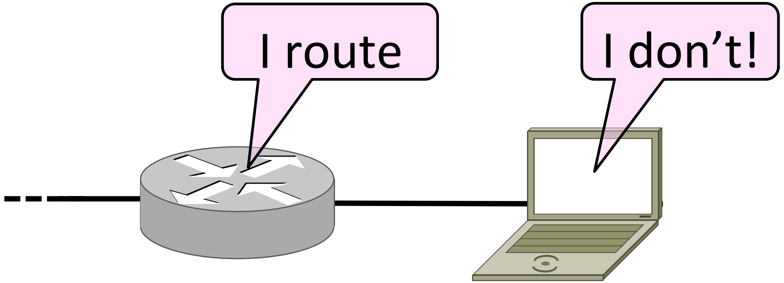
Topic

- How routing protocols work with IP
  - The Host/Router distinction

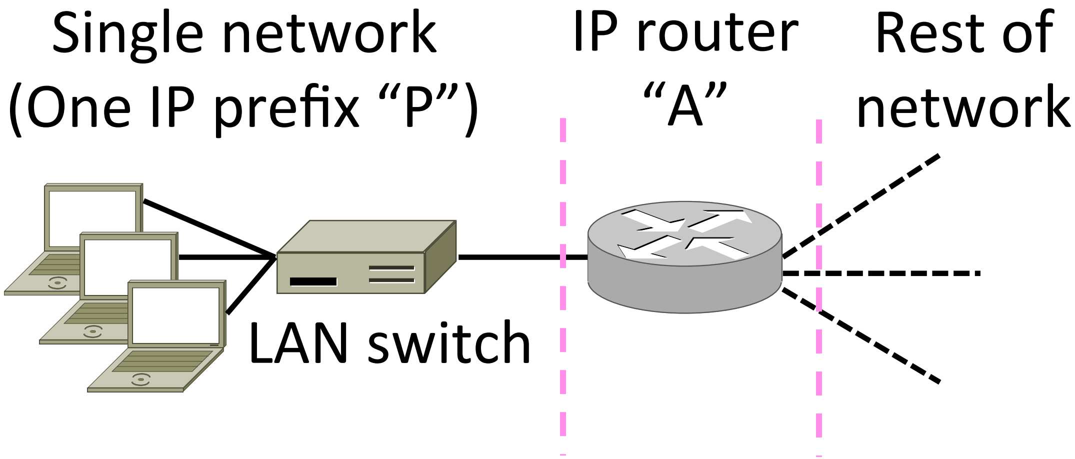


Recap

- In the Internet:
  - Hosts on same network have IP addresses in the same <u>IP prefix</u>
  - Hosts just send off-network traffic to the nearest router to handle
  - Routers discover the routes to use
  - Routers use <u>longest prefix matching</u> to send packets to the right next hop

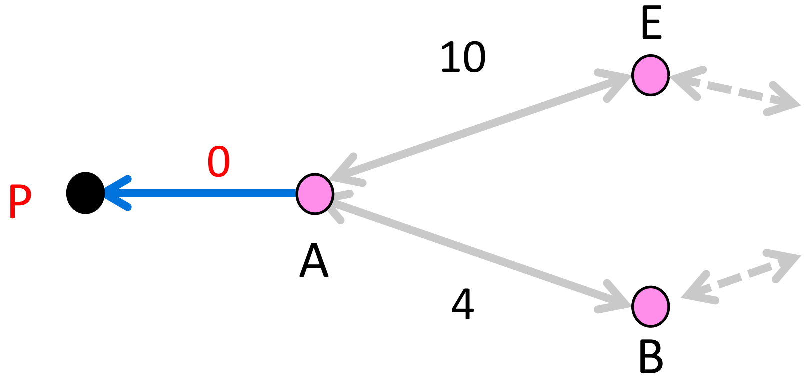
Host/Router Combination

- Hosts attach to routers as IP prefixes
  - Router needs table to reach all hosts



Network Topology for Routing

- Group hosts under IP prefix connected directly to router
  - One entry for all hosts

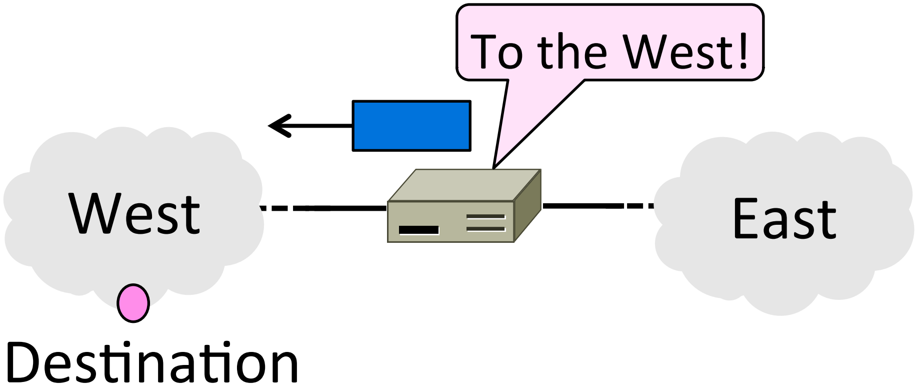


Network Topology for Routing (2)

- Routing now works as before!
  - Routers advertise IP prefixes for hosts
  - Router addresses are “/32” prefixes
  - Lets all routers find a path to hosts
  - Hosts find by sending to their router

