CS419: Computer Networks

Lecture 5, Part 5: Mar 2, 2005

Internet Routing:

AS’s, igp, and BGP

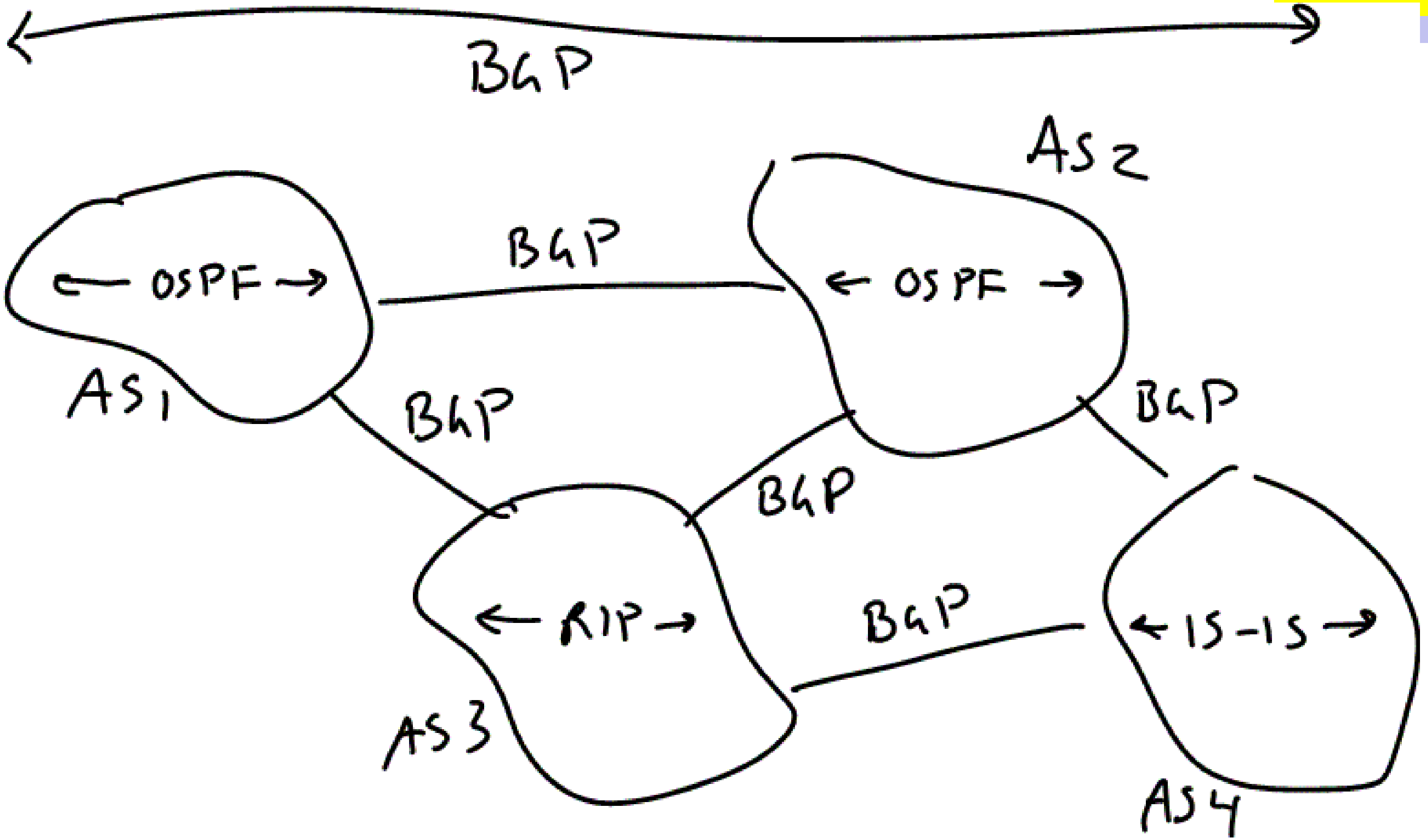
- As we said earlier, the Internet is composed of Autonomous Systems (ASs)
  - Where each AS is a set of routers, links, and hosts
  - And is controlled by a single administration (autonomous)
  - An ISP, a large enterprise (Cornell), etc.

AS’s, igp, and BGP

- Internally, each AS can run any routing protocol it wants
  - “interior gateway protocol”, or igp
  - Examples: RIP, OSPF, IS-IS
- ASs run BGP between them
  - Border Gateway Protocol
- Though many stub ASs don’t run BGP, but simply default to their ISP
  - The ISP runs BGP “on their behalf”

