CSE160: Computer Networks

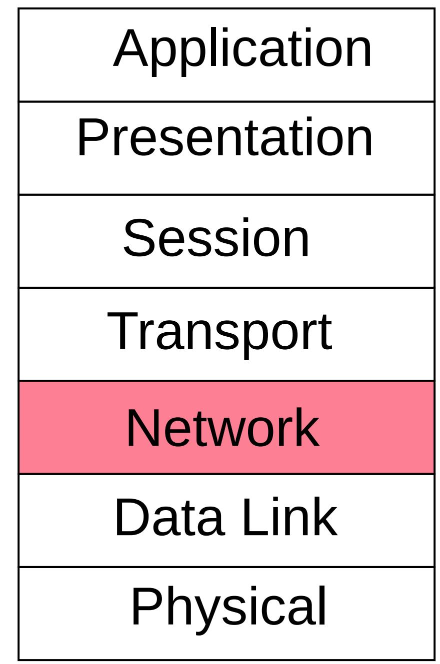
Lecture #11 – Inter-Domain Routing

2020-10-01

Professor
Alberto E. Cerpa

Last Time

- Focus
  - How do we calculate routes for packets over single and multiple paths?
  - How do we get IP addresses and MAC addresses for a destination IP?
  - How do we get more IP addresses?
- Topics
  - Distance Vector routing (RIP)
  - Equal-Cost Multipath routing (ECMP)
  - Dynamic Host Configuration Protocol (DHCP)
  - Address Resolution Protocol (ARP)
  - IPv6



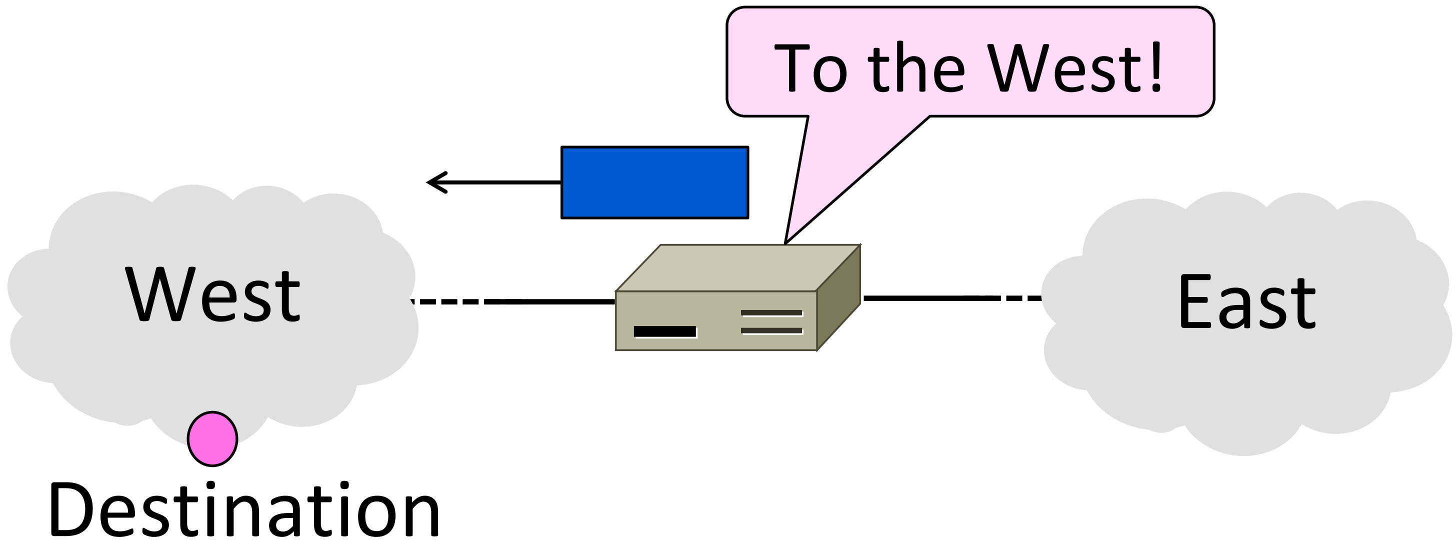
This Lecture

- Focus
  - How do we make routing scale?
- Inter-domain routing
  - Hierarchical Routing
  - ASes and BGP
  - Routing Policies

Application
Presentation
Session
Transport
Network
Data Link
Physical

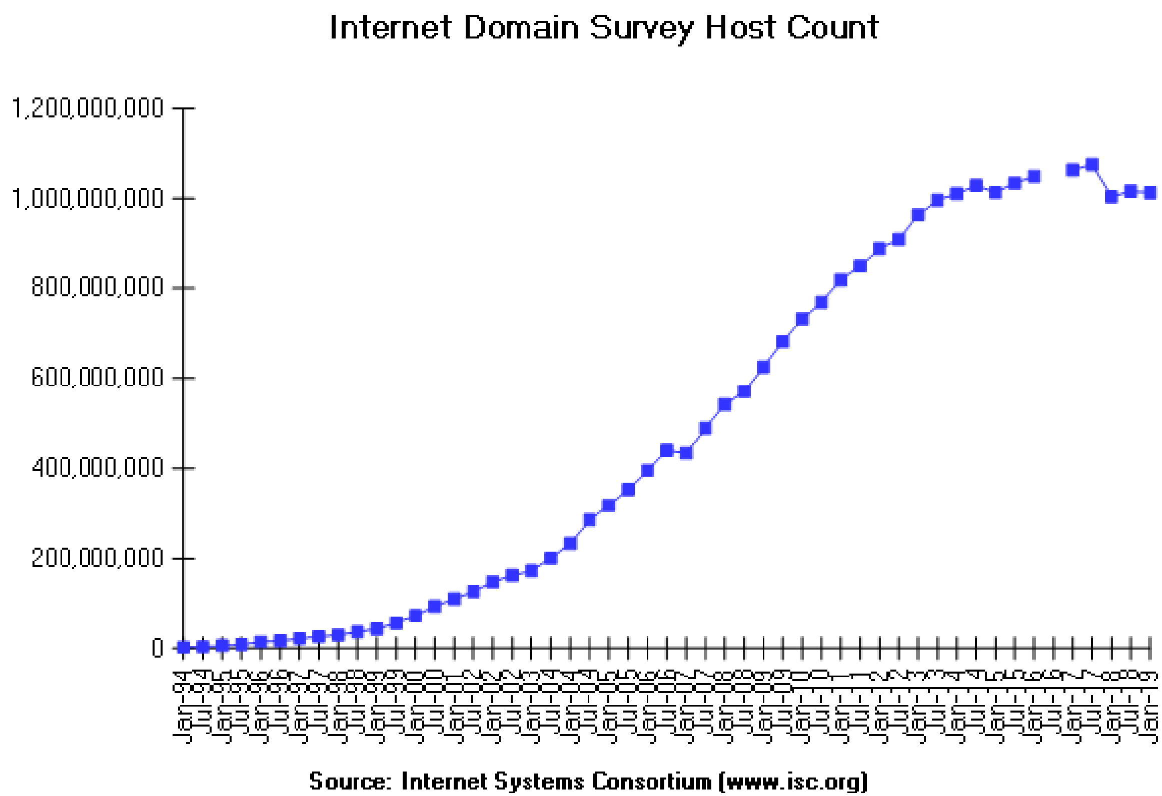
Hierarchical Routing

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



Internet Growth

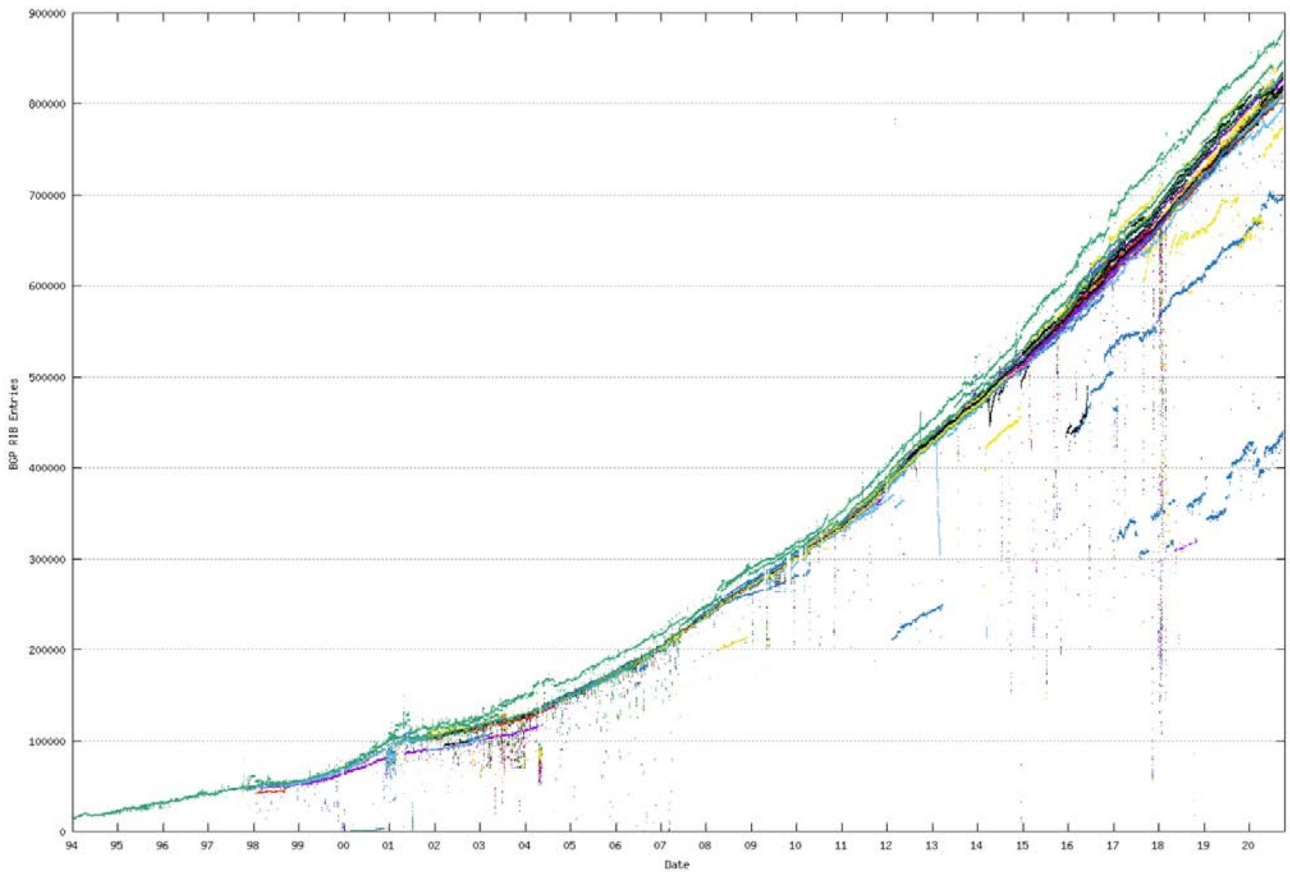
- At least a billion+ Internet hosts and growing…
- Considering both IPv4 and IPv6 addresses more like ~1.7 billion
- If we count total number of devices behind NAT boxes closer to 2+ billions



