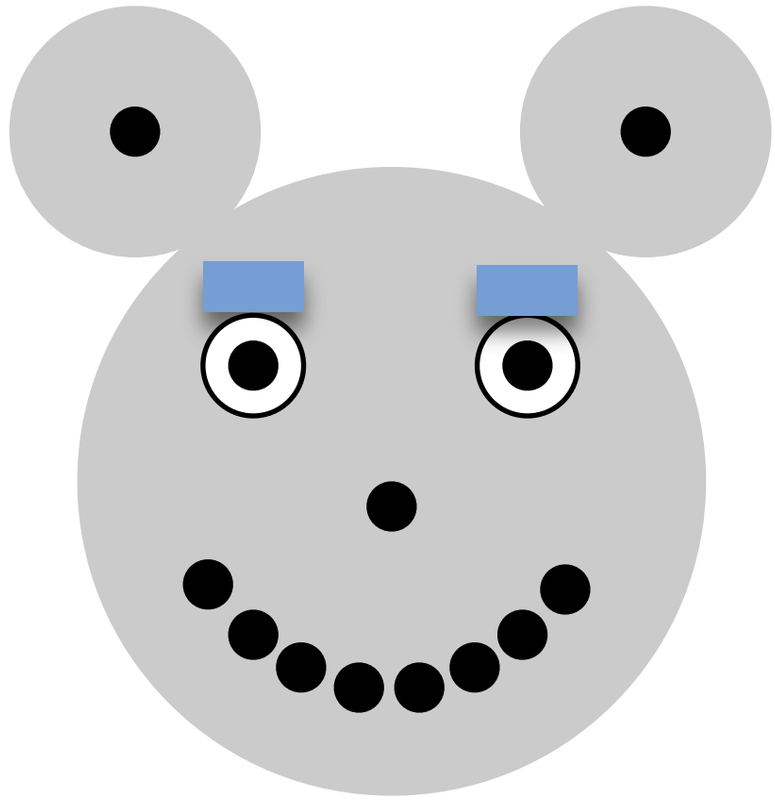
designated ingress vnode for each neighbor



vnodes

vnode: virtual node within an AS

designated ingress vnode for each neighbor

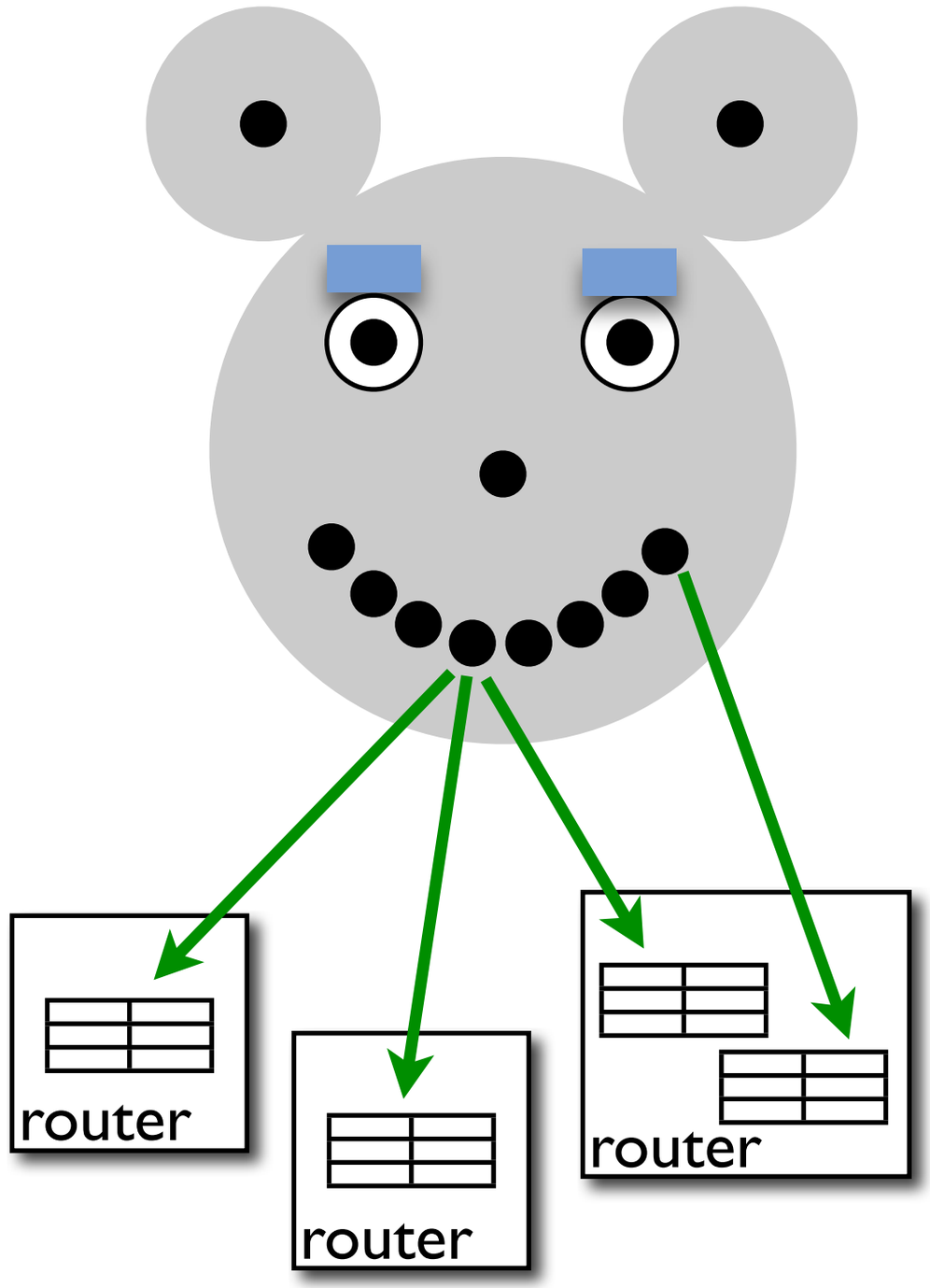


vnodes

vnode: virtual node within an AS

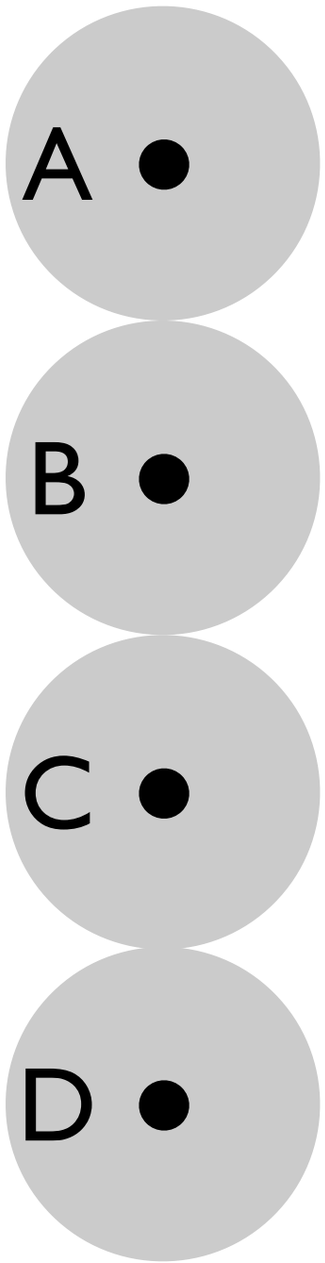
designated ingress vnode for each neighbor