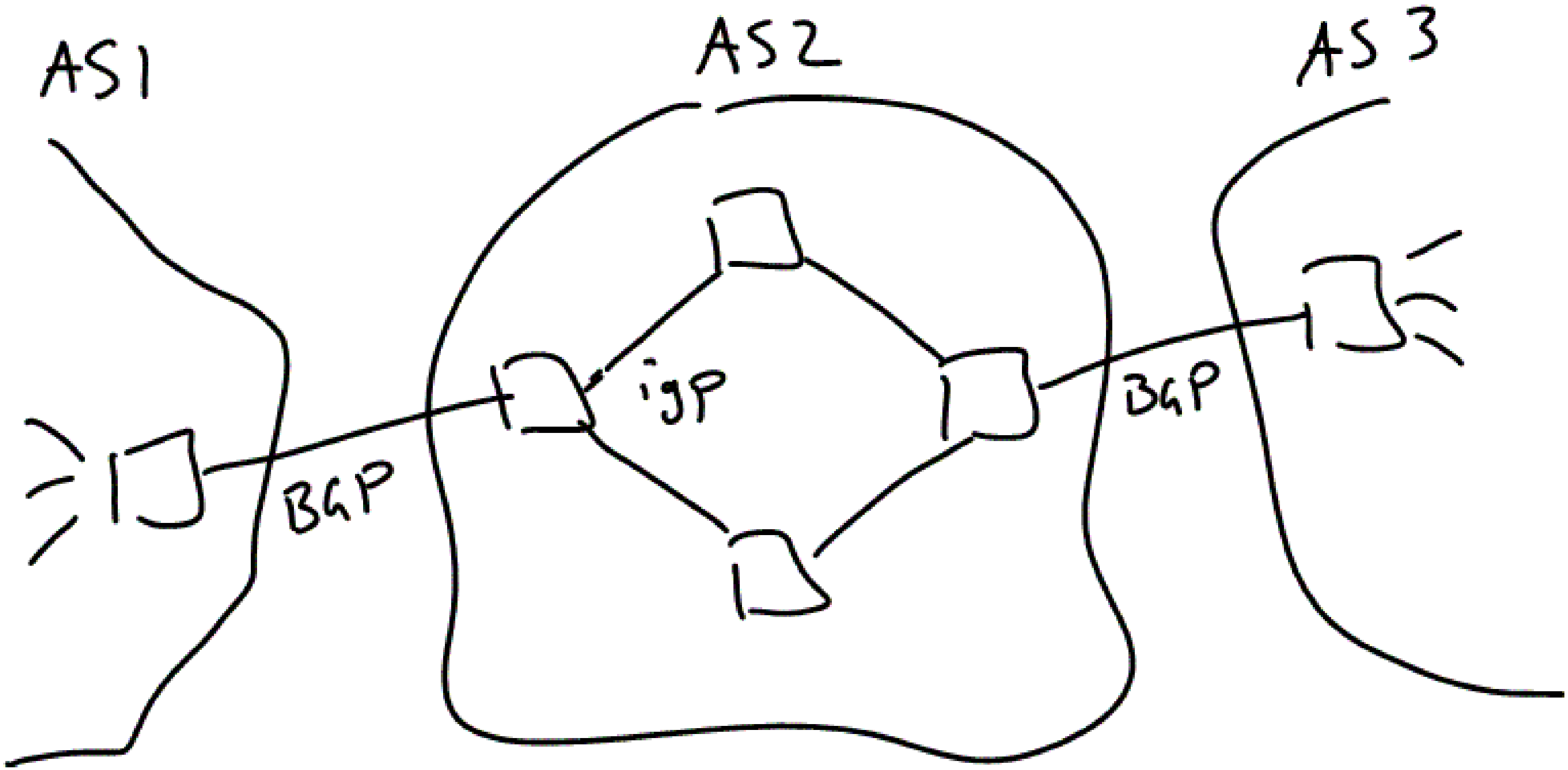
AS's, igp, and BGP

CS419



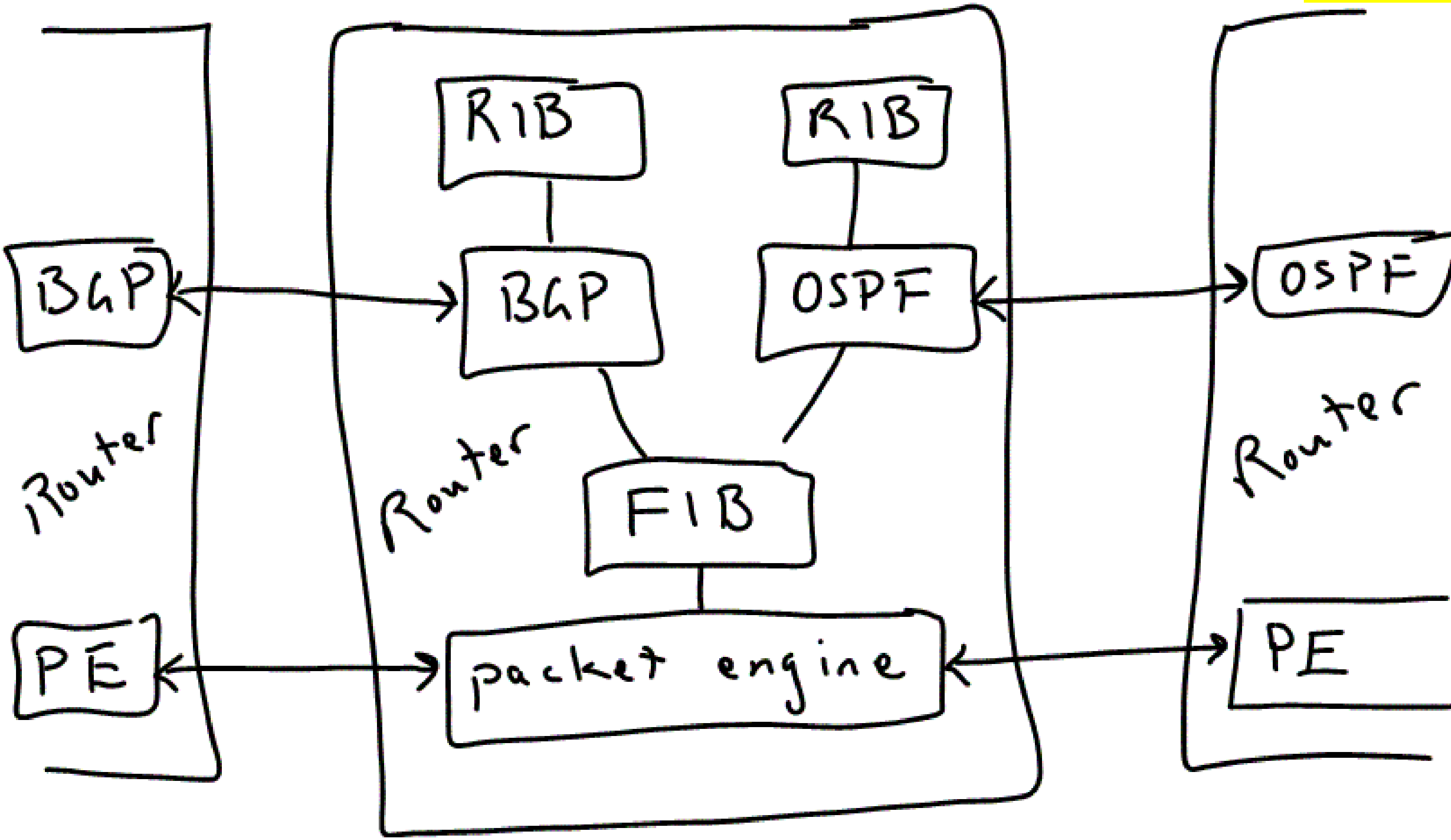
Border routers run both BGP and the igp

CS419



Two routing protocols, one FIB?

CS419

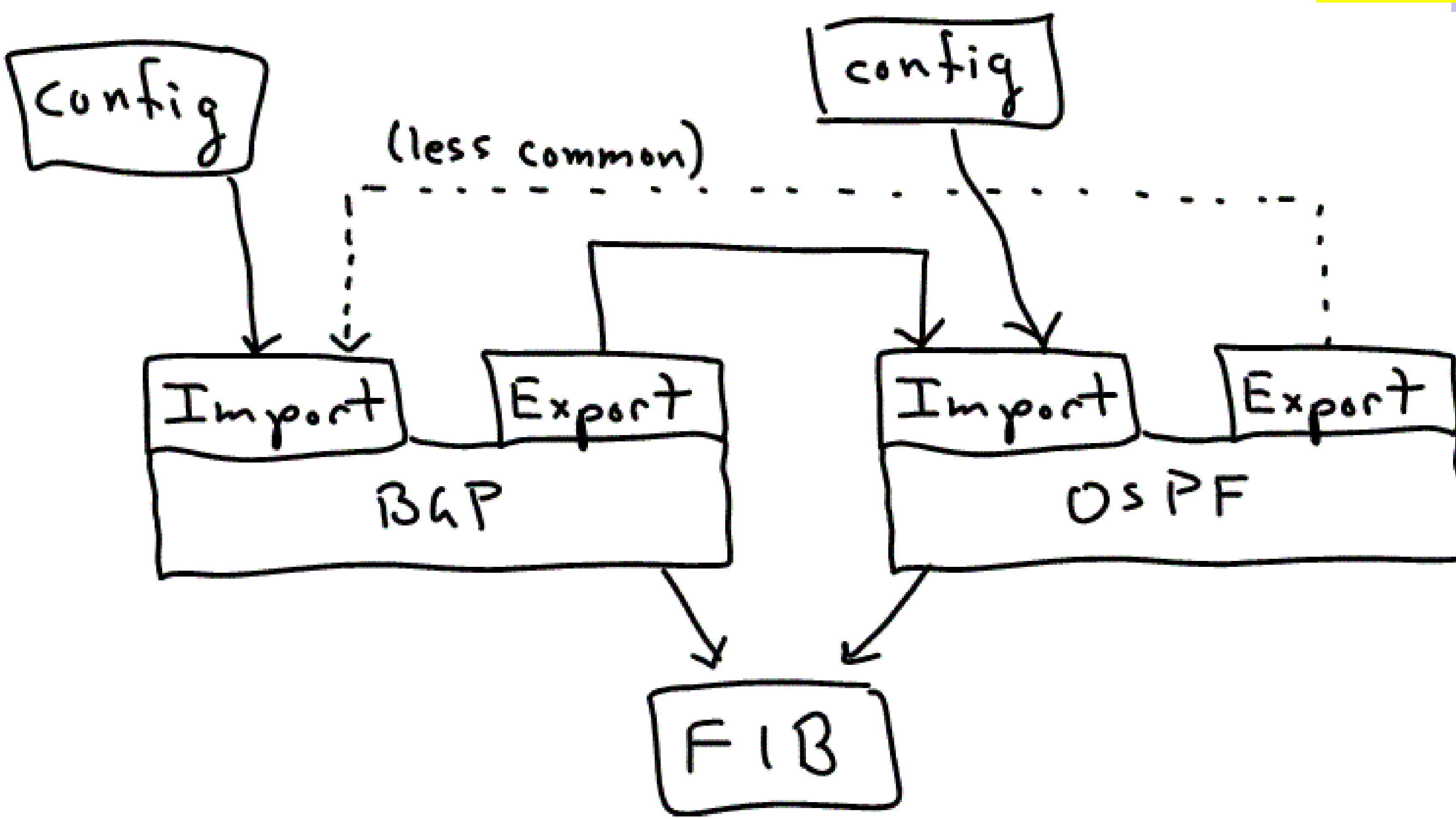


Importing and exporting routes

- Routing algorithms have to "originate" routes
  - Which means that the routing algorithm has to "import" the route in some way other than getting it from a neighbor router
- Two ways:
  - From configuration (iface1 has prefix P . . .)
  - From another routing algorithm!
- Likewise, routing algorithms can "export" routes