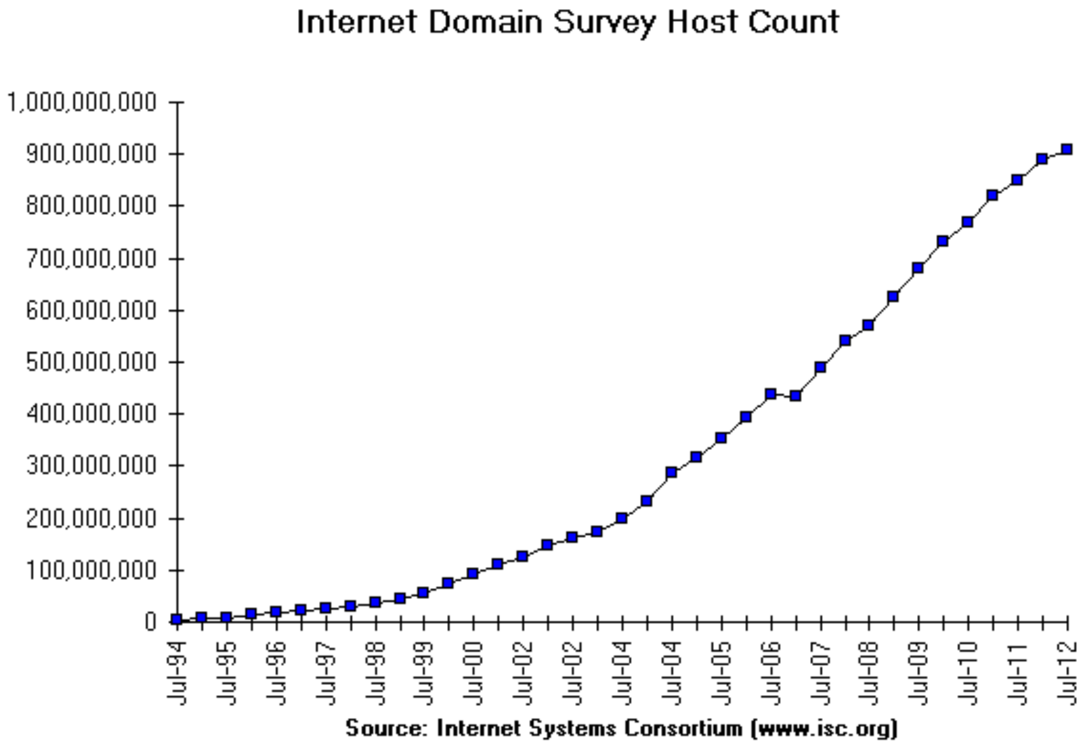
Topic

- How to scale routing with hierarchy in the form of regions
  - Route to regions, not individual nodes



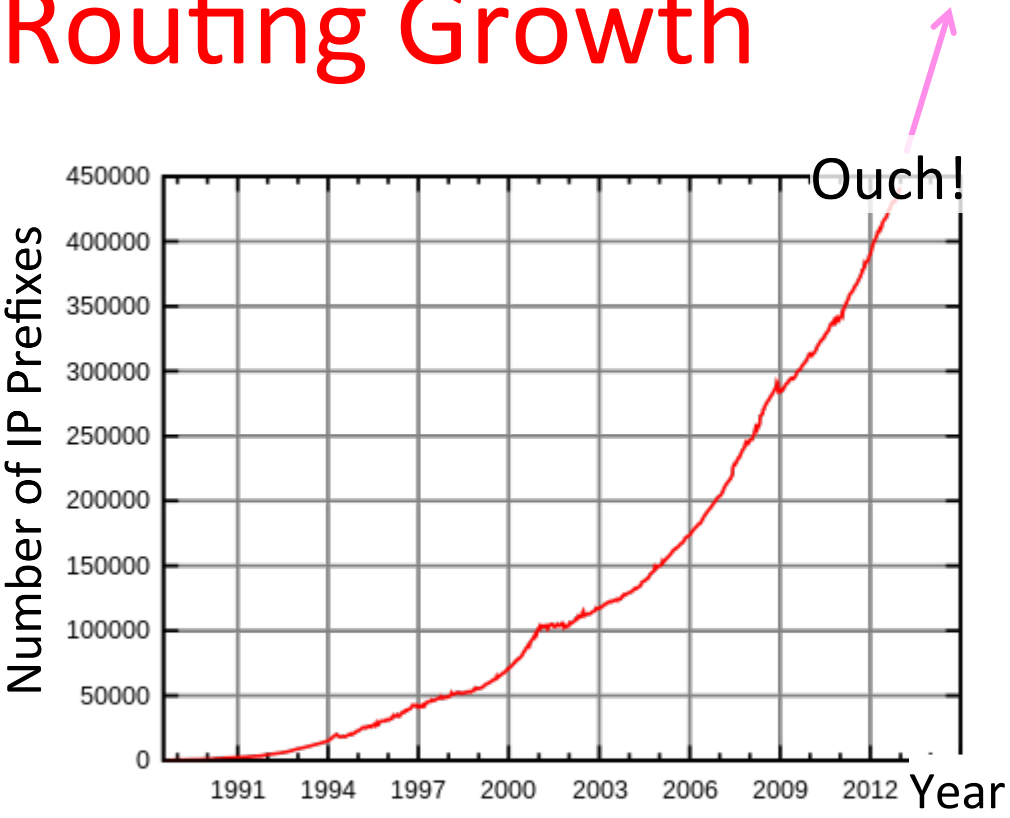
Internet Growth

- At least a billion Internet hosts and growing …



Internet Routing Growth

- Internet growth translates into routing table growth (even using prefixes) …



Source: By Mro (Own work), CC-BY-SA-3.0 , via Wikimedia Commons

Impact of Routing Growth

1. Forwarding tables grow
   - Larger router memories, may increase lookup time
2. Routing messages grow
   - Need to keeps all nodes informed of larger topology
3. Routing computation grows
   - Shortest path calculations grow faster than the size of the network

Techniques to Scale Routing

1. IP prefixes
   – Route to blocks of hosts

Last time

2. Network hierarchy
   – Route to network regions

This time

3. IP prefix aggregation
   – Combine, and split, prefixes

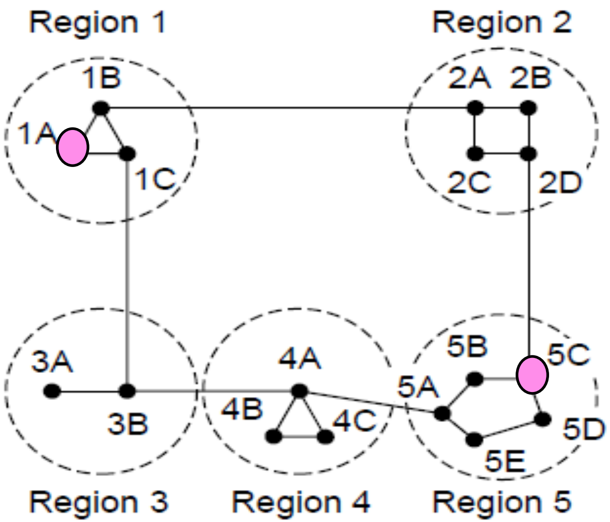
Next time

Hierarchical Routing

- Introduce a larger routing unit
  - IP prefix (hosts) ← from one host
  - Region, e.g., ISP network

- Route first to the region, then to the IP prefix within the region
  - Hide details within a region from outside of the region