Internally: a forwarding table at one or more routers



Pathlets

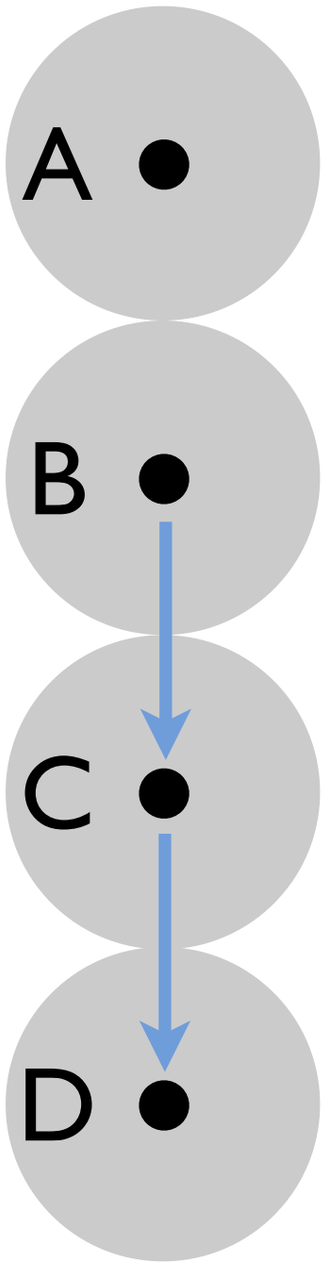
Packet route field Forwarding table



Pathlets

Packet route field

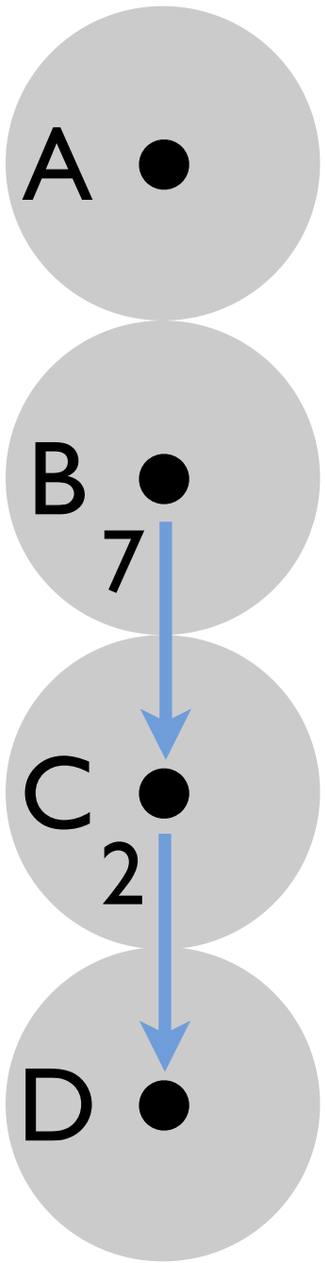
Forwarding table



Pathlets

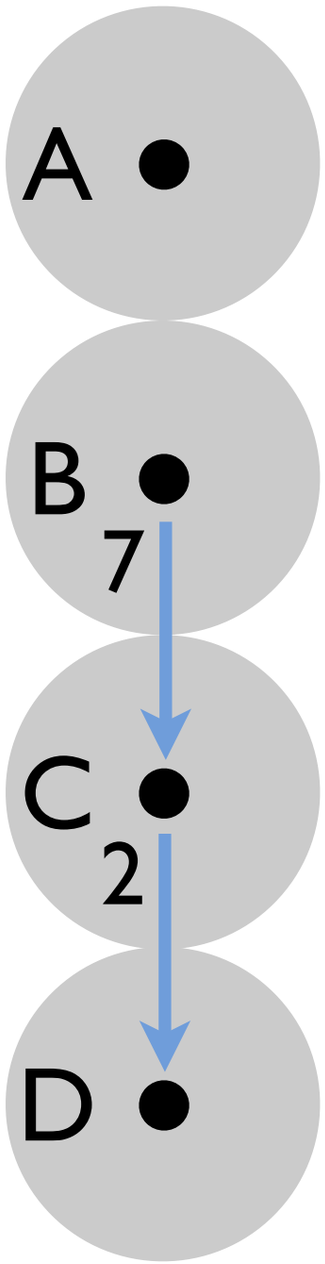
Packet route field

Forwarding table



Pathlets

Packet route field



Forwarding table

...	...
7	fwd to C

...	...
2	fwd to D

Pathlets

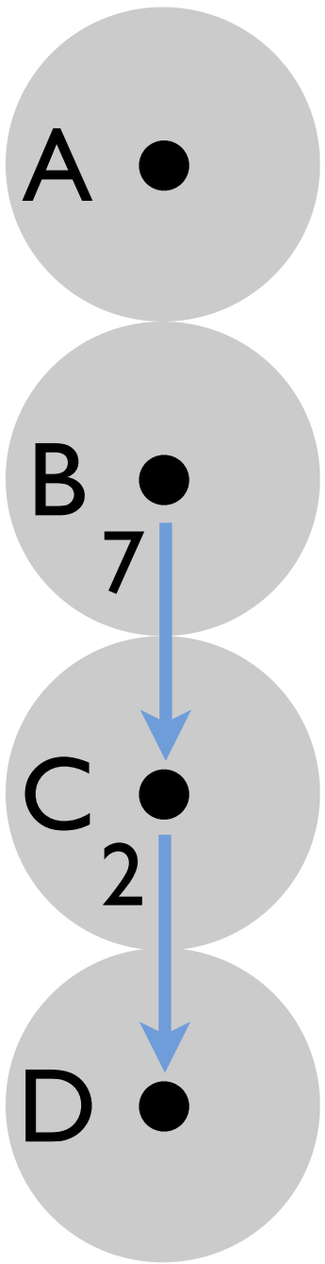
Packet route field

7,2

Forwarding table

...	...
7	fwd to C

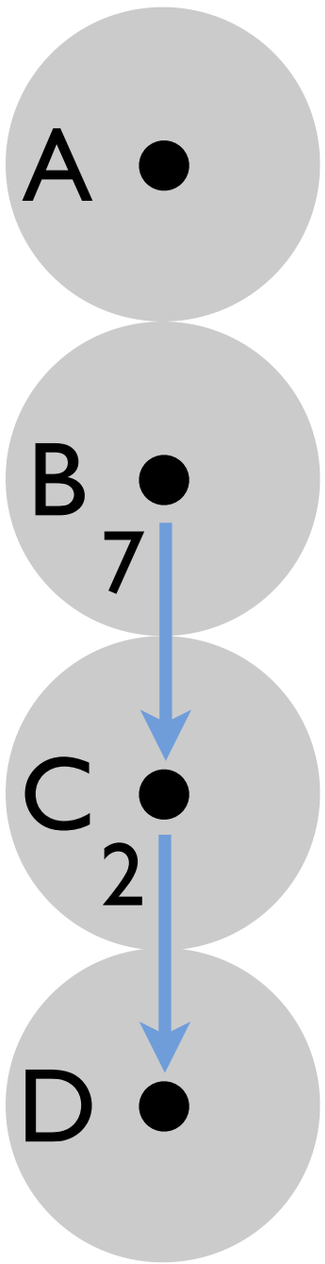
...	...
2	fwd to D



Pathlets

Packet route field

Forwarding table



7,2

2

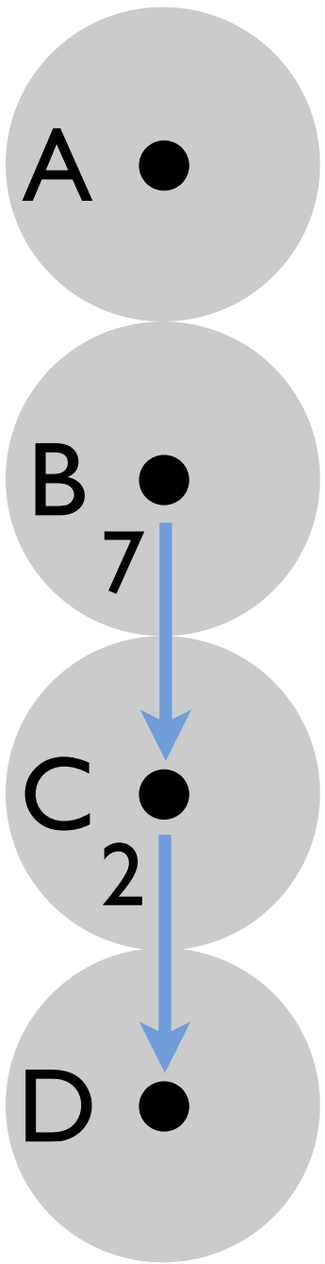
...	...
7	fwd to C

...	...
2	fwd to D

Pathlets

Packet route field

Forwarding table



7,2

2

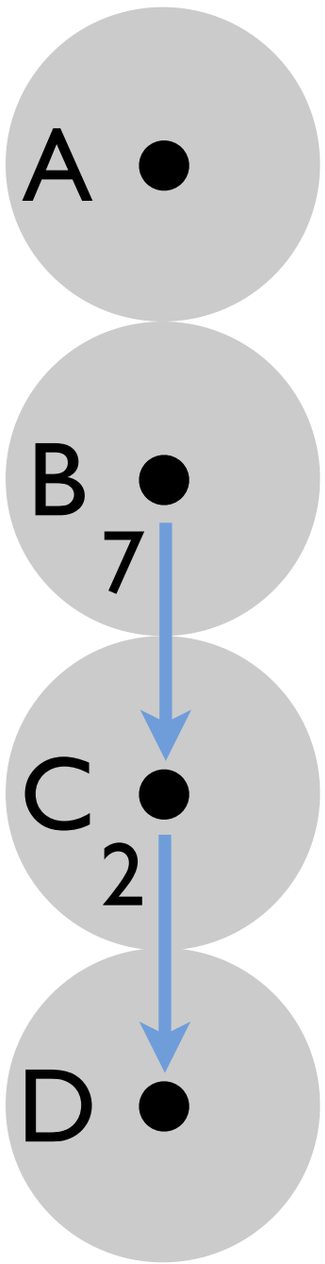
...	...
7	fwd to C

...	...
2	fwd to D

Pathlets

Packet route field

Forwarding table



7,2

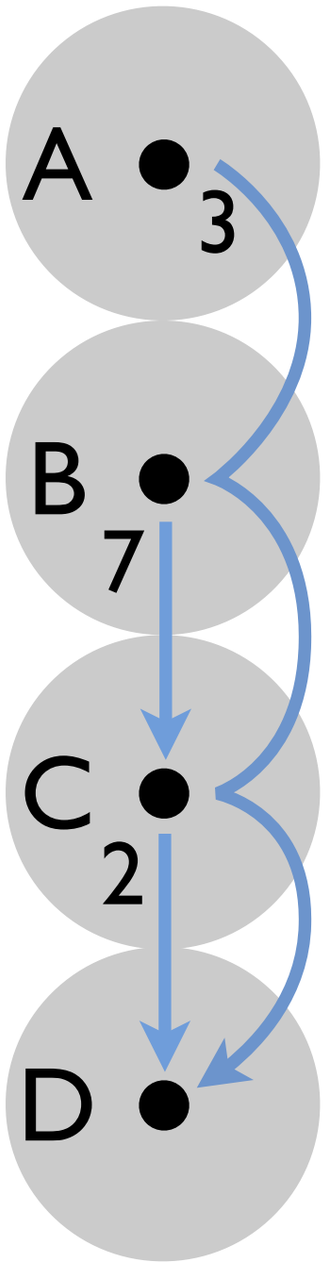
2

...	...
7	fwd to C

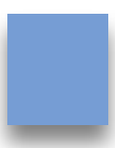
...	...
2	fwd to D

delivered!

Pathlets



Packet route field



Forwarding table

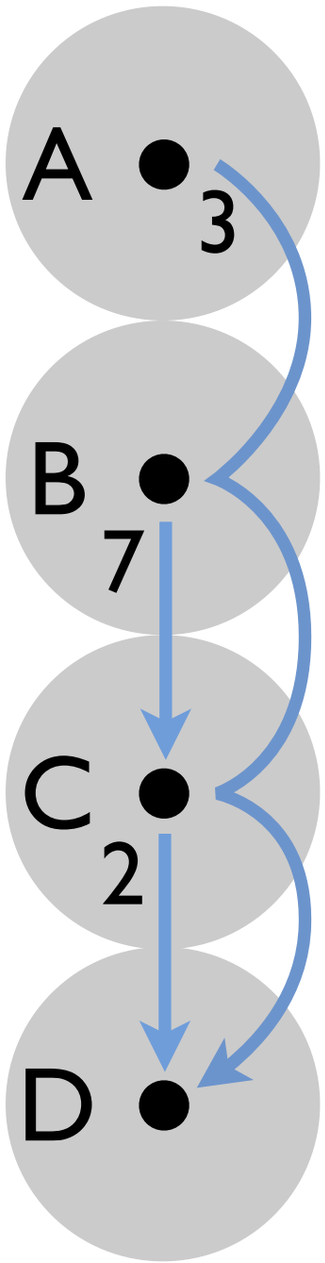
...	...
3	push 7,2; fwd to B

...	...
7	fwd to C

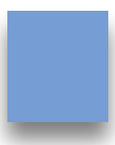
...	...
2	fwd to D

delivered!

Pathlets



Packet route field



Forwarding table

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

...	...
7	fwd to C

...	...
2	fwd to D

delivered!