Internet Routing Growth

- Internet growth translates into routing table growth (even using prefixes, more on this later)

BGP Statistics from Route-Views Data



Report Date: 28 Sep 2020 18:02 UTC+1000
Source: http://bgp.potaroo.net/bgprpts/rva-index.html

Impact of Routing Growth

1. Forwarding tables grow
   – Larger router memories, may increase lookup time
2. Routing messages grow
   – Need to keep 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. Network Hierarchies
   – Route to network regions
2. IP prefixes
   – Route to blocks of hosts
3. IP prefix aggregation
   – Combine and split prefixes

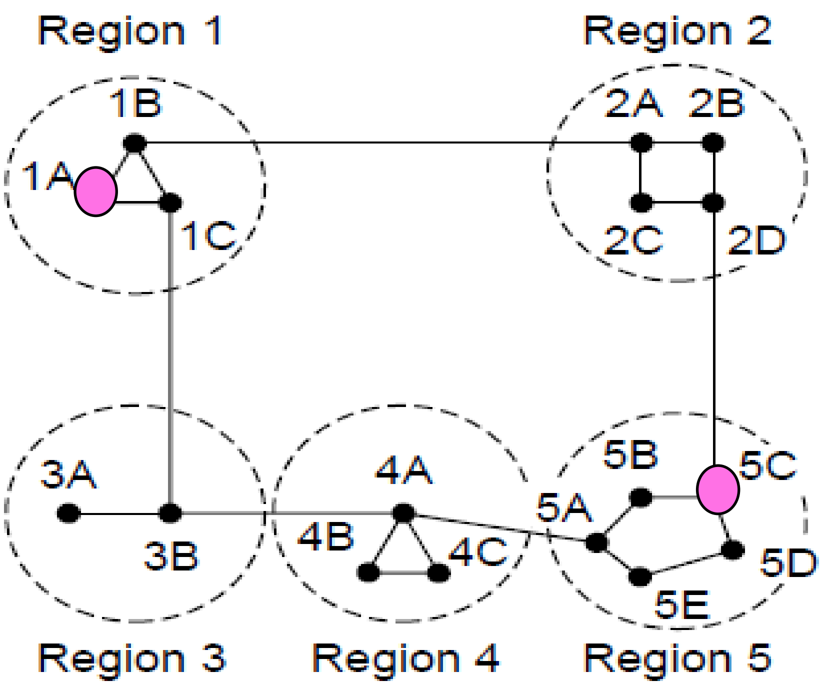
This time

Next time

Hierarchical Routing

- Introduce a larger routing unit
  - IP prefix (hosts) ← from one host (next class)
  - Region, e.g., ISP network
- Route first to the region, then to the IP prefix within a region
  - Hide details within a region from outside of the region
- A routing hierarchy effectively reduces the size of the internetwork:
  - Divide and conquer
  - Allows scaling of routing algorithms
  - It helps mitigate: (a) Volume of routing messages at edges; (b) Amount of routing computation
- But this alone is not enough, it's just part of the solution (more on this later)

Hierarchical Routing Example



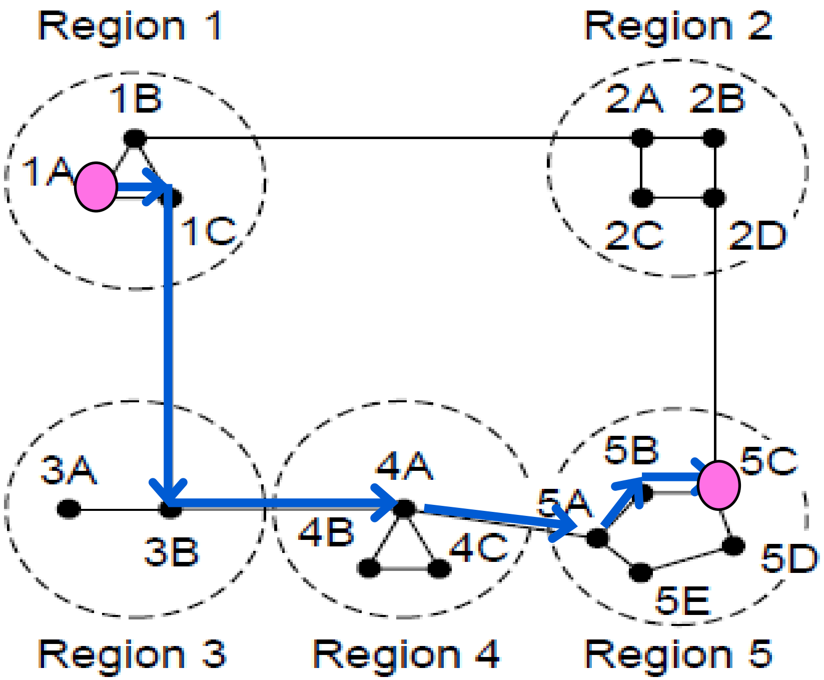
Full table for 1A

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 Example (2)



Full table for 1A

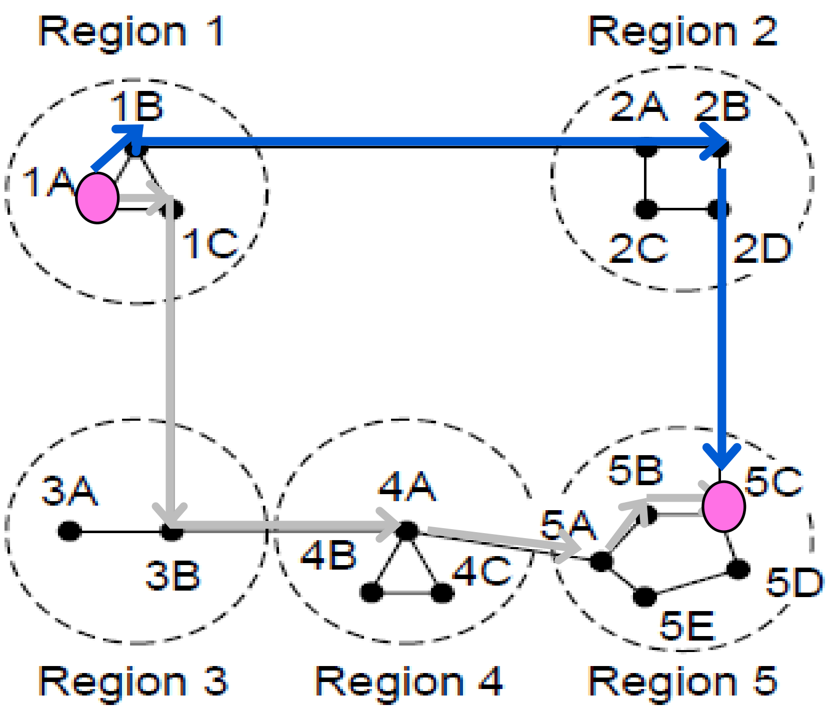
Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

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 Example (3)

- 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

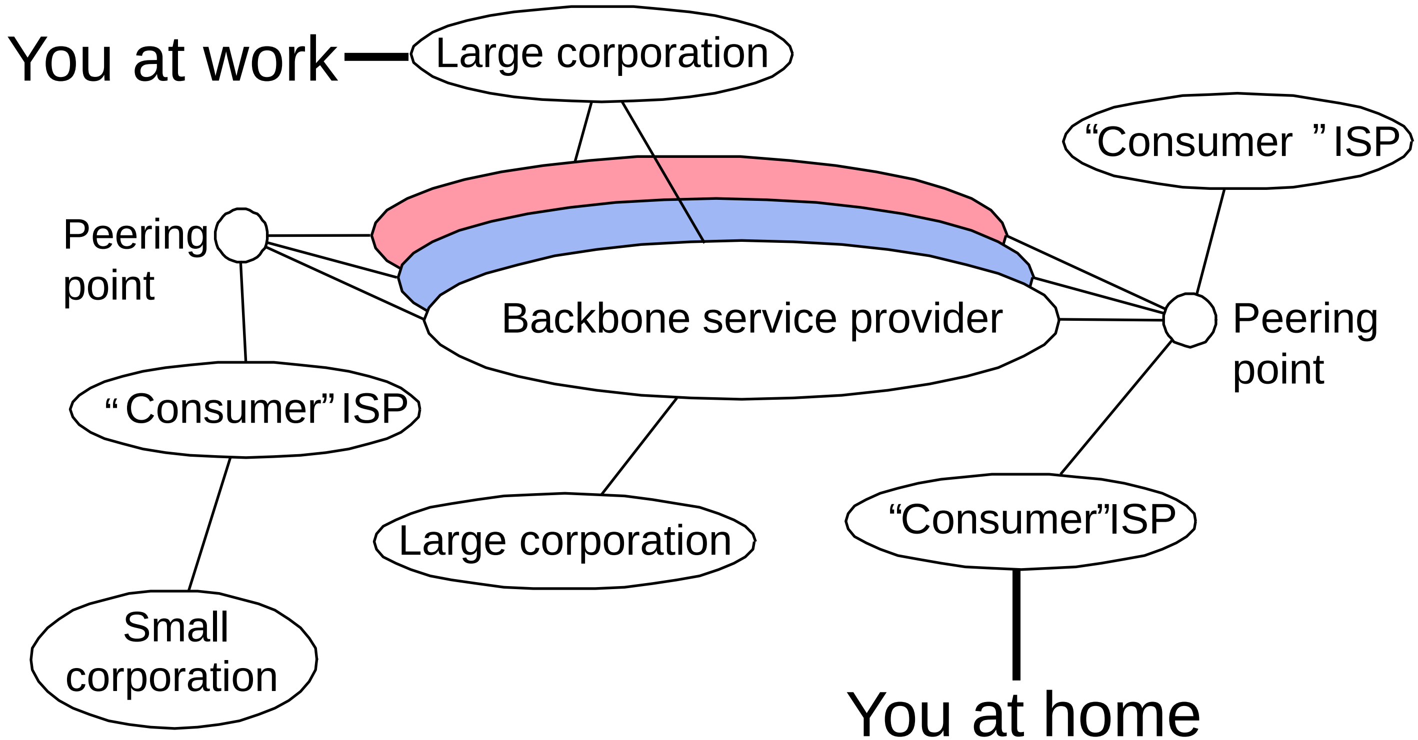
1C is best route to region 5, except for destination 5C

