Equal-Cost Multi-Path Routing

CSE 461

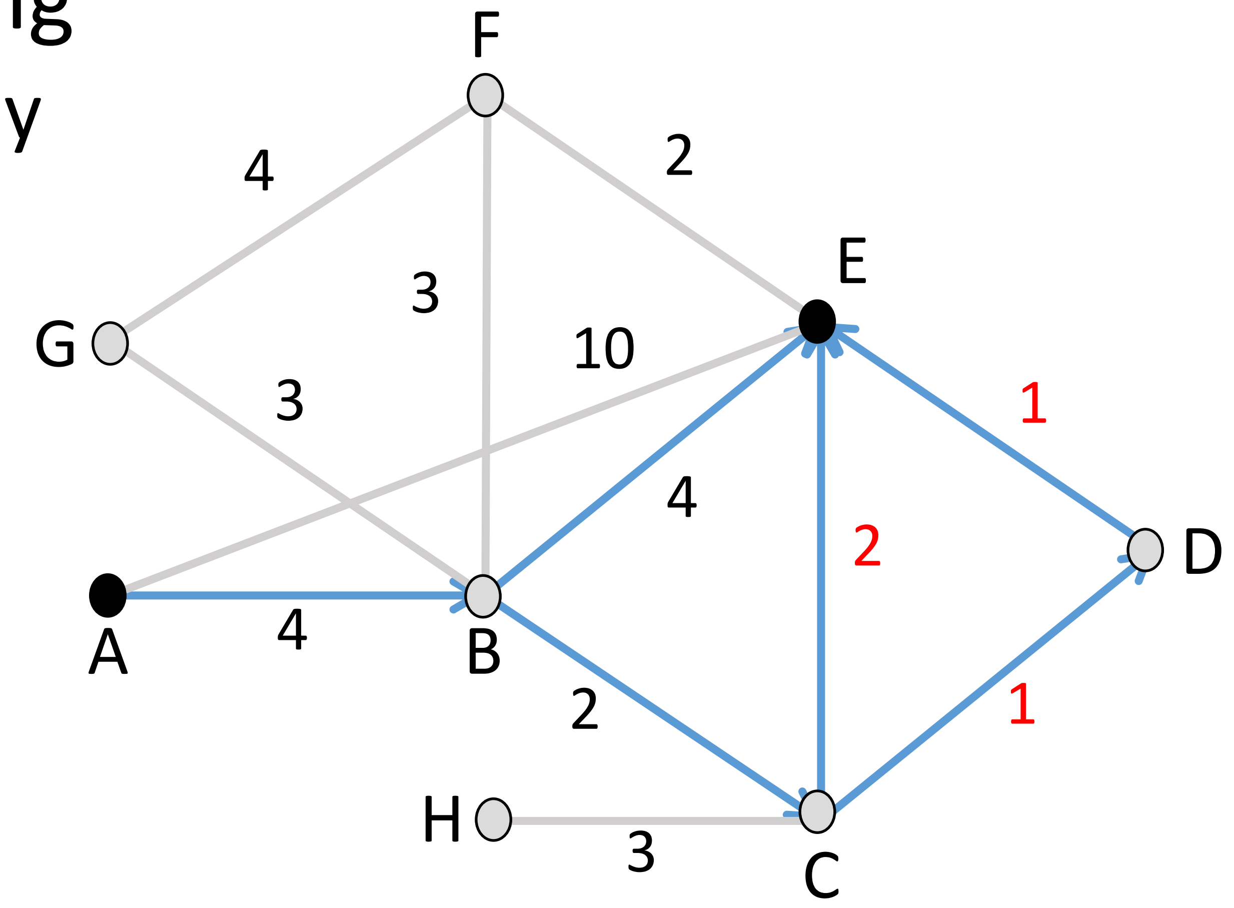
Ratul Mahajan

Multipath Routing

- Allow multiple routing paths from node to destination be used at once
  - Topology has them for redundancy
  - Using them can improve performance
- Questions:
  - How do we find multiple paths?
  - How do we send traffic along them?

Equal-Cost Multipath Routes

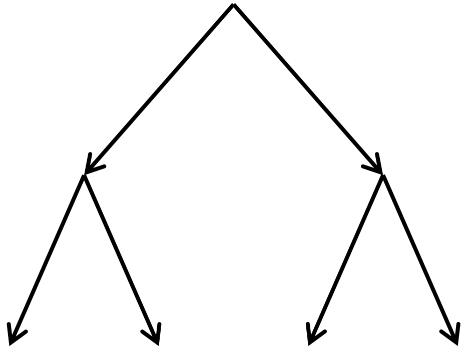
- One form of multipath routing
  - Extends shortest path model by keeping set if there are ties
- Consider A→E
  - ABE = 4 + 4 = 8
  - ABCE = 4 + 2 + 2 = 8
  - ABCDE = 4 + 2 + 1 + 1 = 8
  - Use them all!



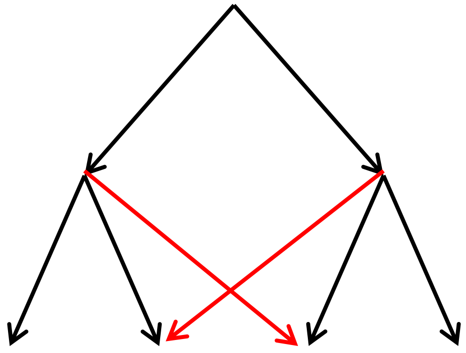
Source “Trees”

- With ECMP, source/sink “tree” is a directed acyclic graph (DAG)
  - Each node has set of next hops
  - Still a compact representation