Dissemination

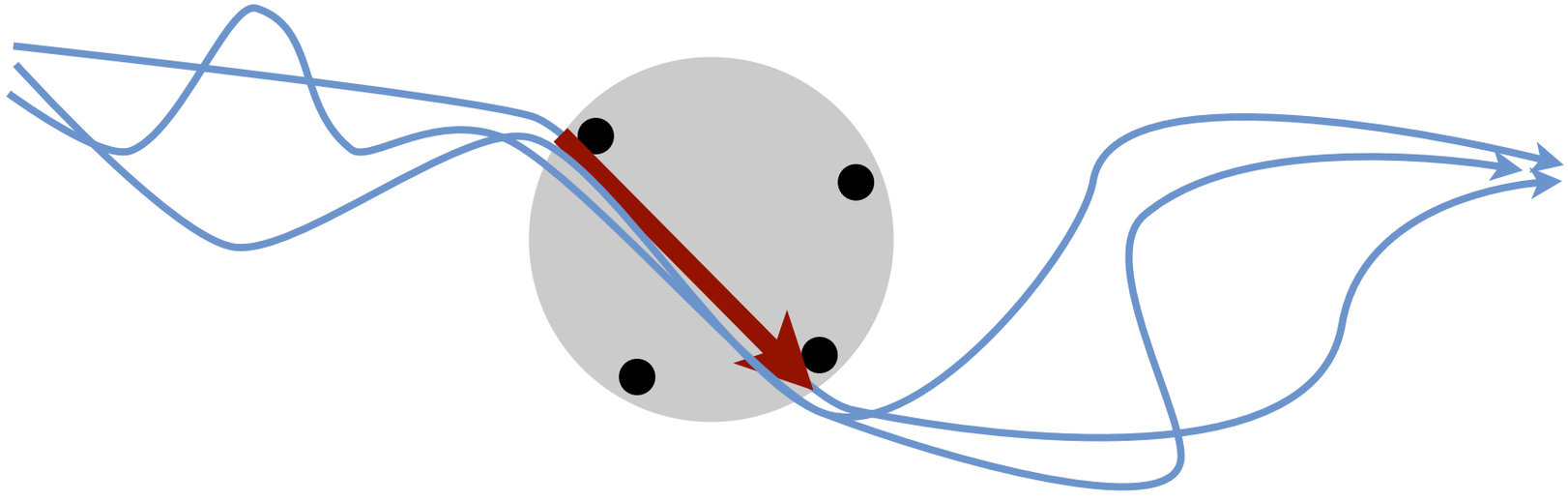
- Global gossip fine, except for scalability
- So, let routers choose not to disseminate some pathlets
- Leads to (ironic) use of path vector — only for pathlet dissemination, not route selection

Outline

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

Local transit policies

Each ingress → egress pair
is either allowed or disallowed.

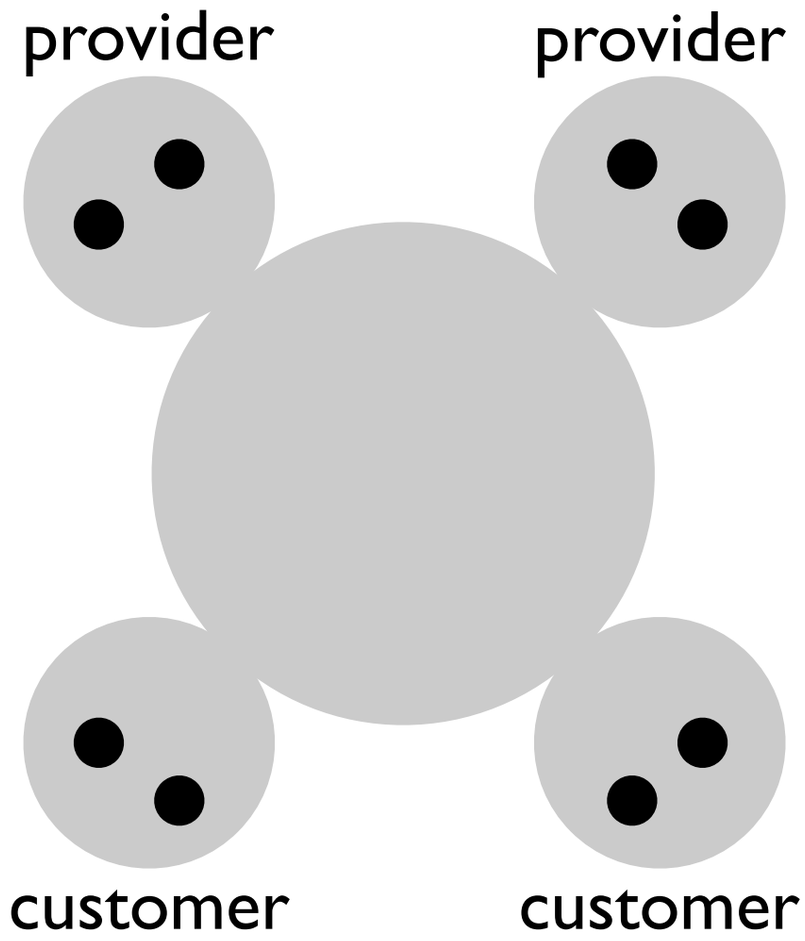


Subject to this, any path allowed!

