CSEP561 – Interdomain routing

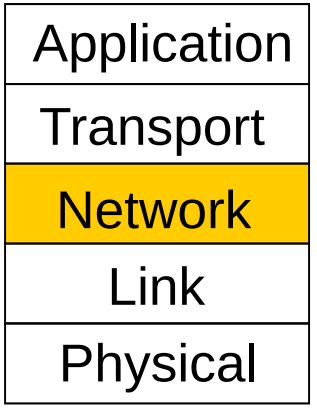
David Wetherall
djw@cs.washington.edu

Interdomain routing

- Focus:
  - Routing across internetworks made up of different parties

- Route scaling
- Route policy

- The protocol: BGP

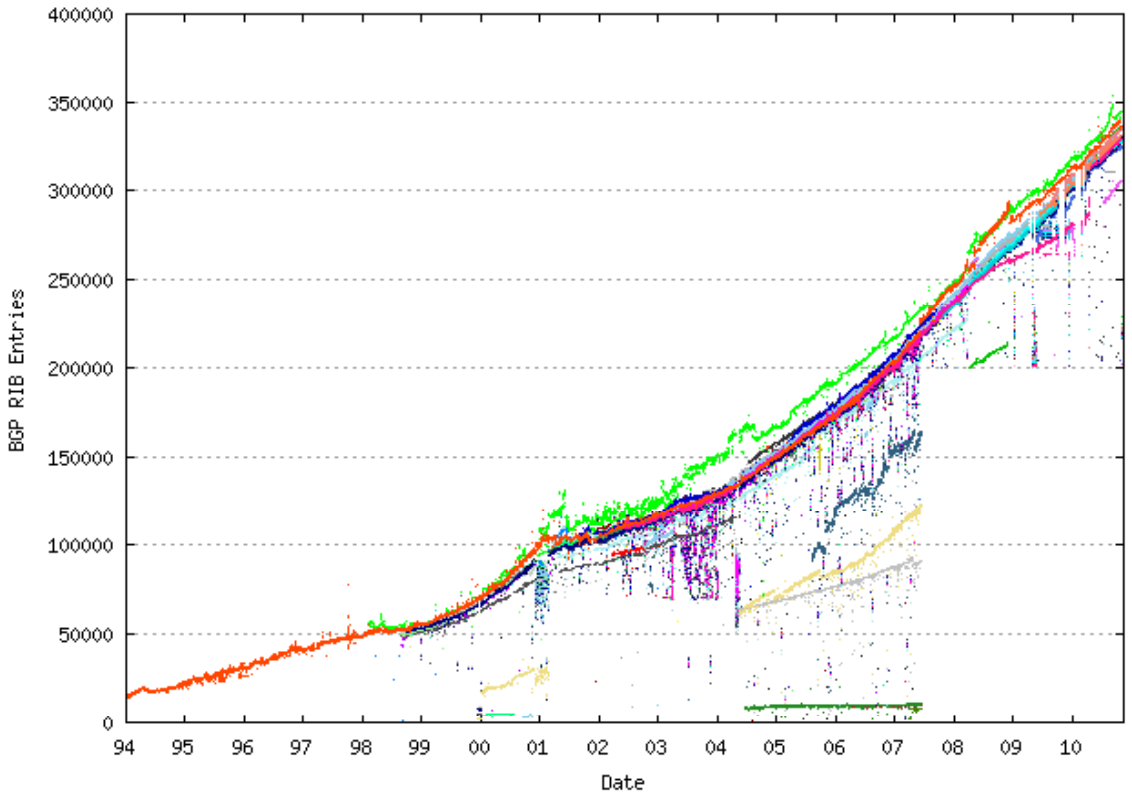


Two key problems beyond intradomain

- Scale
  - Size of routing tables, computation, messages
  - All grow with the size of the network
- Policy
  - Different parties with different goals make different decisions
  - ISPs are out to make money (locally good paths), not save the world (global shortest path)