Tree

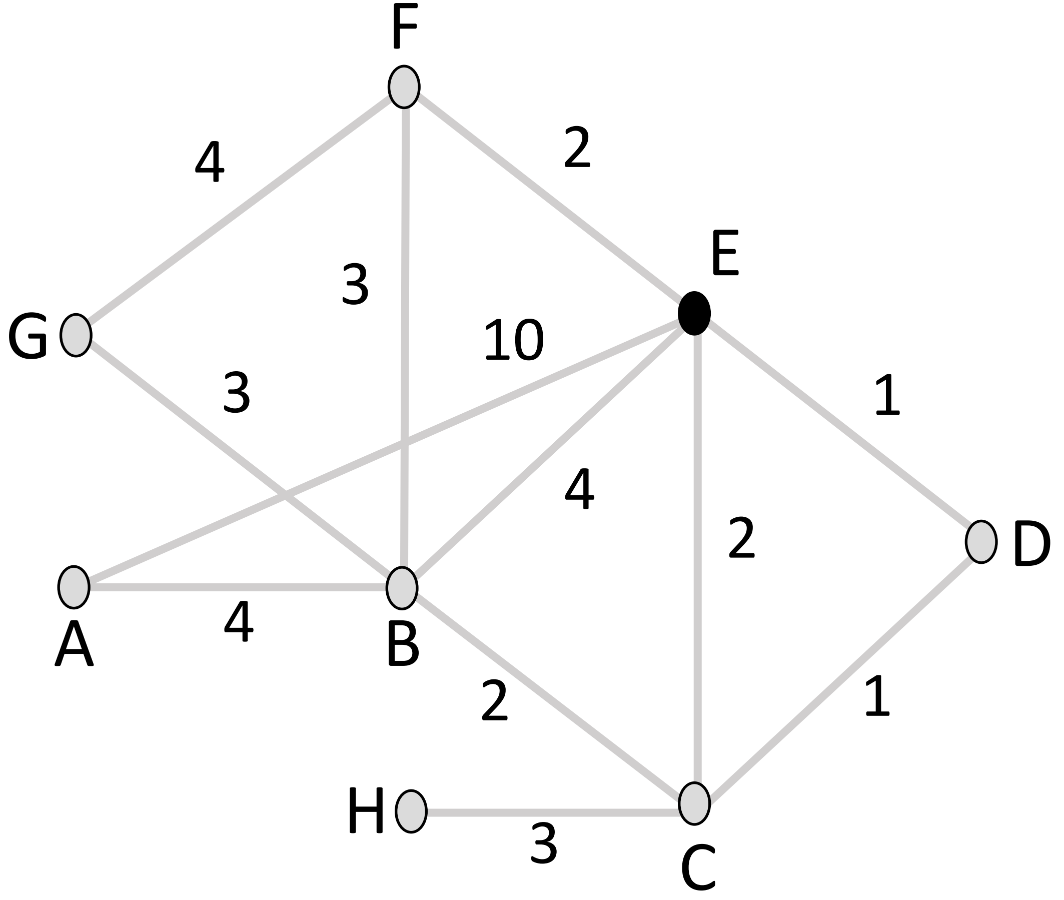


DAG

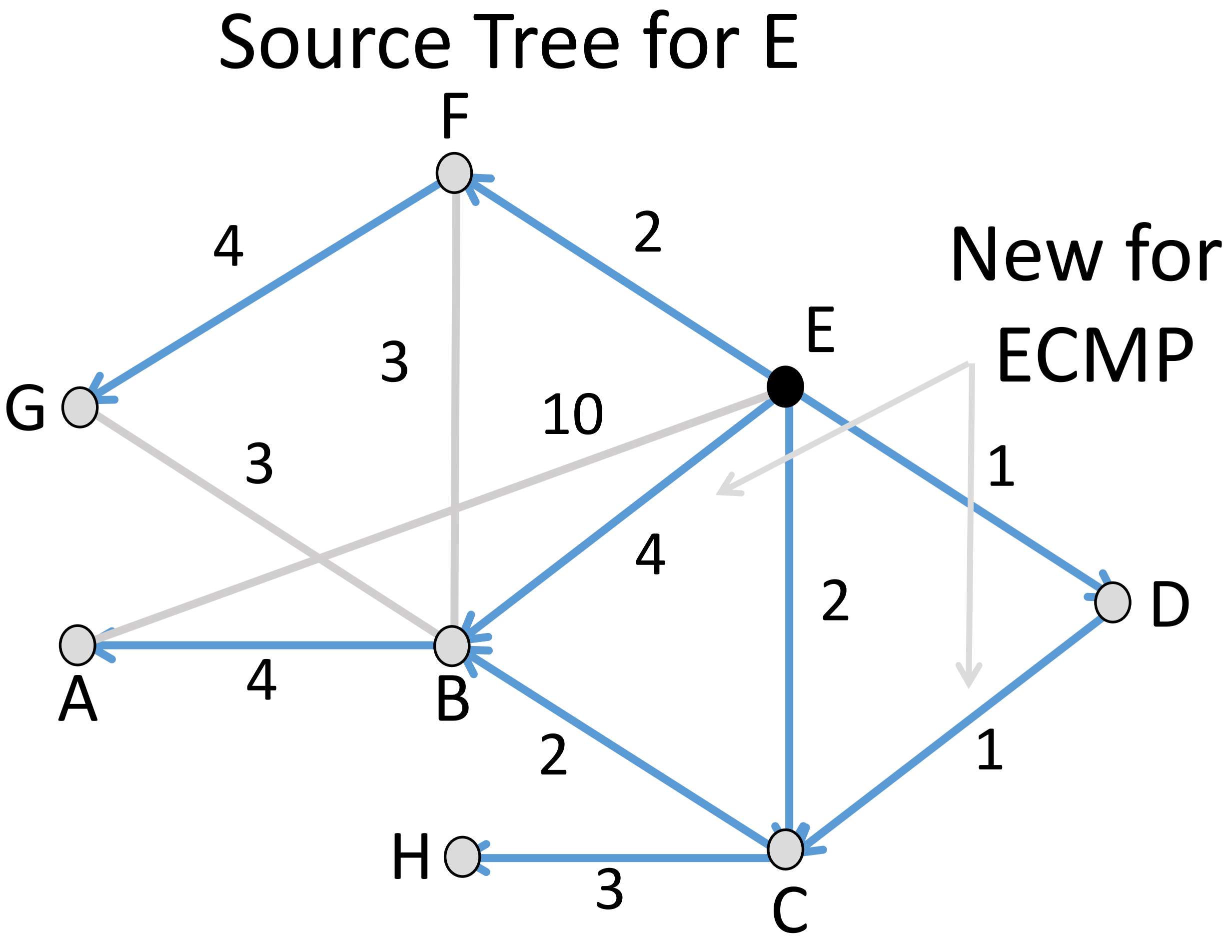


Source “Trees” (2)

- Find the source “tree” for E
  - Procedure is Dijkstra, simply remember set of next hops
  - Compile forwarding table similarly, may have set of next hops
- Straightforward to extend DV too
  - Just remember set of neighbors



Source “Trees” (3)



E’s Forwarding Table

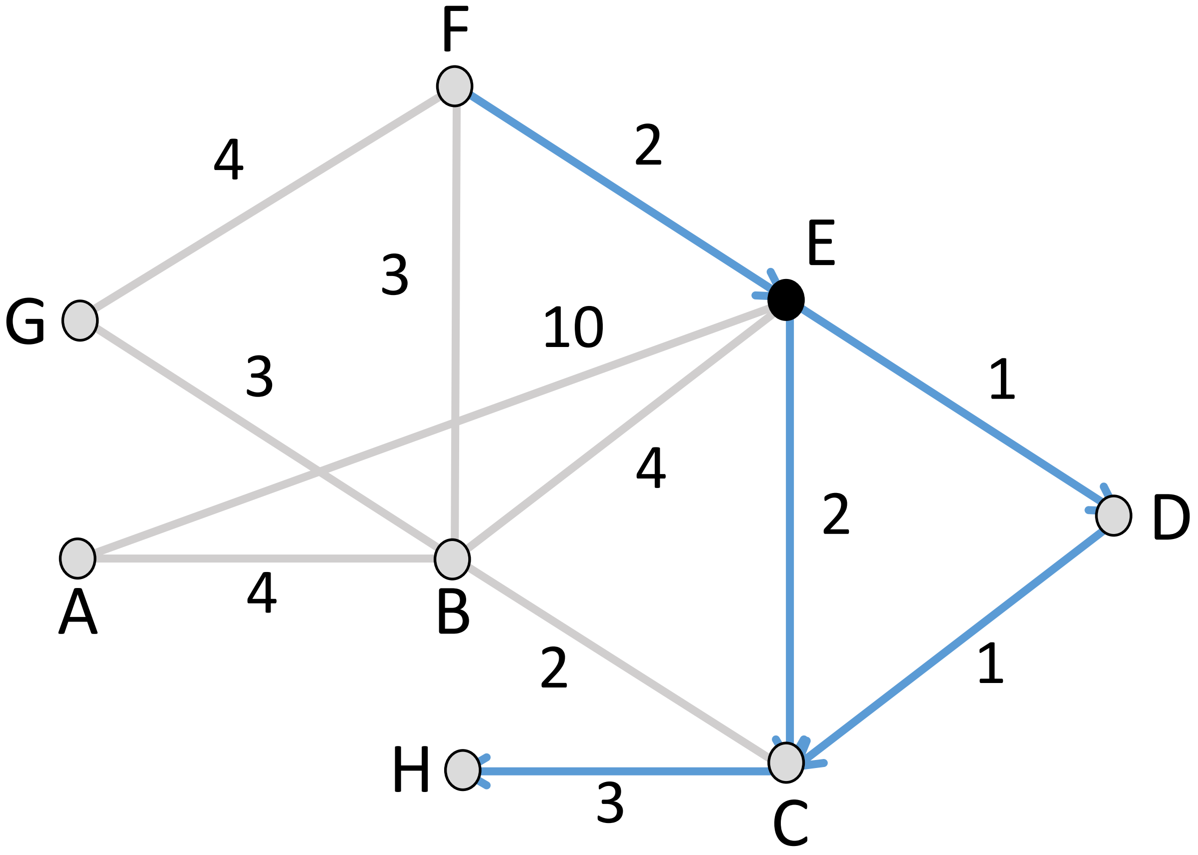
Node	Next hops
A	B, C, D
B	B, C, D
C	C, D
D	D
E	--
F	F
G	F
H	C, D

Forwarding with ECMP

- Could randomly pick a next hop for each packet based on destination
  - Balances load, but adds jitter
- Try sending packets from a flow on the same path
  - Flow identified using 5-tuple
  - Map flow identifier to single next hop
  - No jitter within flow, but less balanced

Forwarding with ECMP (2)

Multipath routes from F/E to C/H



E's Forwarding Choices

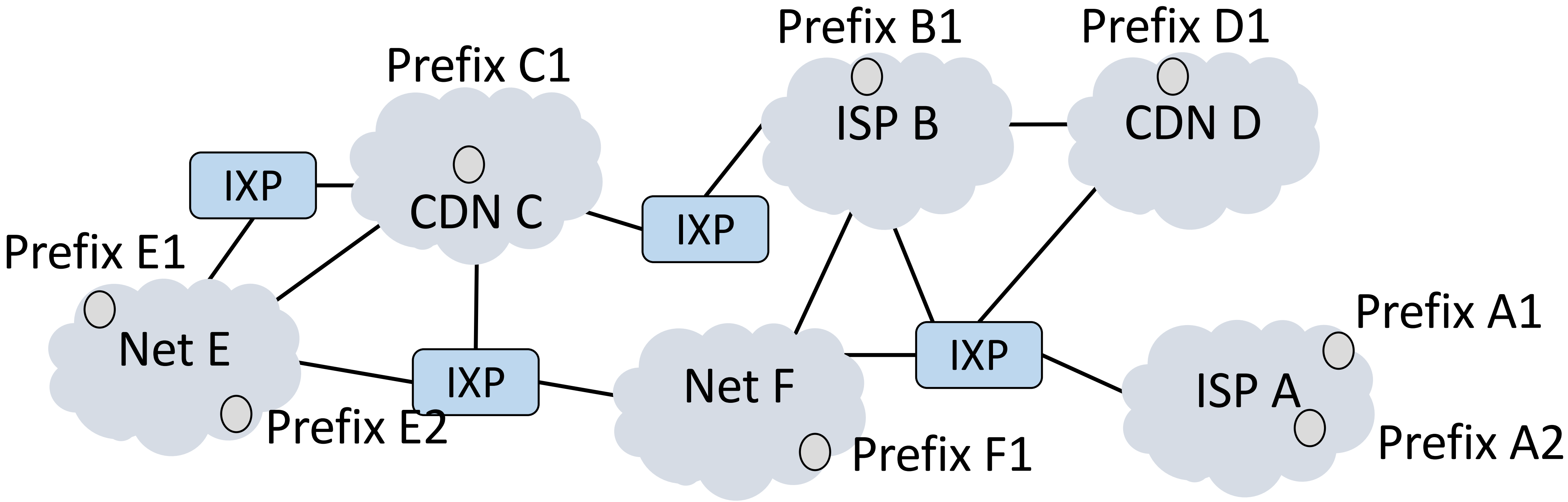
Flow	Possible next hops	Example choice
F → H	C, D	D
F → C	C, D	D
E → H	C, D	C
E → C	C, D	C