Observations

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

Structure of the Internet

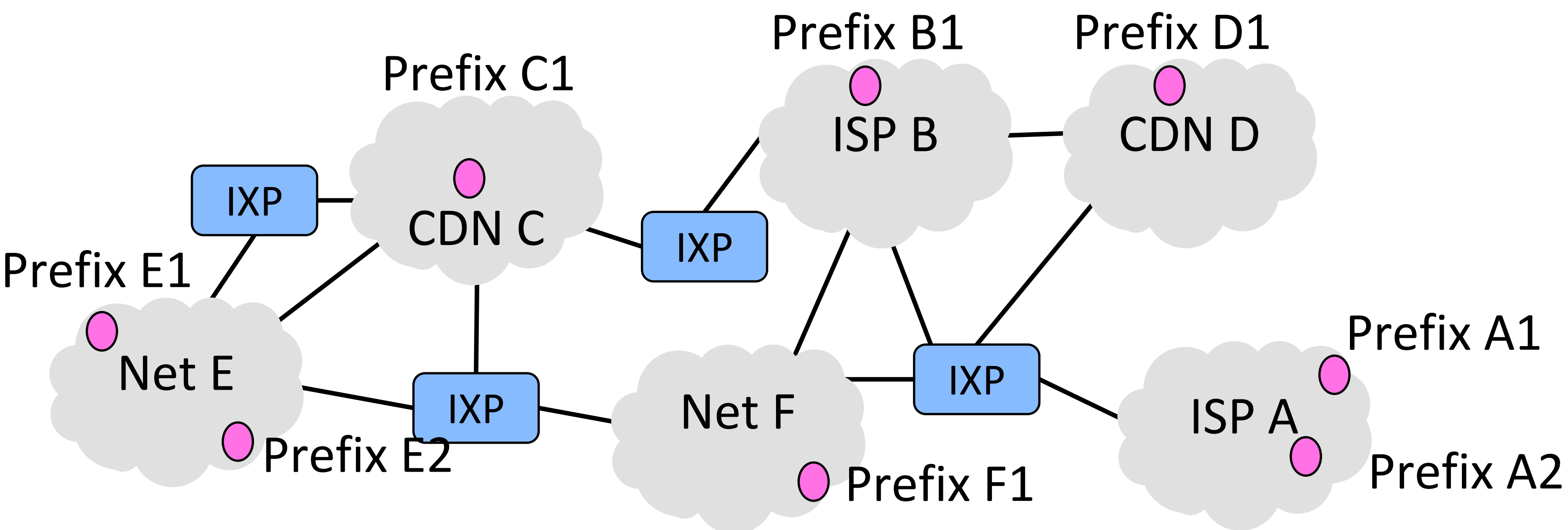
- Inter-domain versus intra-domain routing



- **Tier 1** (e.g. Seabone, Sprint): international operator interconnecting major towns by long-distance, broadband links and big traffic flows;
- **Tier 2** (e.g. Telecom Argentina): national operator collecting traffic from single users through a lot of access points;
- **Tier 3**: local operator serving a very restricted geographical area.

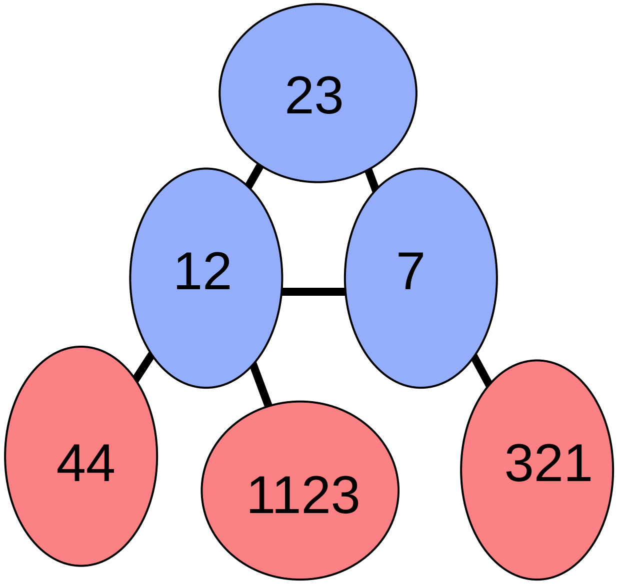
Structure of the Internet (2)

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



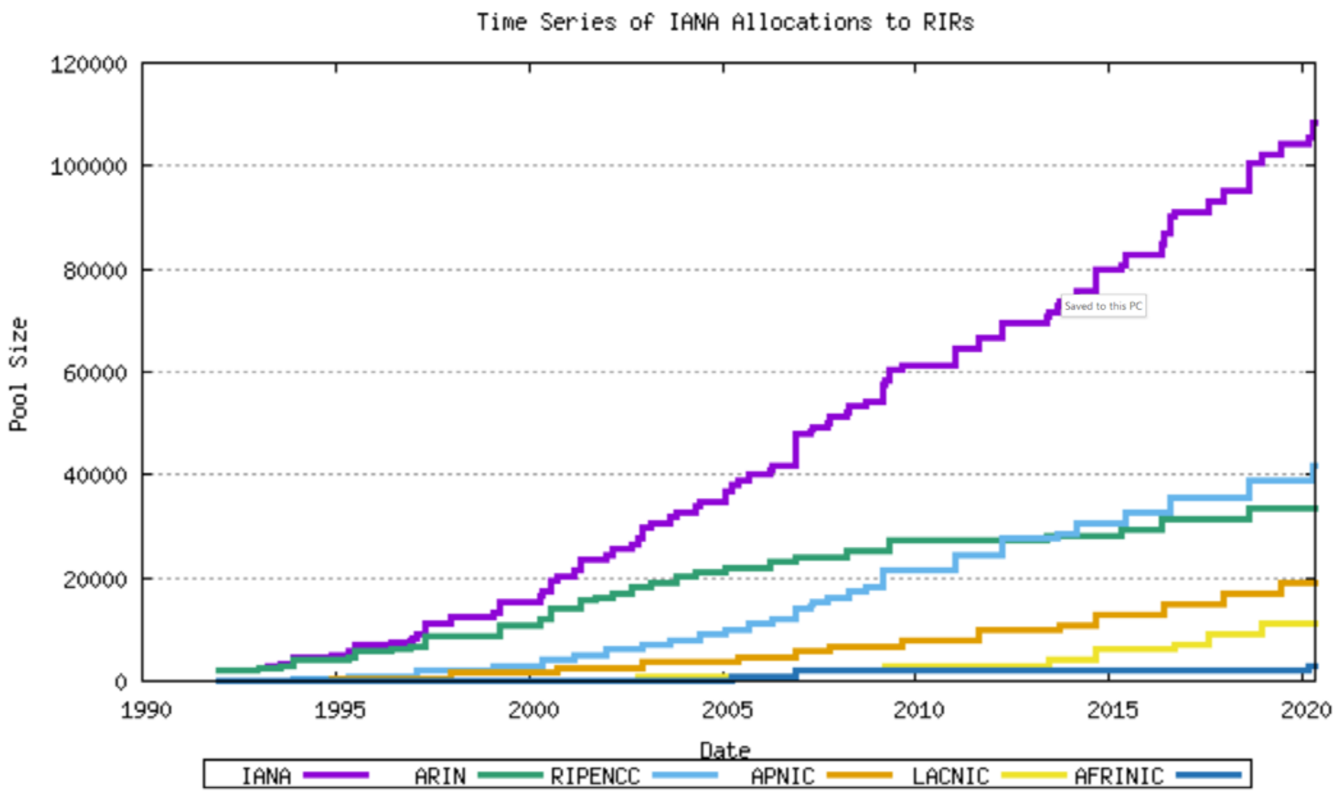
Inter-Domain Routing

- Network comprised of many Autonomous Systems (ASes) or domains
- To scale, use hierarchy: separate inter-domain and intra-domain routing
- Also called interior vs exterior gateway protocols (IGP/EGP)
  - IGP = RIP, OSPF
  - EGP = EGP, BGP



How many ASes are in the planet?

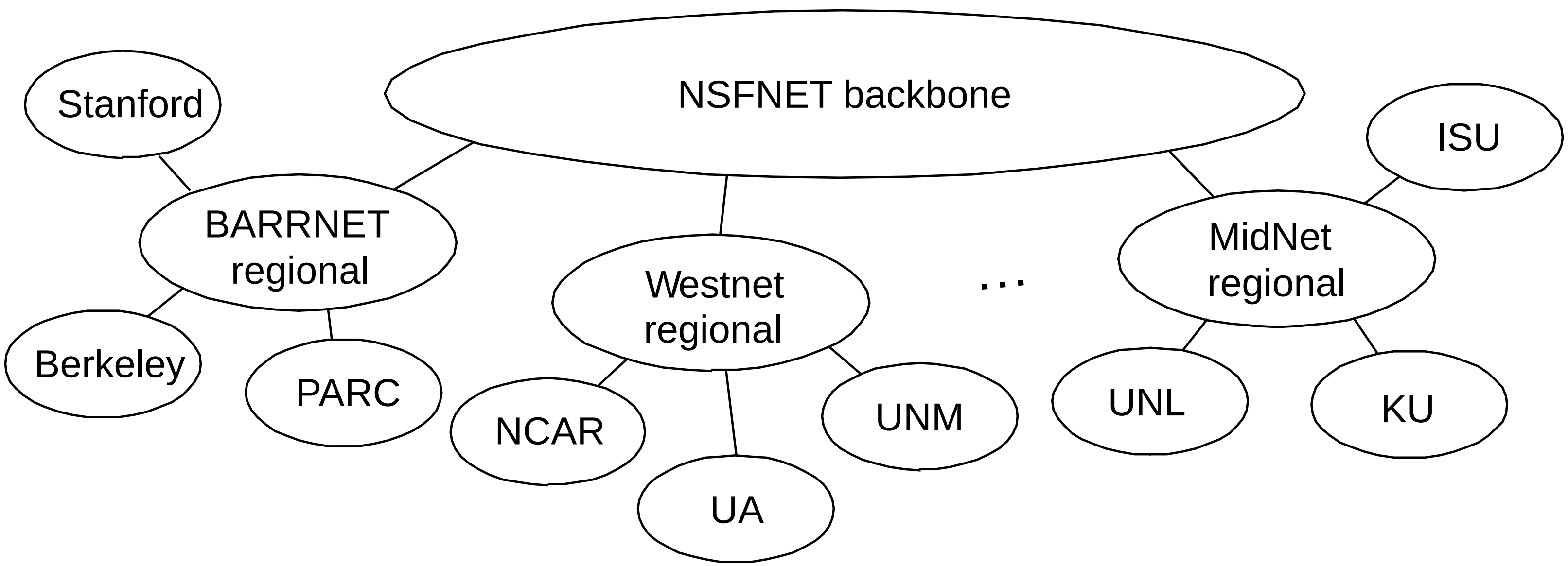
- As of this week: 110,589
- Europe: 42,528; US/Canada: 31,561; Asia & Oceania: 19,095; Latin America: 14,079; Africa: 3,326
- AS numbers assigned by regional Internet Assigned Numbers Authority (IANA) like APNIC