Represented with few pathlets: small FIB

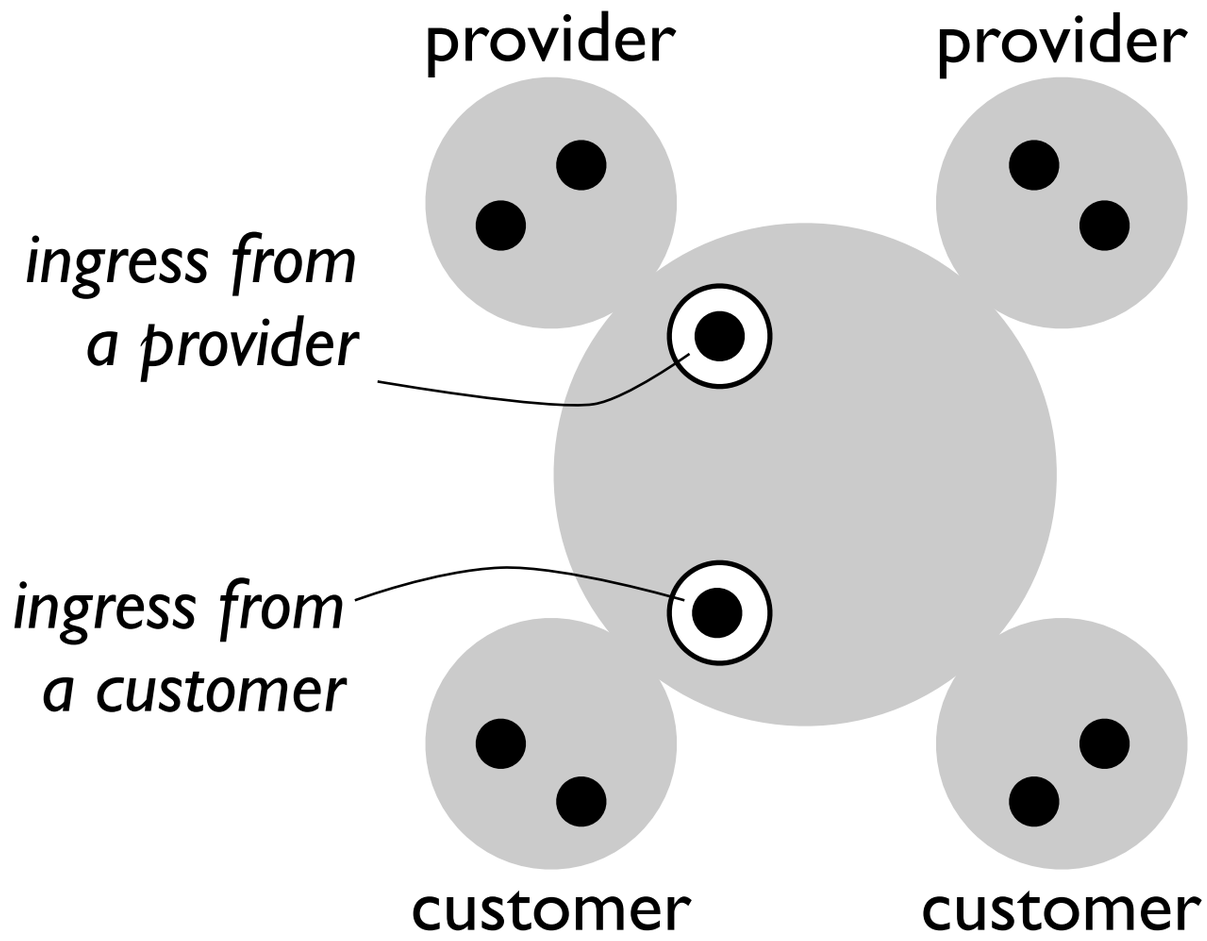
“All valley-free” is local

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



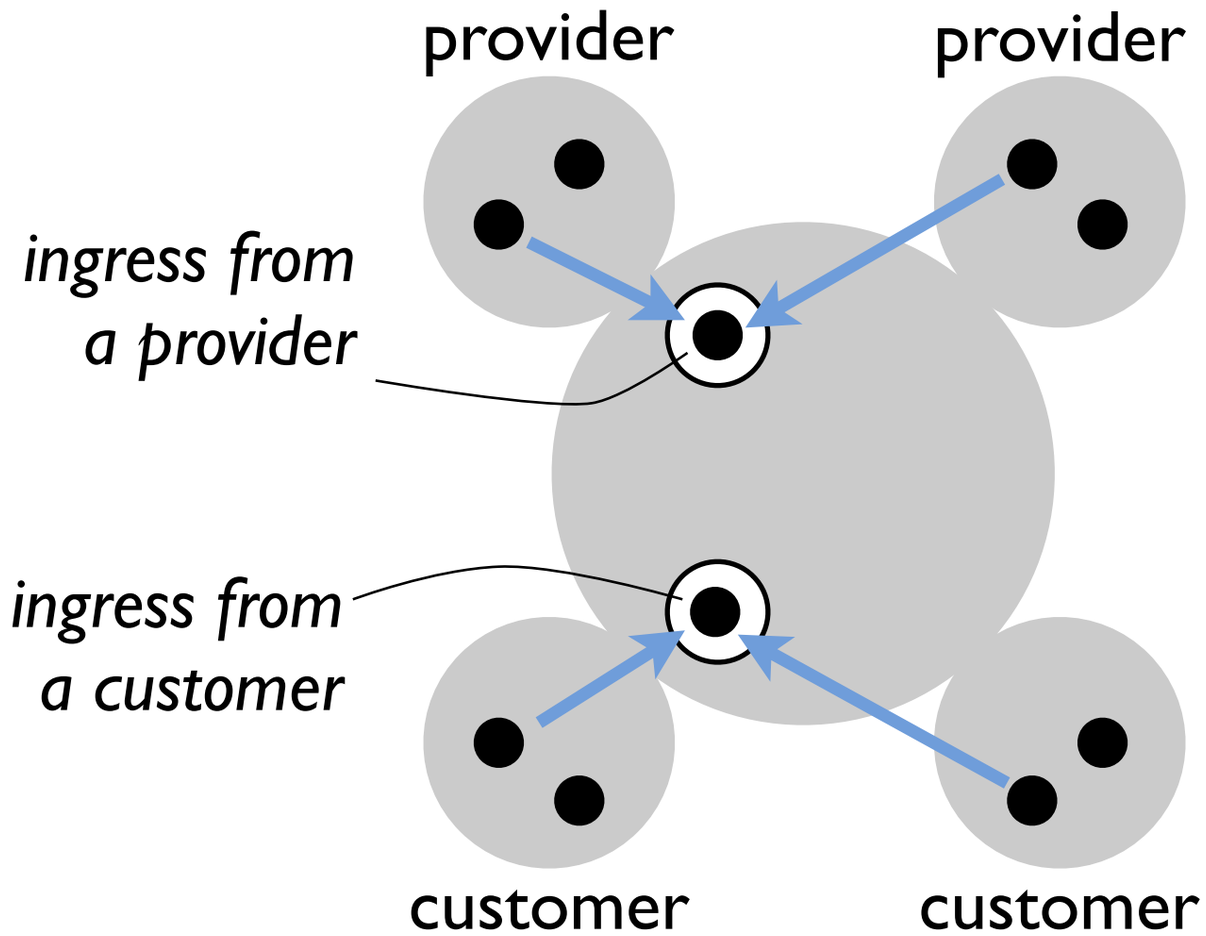
“All valley-free” is local

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



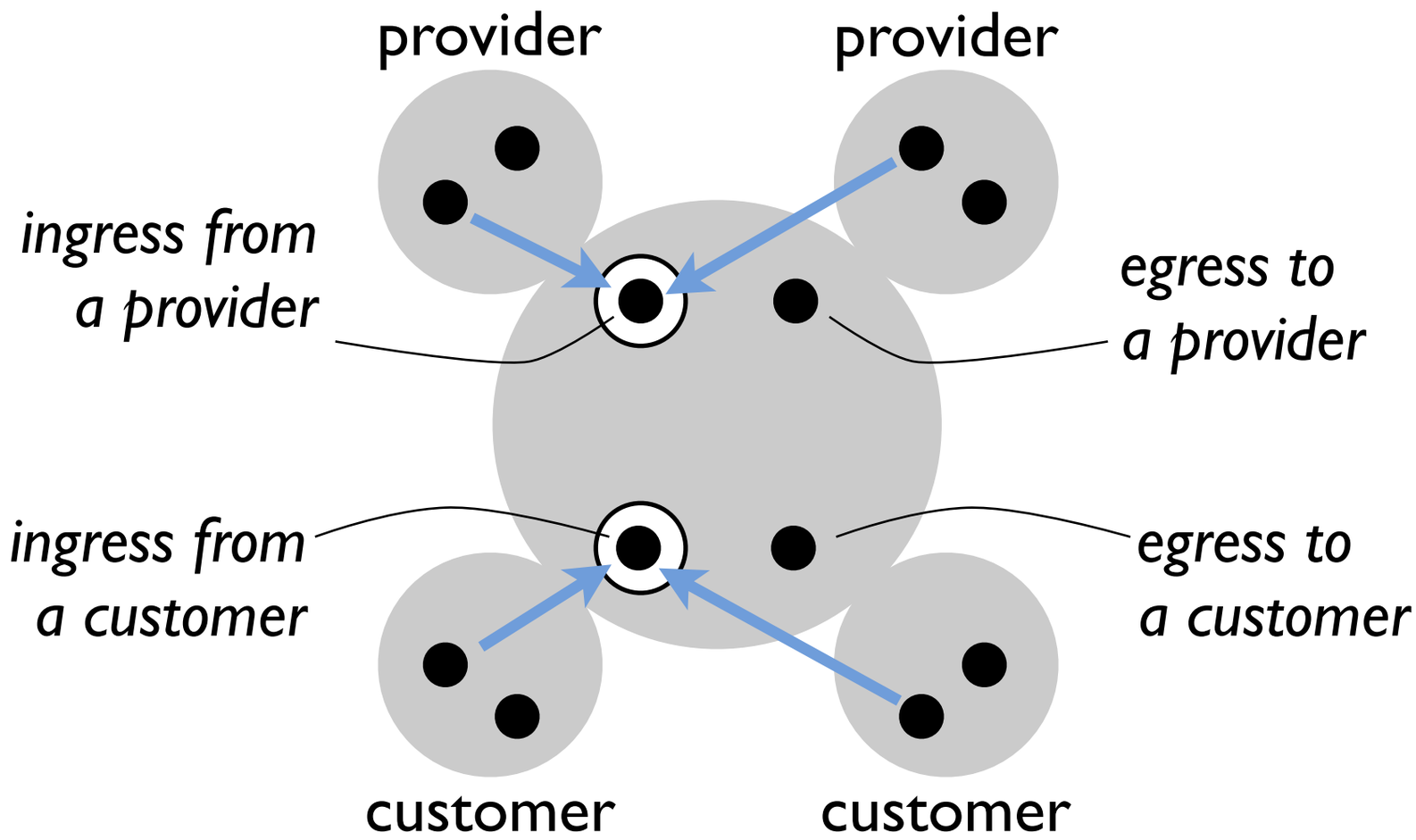
“All valley-free” is local

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



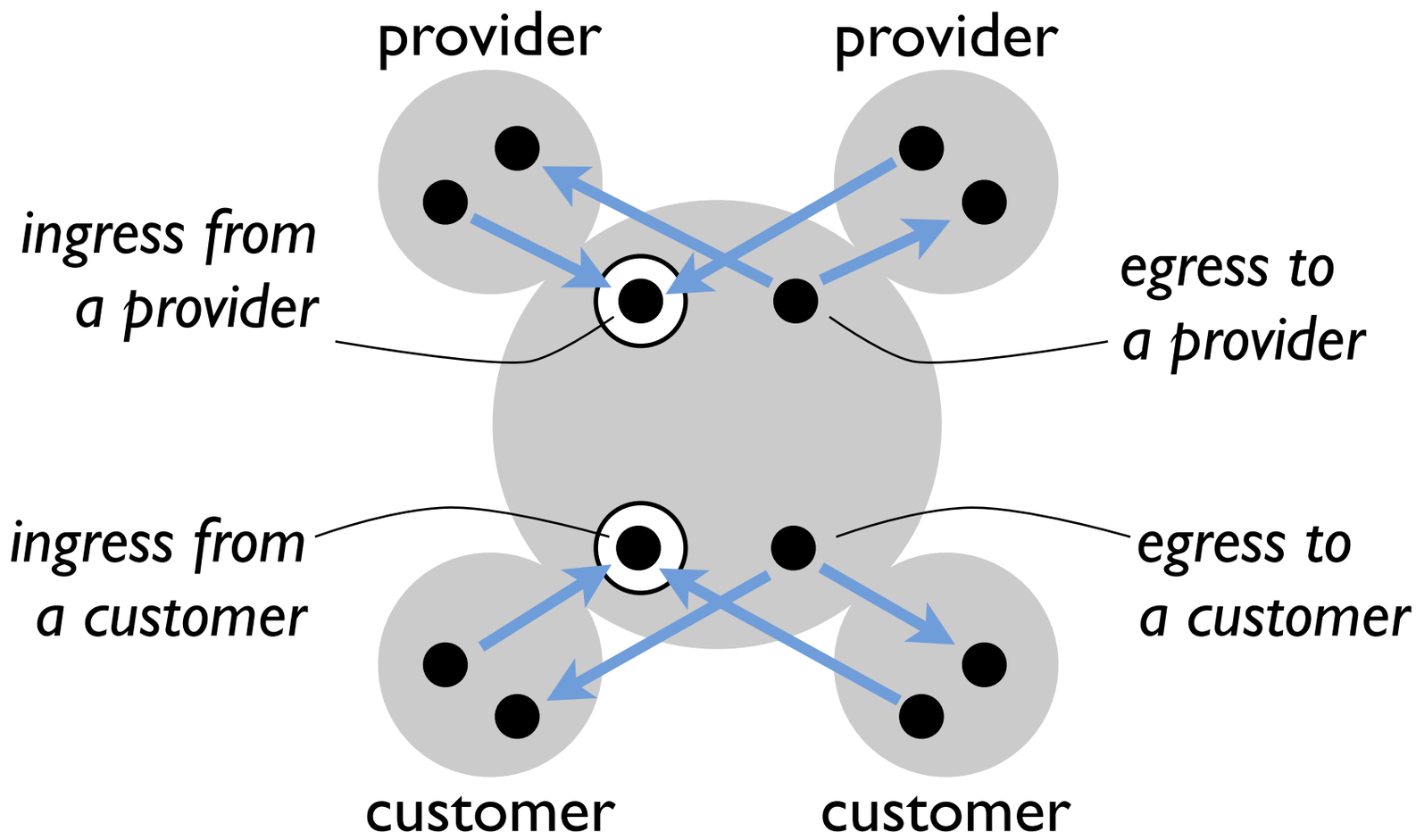
“All valley-free” is local

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