Hierarchical Routing (2)



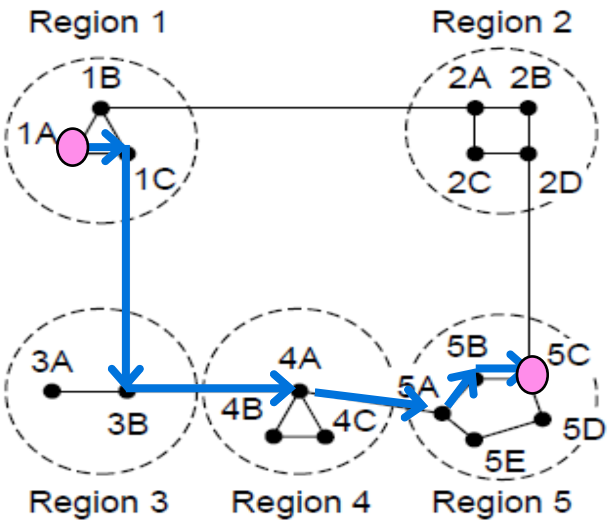
Full table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

Hierarchical Routing (3)



Full table for 1A

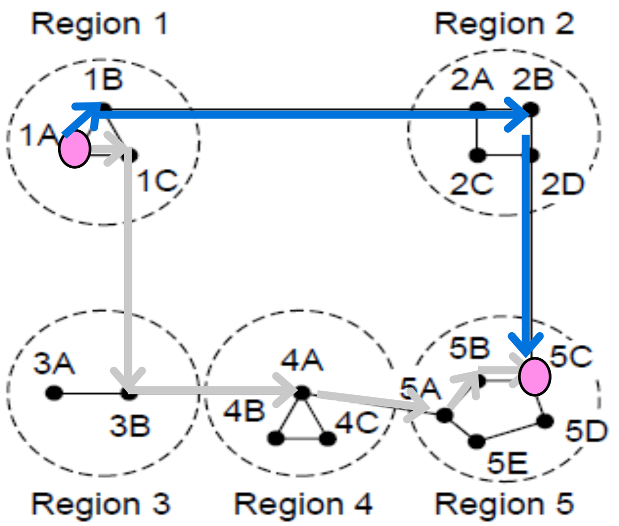
Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

Hierarchical Routing (4)

- Penalty is longer paths



Full table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

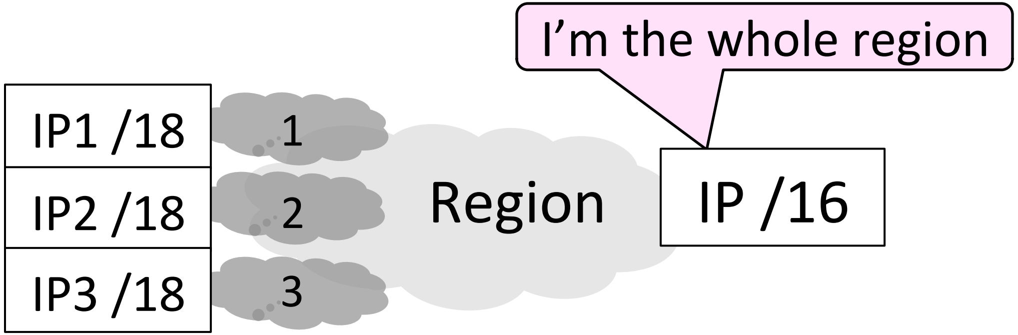
1C is best route to region 5, except for destination 5C

Observations

- Outside a region, nodes have one route to all hosts within the region
  - This gives savings in table size, messages and computation
- However, each node may have a different route to an outside region
  - Routing decisions are still made by individual nodes; there is no single decision made by a region

Topic

- How to help scale routing by adjusting the size of IP prefixes
  - Split (subnets) and join (aggregation)



Recall

- IP addresses are allocated in blocks called IP prefixes, e.g., 18.31.0.0/16
  - Hosts on one network in same prefix
- A "/N" prefix has the first N bits fixed and contains 2^{32-N} addresses
  - E.g., a "/24" has 256 addresses
- Routers keep track of prefix lengths
  - Use it as part of longest prefix matching

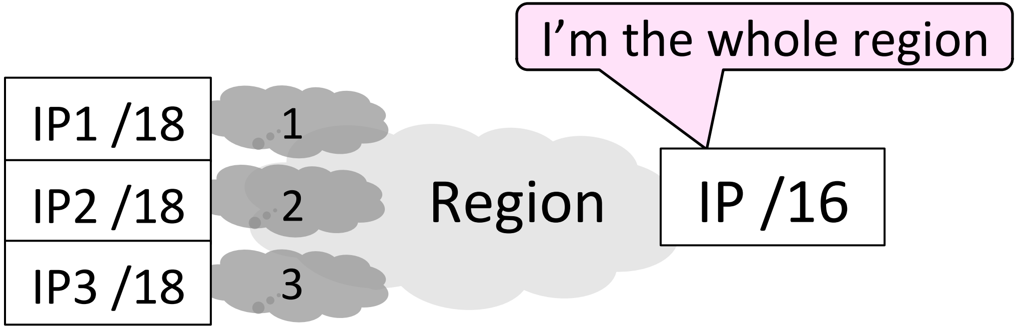
Recall (2)

- IP addresses are allocated in blocks called IP prefixes, e.g., 18.31.0.0/16
  - Hosts on one network in same prefix
- A "/N" prefix has the first N bits fixed and contains 2^{32-N} addresses
  - E.g., a "/24" has 256 addresses
- Routers keep track of prefix lengths
  - Use it as part of longest prefix matching

Routers can change prefix lengths without affecting hosts

Prefixes and Hierarchy

- IP prefixes already help to scale routing, but we can go further
  - We can use a less specific (larger) IP prefix as a name for a region