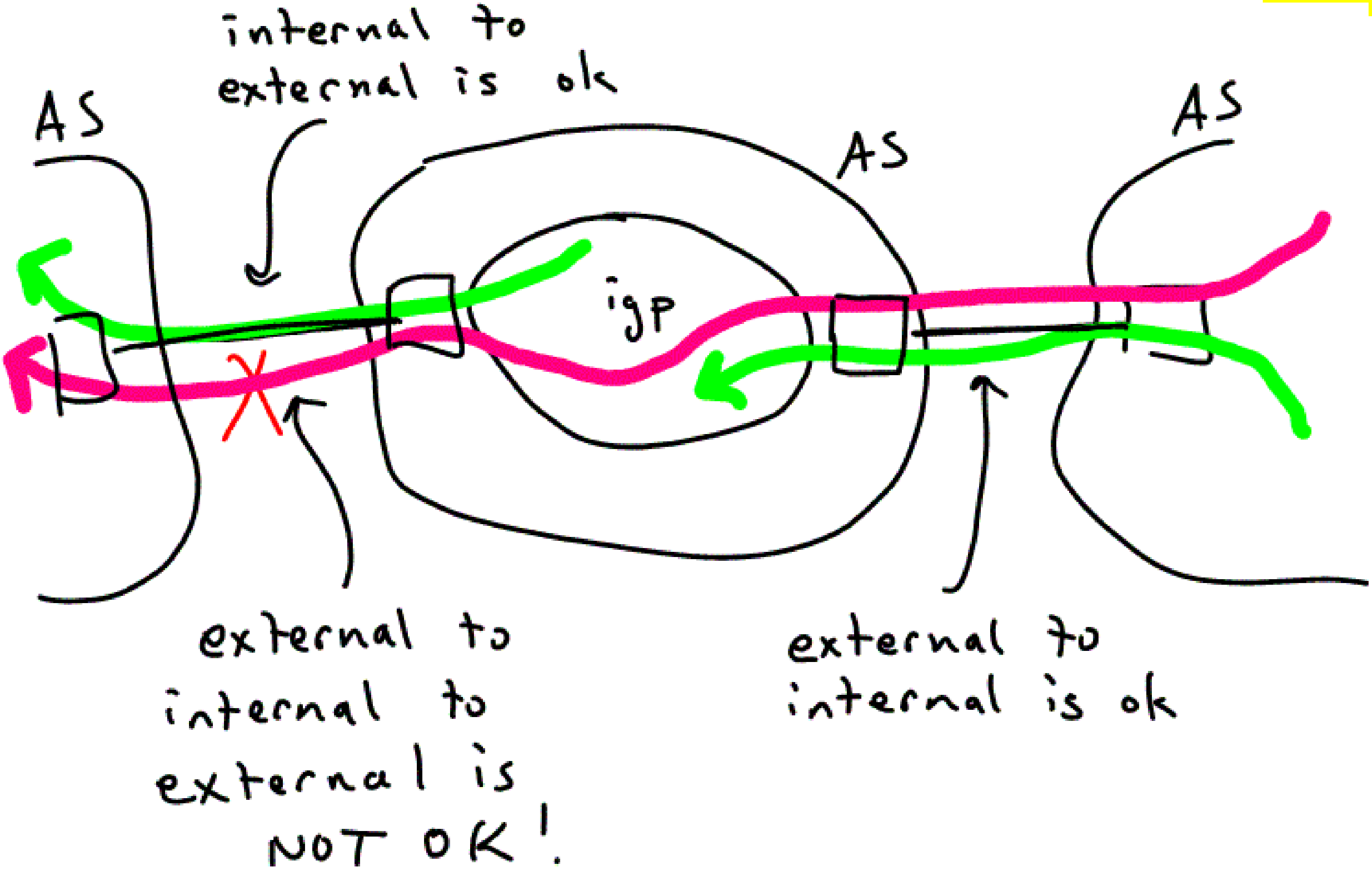
Importing and exporting routes

CS419



Limits to importing and exporting

CS419



Why this limitation?

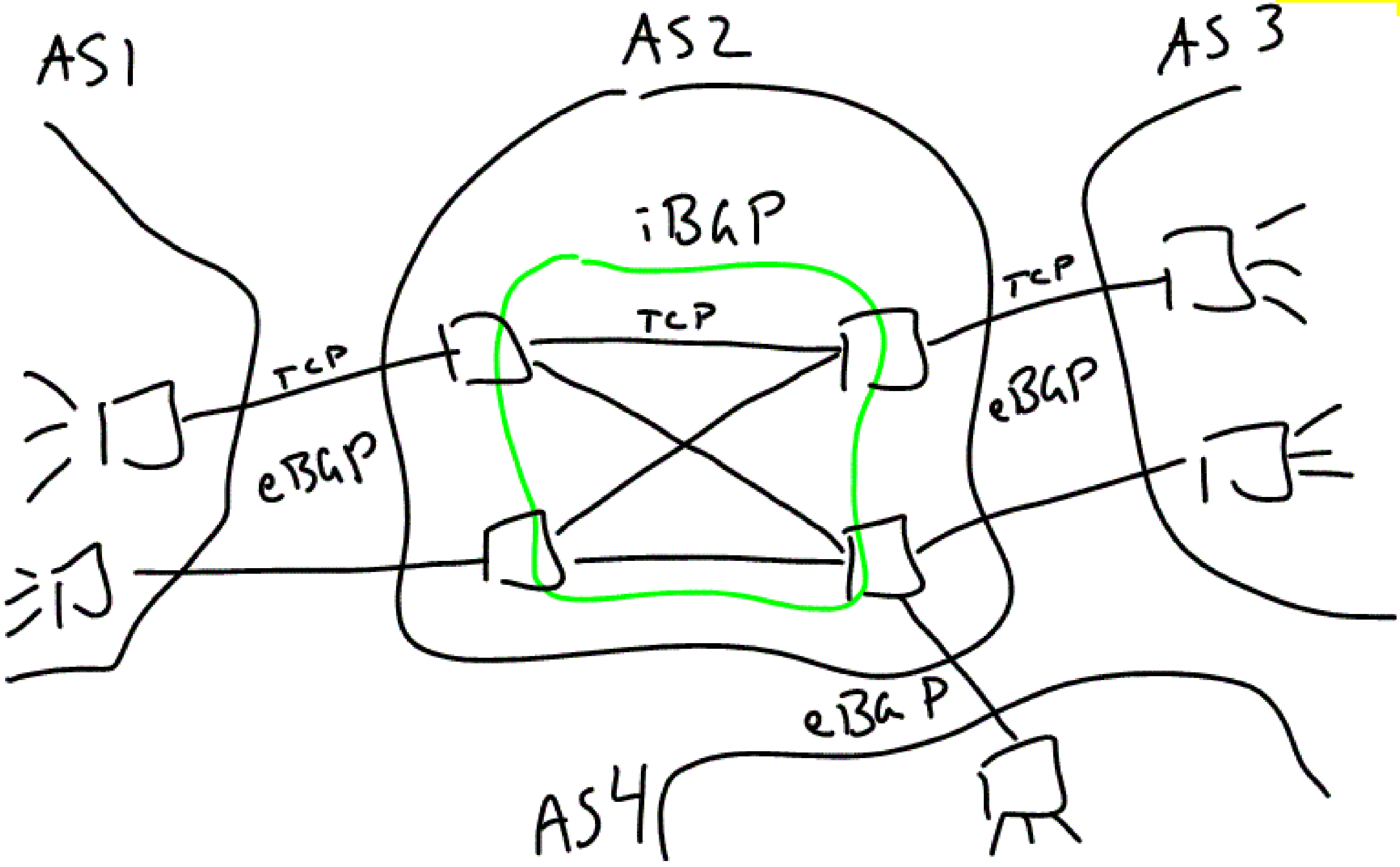
- Semantic mismatch between BGP and igp
  - BGP is path-vector, and requires the AS-path to the destination prefix
  - igp's don't require this AS-path, and can't be expected or forced to carry it
  - Want to maintain independence between igp and BGP
- Also, igp convergence may be slow…

iBGP and eBGP (interior and exterior)

- BGP avoids dependence on igp by running both between ASs (exterior, eBGP) and within ASs (interior, iBGP)
- iBGP runs over TCP