Problem: Core BGP Table Growth 1994-2010

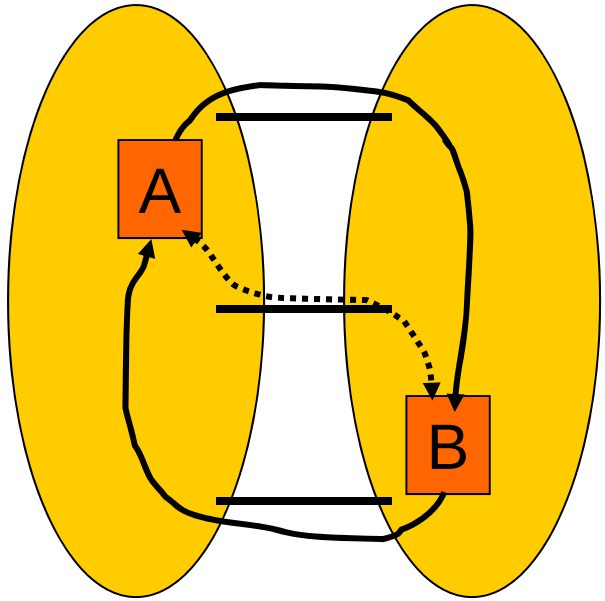
- Growth roughly indicates routing/forwarding workload



www.cidr-report.org

Problem: Independent decisions

- Early Exit / Hot Potato policy
  - "if it's not for you, get rid of it"
- Combination of best local policies is not globally best
  - Shorter paths exist
- Side-effect: route asymmetry

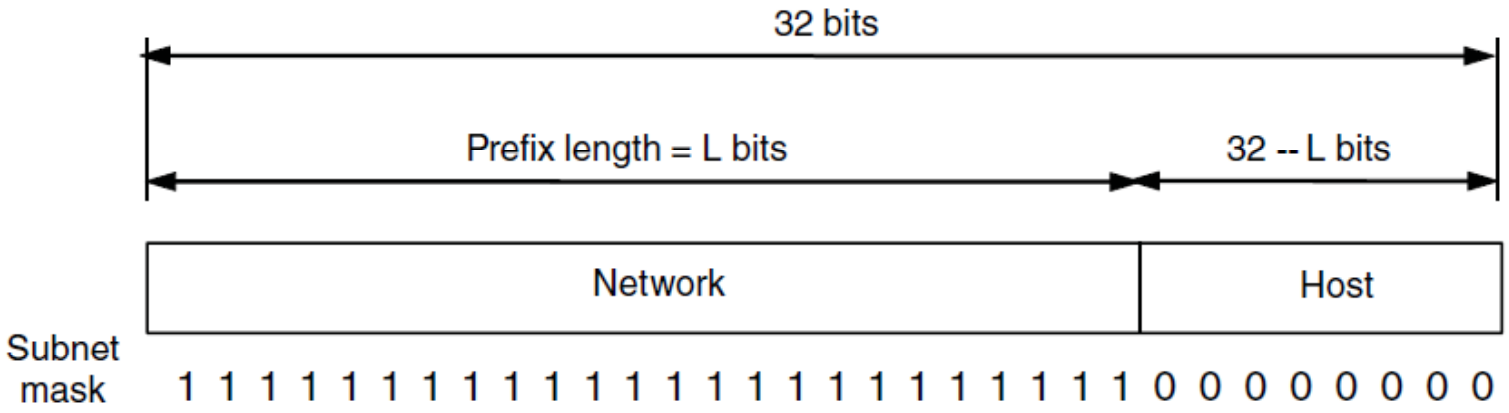


Solutions?

- Scale solution
  - Standard approach of hierarchy / information hiding
  - In the forms of prefixes and ASes
- Policy solution
  - No great solutions here!
  - Let everyone make their own decisions to the extent possible
  - Economic model gives rise to common commercial policies, e.g, transit vs peering