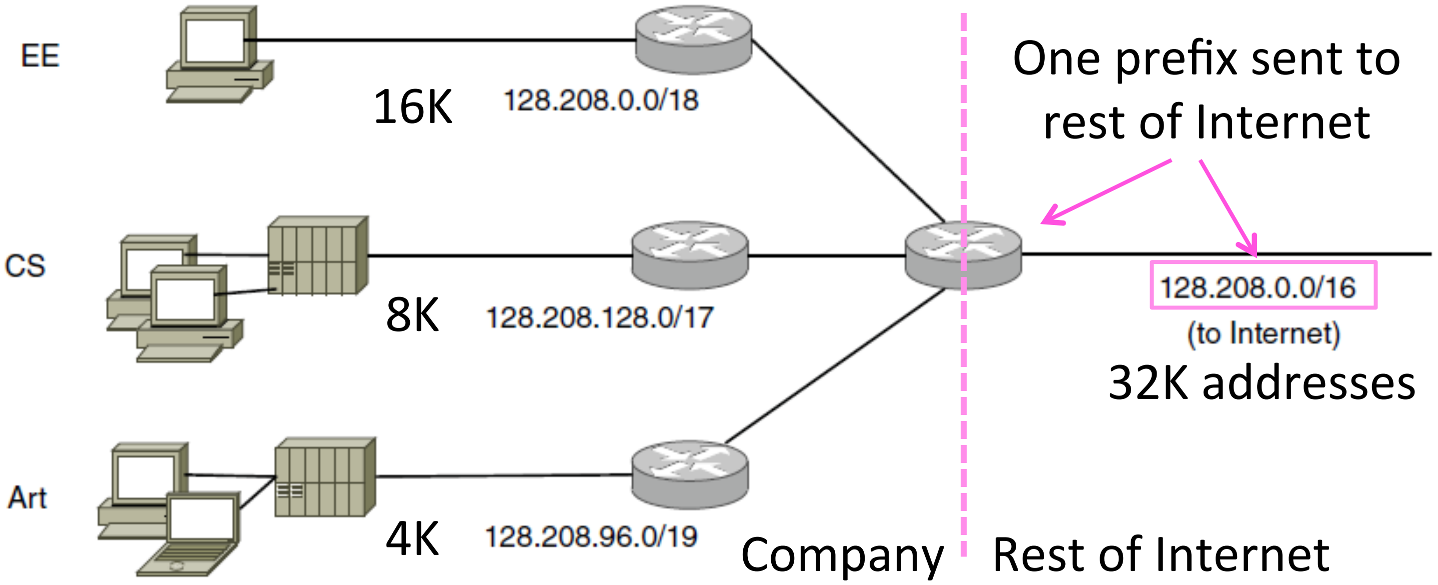
Subnets and Aggregation

- Two use cases for adjusting the size of IP prefixes; both reduce routing table

1. Subnets
   - Internally split one large prefix into multiple smaller ones
2. Aggregation
   - Externally join multiple smaller prefixes into one large prefix