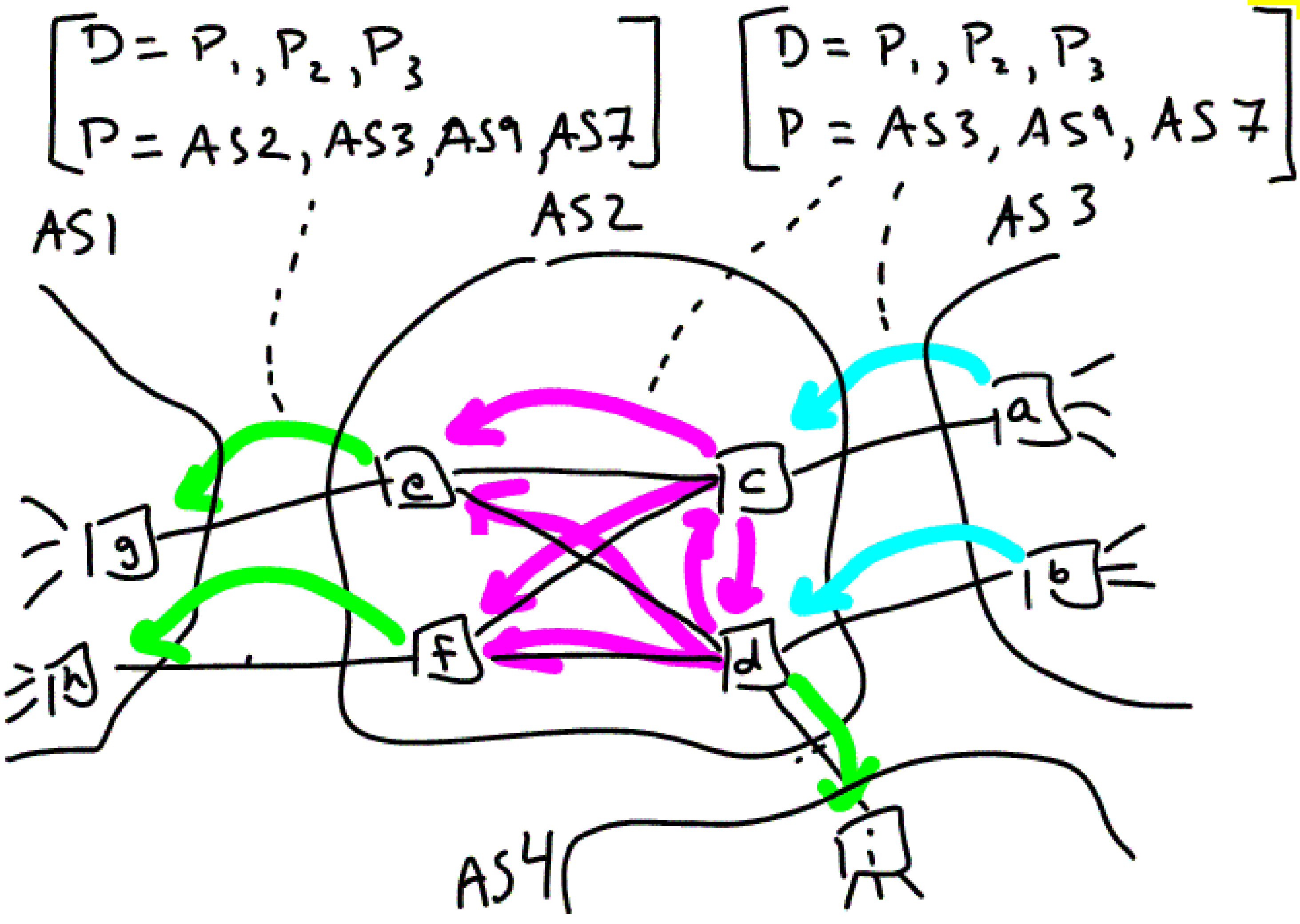
iBGP and eBGP (interior and exterior)

CS419



iBGP and eBGP (interior and exterior)

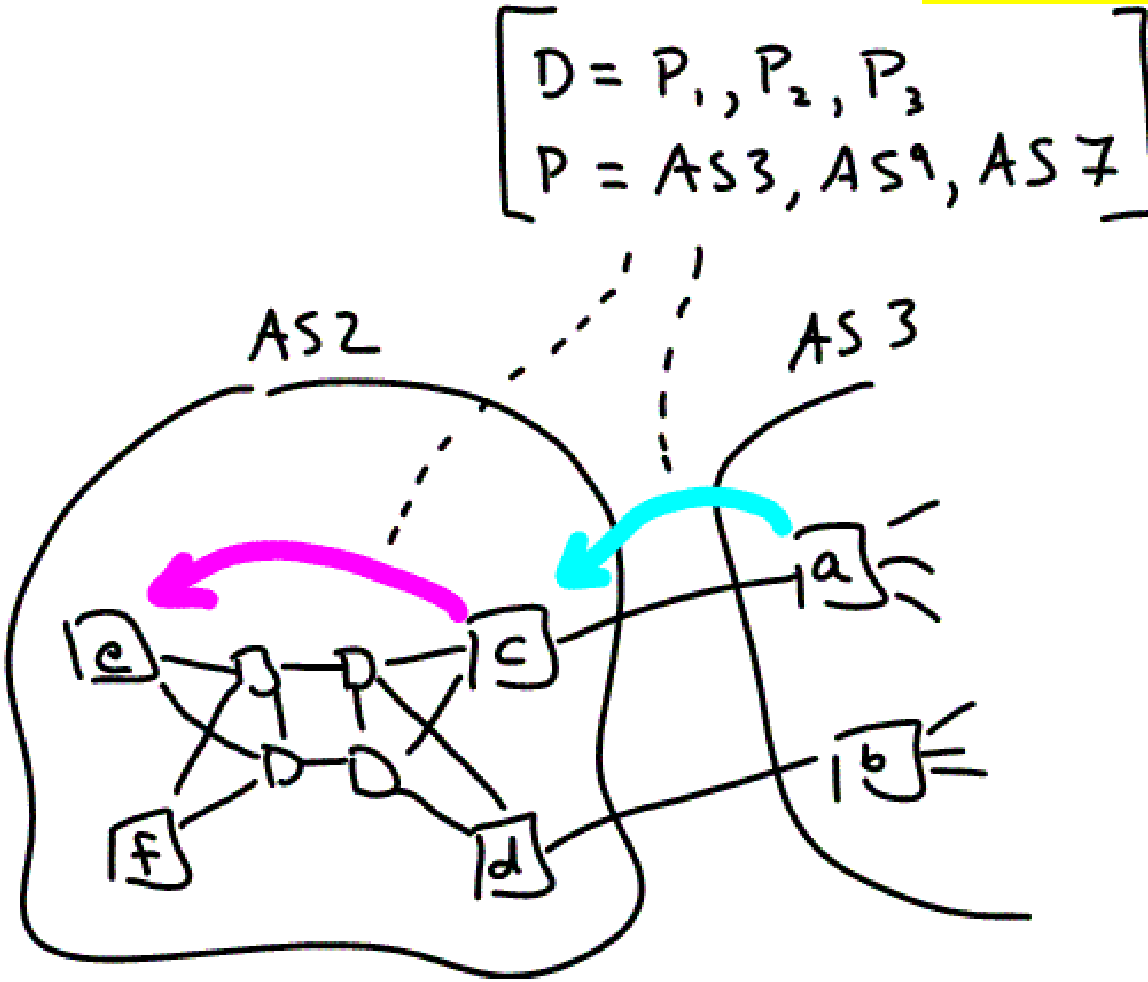
CS419



Next hops are weird in iBGP

CS419

What does it mean for e to have c as its next hop to P1, when there are multiple routers between e and c???