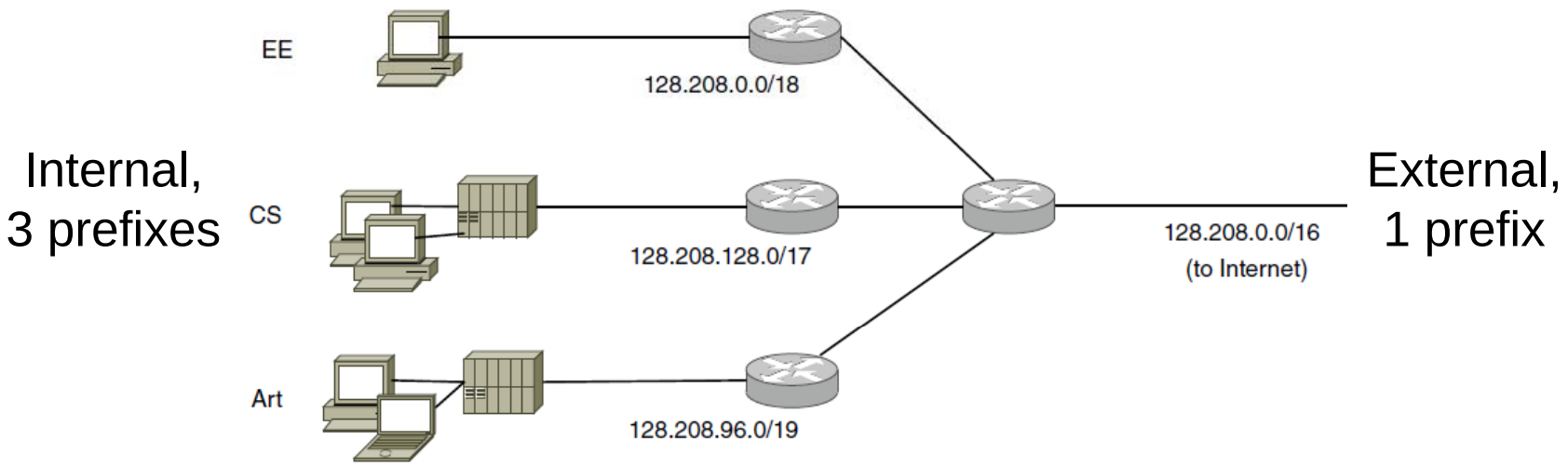
Scaling with IP prefixes

- Route to blocks of addresses called “prefixes”
  - Written as IP prefix “x.x.x.x/length” for $2^{(32\text{-length})}$ addresses
  - Replaces old fixed blocks of lengths 8, 16 and 24
  - Only store one entry for a prefix in the routing table
  - Can’t tell from an address which prefix it belongs to, so forwarding uses “longest matching prefix” rule