Use both paths to get to one destination

Border Gateway Protocol (BGP)

Structure of the Internet

- Networks (ISPs, CDNs, etc.) group with IP prefixes
- Networks are richly interconnected, often using IXPs



Internet-wide Routing Issues

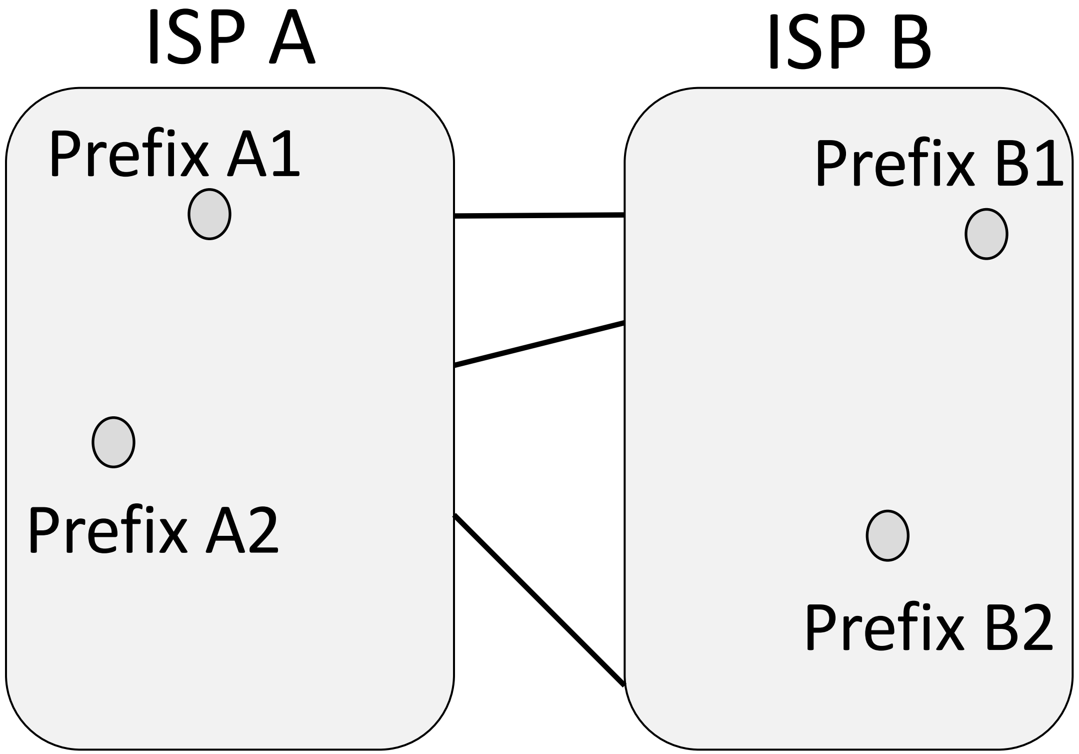
- Two problems beyond routing within a network

1. Scaling to very large networks
    - Techniques of IP prefixes, hierarchy, prefix aggregation
2. Incorporating policy decisions
    - Letting different parties choose their routes to suit their own needs

Yikes!

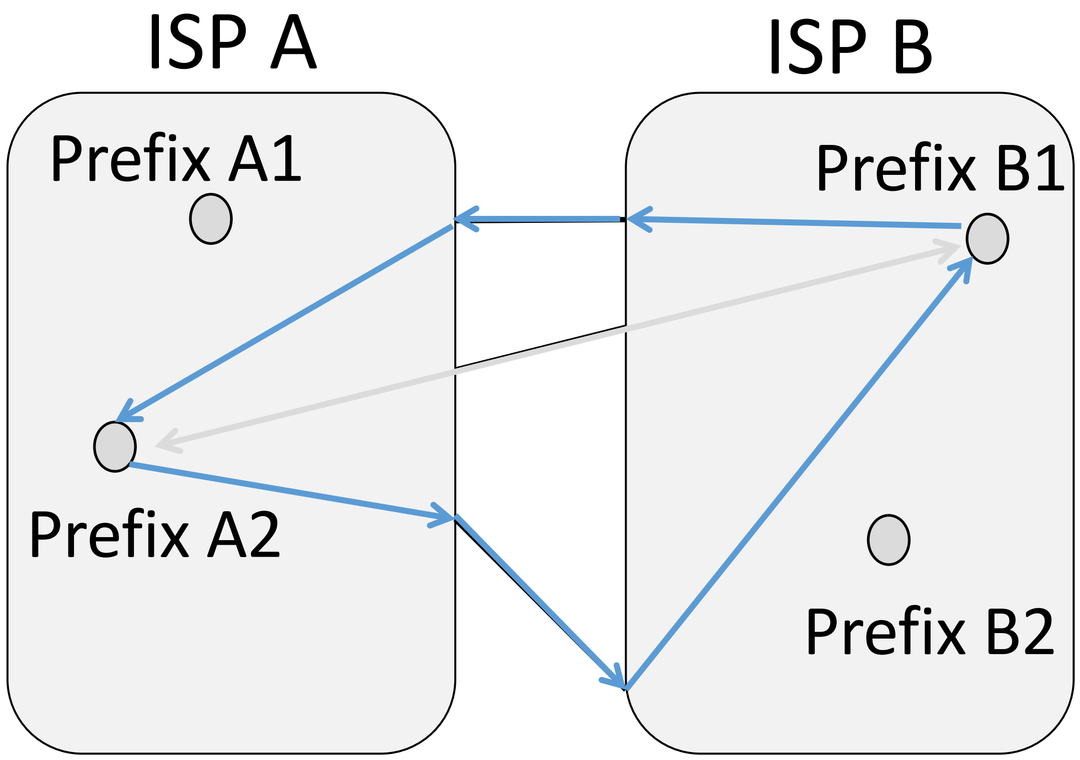
Effects of Independent Parties

- Each party selects routes to suit its own interests
  - e.g, shortest path in ISP
- What path will be chosen for A2→B1 and B1→A2?
  - What is the best path?



Effects of Independent Parties (2)

- Selected paths are longer than overall shortest path
  - And asymmetric too!
- Consequence of independent goals and decisions, not hierarchy

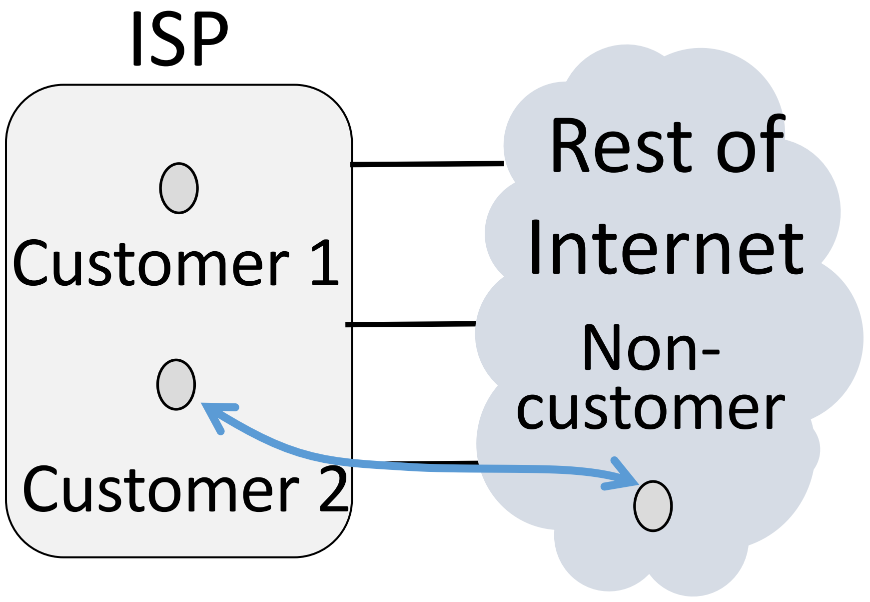


Routing Policies

- Capture the goals of different parties
  - Could be anything
  - E.g., Internet2 only carries non-commercial traffic
- Common policies we'll look at:
  - ISPs give TRANSIT service to customers
  - ISPs give PEER service to each other