Source: http://www.potaroo.net/tools/asn32/
Date: 28-Sep-2020 07:55 UTC

Exterior Gateway Protocol (EGP)

- First major inter-domain routing protocol
- Constrained Internet to tree structure; no longer in use



Border Gateway Protocol (BGP-4)

- EGP used in the Internet backbone today
- Features:
  - Path vector routing
  - Operates over reliable transport (TCP)
  - Application of policy
  - Uses route aggregation (CIDR) (next class)

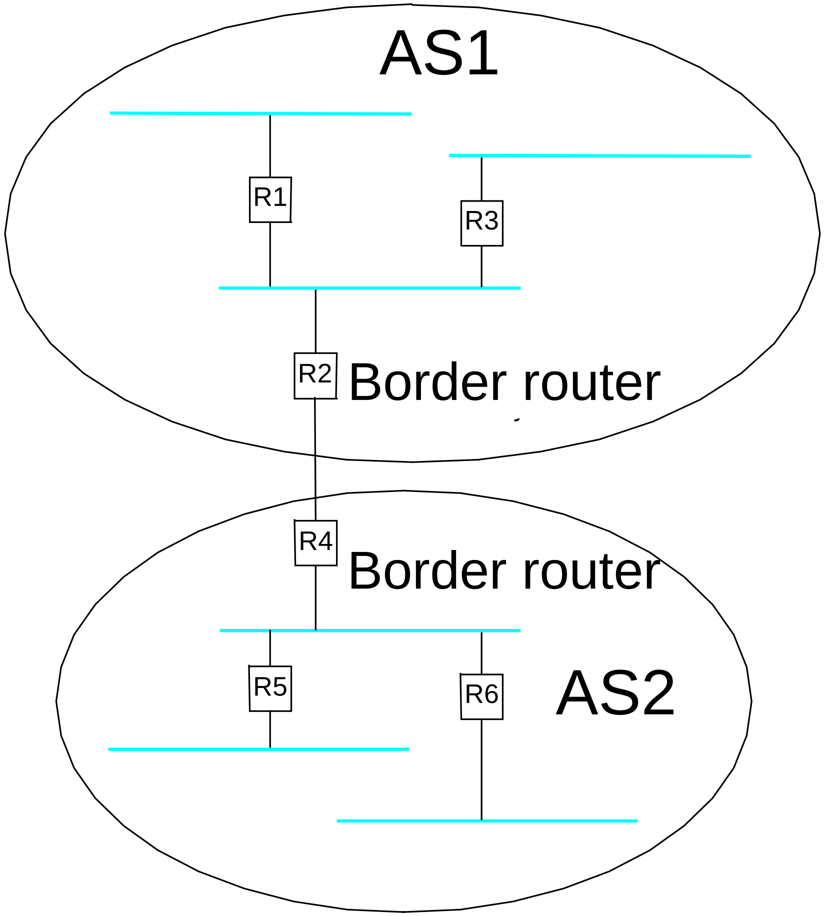
Inter-Domain Routing

- Border routers *summarize* and advertise internal routes to external neighbors and vice-versa

- Border routers apply <u>policy</u>

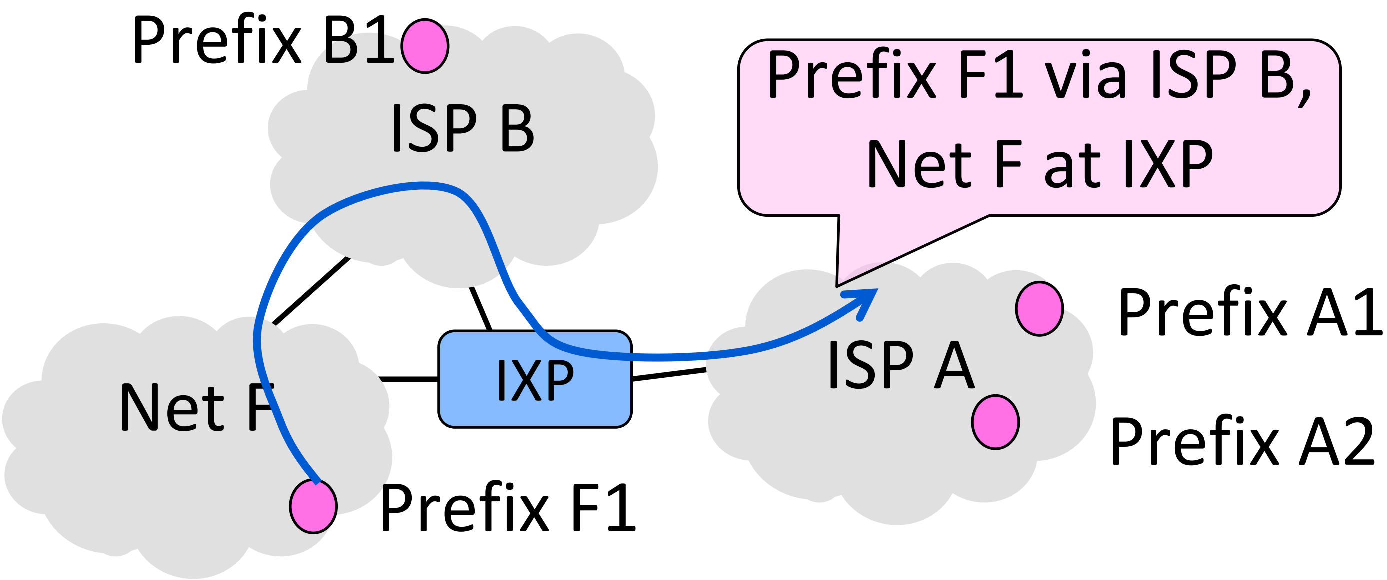
- Internal routers can use notion of default routes

- Core is “default-free”; routers must have a route to all networks in the world



Routing with BGP

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



Routing with BGP (2)

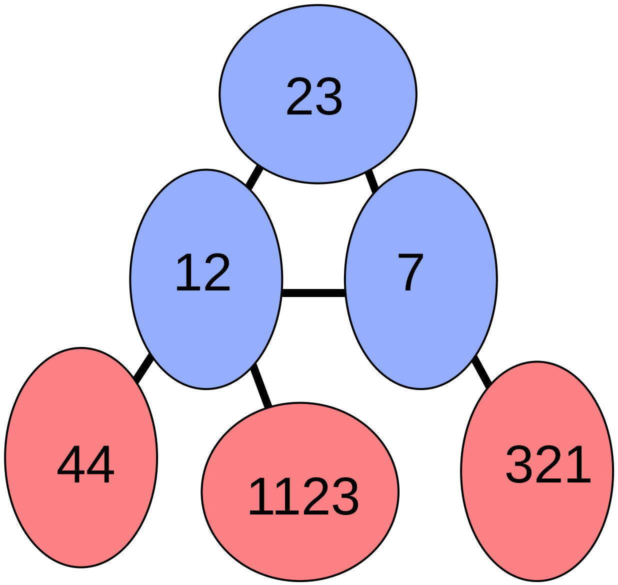
- Parties like ISPs are called AS (Autonomous Systems)
- ASes MANUALLY configure their internal BGP routes
- External routes go through complicated filters for forwarding/filtering (more on this with routing policy)
- AS BGP routers communicate with each other to keep consistent routing rules
- Border 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 prefix; list is to find loops
- Route announcements move in the opposite direction to traffic

Path Vectors

- Similar to distance vector, except send entire paths
  - e.g. 321 hears [7,12,44]
  - stronger avoidance of loops
  - supports policies (later)

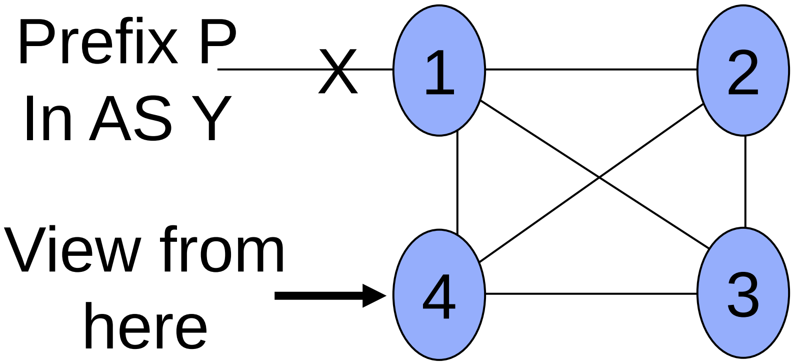
- Modulo policy, shorter paths are chosen in preference to longer ones

- Reachability only – no metrics



An Ironic Twist on Convergence

- Research has shown that BGP convergence can undergo a process analogous to count-to-infinity!



- AS 4 uses path 4 1 Y. A link fails and 1 withdraws 4 1 Y.
- So 4 uses 4 2 1 Y, which is soon withdrawn, then 4 3 2 1 Y, …
- Result is many invalid paths can be explored before convergence
- Why?

Operation over TCP

- Most routing protocols operate over UDP/IP
- BGP uses TCP
  - TCP handles error control; reacts to congestion
  - Allows for incremental updates
- Issue: Data vs. Control plane
  - Shouldn't routing messages be higher priority than data?

Internet-wide Routing Issues

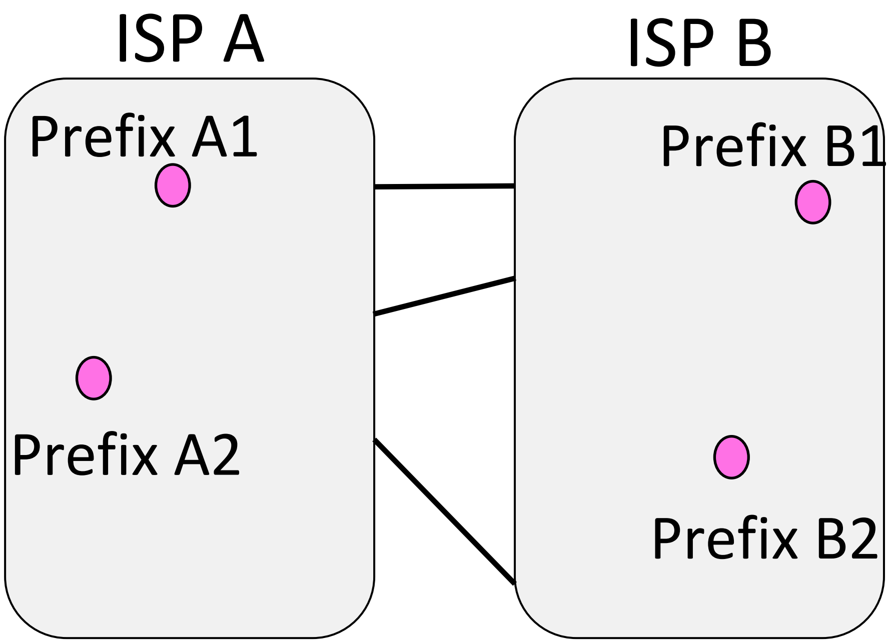
- Two problems beyond routing within an individual network

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

Yikes!