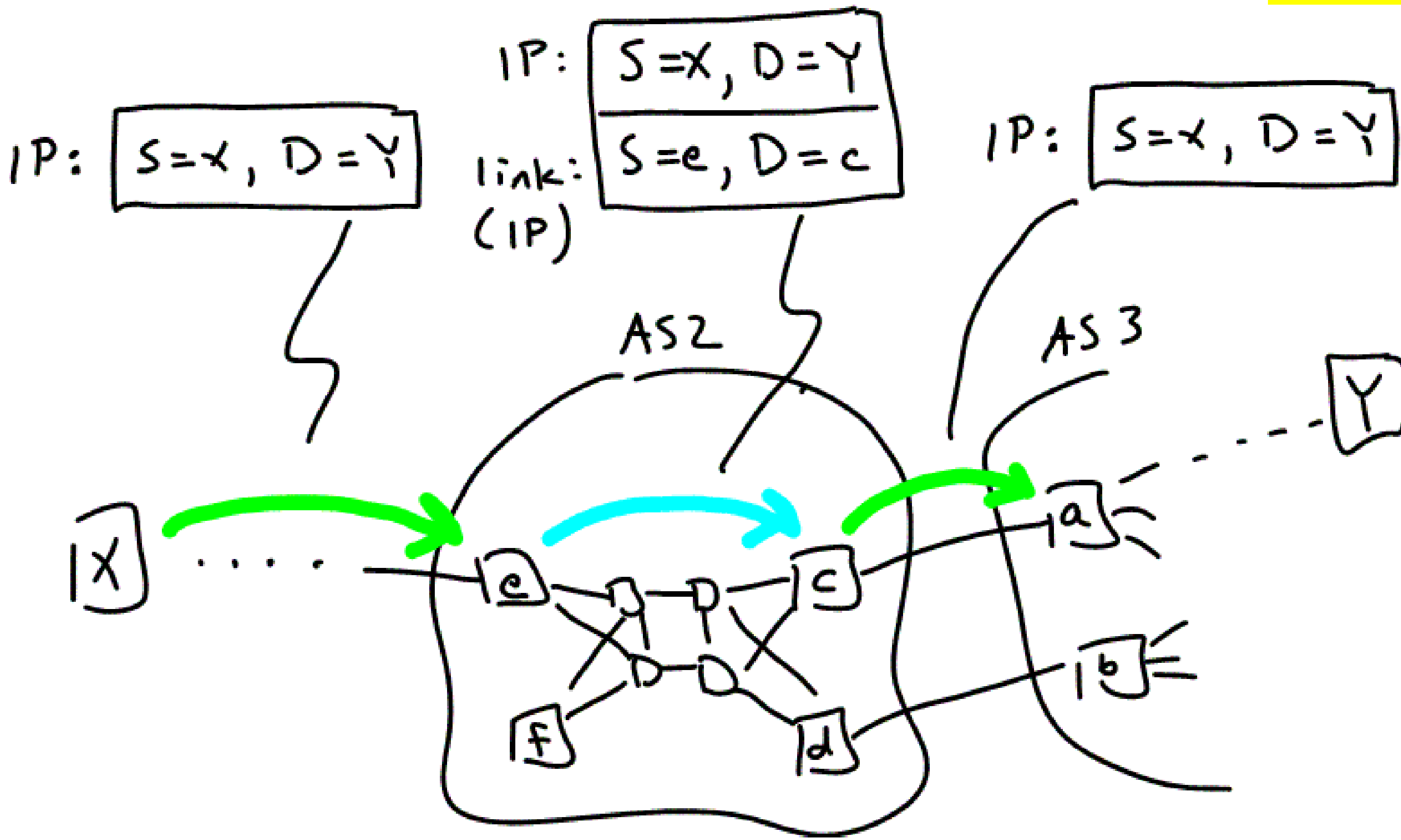
One option: tunnel across AS

CS419

- iBGP speakers form IP tunnels across the AS
  - IP over IP
    - (perhaps with GRE between them, but lets not get into this now)
  - This creates a “link” between the two iBGP speakers
  - Remember, IP doesn’t care what subnet technology it runs over, even if that subnet is IP!!!

iBGP next hop using IP-in-IP tunneling

CS419

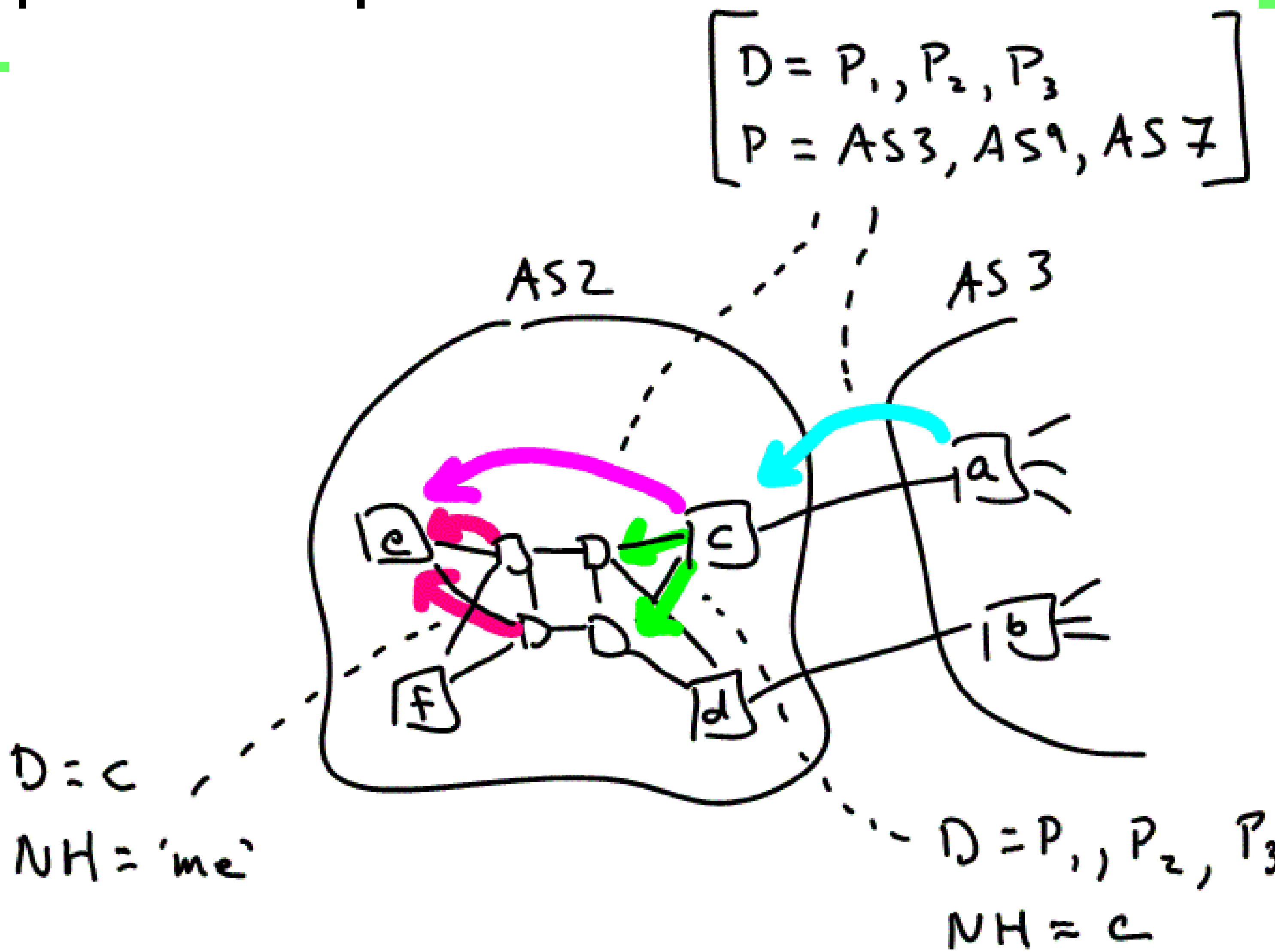


Another option, use both BGP and igp RIBs

CS419

- iBGP “resolves” the iBGP next hop to its igp next hop
  - iBGP computes its next hop
  - iBGP looks into igp RIB to determine igp next hop to iBGP next hop
  - This becomes the actual next hop
- iBGP must advertise external prefixes into the igp

iBGP using igp RIB to resolve next hop



BGP security model

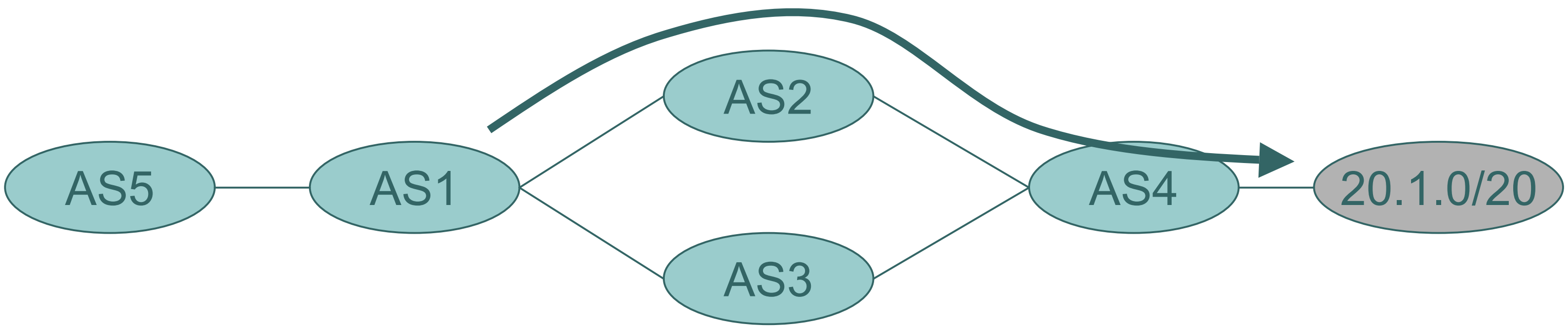
- Authentication is hop-by-hop, like OSPF
- But threat is much worse, because no single organization controls all of BGP
- So, BGP uses policy to help prevent bogus routes
  - BGP routers have an expectation of what they should hear from where