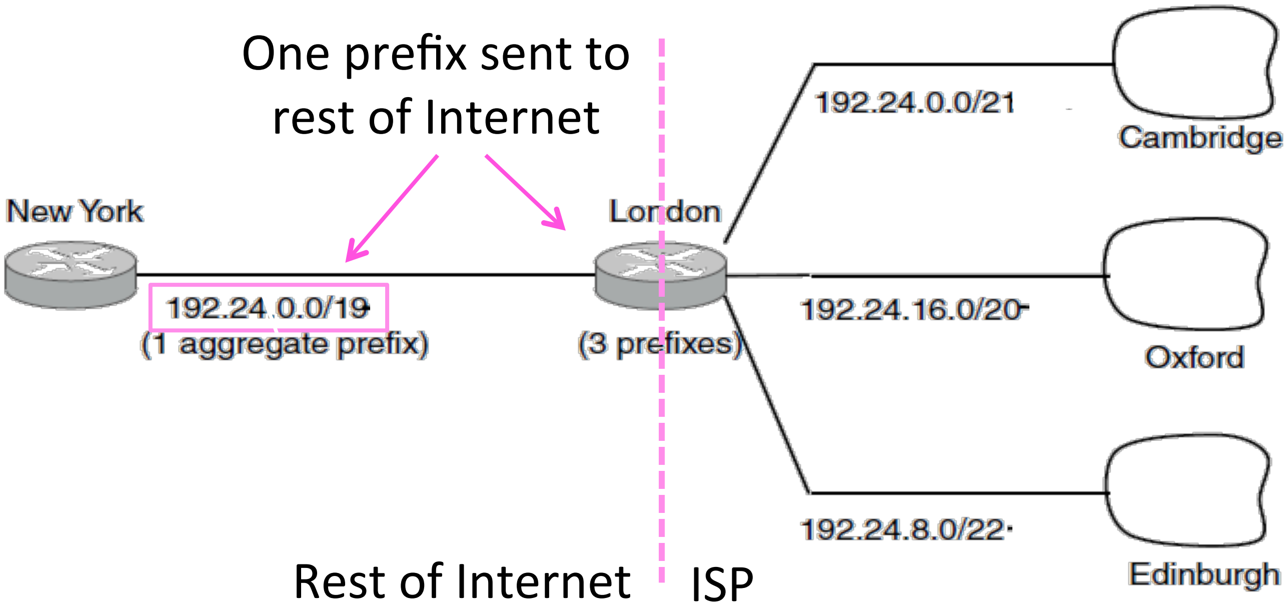
Subnets

- Internally split up one IP prefix



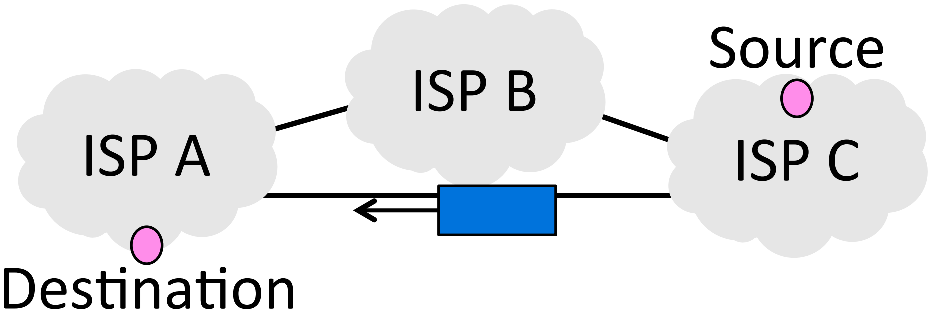
Aggregation

- Externally join multiple separate IP prefixes



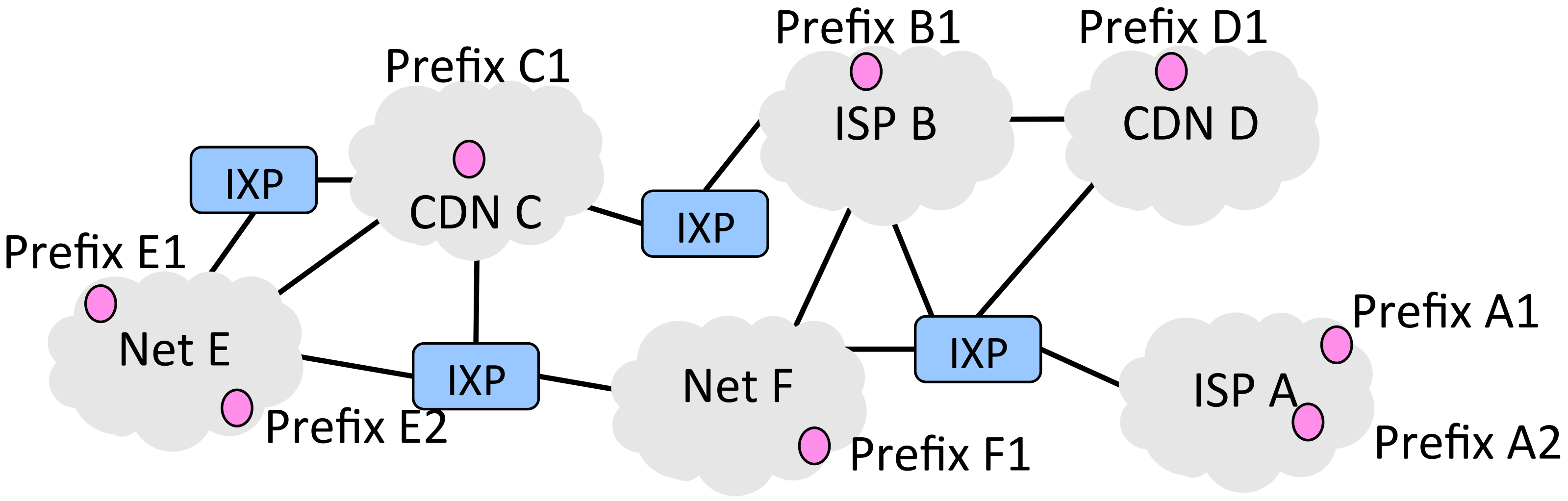
Topic

- How to route with multiple parties, each with their own routing policies
  - This is Internet-wide BGP routing



Structure of the Internet

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



Internet-wide Routing Issues

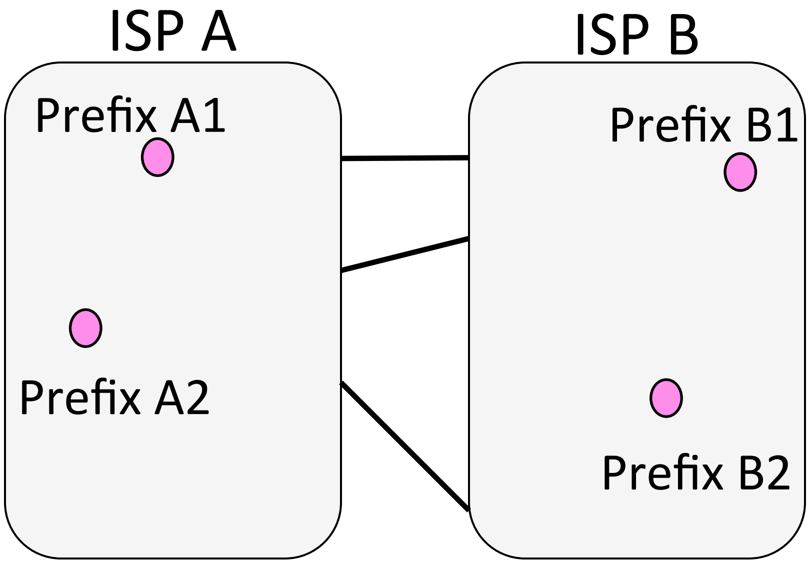
- Two problems beyond routing within an individual network

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!

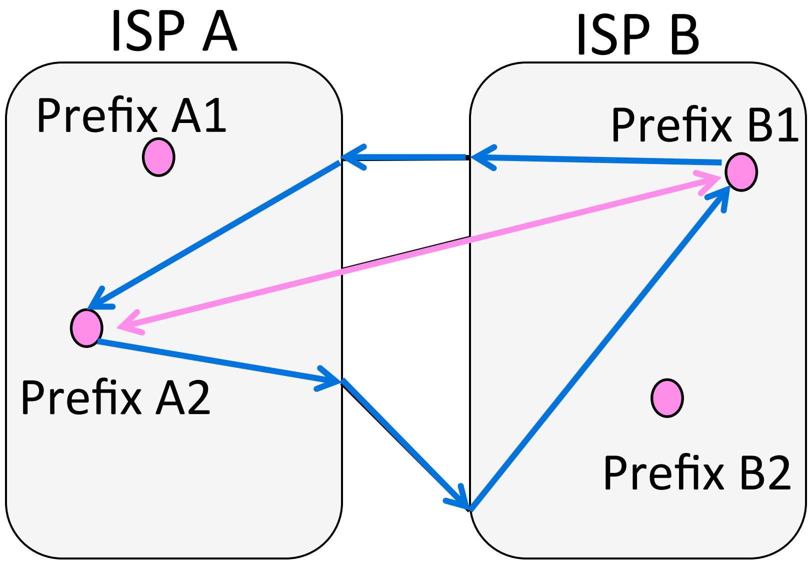
Effects of Independent Parties

- 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 symmetric too!
- This is a consequence of independent goals and decisions, not hierarchy