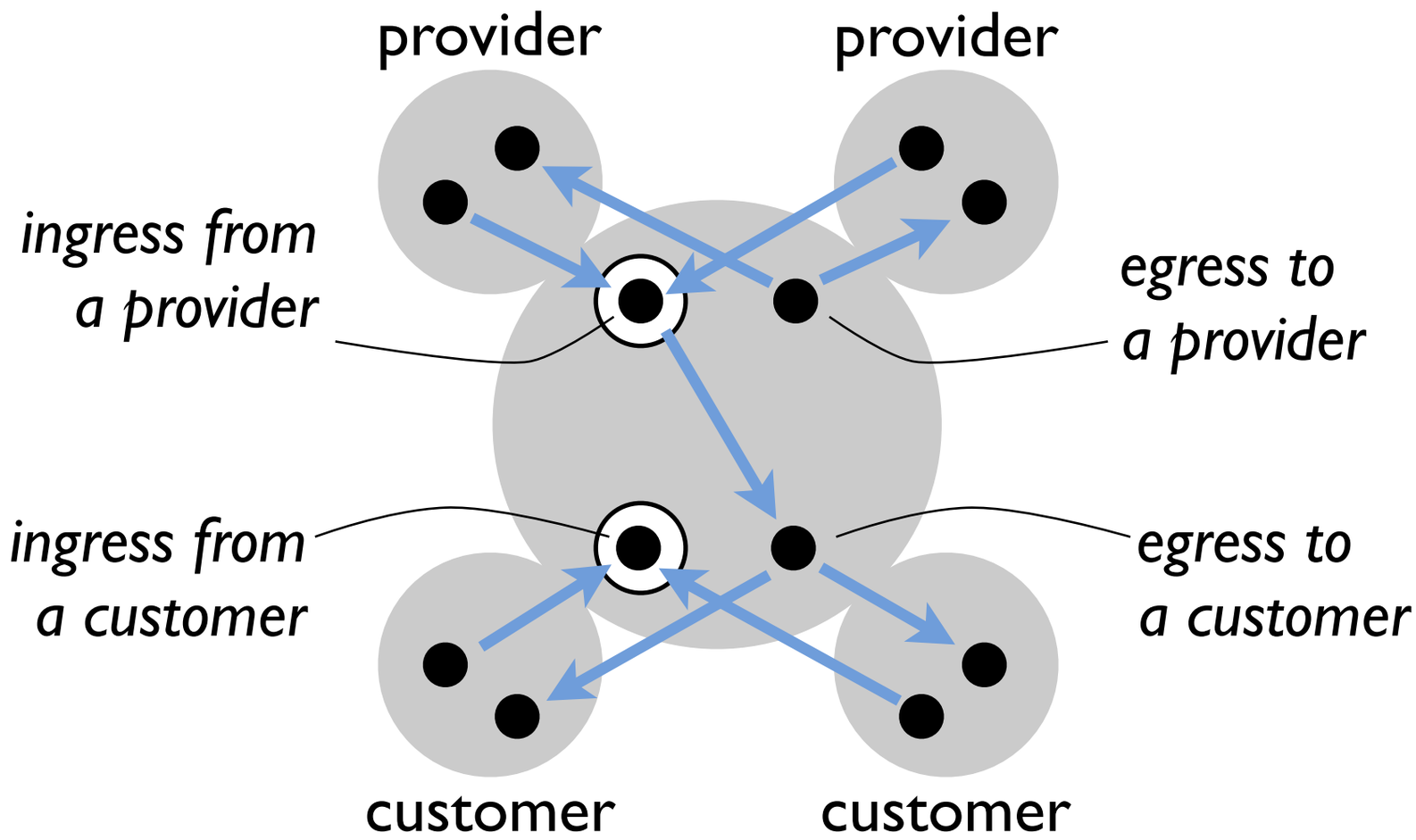
“All valley-free” is local

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



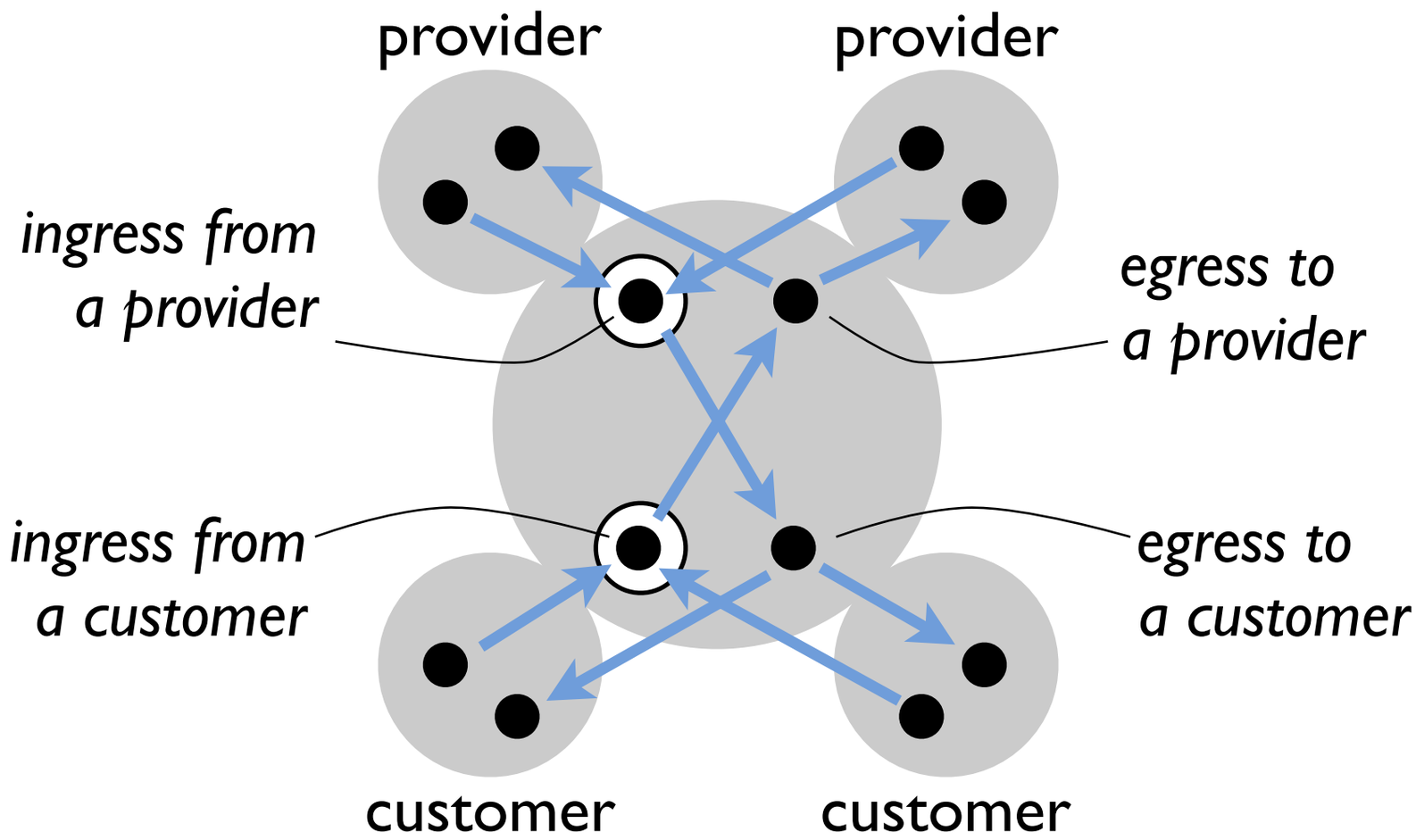
“All valley-free” is local

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



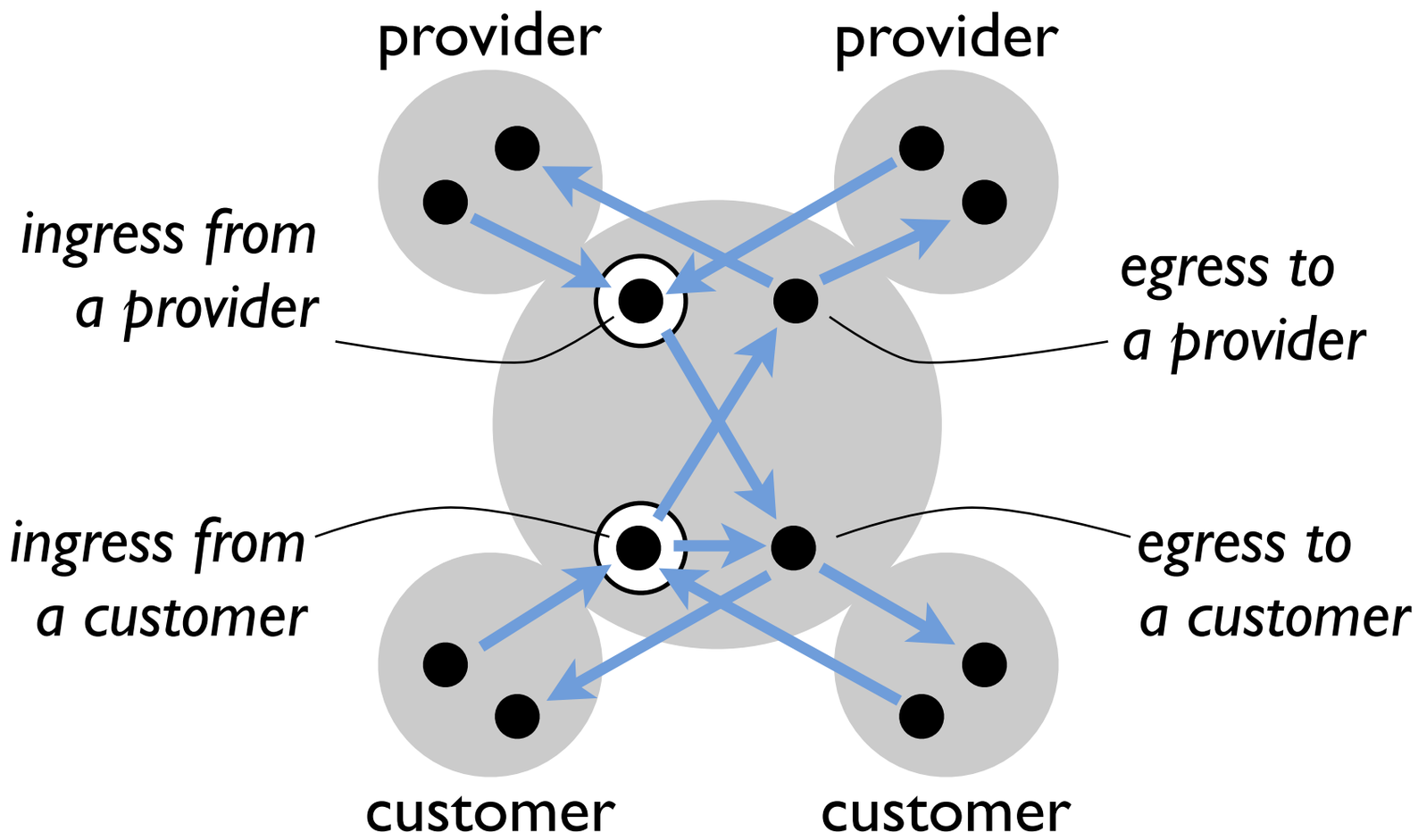
“All valley-free” is local

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



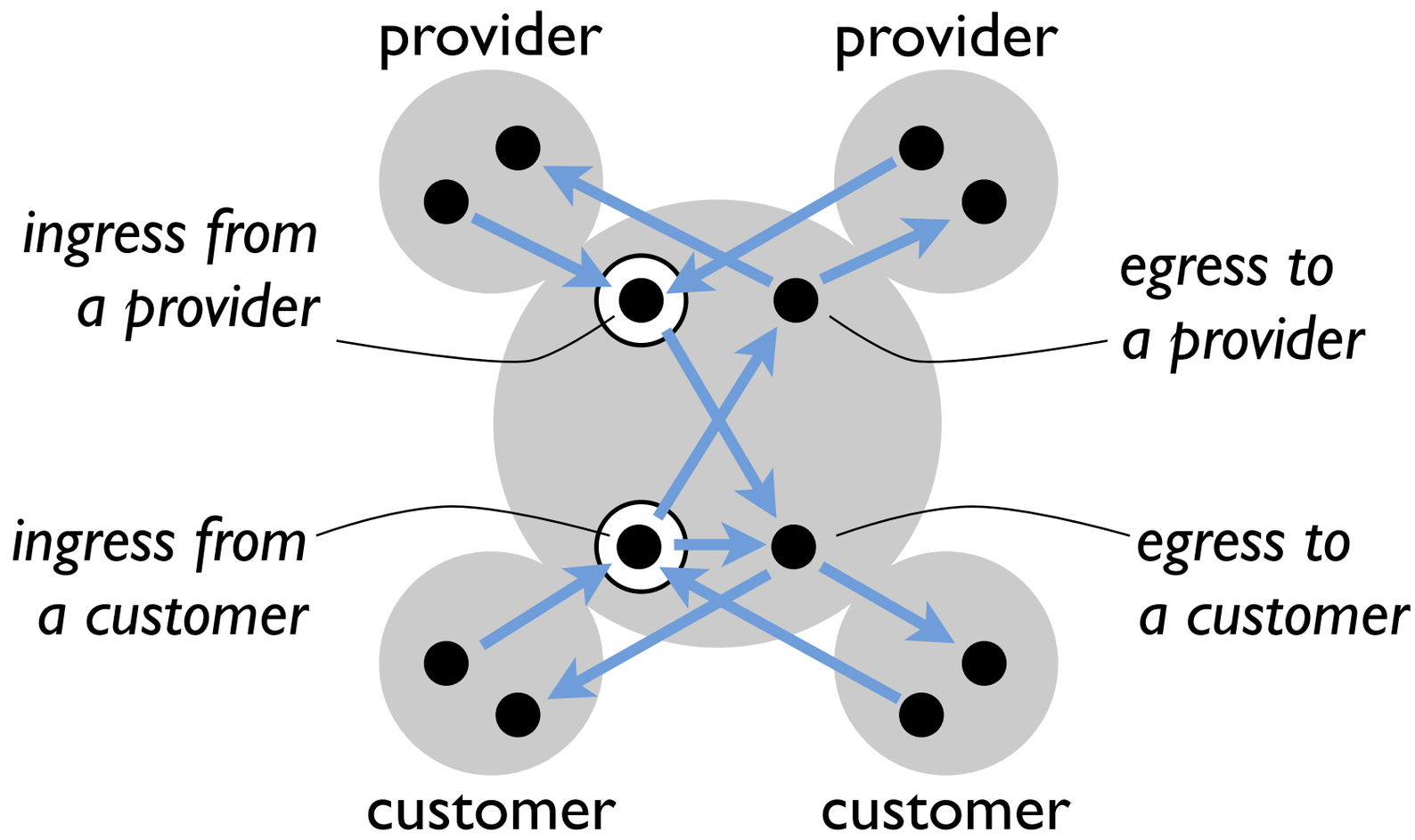
“All valley-free” is local

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