BGP policies

- Who to peer with (which ASs)
- What routes to originate
- What routes to import (prevent bogus advertisements)
- What routes to export (and how to aggregate them)
- What paths to prefer
  - Shorter AS paths
  - Some ASs preferred over others
    - The big ASs (UUnet, AT&T, etc.)
    - Primary versus backup transit AS

BGP policy limitation (hop by hop policy decisions)

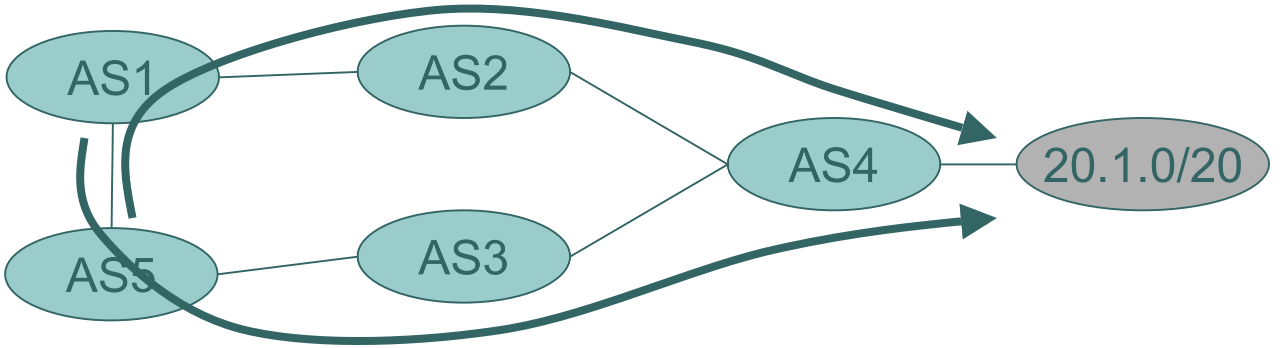
CS419



AS1 chooses AS2 as the path to 20.1.0/20.
AS5 is forced to accept the choice of AS1
(If AS5 <u>really</u> doesn't like it, it should find a new peer)

BGP policy conflict

CS419

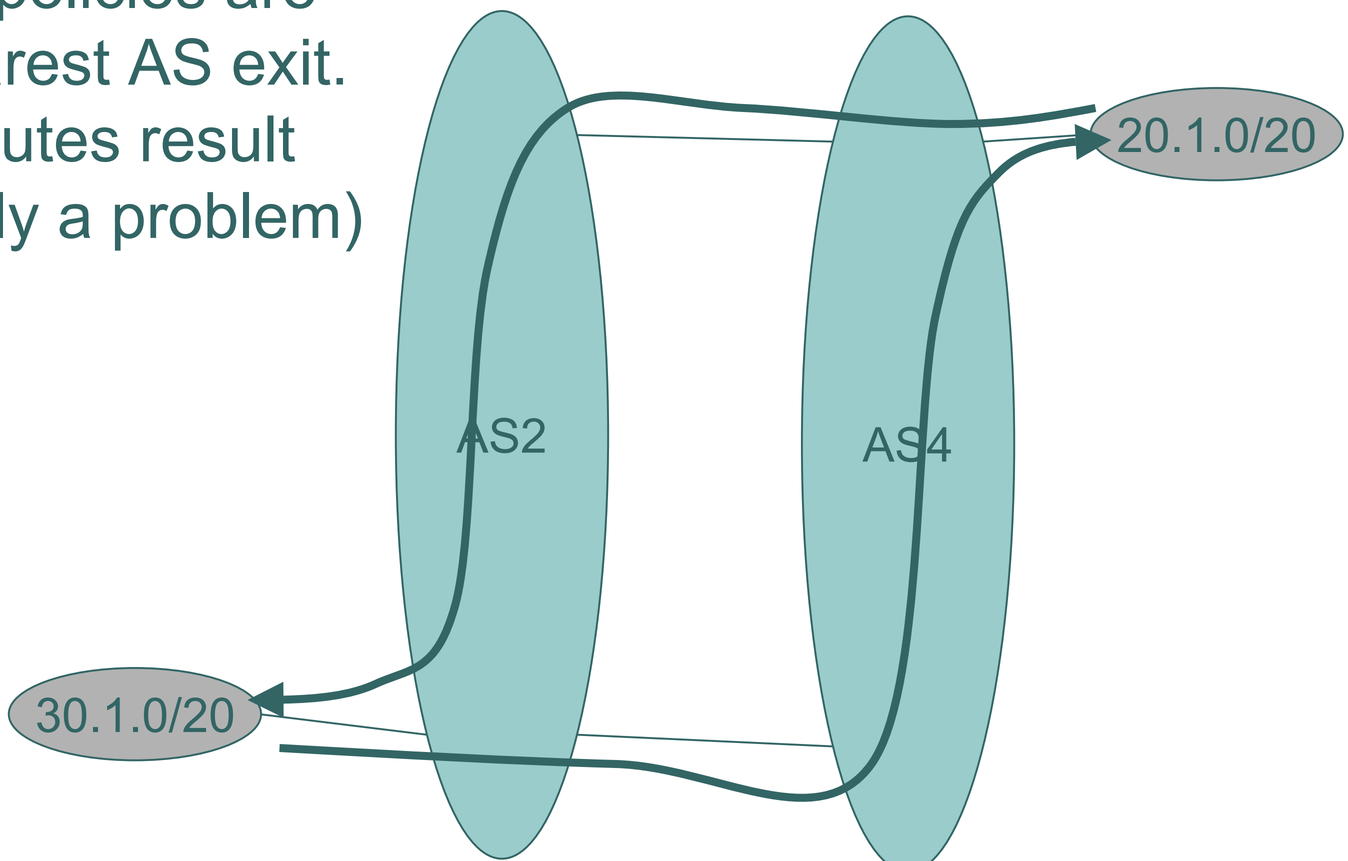


AS5 policy is to prefer route to AS4 via AS2
AS1 policy is to prefer route to AS4 via AS3
<u>Both policies cannot be satisfied</u>

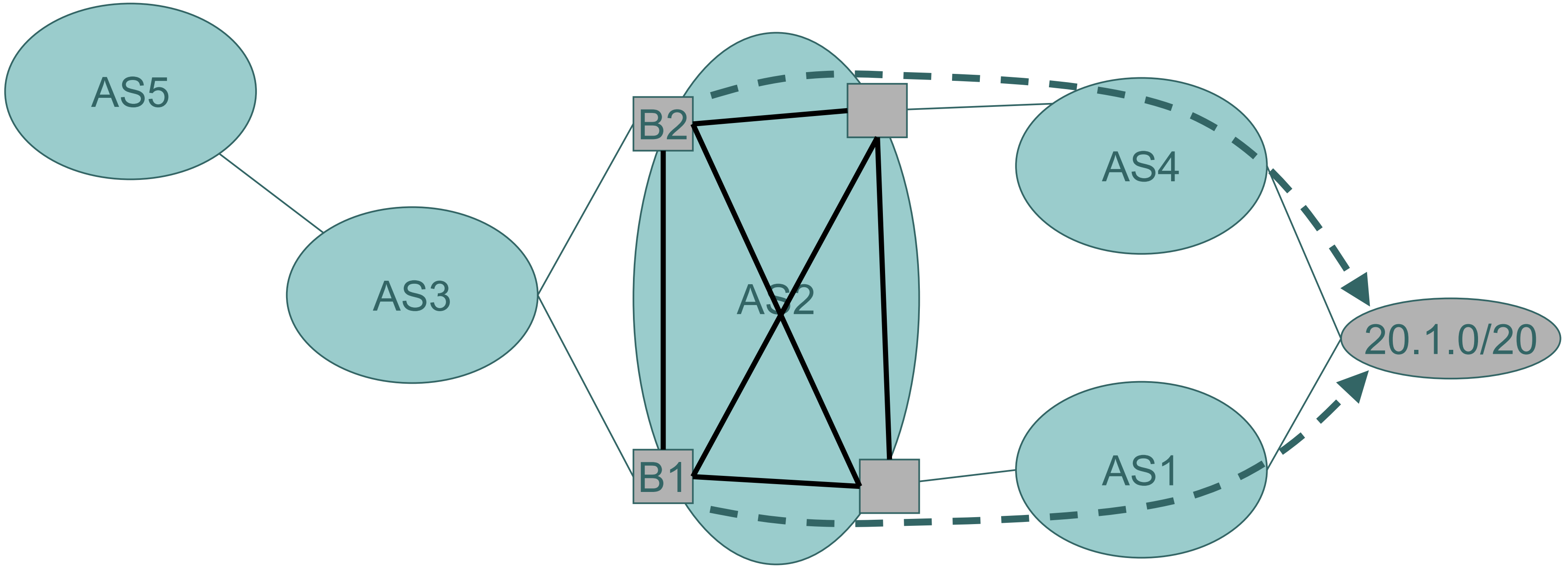
Hot potato routing

CS419

AS2 and AS4 policies are to route to nearest AS exit. Asymmetric routes result (not necessarily a problem)