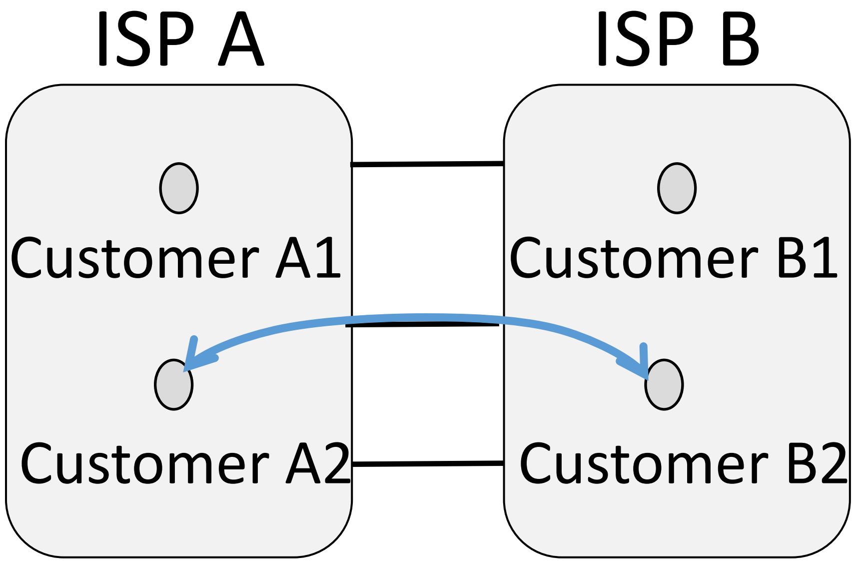
Routing Policies – Transit

- One party (customer) gets TRANSIT service from another party (ISP)
  - ISP accepts traffic for customer from the rest of Internet
  - ISP sends traffic from customer to the rest of Internet
  - Customer pays ISP for the privilege



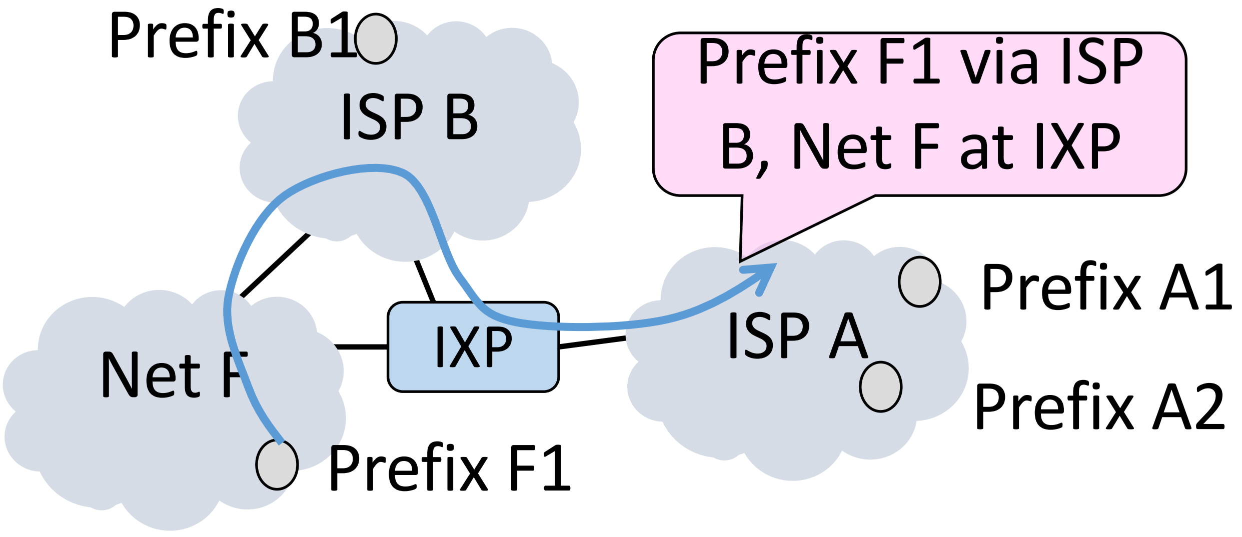
Routing Policies – Peer

- Both party (ISPs in example) get PEER service from each other
  - Each ISP accepts traffic from the other ISP only for their customers
  - ISPs do not carry traffic to the rest of the Internet for each other
  - ISPs don’t pay each other



Routing with BGP

- iBGP is for internal routing
- eBGP is <u>interdomain</u> routing for the Internet
  - Path vector, a kind of distance vector



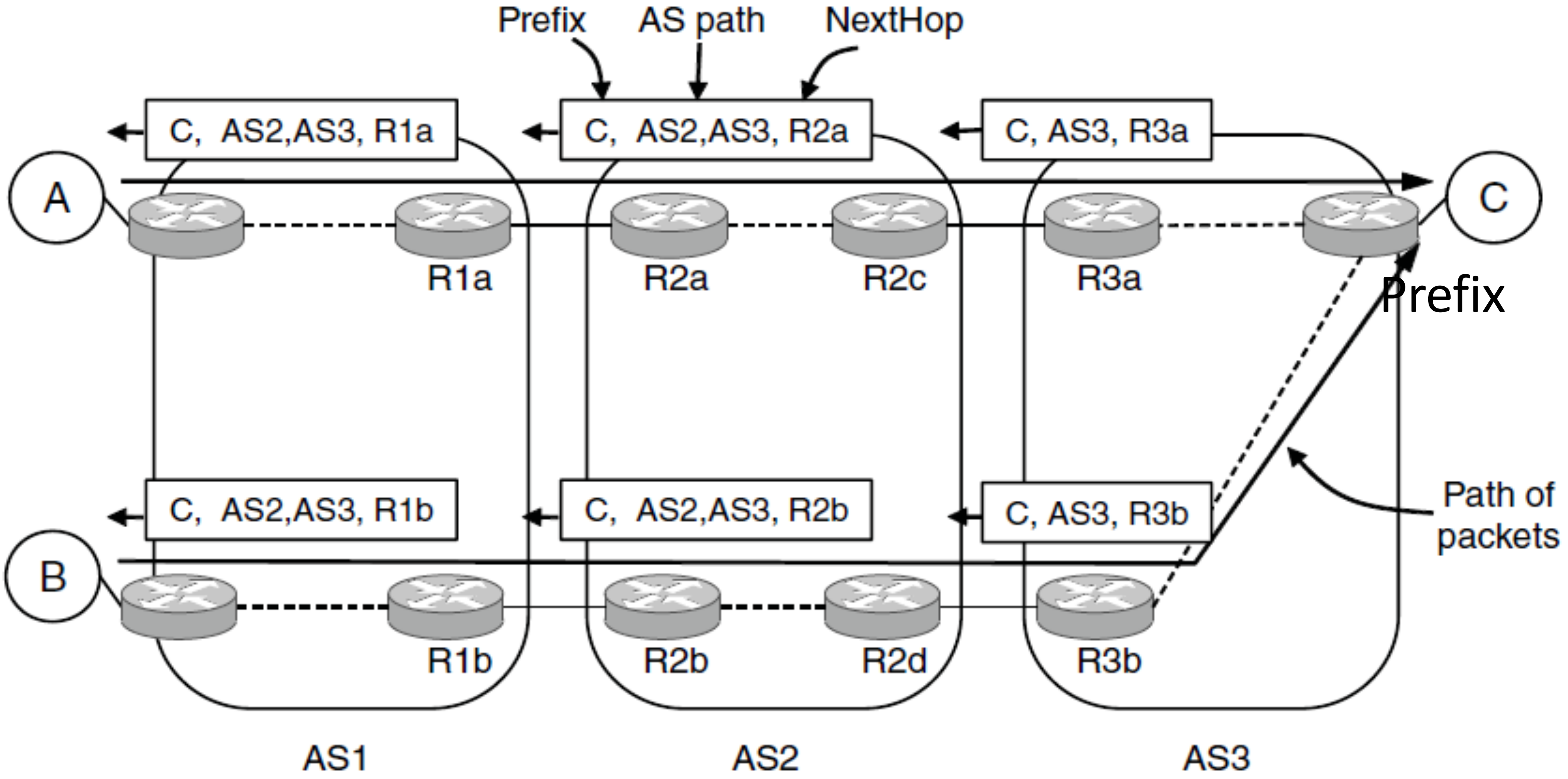
Routing with BGP (2)