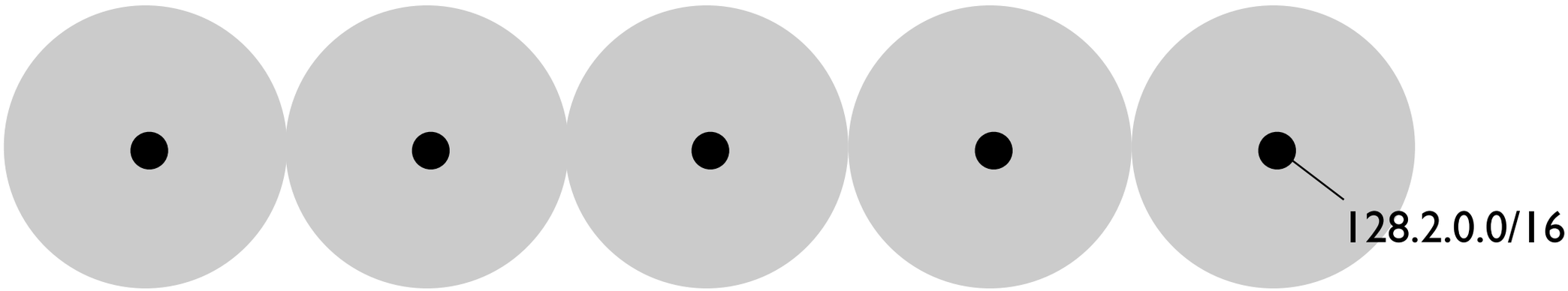
“All valley-free” is local

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

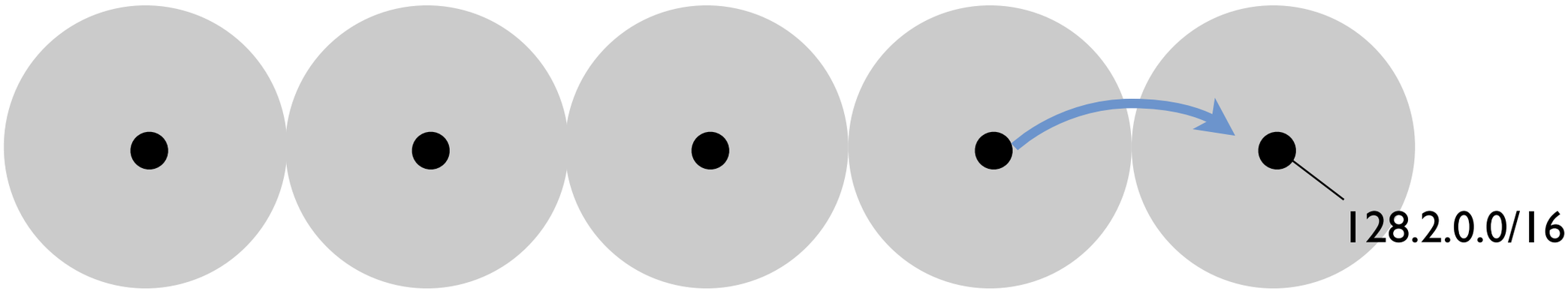


Forwarding table size: 3 + #neighbors

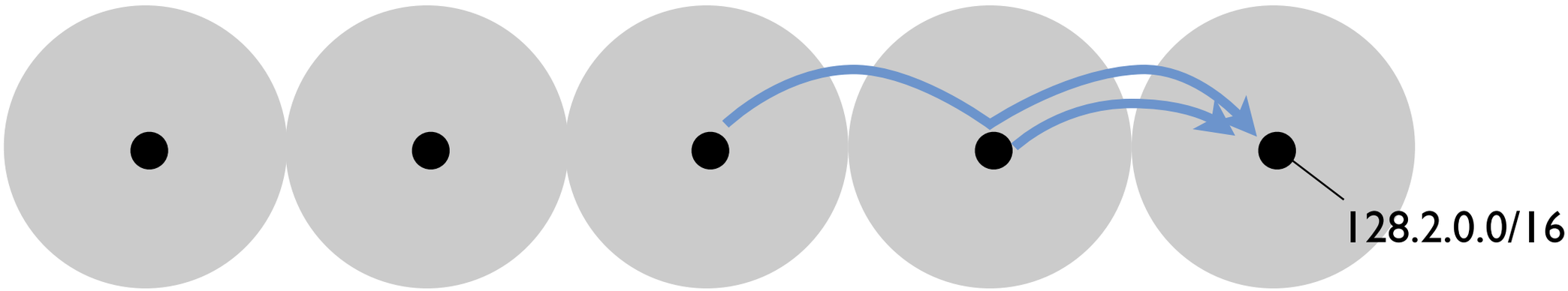
Emulating BGP



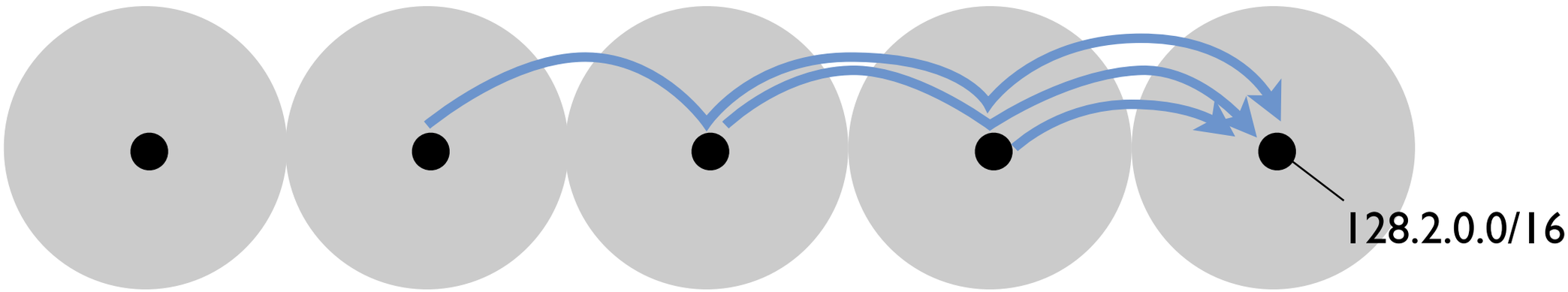
Emulating BGP



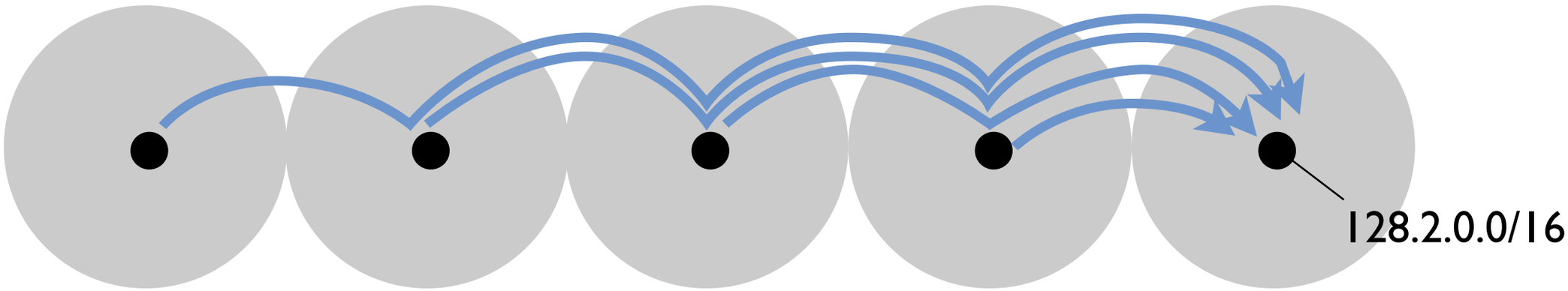
Emulating BGP



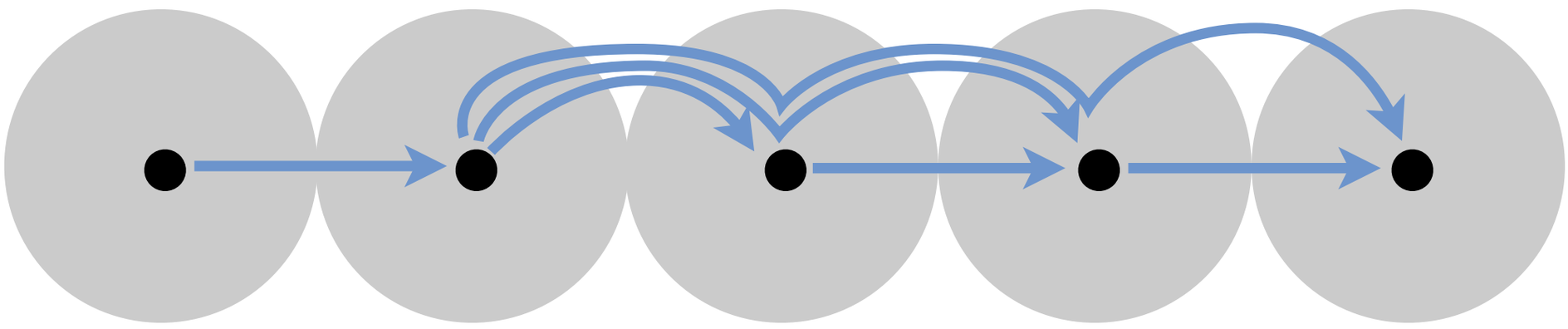
Emulating BGP



Emulating BGP



Mixed policies

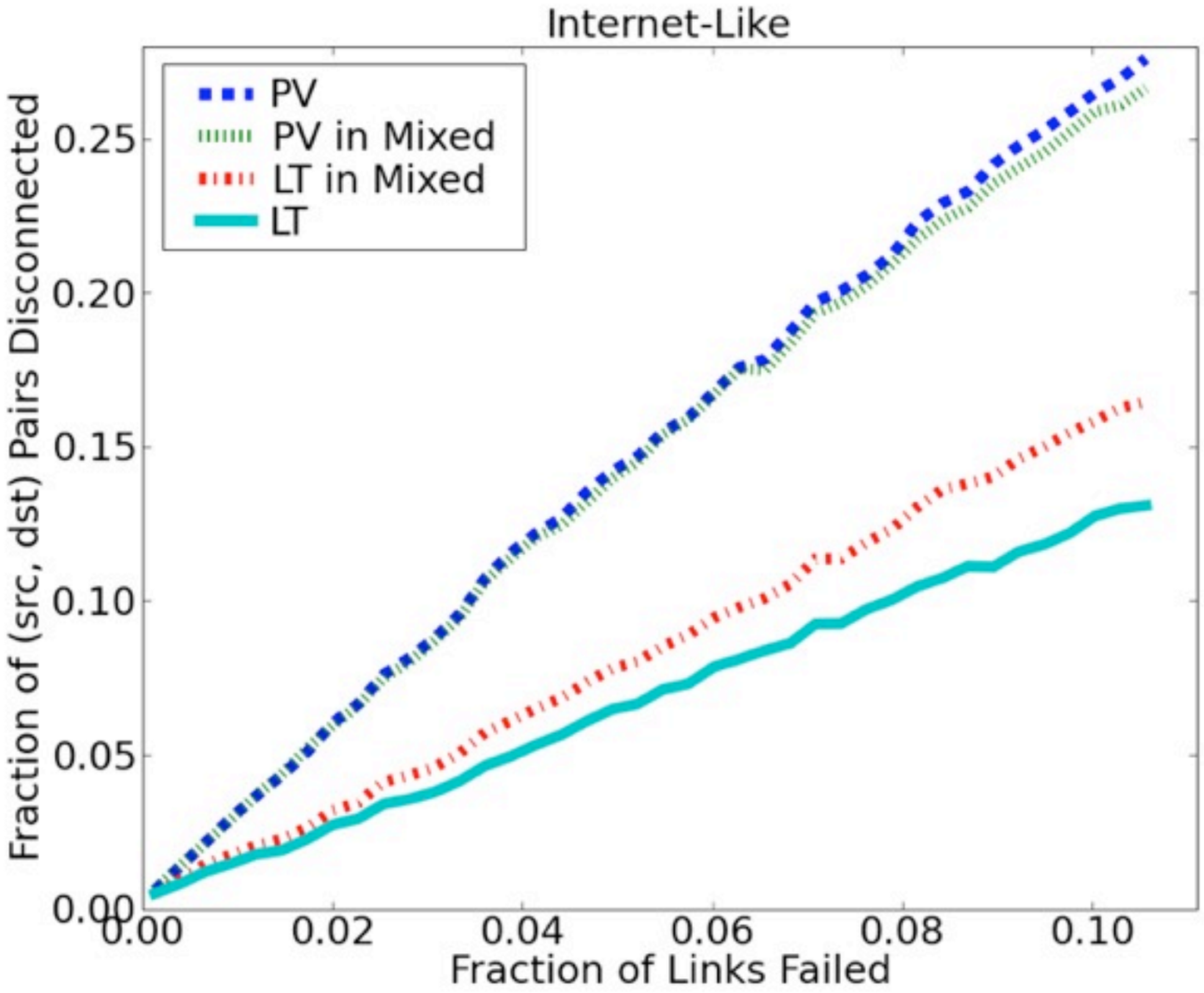


local BGP-like local local local

Outline