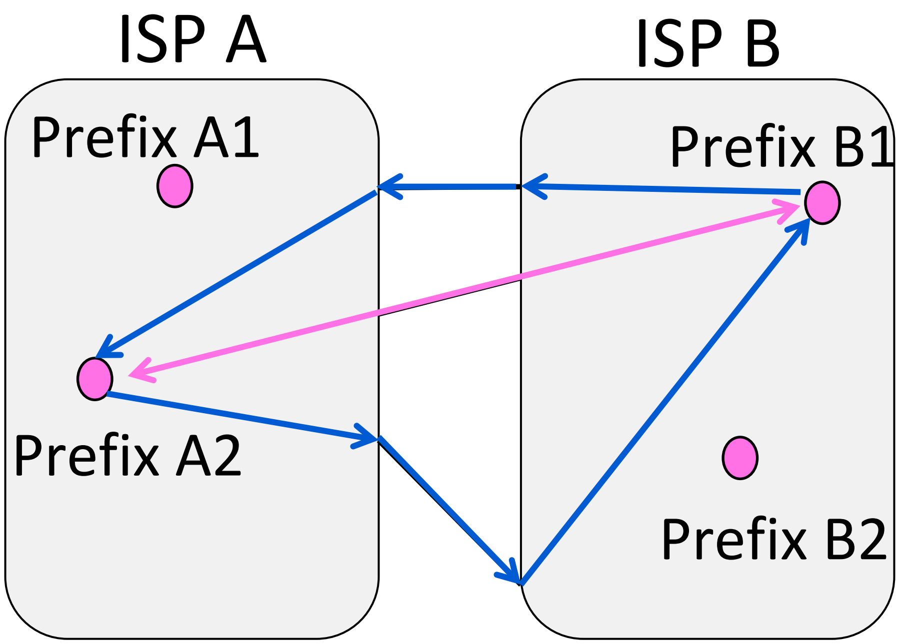
Effect of Independent Parties

- 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?



Effect of Independent Parties (2)

- Selected paths are longer than overall shortest paths
  - And asymmetric too!
- This is a consequence of independent goals and decisions, not hierarchy



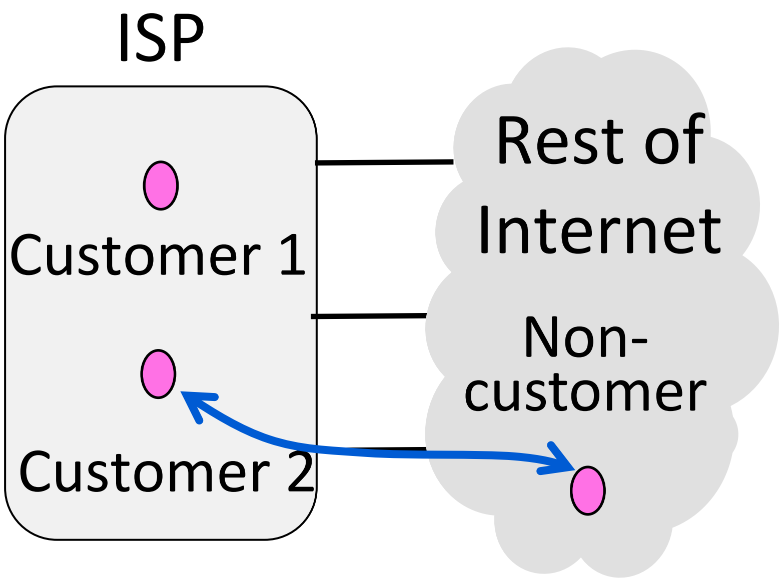
Routing Policies

- 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

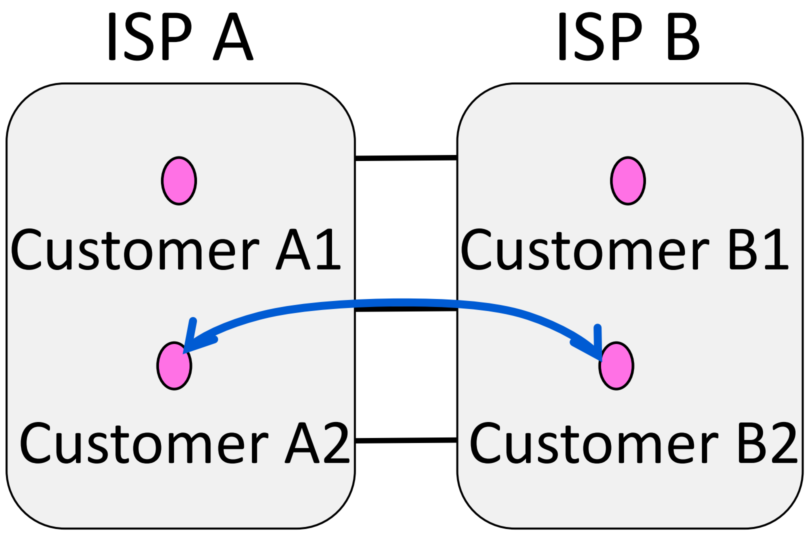
Routing Policies – Transit

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



Routing Policies – Peer

- Both party (ISP 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



Simplified Policy Roles

- Providers sell Transit to their customers
  - Customer announces path to their prefixes to providers in order for the rest of the Internet to reach their prefixes
  - Providers announces path to all other Internet prefixes to customer C in order for C to reach the rest of the Internet
- Additionally, parties Peer for mutual benefit
  - Peers A and B announce path to their customer's prefixes to each other but do not propagate announcements further
  - Peering relationships aren't transitive
  - Tier 1s peer to provide global reachability

Transit Agreement

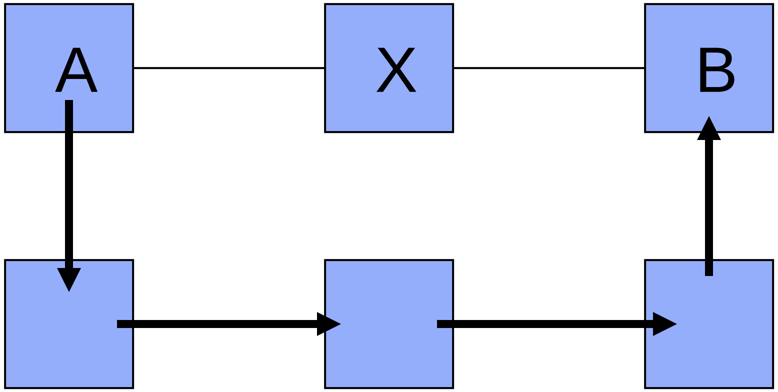
- ISP has to pay another ISP to connect to its AS. The economic agreement may establish:
  - Payment method:
    - Fee by volume: bytes/months or bytes/day + extra fees for exceeding traffic
    - Flat fee: monthly fee for maximum bandwidth (bytes/sec)
  - Destinations reachability:
    - Full route: all destinations around the world must be reachable
    - Geographical area: only destinations in a certain area (e.g. USA), packets to other destinations dropped.
- Price influenced by “importance” of ISP selling transit
  - An US ISP is important because inside its AS there are the most visited web servers in the world
  - A very large ISP can offer good reachability world wide

Peering Agreement

- When two peer ISPs agree to exchange traffic between themselves without having to pay each other.
  - The costs for direct interconnection are lower than the costs for buying transit from each other through a higher tier operator
  - Costs for setup and maintenance of the direct link between the ASes are equally split by the two ISPs
  - They can send data at the full speed allowed by the link.
- Tier-1 operators work in a very competitive market:
  - Tier-2 operators can establish new peering agreements among themselves as soon as they become more convenient than transit
  - A Tier-2 operator can shortly move to a more convenient Tier-1 operator
  - A dominant operator may be forced by the market guarantor to offer peering connections with minor ISPs

Policies

- Choice of routes may depend on owner, cost, etc…
  - Business considerations
- Local policy dictates what route will be chosen and what routes will be advertised!
  - e.g., X doesn't provide transit for B, or A prefers not to use X

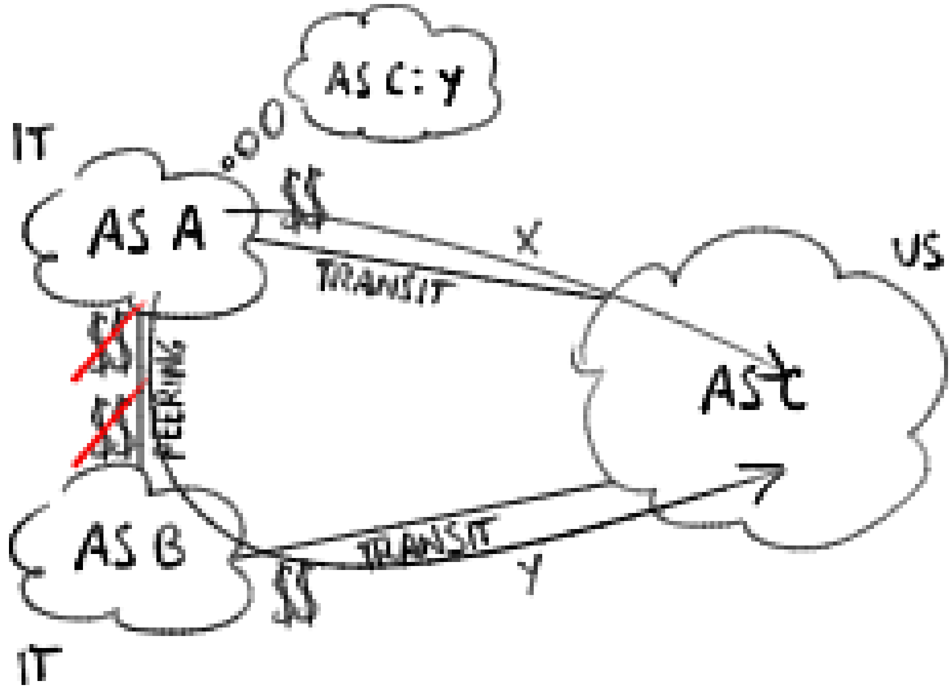


Routing Policies

- Requirements:
  - Economic (who pays for the bandwidth?): sometimes longer paths may be preferred to best paths;
  - Administrative (is it allowed to go?): sometimes some paths are omitted to the other party;
  - Security (is that administrative domain trusted?): sometimes safer (and longer) paths may be preferred to best paths.
- The path chosen is:
  - best path among the ones which satisfy the constraints established by routing policies
  - Configured by the network administrator, which reflect commercial agreements among ASes