- Parties like ISPs are called AS (Autonomous Systems)
  - AS numbers are unique identifiers
- AS’s configure their internal BGP routes
- External routes go through complicated filters
- Intra-AS BGP routers communicate to keep consistent routing information

Routing with BGP (3)

- Border routers of ASes announce BGP routes
- Route announcements have IP prefix, path vector, next hop
    - Path vector is list of ASes on the way to the prefix
    - List is to find loops
- Route announcements move in the opposite direction to traffic

Routing with BGP (4)



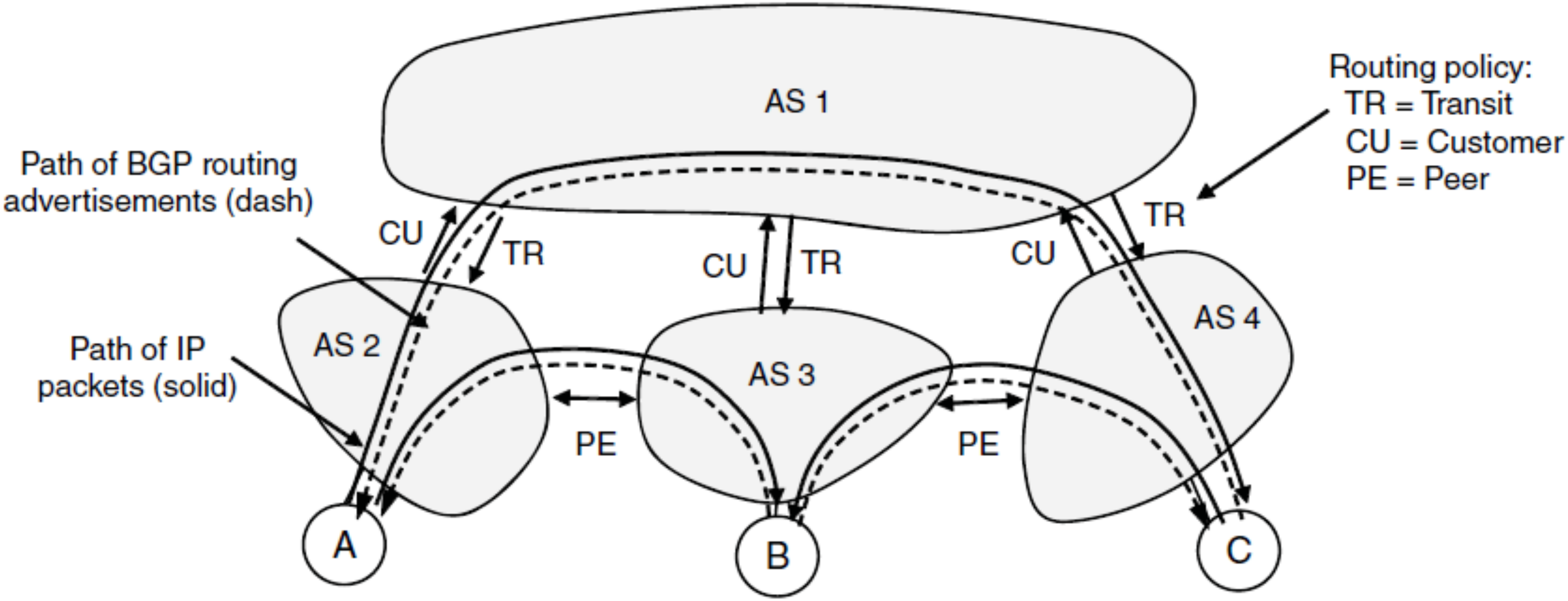
Routing with BGP (5)

Policy is implemented in two ways:

1. Border routers of ISP announce paths only to other parties who may use those paths
   - Filter out paths others can’t use
2. Border routers select the best path of the ones they hear in any way (not necessarily shortest)

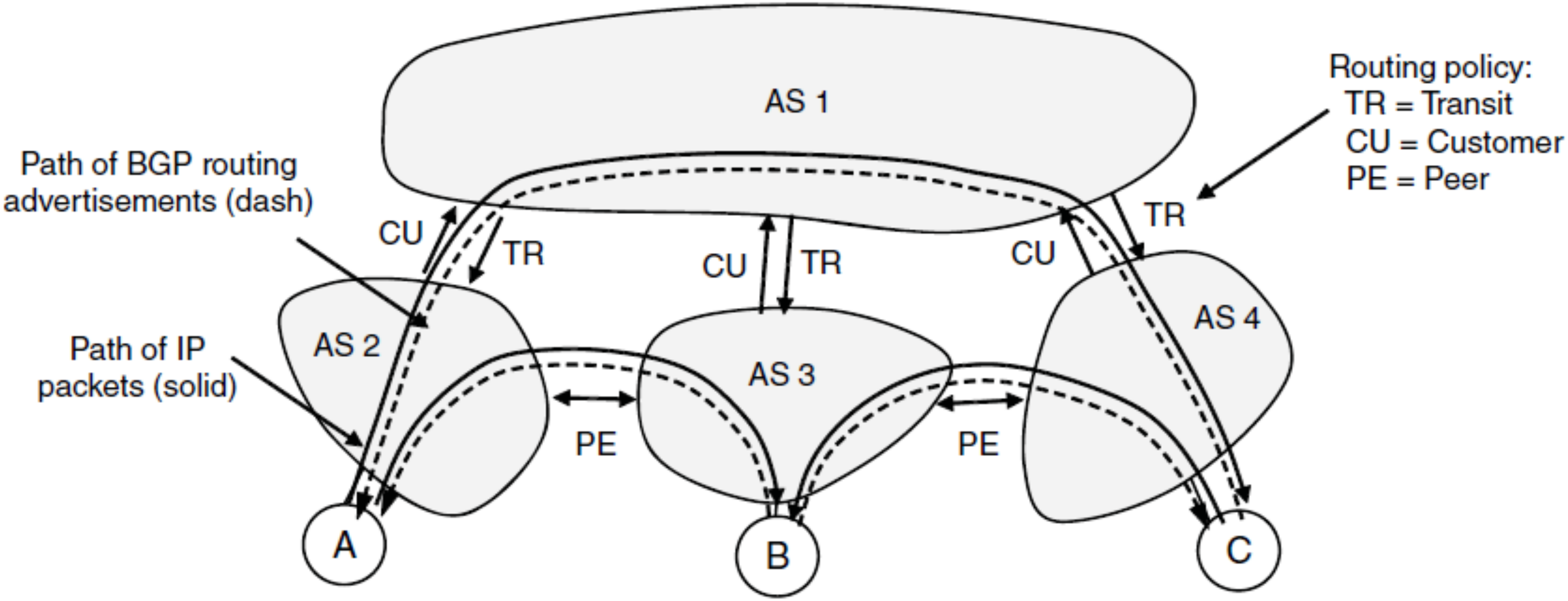
Routing with BGP (6)

- TRANSIT: AS1 says [B, (AS1, AS3)], [C, (AS1, AS4)] to AS2



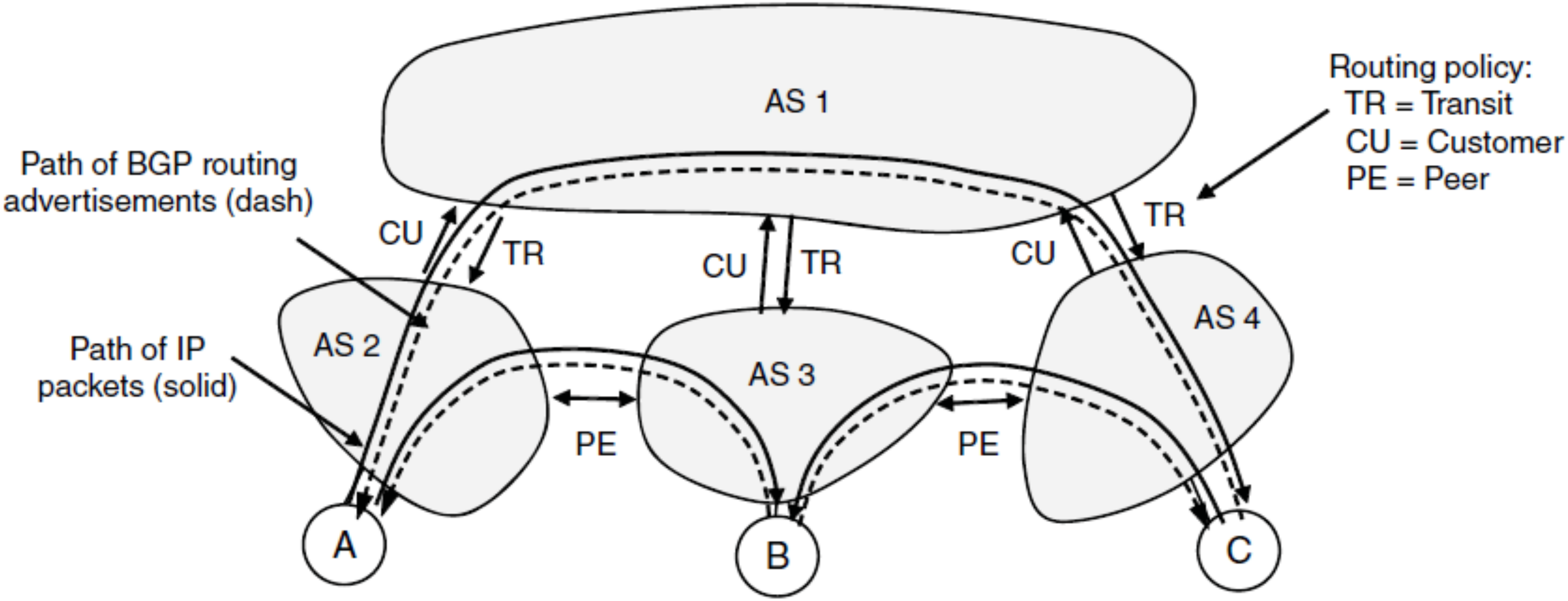
Routing with BGP (7)