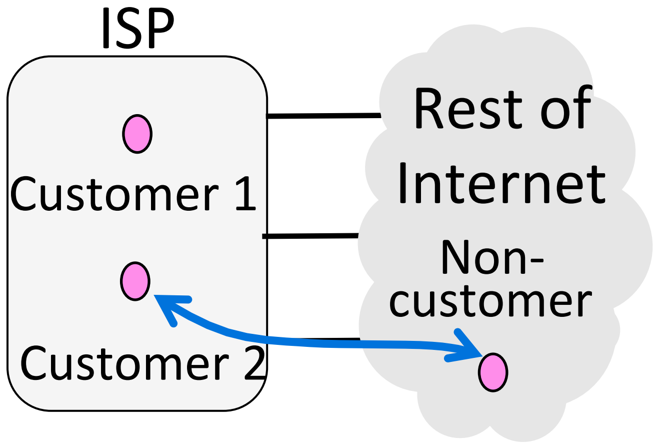


Routing Policies

- 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

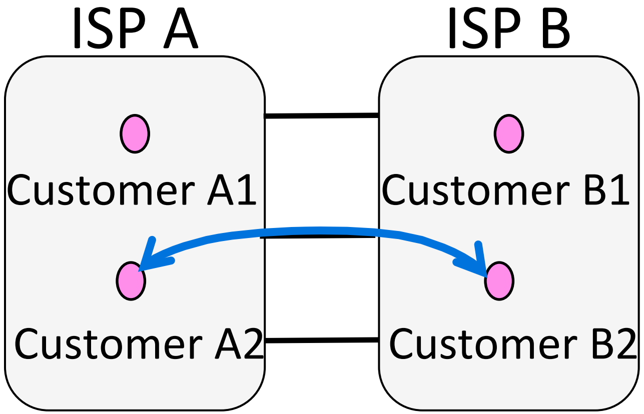
Routing Policies – Transit

- 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



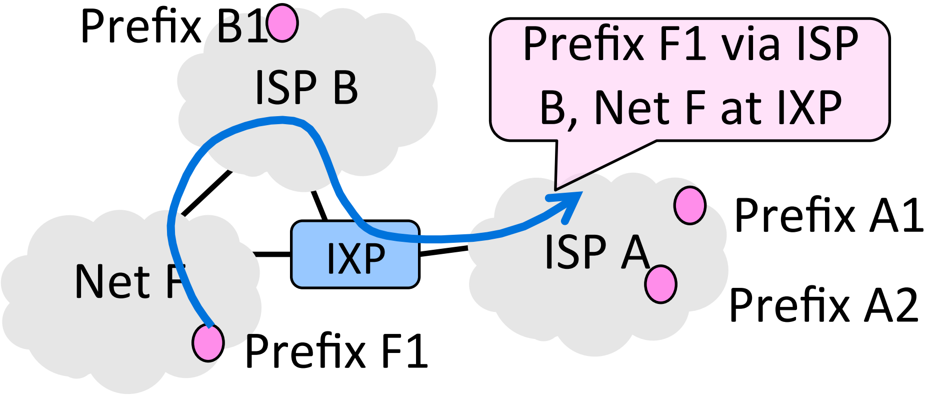
Routing Policies – Peer

- 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 (Border Gateway Protocol)

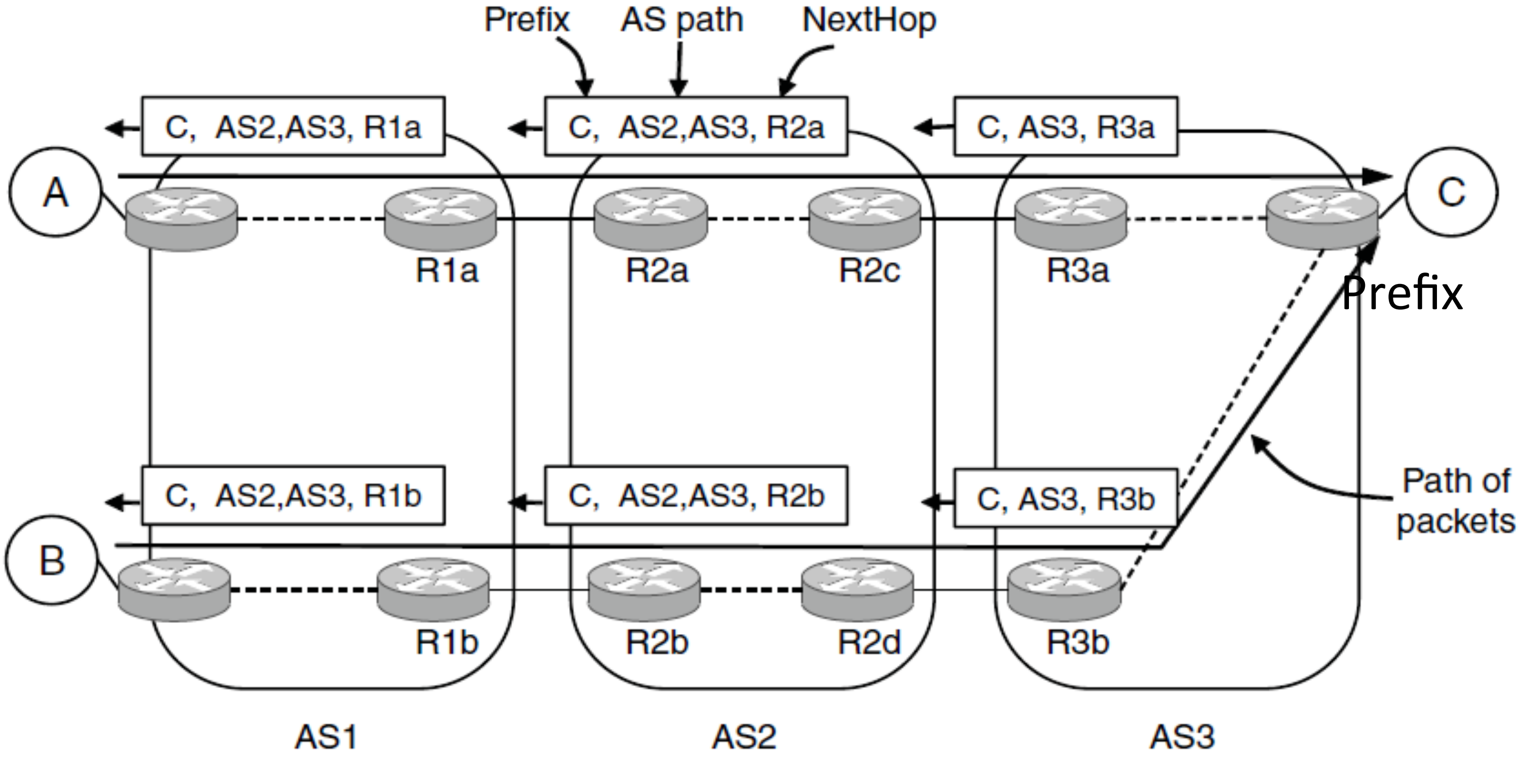
- BGP is the interdomain routing protocol used in the Internet
  - Path vector, a kind of distance vector



Routing with BGP (2)

- Different parties like ISPs are called AS (Autonomous Systems)
- Border routers of ASes announce BGP routes to each other
- Route announcements contain an 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 (3)