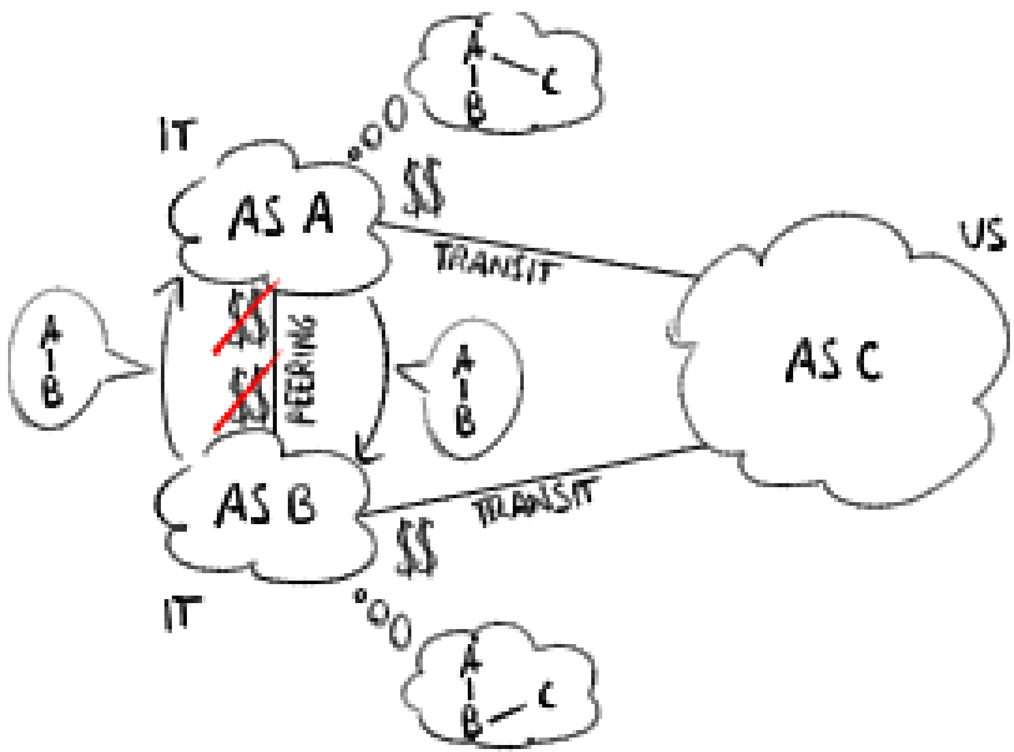
Economic Requirements

- Freeriding:
  - Sending traffic on a transit link costs money → an AS can take advantage of a peering link, even if it is not a direct link, to make the other peer AS pay the transit cost.



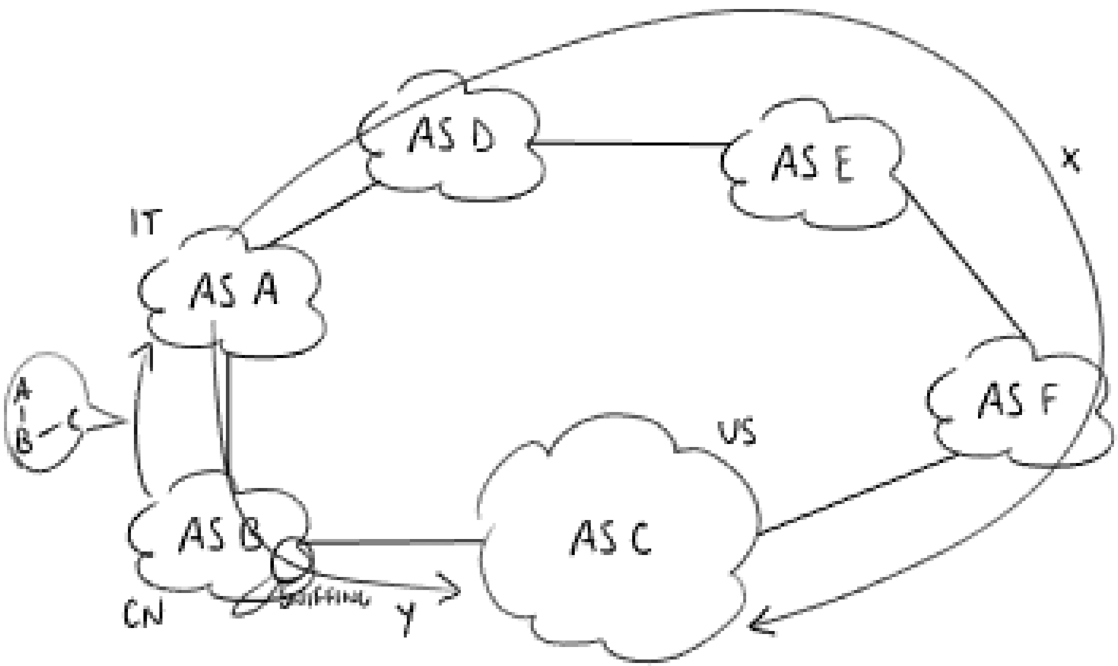
Administrative Requirements

- Route Hiding:
  - An AS can set a routing policy in order not to announce connectivity with other ASes to an AS.



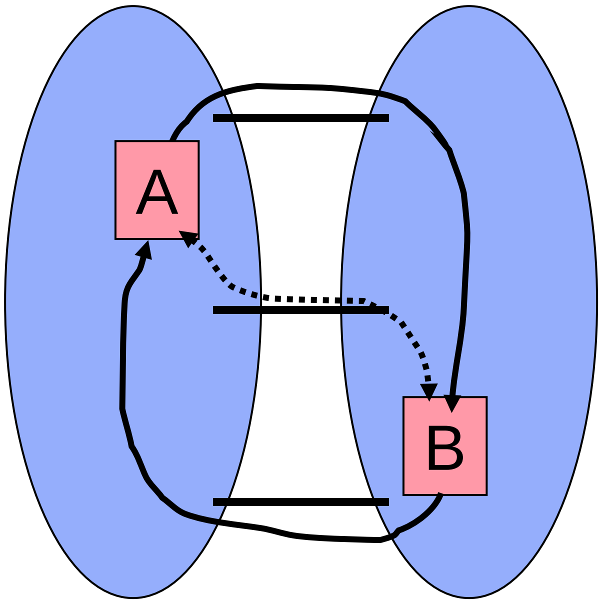
Security Requirements

- Trusted operators:
  - A network operator (e.g. AS B) makes sniffing actions on traffic crossing its AS → an AS would like to avoid B and has its traffic directed to other ASes instead of going through that untrusted operator.

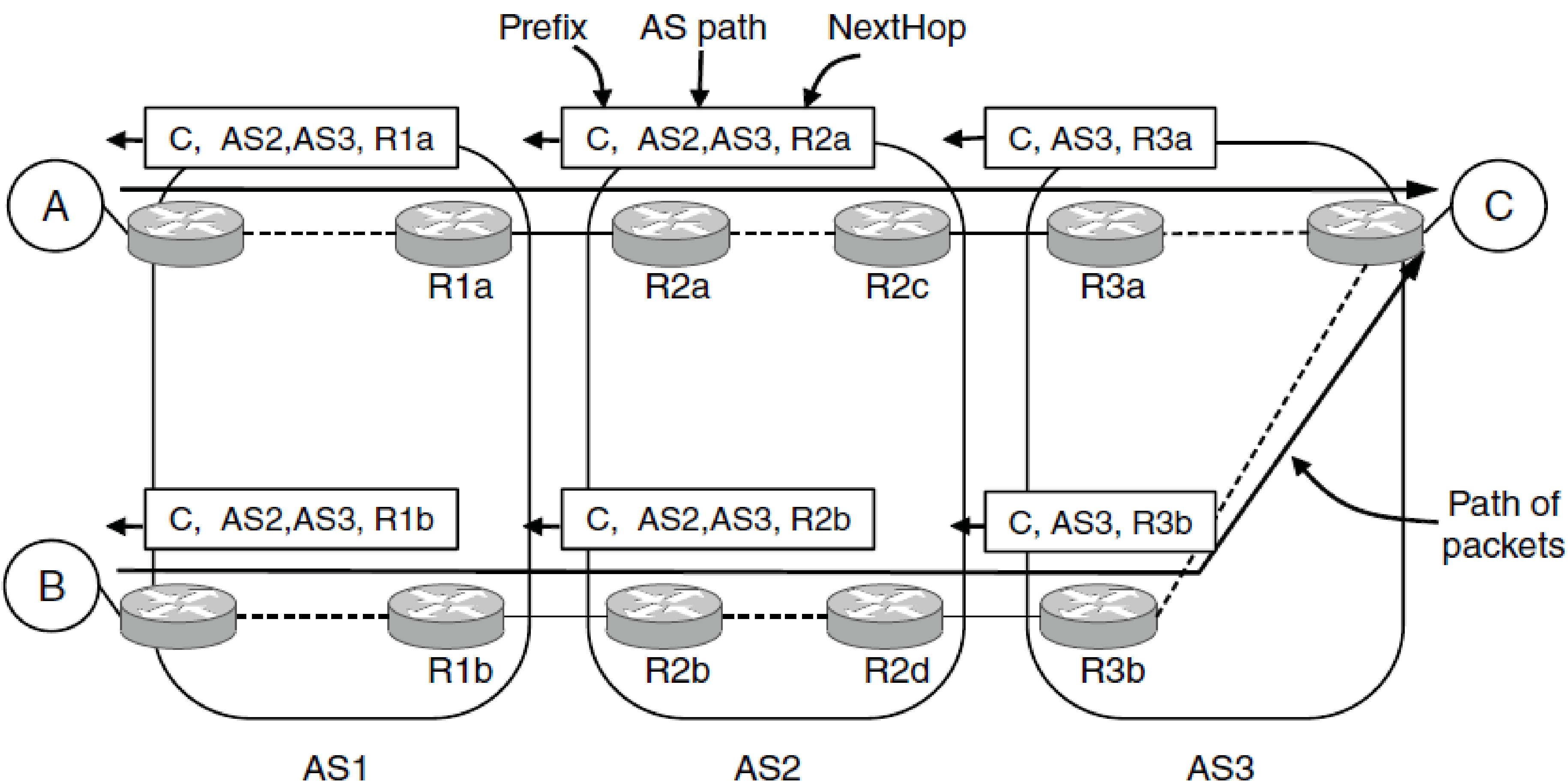


Impact of Policies – Example

- Early Exit / Hot Potato
  - "if it's not for you, bail"
- Combination of best local policies not globally best
- Side-effect: asymmetry