- CUSTOMER (other side of TRANSIT): AS2 says [A, (AS2)] to AS1



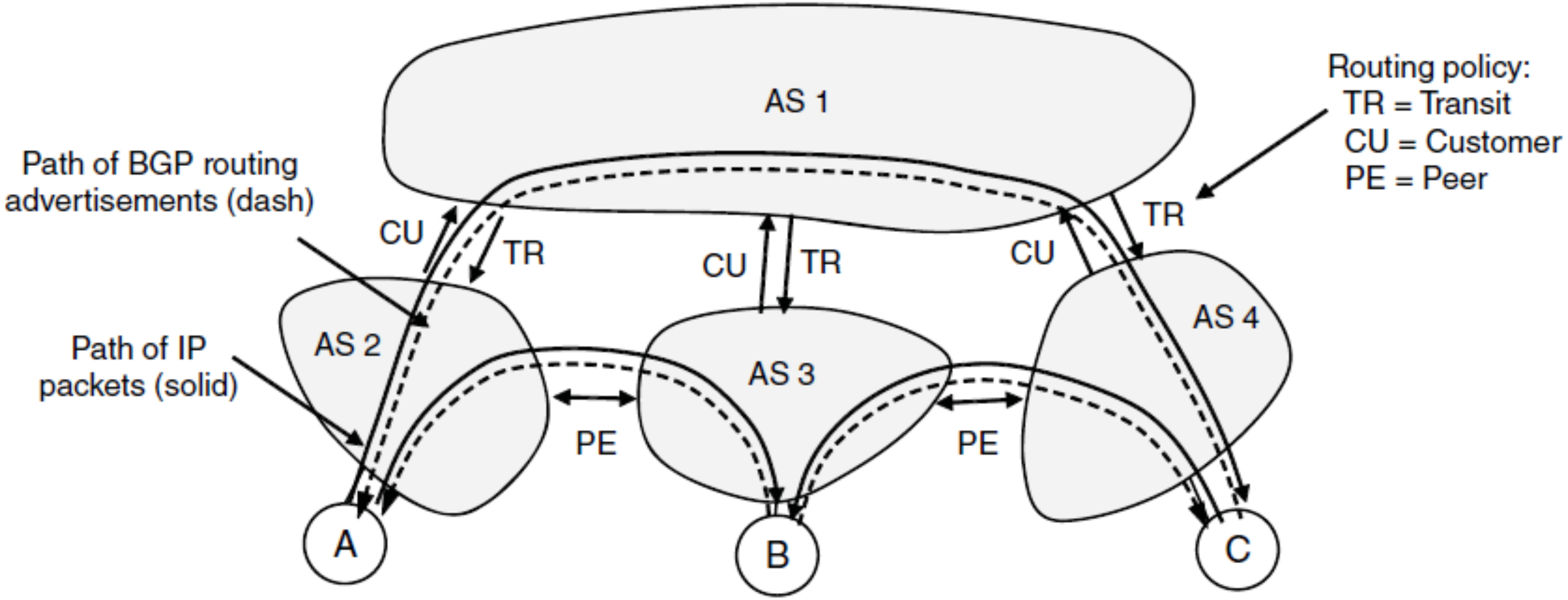
Routing with BGP (8)

- PEER: AS2 says [A, (AS2)] to AS3, AS3 says [B, (AS3)] to AS2



Routing with BGP (9)

- AS2 has two routes to B (AS1, AS3) and chooses AS3 (Free!)



BGP Thoughts

- Much more beyond basics to explore!
- Policy is a substantial factor
  - Can independent decisions be sensible overall?
- Other important factors:
  - Convergence effects
  - Security
  - Integration with intradomain routing
  - …

