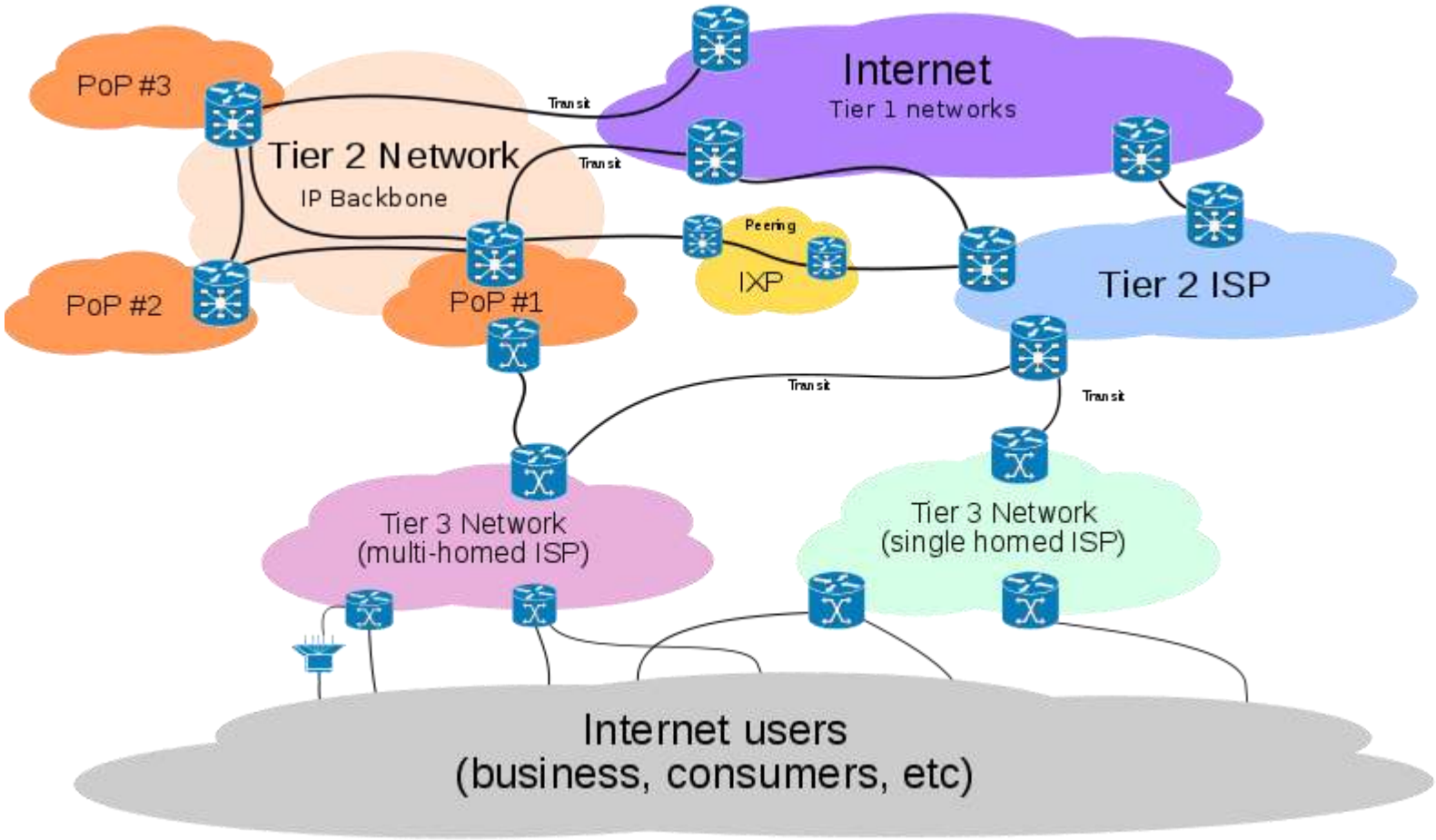
Interdomain routing

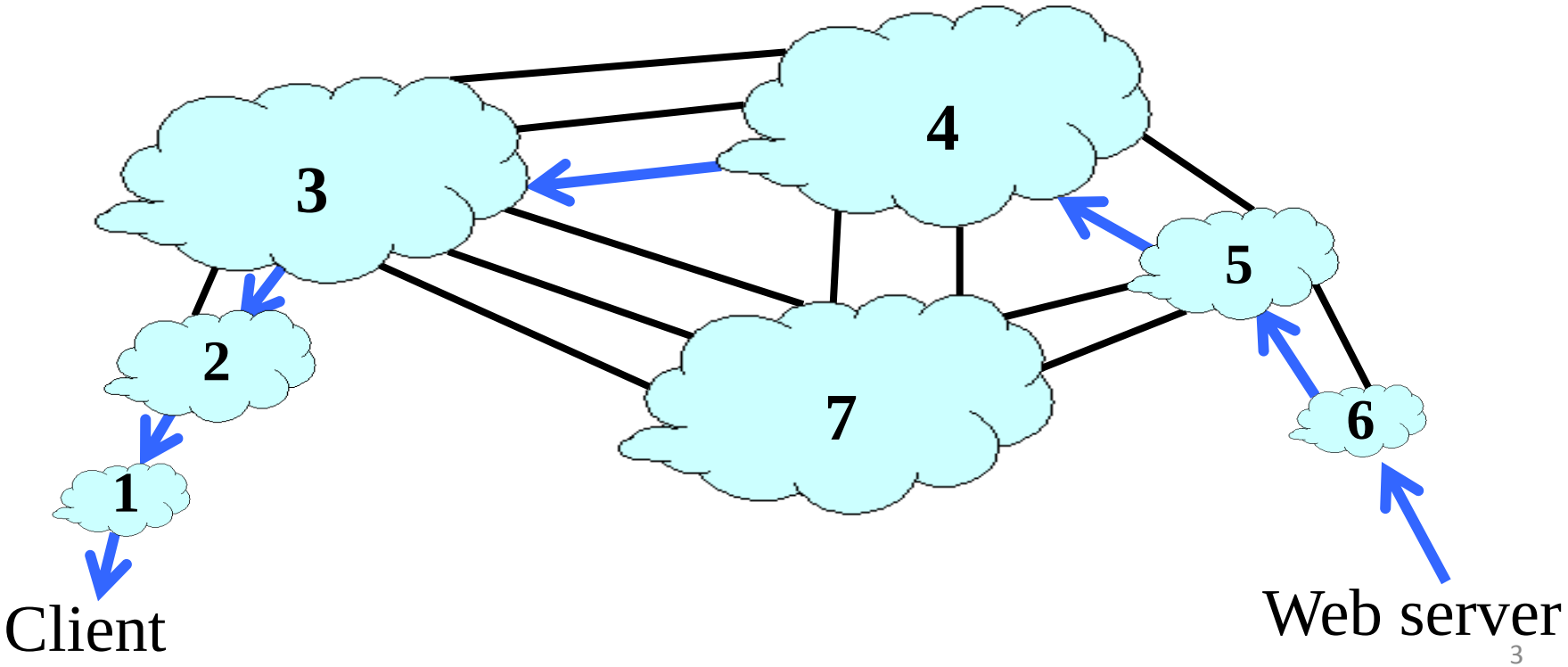


Overview

- Business relationships between ASes
- Interdomain routing using BGP
  - Advertisements
  - Routing policy
  - Integration with intradomain routing
- Routing security
  - Prefix hijacking
  - Secure BGP

Autonomous systems (ASes)

- AS-level topology
  - Destinations are IP prefixes (e.g., 12.0.0.0/8)
  - Nodes are Autonomous Systems (ASes)
  - Edges are links and business relationships