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

Improved connectivity



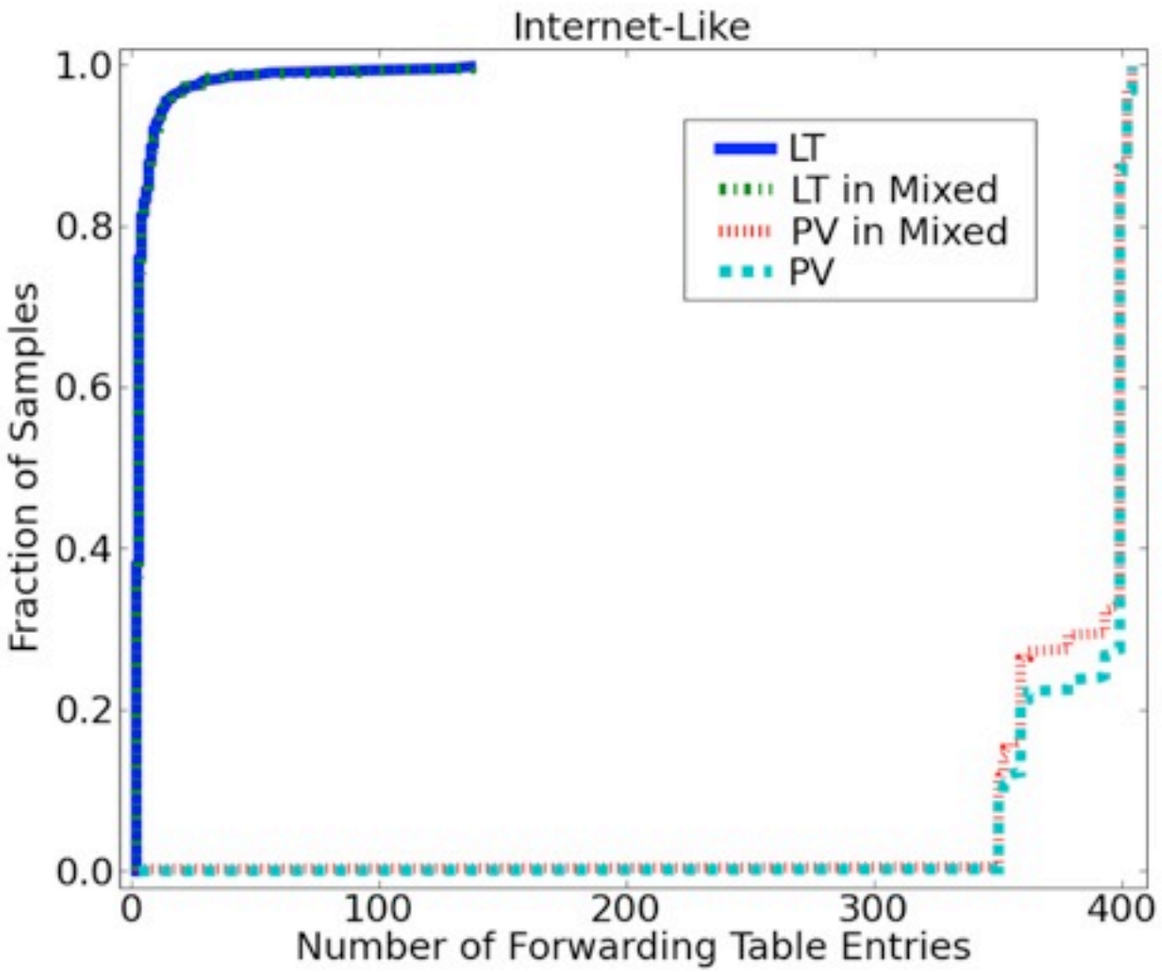
BGP-style

Mixed

LT policies

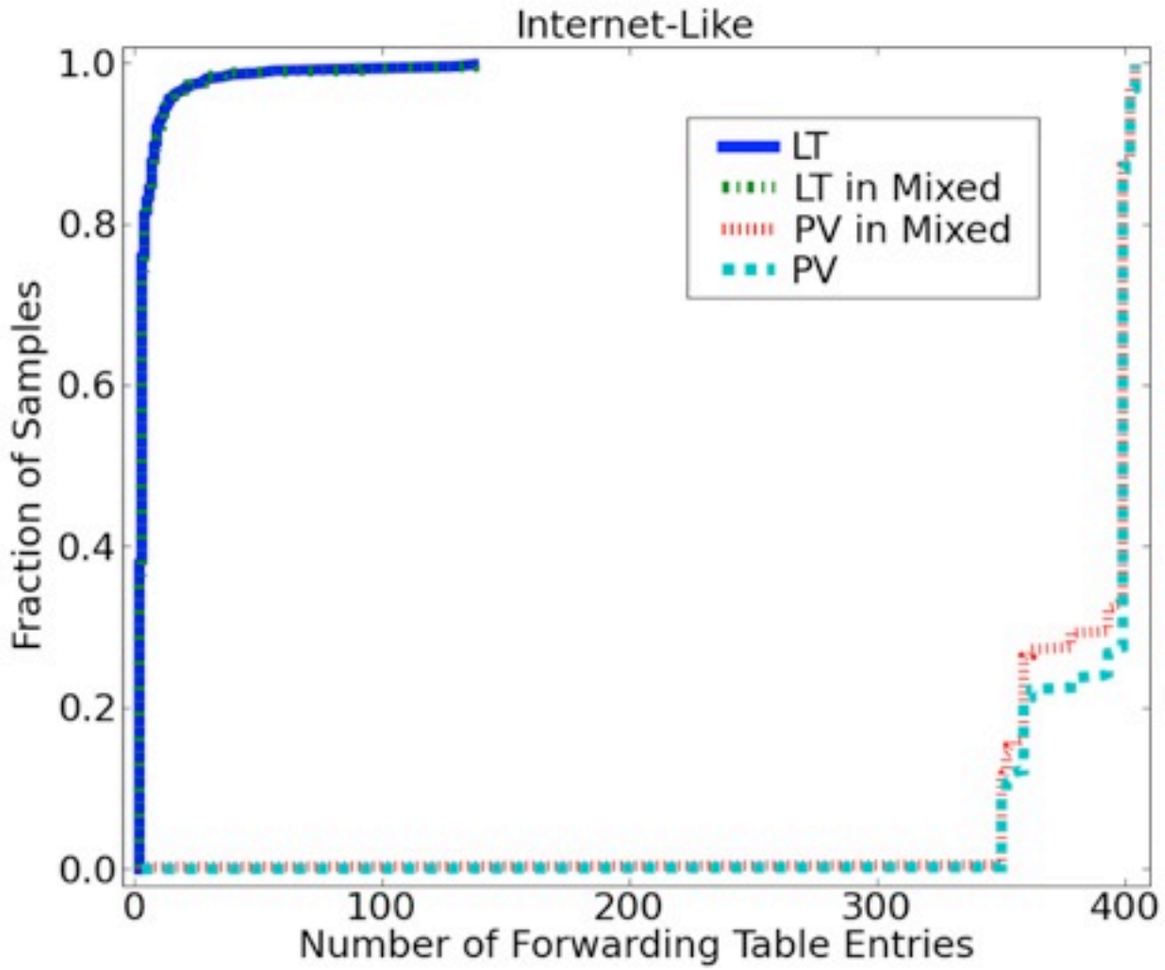
Tiny forwarding tables

Forwarding table size CDF



Tiny forwarding tables

Forwarding table size CDF

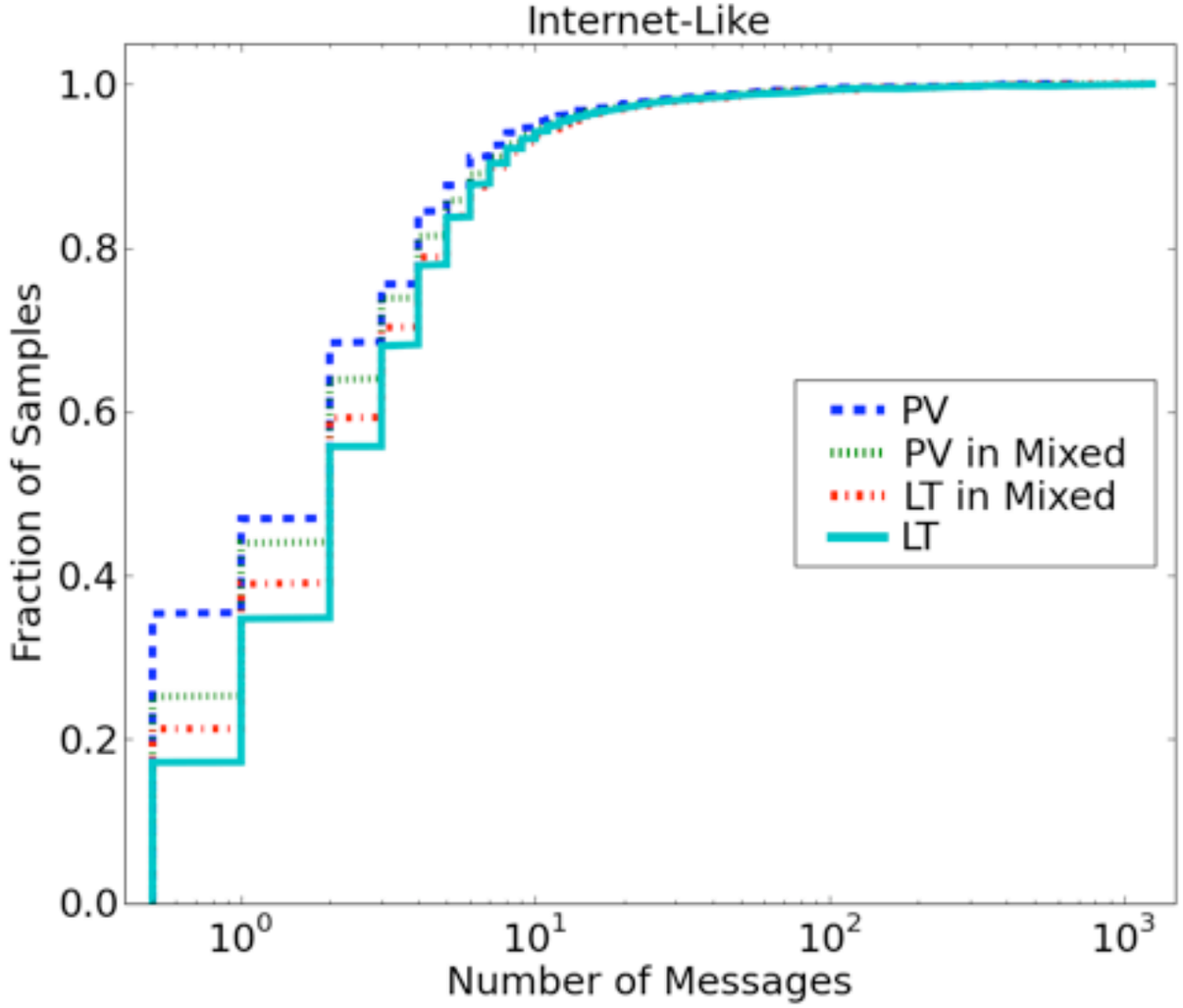


current Internet (CAIDA/APNIC):

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.

Outline

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

Comparing protocols

Comparing protocols