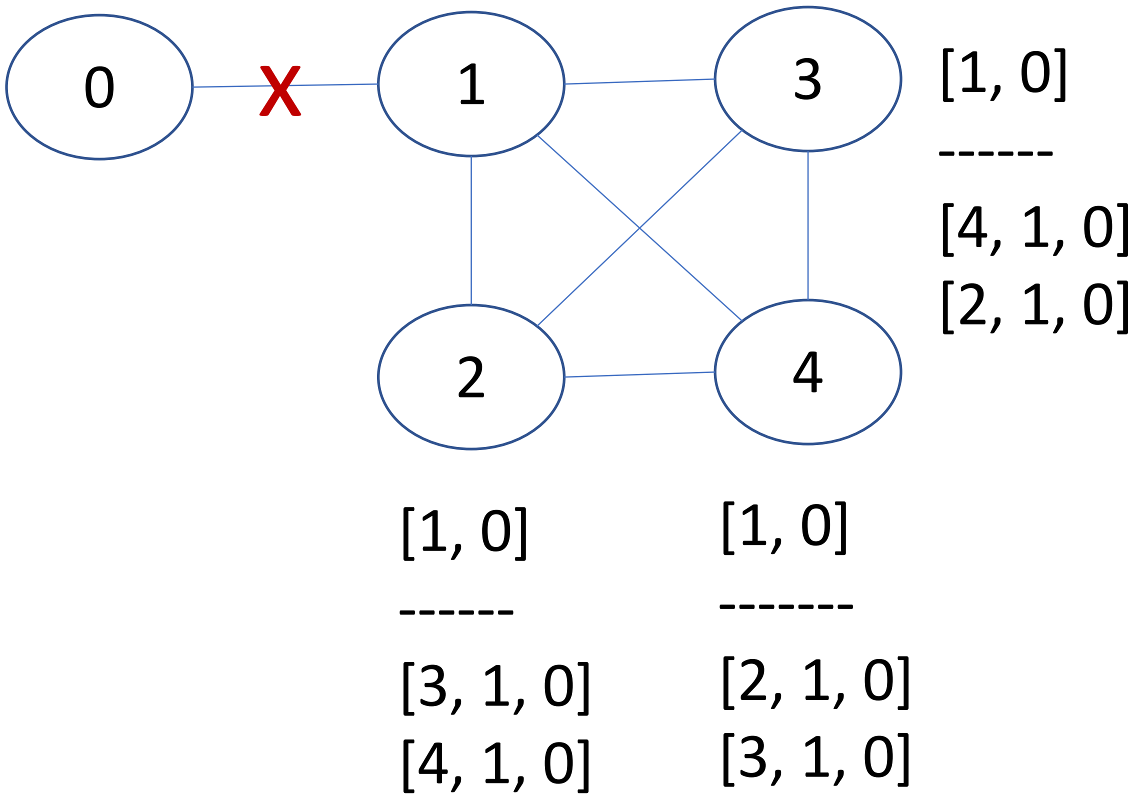
BGP convergence

Path vector protocols have a version of count to infinity problem

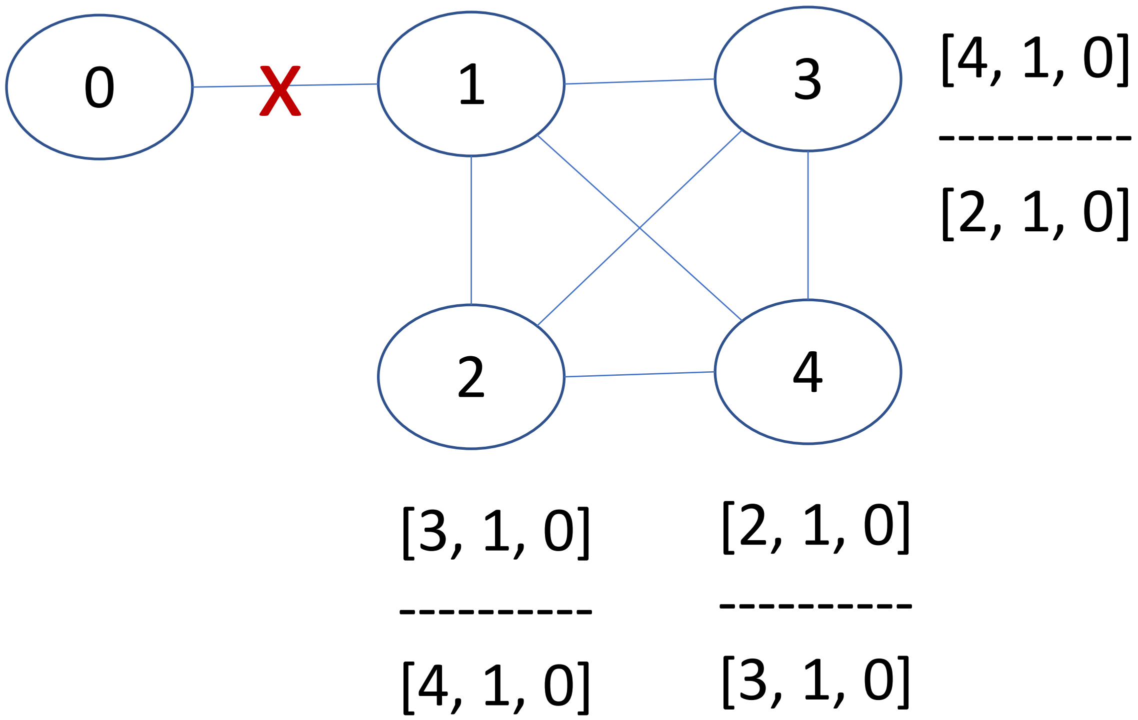
- Explore many non-existent paths

Worse, uncoordinated policies can lead to never converging

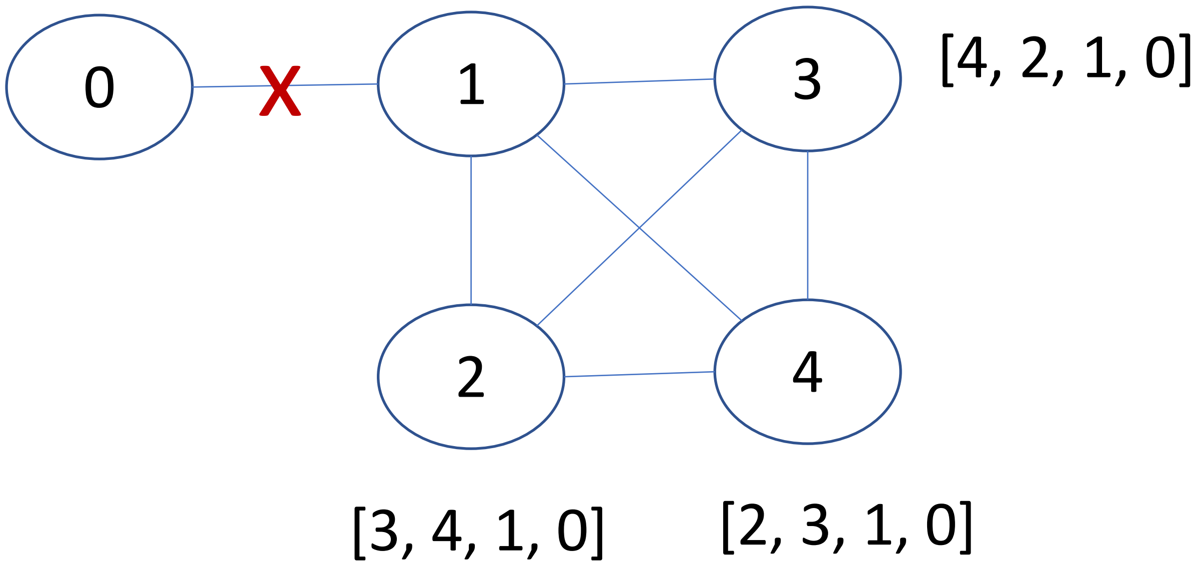
BGP slow convergence



BGP slow convergence

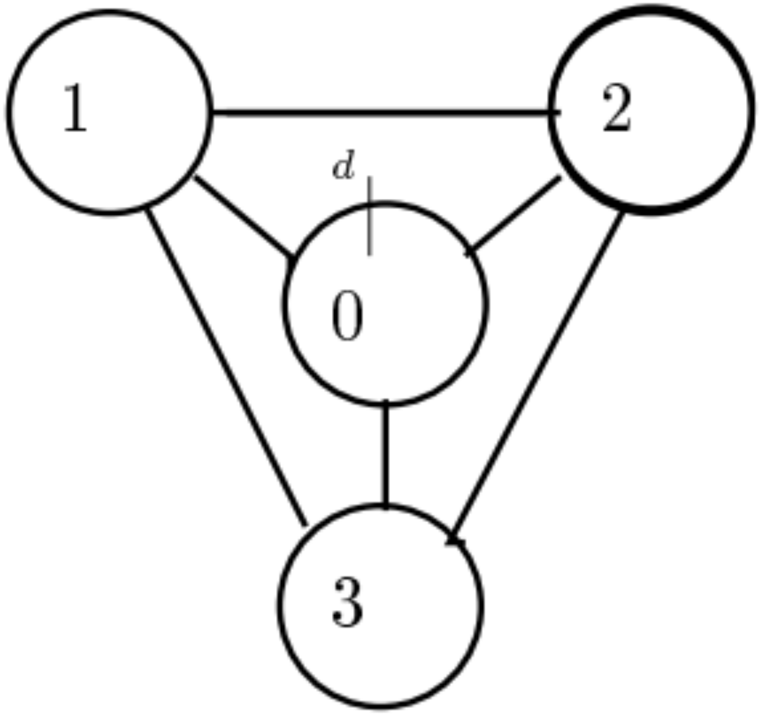


BGP slow convergence



BGP “bad gadget”: Non-convergence

[2, 0] > [0] > [2, 3, 0]



[3, 0] > [0] > [3, 1, 0]

[1, 0] > [0] > [1, 2, 0]

BGP security

Anyone can announce anything

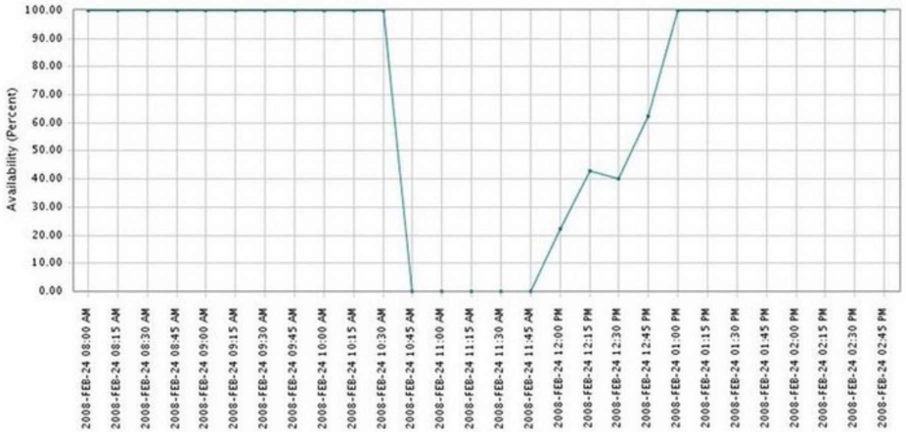
- By accident
- By malice

CULTURE

How Pakistan knocked YouTube offline (and how to make sure it never happens again)

YouTube becoming unreachable isn't the first time that Internet addresses were hijacked. But if it spurs interest in better security, it may be the last.

BY DECLAN MCCULLAGH | FEBRUARY 25, 2008 4:28 PM PST



This graph that network-monitoring firm Keynote Systems provided to us shows the worldwide availability of YouTube.com dropping dramatically from 100 percent to 0 percent for over an hour. It didn't recover completely until two hours had elapsed.

Keynote Systems

A high-profile incident this weekend in which Pakistan's state-owned telecommunications company managed to cut YouTube off the global Web highlights a long-standing security weakness in the way the Internet is managed.