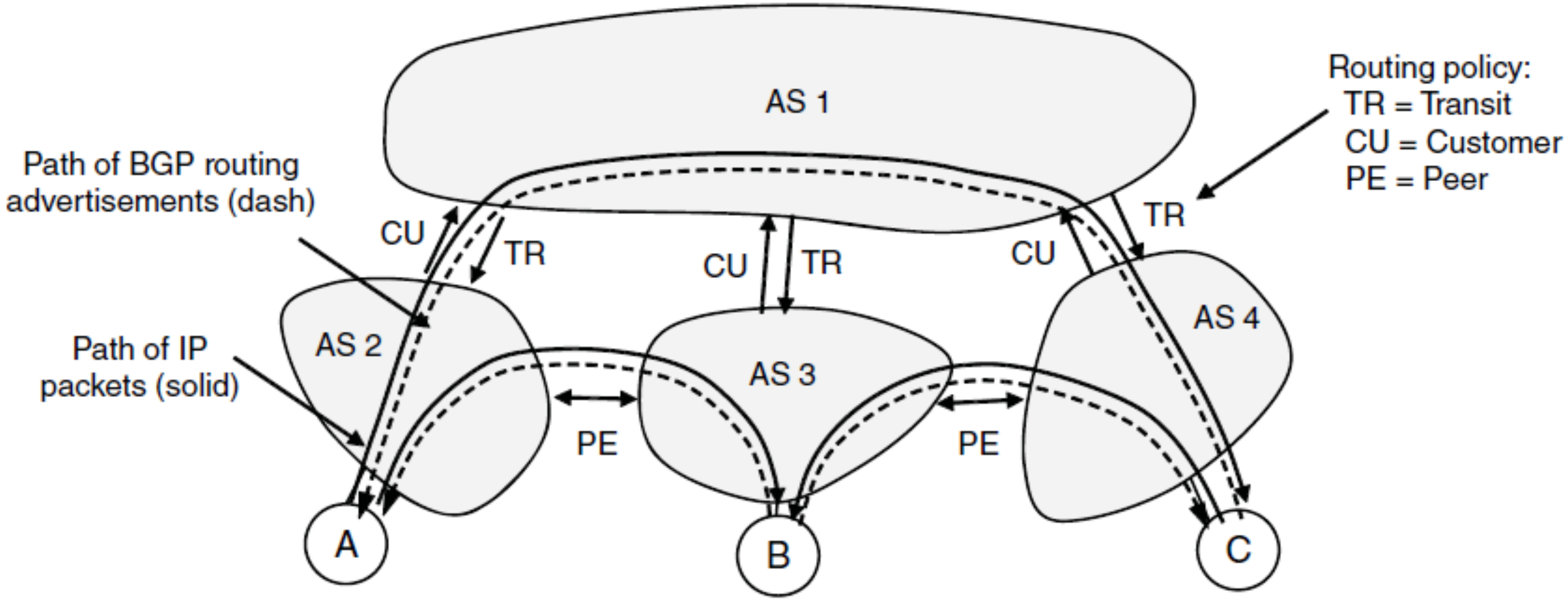
Routing with BGP (4)

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 of ISP select the best path of the ones they hear in any, non-shortest way

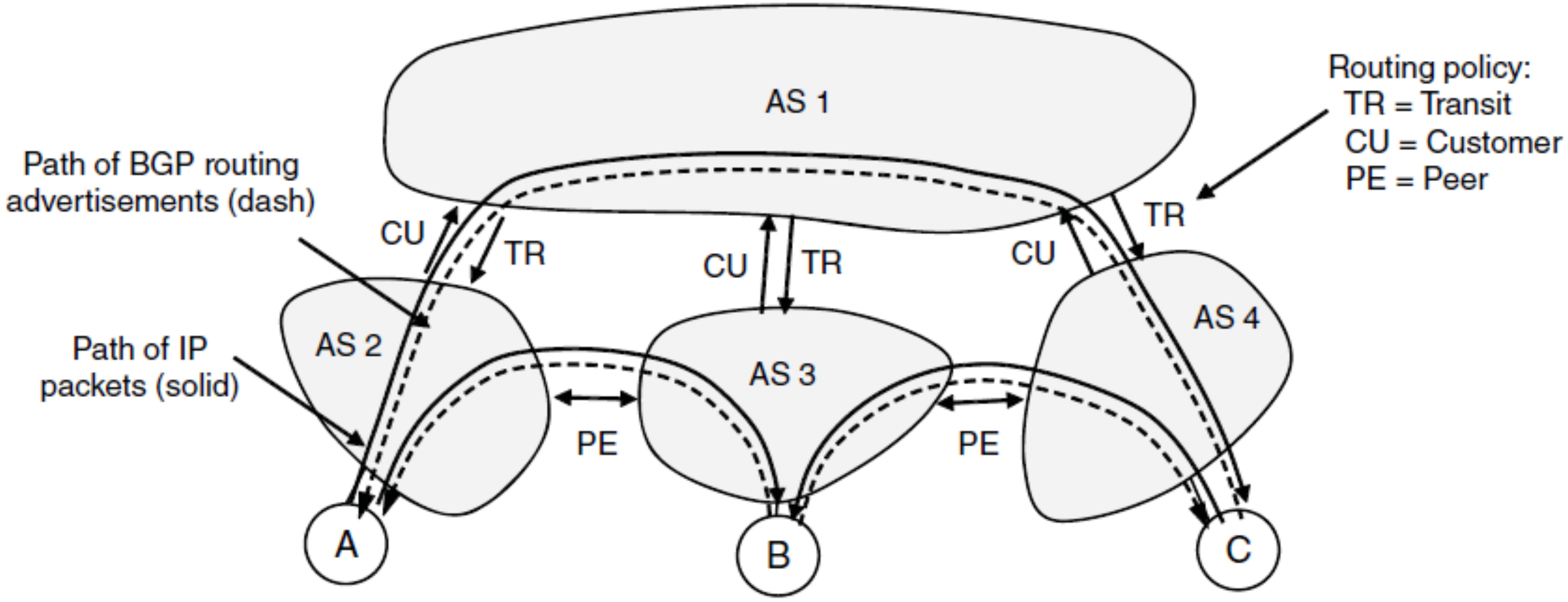
Routing with BGP (5)