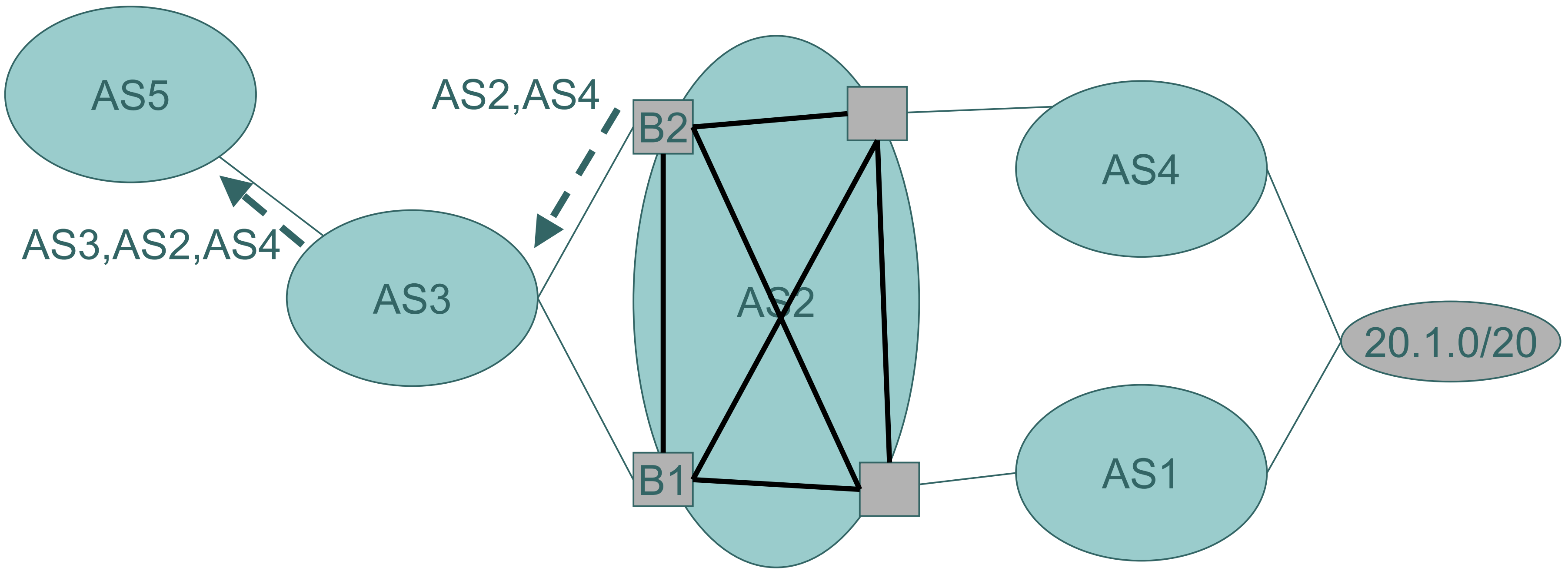
Misconfigured policies may lead to oscillation



B2 configured to prefer AS4
B1 configured to prefer AS1

Misconfigured policies may lead to oscillation

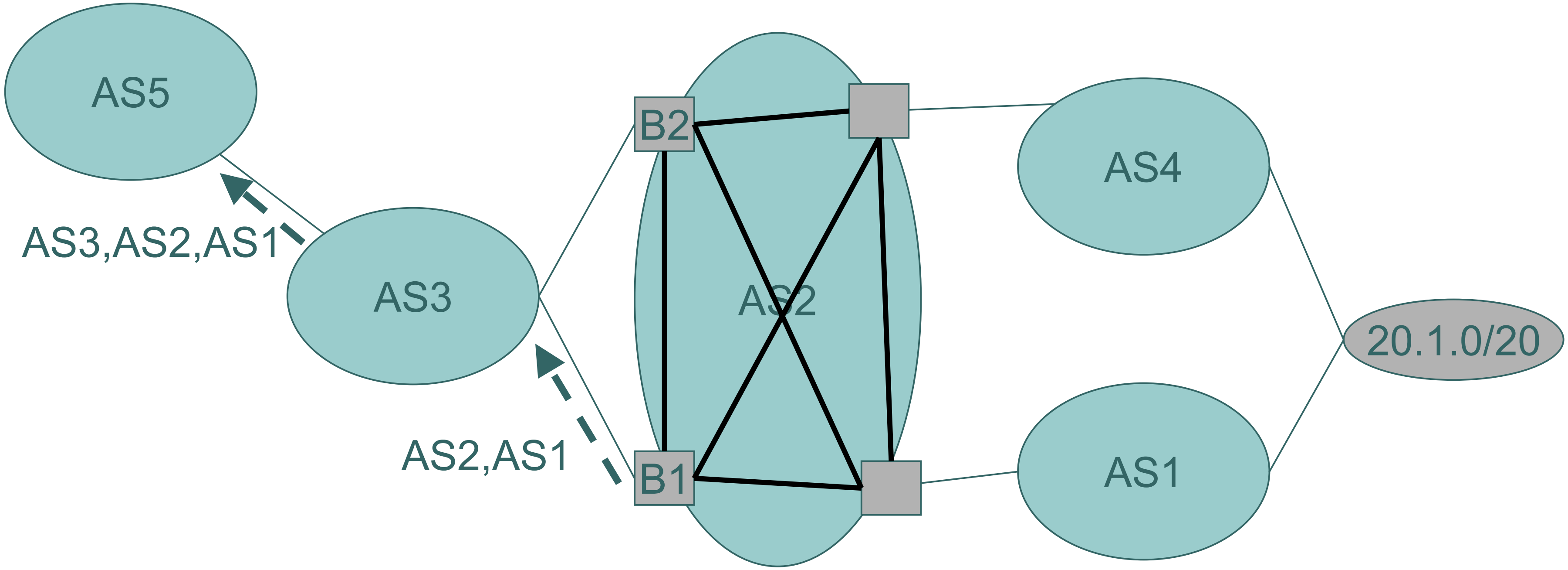
CS419



B2 (periodically) updates AS3 with path AS2,AS4

Misconfigured policies may lead to oscillation

CS419



B1 (periodically) updates AS3 with path AS2,AS1
With each period AS3 advertises a different route

Other route flapping

CS419

- A link continuously goes up and down
  - The update for this is propagated throughout the internet
- Mid-90's these kinds of problems were severe
  - 1996: 45,000 prefixes, 1,500 unique AS paths, 1,300 ASs, 3-6 million BGP update messages/day
    - 6 updates per prefix per hour!
    - (Labovitz et. al.)