Routing with BGP Example



Routing with BGP Example (2)

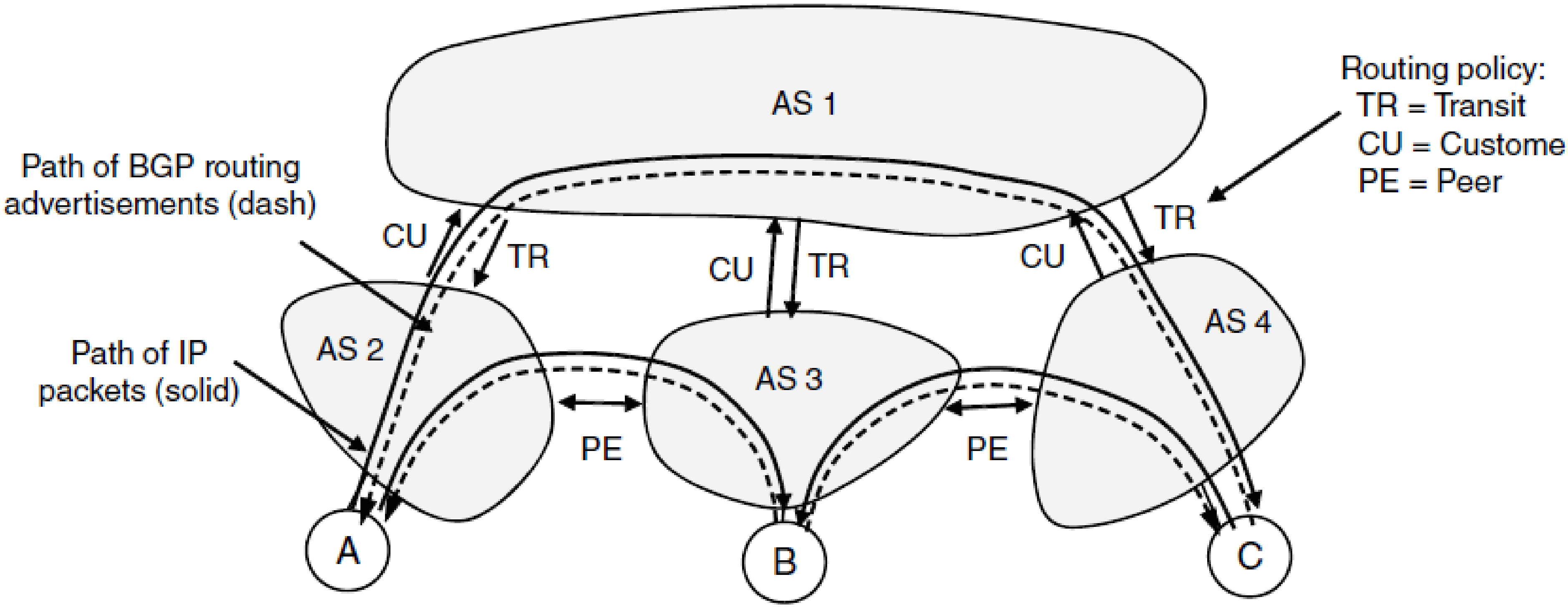
- Policy is implemented in two ways:

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

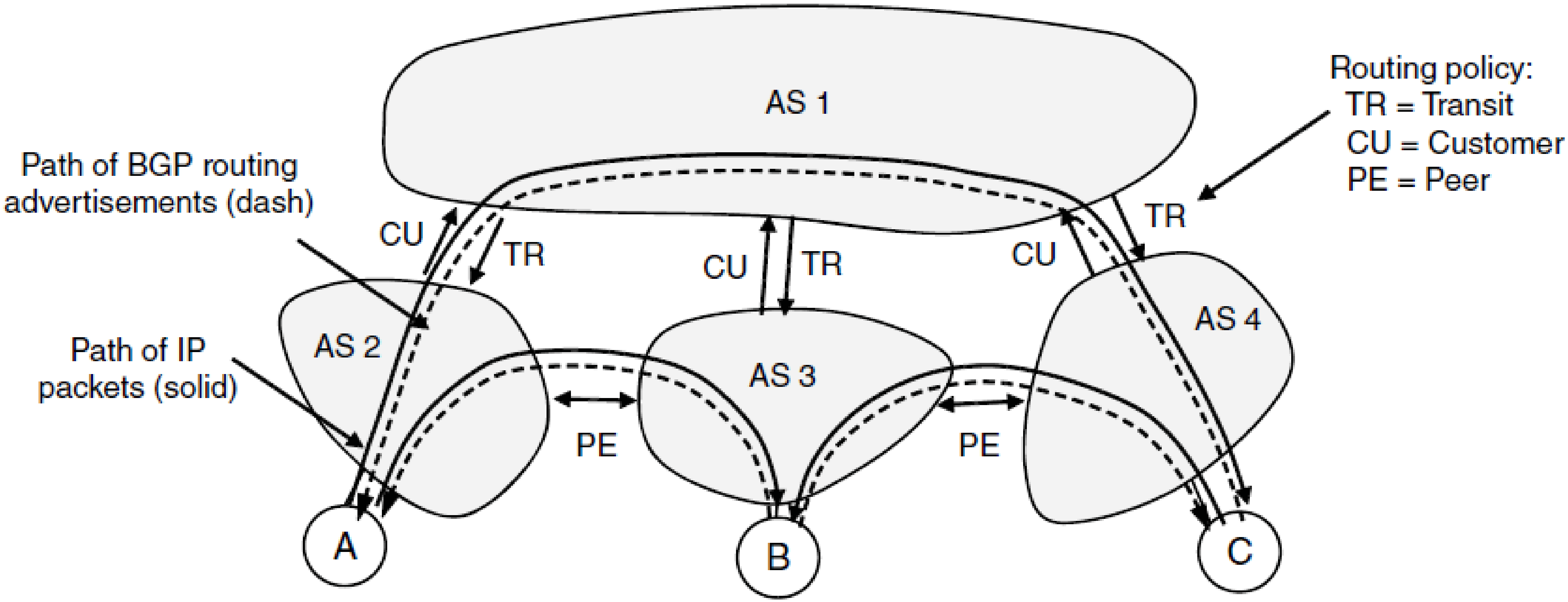
Routing with BGP Example (3)

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



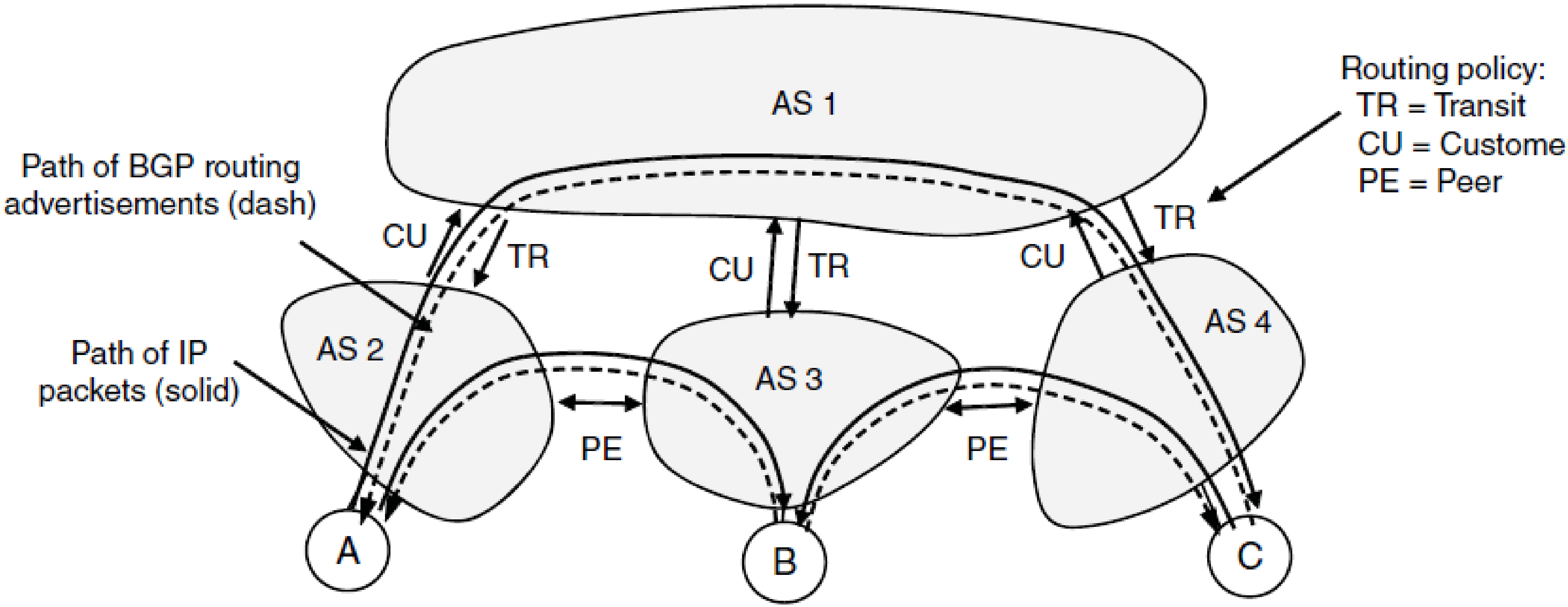
Routing with BGP Example (4)