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



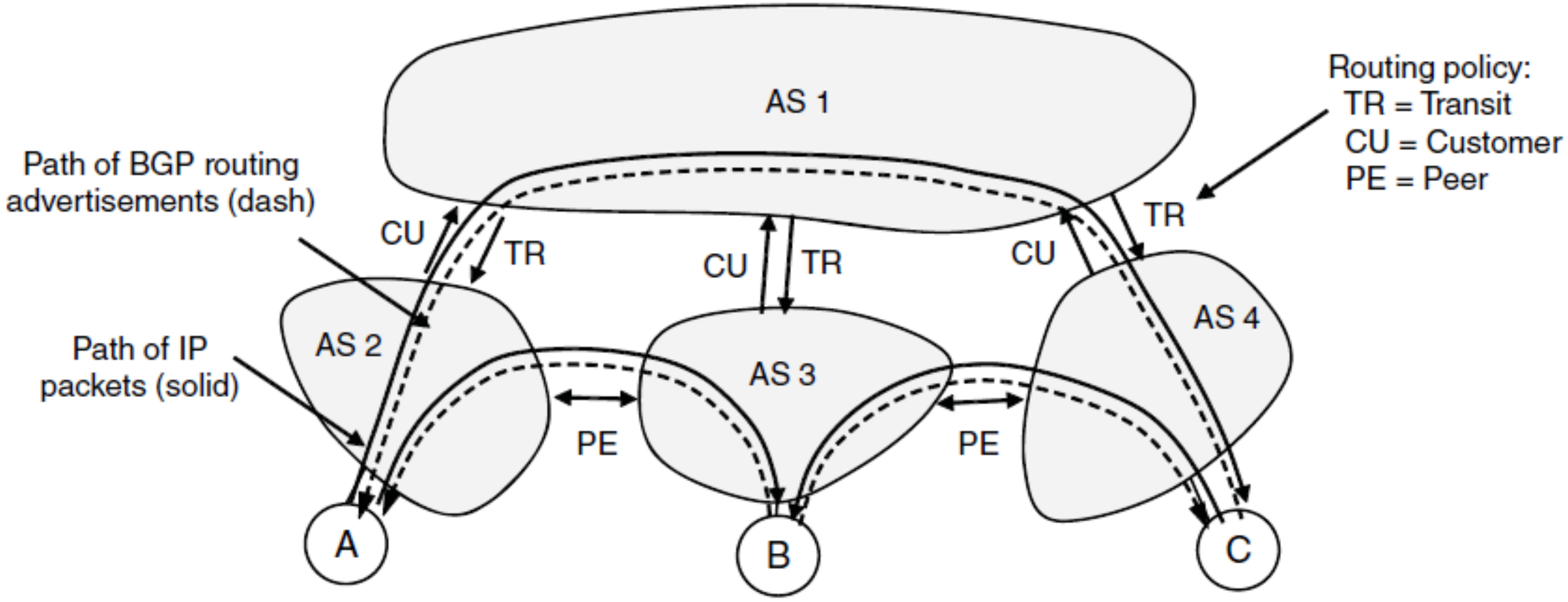
Routing with BGP (6)

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



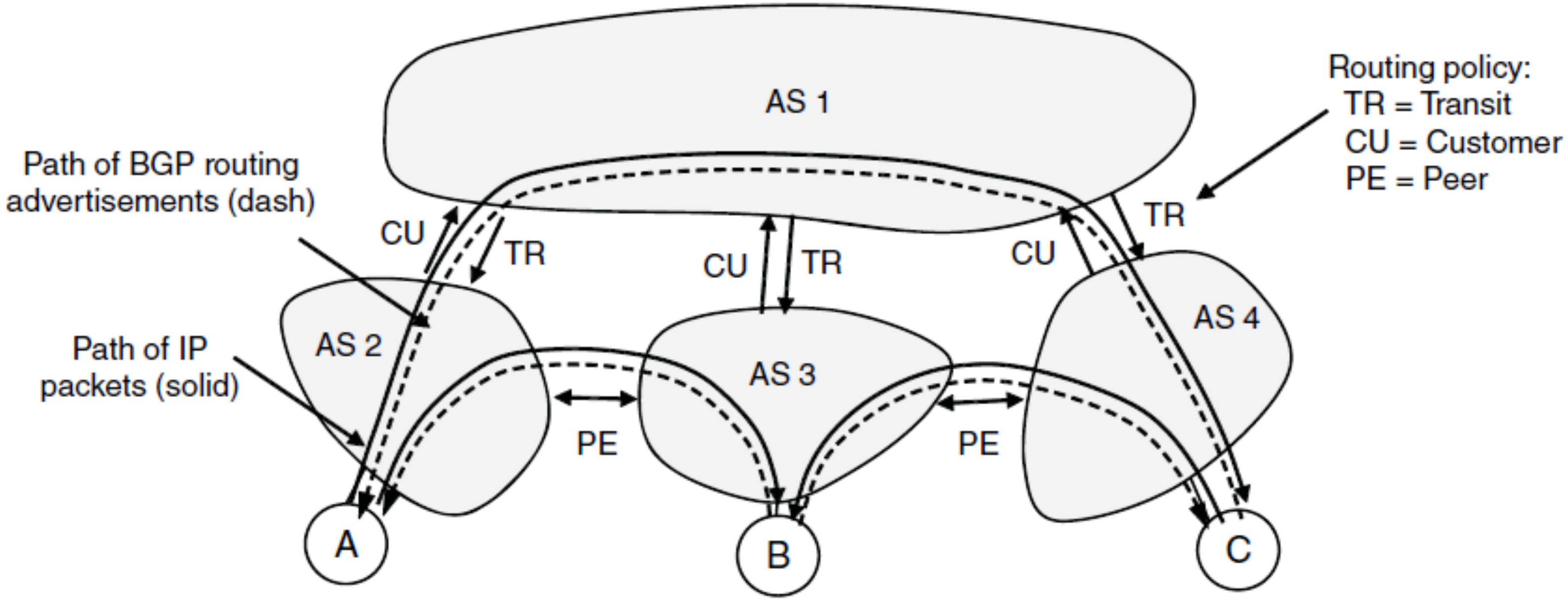
Routing with BGP (7)

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



Routing with BGP (8)

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



BGP Thoughts

- Much more beyond basics to explore!
- Policy is a substantial factor
  - Can we even be independent decisions will be sensible overall?
- Other important factors:
  - Convergence effects
  - How well it scales
  - Integration with intradomain routing
  - And more …