Pathlet routing

Feedback-based routing

NIRA

MIRO

Loose source routing

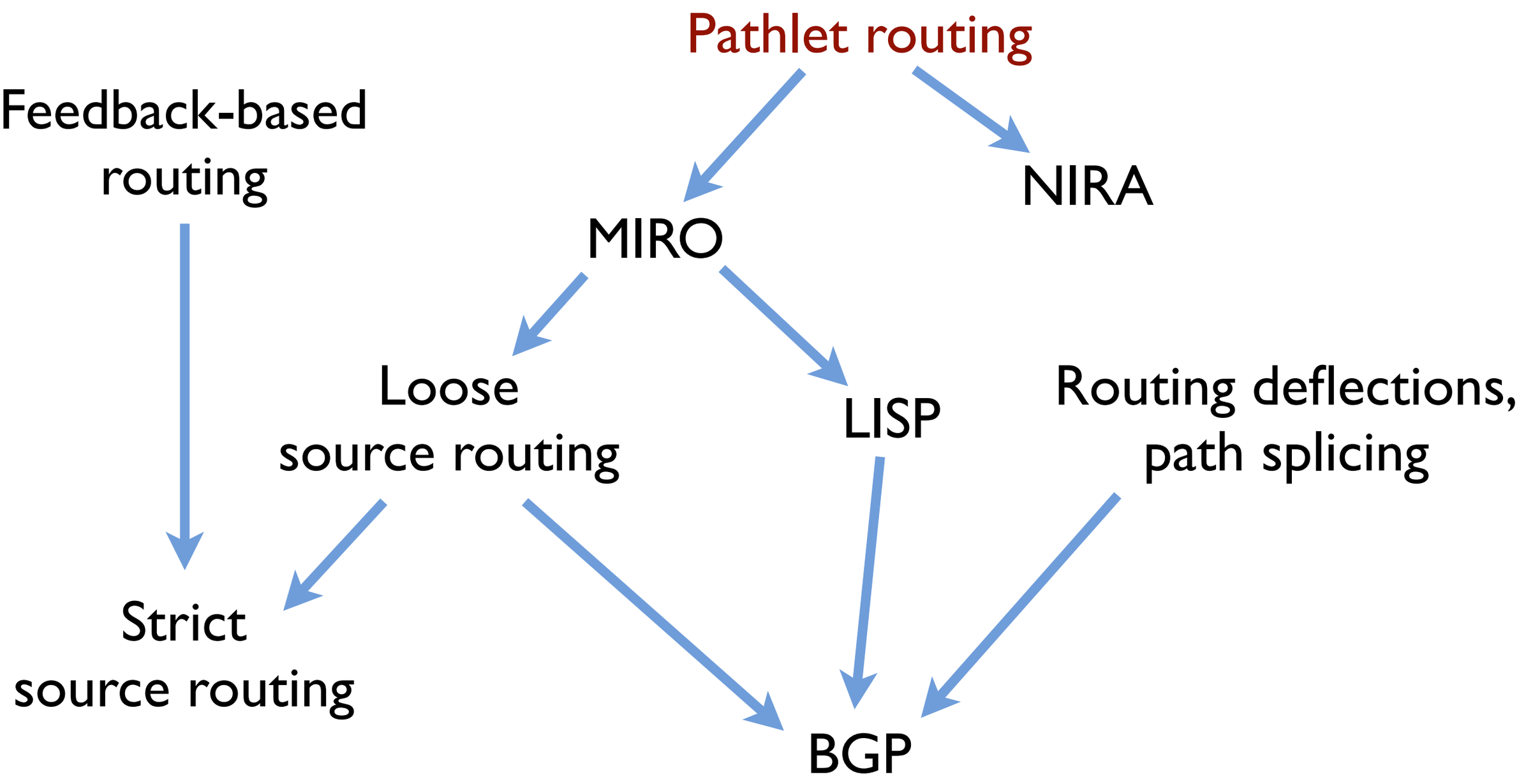
LISP

Routing deflections, path splicing

Strict source routing

BGP

Comparing protocols



Conclusion

- Pathlet routing: source routing over a virtual topology formed by pathlets and vnodes
- Highly flexible; supports both “local” policies with small forwarding tables and many paths, and complex BGP policies
- Challenges for source routing: Incentives to provide multiple paths; selecting paths; security; ...