Today much improved

CS419

- Better policy tools
- Better software
- Lots of damping
- But still, advances in BGP lead to new policy bugs
  - Route reflectors published in 2000 (RFC2796)
  - Inconsistent route reflectors problem published in 2002 (RFC3345)

Policy Tools

- Routing Policy Specification Language (RPSL) (RFC 2280)
  - Earlier policy languages exist
- Language to define BGP policies
  - Peers, import, export, route preference, aggregation
- Posted at Routing Registries (RIPE, RADB, etc.)
- Tools created to look for policy inconsistencies (within AS and across ASs)
- Tools created to match measured reality (BGP tables, traceroute) with policy expectations
  - RAToolSet, USC/ISI

Lots of Damping

- Stop advertising certain prefixes if they go up and down a lot
  - Improve stability
  - Lower overhead
- RIPE guidelines:
  - Don't dampen until after 4th flap in a row (in 50 minutes)
  - /24: dampen 60 minutes
  - /22,/23, dampen 30-45 minutes
  - </22, dampen 10-30 minutes

Lots of Damping

- Helps the internet, but means that you can go away for a long time
  - Because of some problem in the middle!
- Most damping is done on routes that you don’t care about
  - Poorly managed small ISPs
- Routes through major ISPs tend to be very stable
  - Your favorite web sites

Effect of BGP policies on path quality

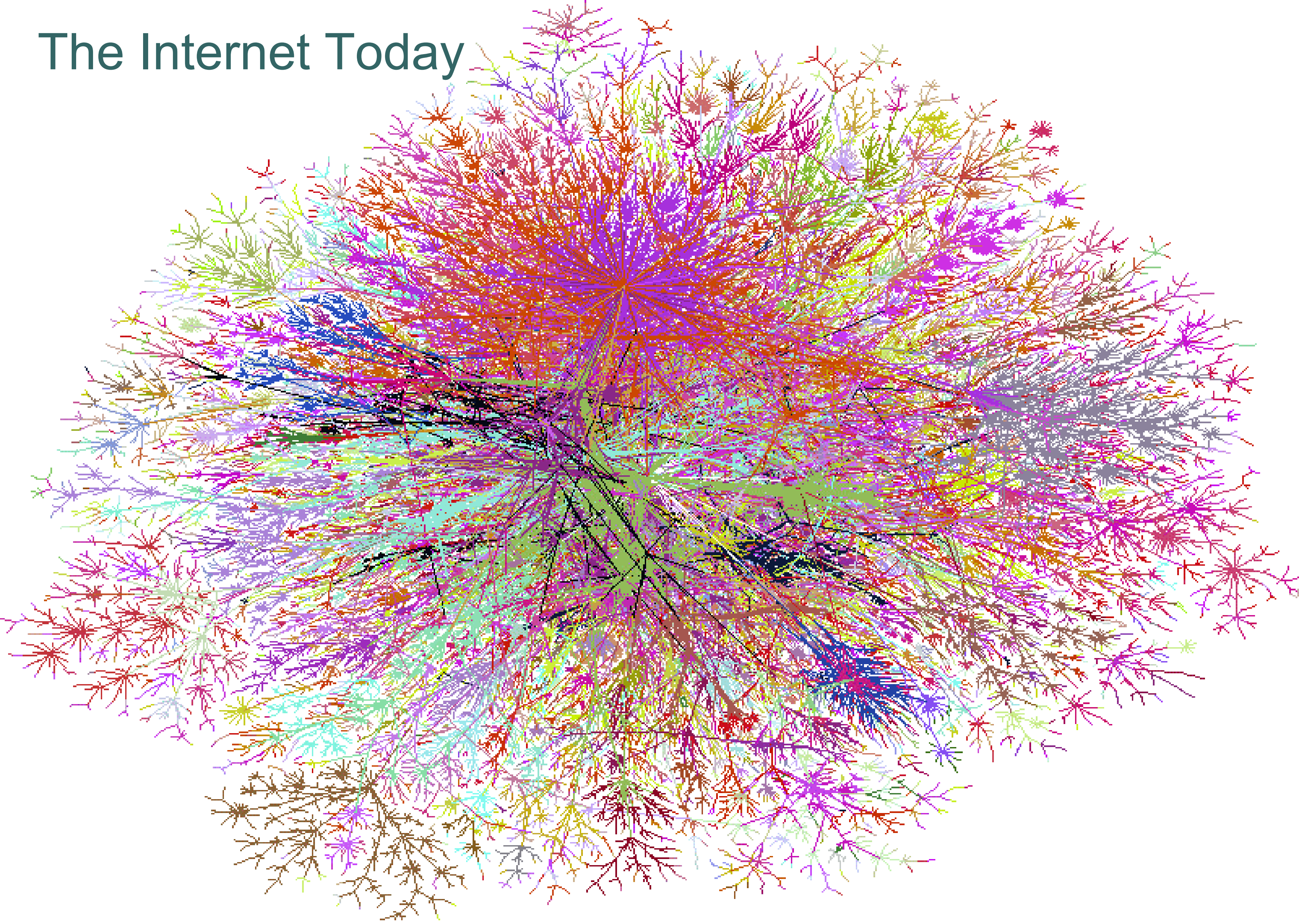
- Ramesh Govindan study (USC)
- Methodology:
  - Learn real physical topology with traceroutes, deduce actual AS connectivity
    - Imperfect, but not bad
  - Examine used “policy topology” from BGP tables, RADB (routing registry) database
  - Compare the two

Effect of BGP policies on path quality

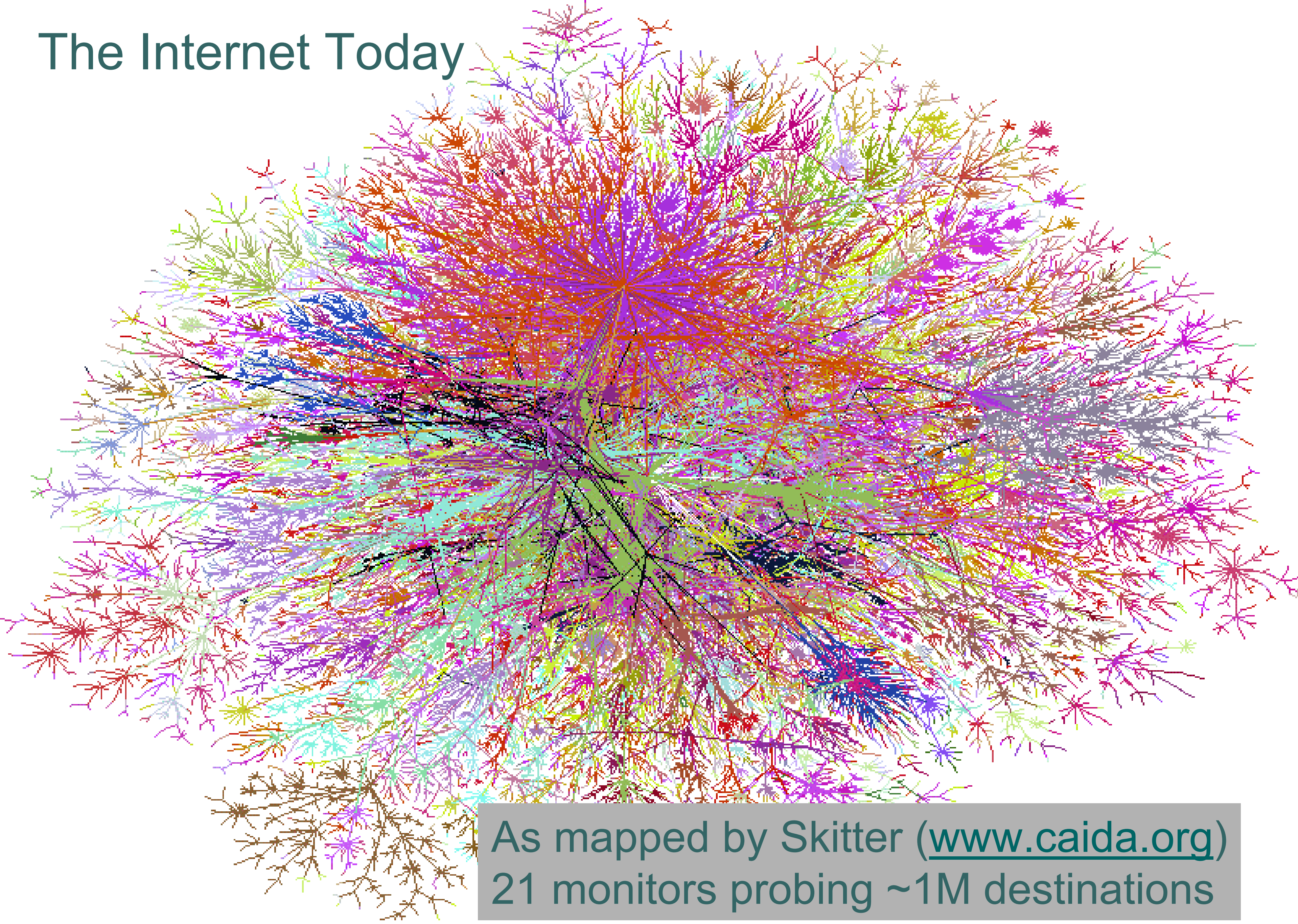
CS419

- Results:
  - About ½ of the paths a longer than shortest path
  - 20% of policy paths are 50% or more longer
  - 20% of policy paths are 5 hops or more longer
  - Policy tends to push paths through major backbones rather than possibly shorter routes
    - (But shorter routes may not be better routes!)

The Internet Today



The Internet Today



As mapped by Skitter (www.caida.org)
21 monitors probing ~1M destinations

BGP Routing Table Growth

CS419