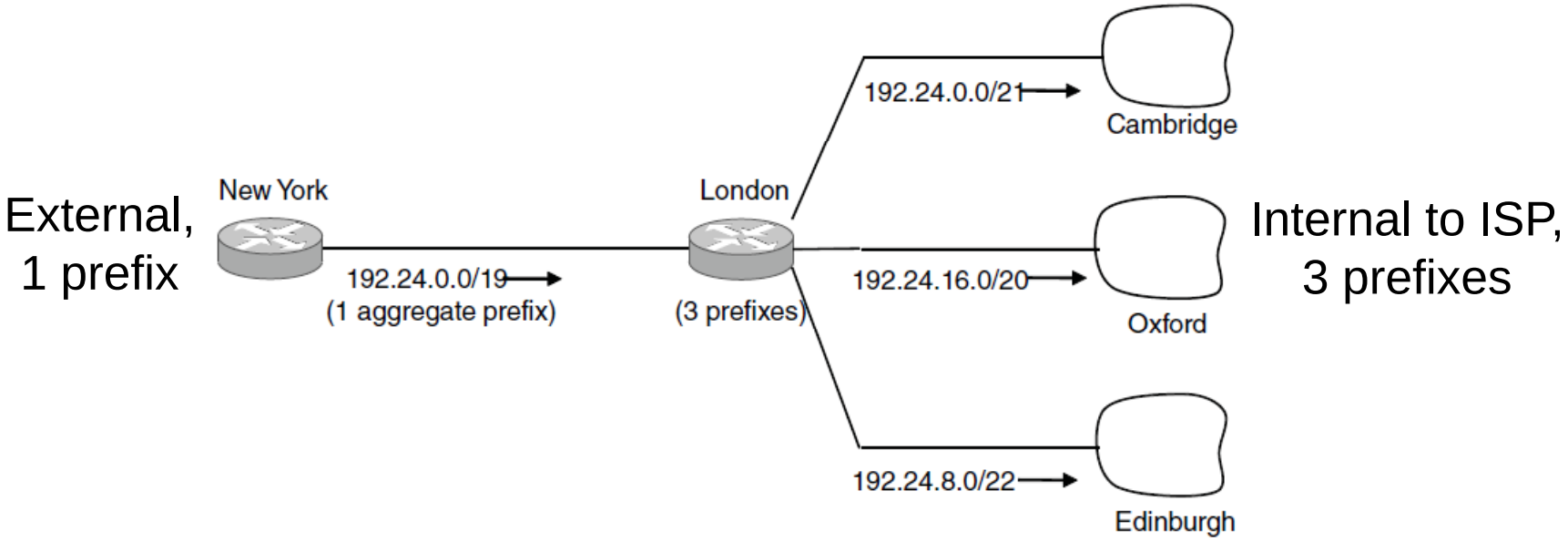
Subnetting

- Can internally divide a prefix
  - Better manageability and efficient allocation



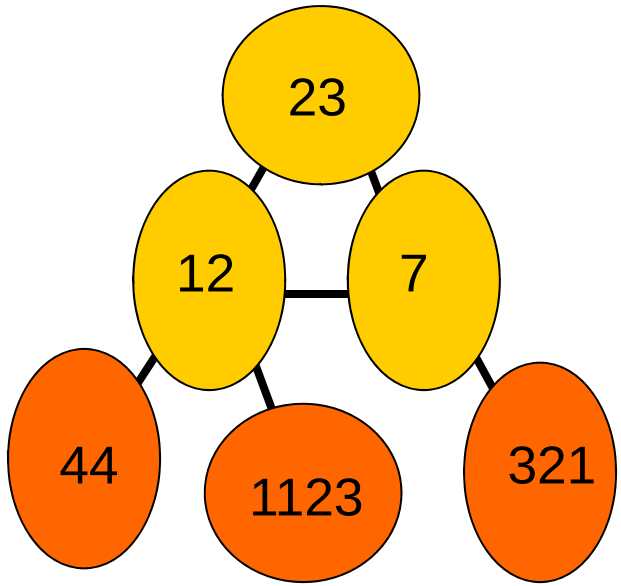
Aggregation (CIDR, supernetting)

- Can externally combine prefixes
  - Same mechanism, different goal -- smaller routing tables
  - Would reduce table size by up to half if widespread



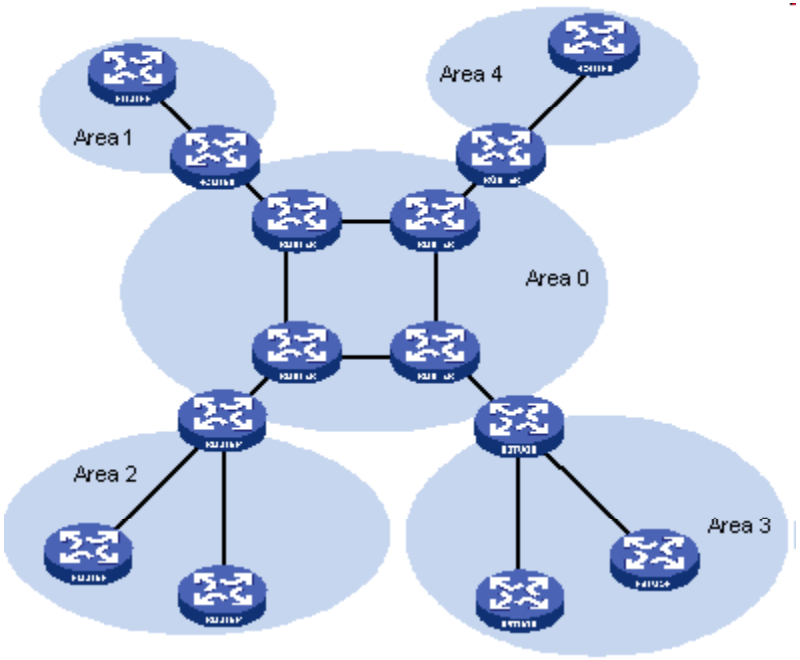
Scaling with ASes

- Network comprised of many Autonomous Systems (ASes) or domains
- To scale could use hierarchy to separate inter-domain (BGP) and intra-domain (OSPF) routing
- But not really how BGP works!