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



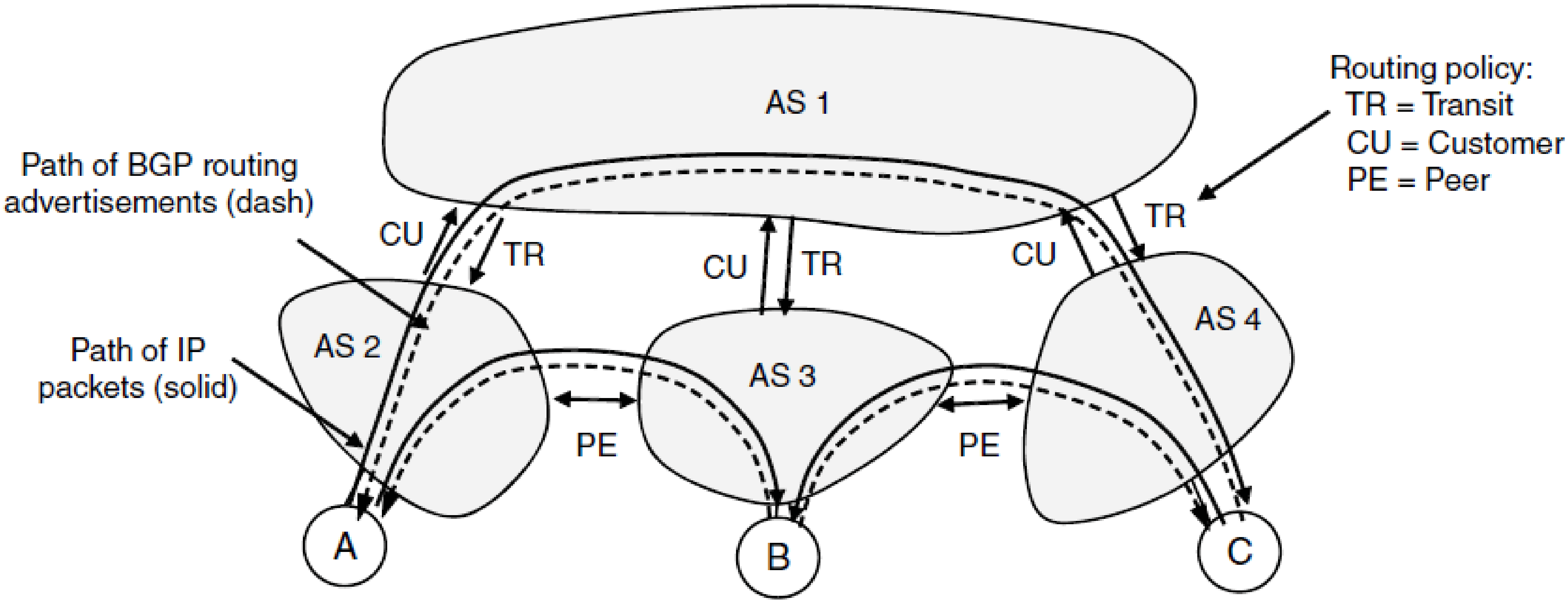
Routing with BGP Example (5)

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



Routing with BGP Example (6)

- 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 sure independent decisions will be sensible overall?
- Other important factors:
  - Convergence effects
  - How well it scales
  - Integration with intra-domain routing
  - And more …

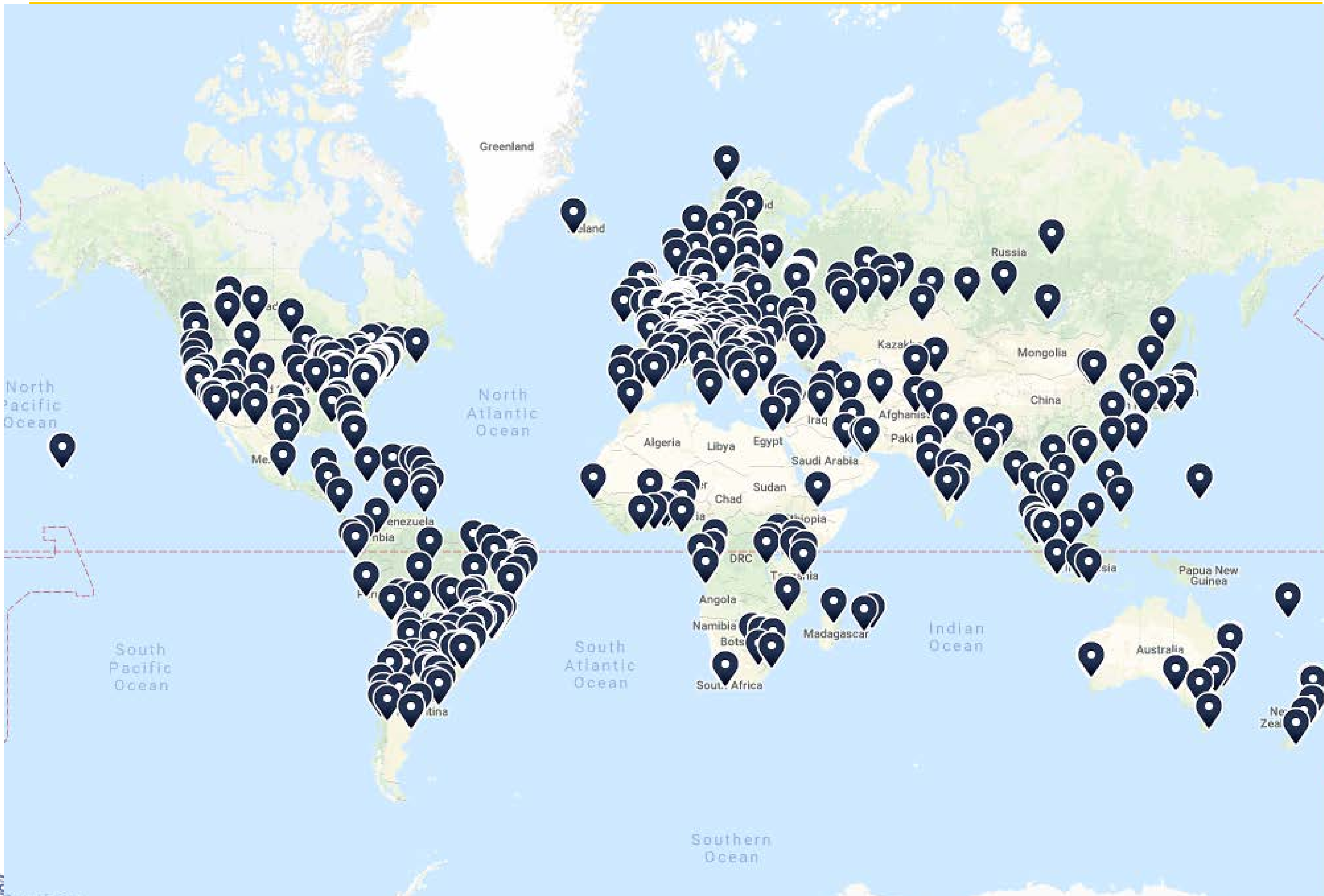
Internet Exchange Point (IXP)

- Interconnecting two ASes by direct connection is not convenient:
  - link cost: may require digging operations;
  - cost of interfaces on routers: send signal over long distances;
  - flexibility: intervention is necessary on the physical infrastructure to create a new interconnection.
- Internet Exchange Point (IXP)
  - allows multiple border routers of different ASes (ISPs) to exchange external routing information in a more dynamic and flexible way.
  - routers are connected through an intermediate data-link-layer LAN: routing policies define interconnections according to commercial agreements among ASes
  - to create a new interconnection, it is sufficient to configure routing policies on single routers without having to change the physical infrastructure.

IXP Services

- Each AS pays a monthly fee, depending on the speed of the connection to the IXP.
- The IXP is in charge of the technical functioning of switches within the intermediate network:
  - single location: often all routers are concentrated inside a room in a datacenter, where they are provided with:
    - high-speed data-link-layer network
    - electrical power, conditioning system
    - monitoring service
    - proximity to optical-fiber backbones
  - distributed infrastructure: multiple access points are available in the main towns over the territory (for example, CABASE runs across the entire Argentina).
- The IXP is also known as Neutral Access Point (NAP): IXP must be neutral and uninvolved in its customers' business.
- An IXP can decide to disallow transit agreements

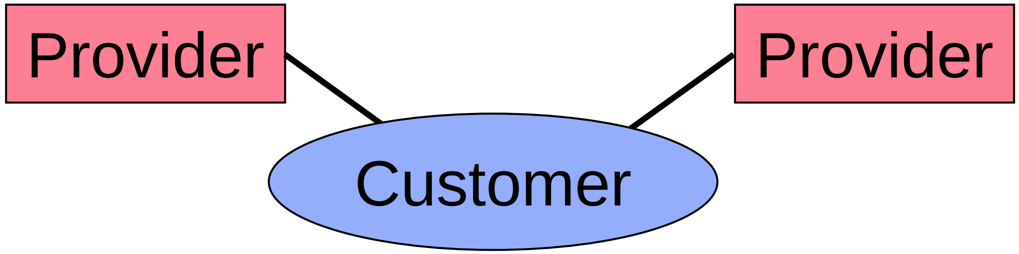
IXP Map



https://www.internetexchangemap.com

Multi-Homing

- Connect to multiple providers for reliability, load sharing



- Customer can choose the best outgoing path from any of the announcements heard from its providers
  - Easy to control outgoing traffic, e.g., for load balancing

- Less control over what paths other parties will use to reach us
  - Both providers will announce that they can reach to the customer
  - Rest of Internet can choose which path to take to customer
    - Hard for the customer to influence this

Network Neutrality

- Network neutrality: all traffic should be treated equally, without privileging or damaging a part of traffic for economic interests.
- Network operators can be tempted to give 'preferential treatment' to portions of traffic:
  - privilege some traffic: offer a better service for a certain kind of traffic (e.g. higher speed);
  - damage some traffic: offer a worse service, or no service at all, for a certain kind of traffic.
- A neutral network guarantees that all entities (e.g. content providers) have the same service, without making some service be killed at the discretion of the network operator.
- Enforcing 'pure' network neutrality implies that traffic control, which may be useful in many cases, is not possible at all.
- If network may not be neutral, the network operator is given the power to privilege some traffic or content.

Network Neutrality (2)

- In an open market the ball is on the user side:
  - if users do not agree that their VoIP traffic is discriminated, they can switch to another network operator (*although in practice this may not always be possible due to cartels among network operators*).
- Examples of non-neutrality:
  - content providers: ISPs would like to have a part of revenues of content providers → an ISP may privilege traffic directed to a content provider with which it stipulated a revenue sharing agreement;
  - peer-to-peer (P2P):
    - end users do not care about destination of their traffic, but P2P traffic can reach every user in every AS around the world making the ISP pay high costs → an ISP may privilege traffic which is generated within the AS (e.g. AdunanzA by Fastweb);
    - P2P traffic is more symmetric because it uses a lot the upload bandwidth, while networks have been sized to support asymmetric traffic → an ISP may privilege asymmetric traffic (e.g. normal web traffic);
  - quality of service (QoS): an ISP may privilege traffic with a higher priority level (e.g. VoIP traffic);
  - security: an ISP may block malicious traffic (e.g. DDoS attack).

Key Concepts

- Internet is a collection of Autonomous Systems (ASes)
  - Policy dominates routing at the AS level
- Structural hierarchy helps make routing scalable
  - BGP routes between autonomous systems (ASes)