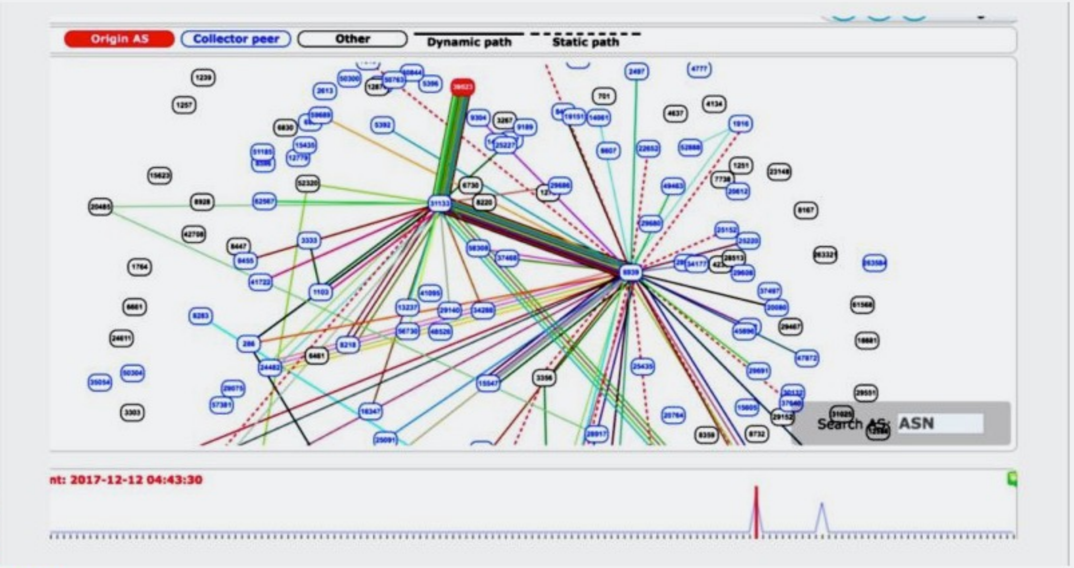
After receiving a censorship order from the telecommunications ministry directing that YouTube.com be blocked, Pakistan Telecom went even further. By accident or

BIZ & IT —

"Suspicious" event routes traffic for big-name sites through Russia

Google, Facebook, Apple, and Microsoft all affected by "intentional" BGP mishap.

DAN GOODIN - 12/13/2017, 2:43 PM



BGPMon

Enlarge

104

Traffic sent to and from Google, Facebook, Apple, and Microsoft was briefly routed through a previously unknown Russian Internet provider Wednesday under circumstances researchers said was suspicious and intentional.

FURTHER READING

Russian-controlled telecom hijacks financial services' Internet traffic

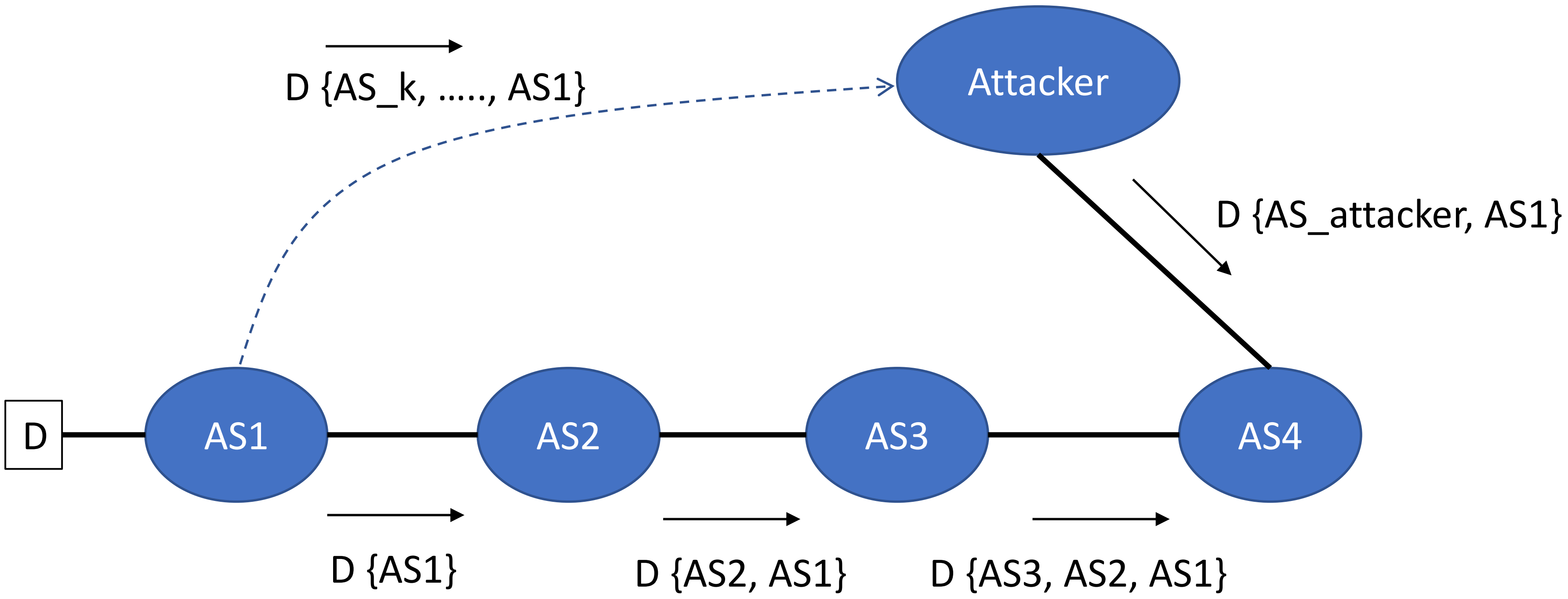
The unexplained incident involving the Internet's Border Gateway Protocol is the latest to raise troubling questions about the trust and reliability of communications sent over the global network. BGP routes large-scale amounts of traffic among Internet backbones, ISPs, and other large networks. But despite the sensitivity and amount of data it controls, BGP's security is often based on trust and word of mouth. Wednesday's event comes eight months after large chunks of network traffic belonging to MasterCard, Visa, and more than two dozen other financial services were briefly routed through a Russian government-

BGP security mechanisms

Validate who can originate what prefix

- Major push for origin validation
- RPKI: https://en.wikipedia.org/wiki/Resource_Public_Key_Infrastructure

Helpful but not enough

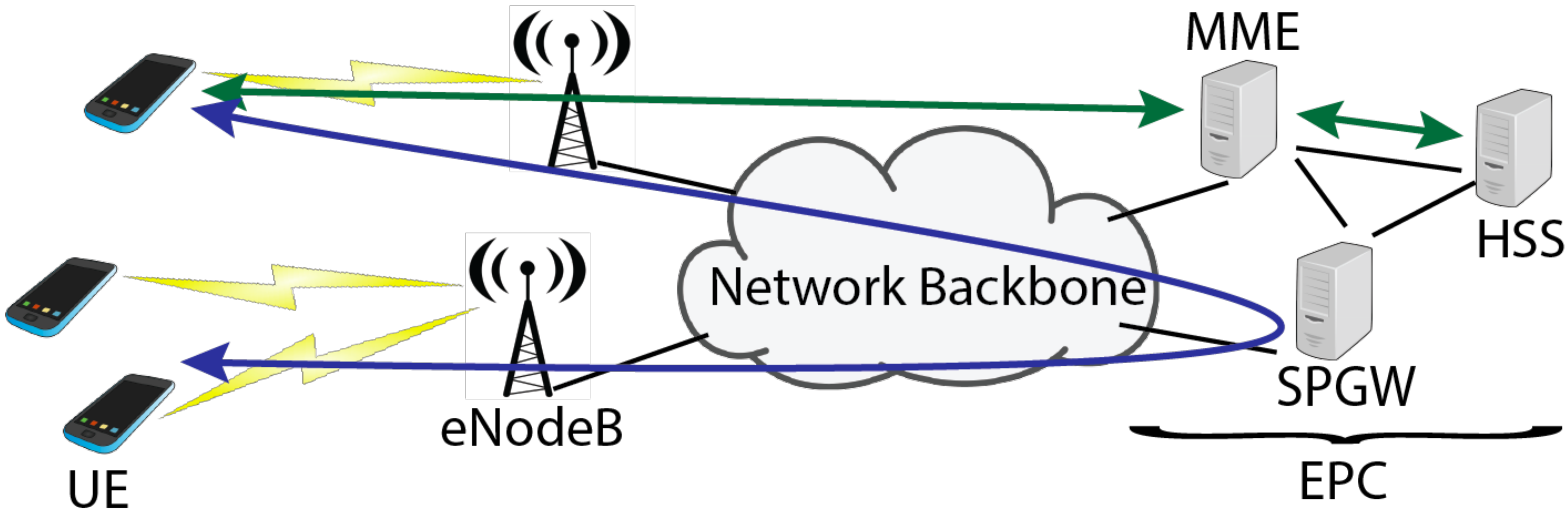

D {AS_k, ….., AS1}
Attacker
D {AS_attacker, AS1}
D
AS1
AS2
AS3
AS4
D {AS1}
D {AS2, AS1}
D {AS3, AS2, AS1}

Cellular Routing

Addressing in Cellular

- Everyone has a unique physical identifier: SIM Card
  - IMSI: International Mobile Subscriber Identity
  - Has associated mobile provider
  - Phone number **not** present
    - Known as “msisdn”

Cellular Core Networks



In-network routing

1. User dials phone number
2. Number is “looked up” in some database
3. If local, we get the associated IMSI
4. Check that sender can send and receiver can receive
5. Look up tower group of IMSIs last registration
6. Page the receiver
7. Bill them both

Out-of-network Routing

- Signaling System No. 7 (SS7)
  - Performs number translation, local number portability, prepaid billing, Short Message Service (SMS), roaming, and other stuff
  - Either directly connected or connected through aggregators such as Cybase
  - Business vs Protocols