Scaling OSPF with Areas

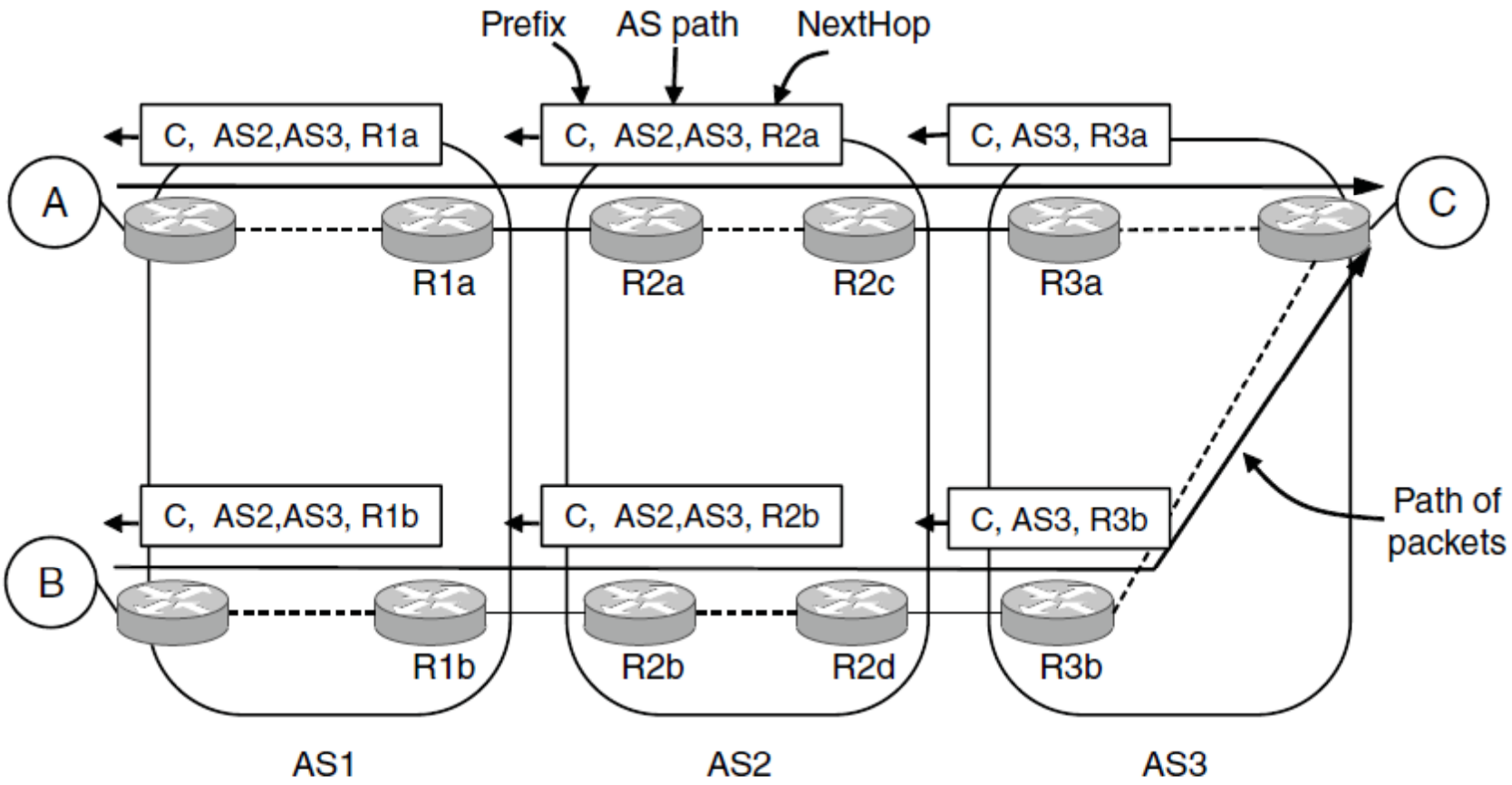
- Split large network into “areas”
  - Connected via border routers
  - Backbone area connects to all
- Border routers send a summary of the area routes to other areas
  - Hides internal area detail
- Example of applying hierarchy



BGP

- Interdomain routing protocol of the Internet
- Each AS tells other ASes the paths it is offering
  - Paths are summaries to prefixes via the sequence of ASes
  - No detailed paths of cost metrics to particular IPs
  - This happens at each border router of the AS
- Each AS picks the paths it wants to use to send traffic
  - Default rule: prefer shortest AS path, then shortest internal path
  - But selection heavily customized by ISPs
  - This happens at each border router of the AS

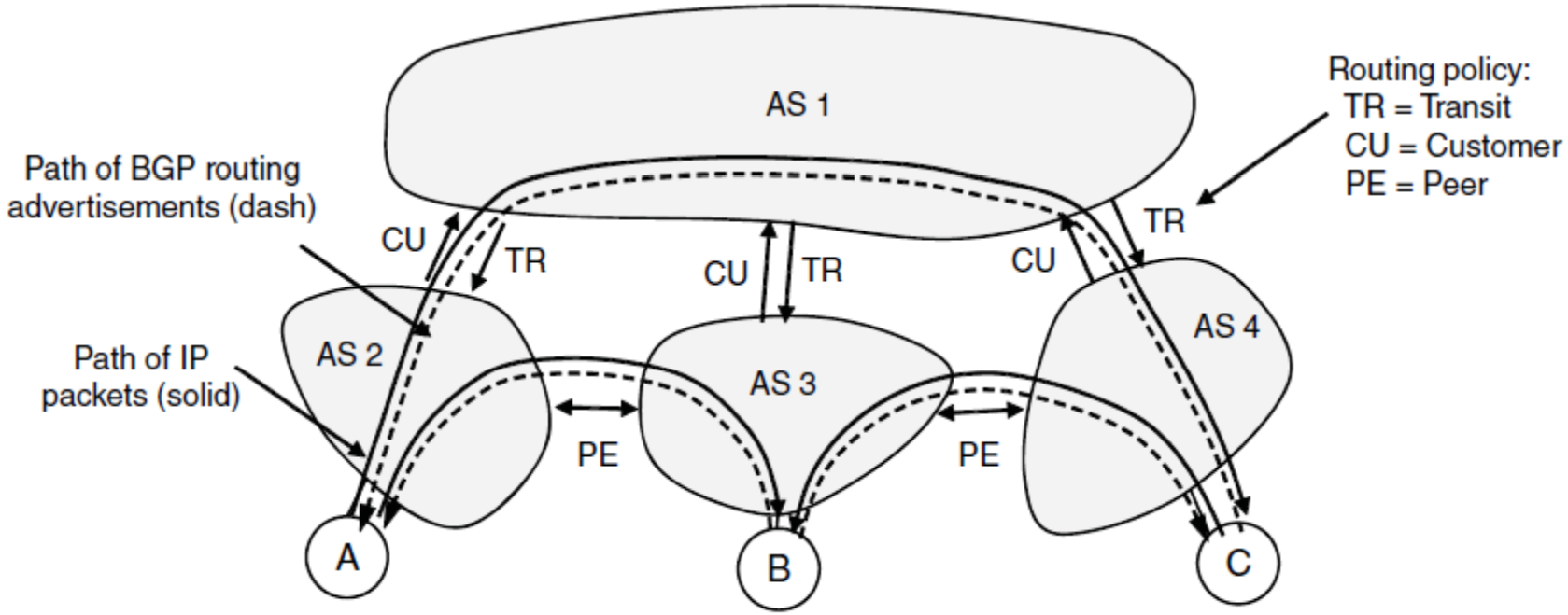
BGP


Prefix
AS path
NextHop
C, AS2,AS3, R1a
C, AS2,AS3, R2a
C, AS3, R3a
A
C
R1a
R2a
R2c
R3a
C, AS2,AS3, R1b
C, AS2,AS3, R2b
C, AS3, R3b
Path of packets
B
R1b
R2b
R2d
R3b
AS1
AS2
AS3

Policies

- Each ISP decides which routes to advertise, which to use
  - Choice of routes may depend on owner, cost, AUP, …
- Example: providers sell Transit to their customers
  - Customer announces their prefixes to provider for the rest of the Internet to reach them; Provider announces all other prefixes to customer for them to reach the rest of the Internet
- Example: parties Peer for mutual benefit
  - Peers announce path to their customer's prefixes to each other but do not propagate announcements further

Policies



- Q: What routing do A, B, and C need to do?