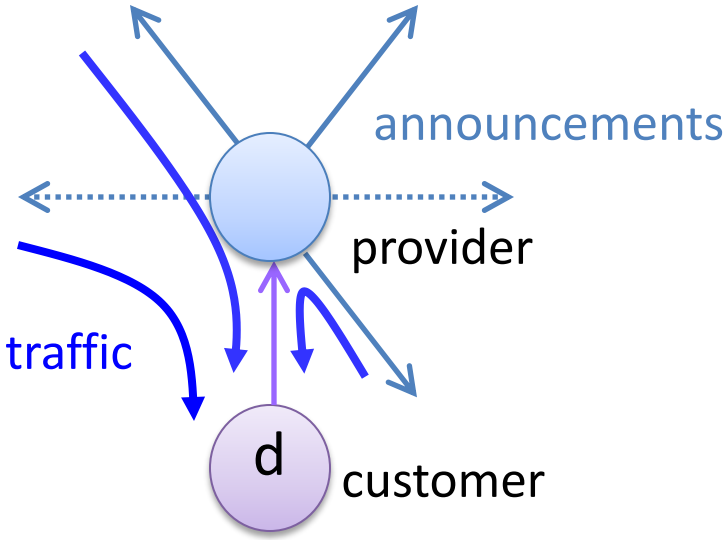
Business relationships

- Neighboring ASes have business contracts
  - How much traffic to carry
  - Which destinations to reach
  - How much money to pay
- Common business relationships
  - Customer-provider: Customer pays provider for transit
    - e.g. Princeton is a customer of USLEC
    - e.g. MIT is a customer of Level3
  - Peer-peer: No money changes hands
    - e.g. UUNET is a peer of Sprint
    - e.g. Harvard is a peer of Harvard Business School

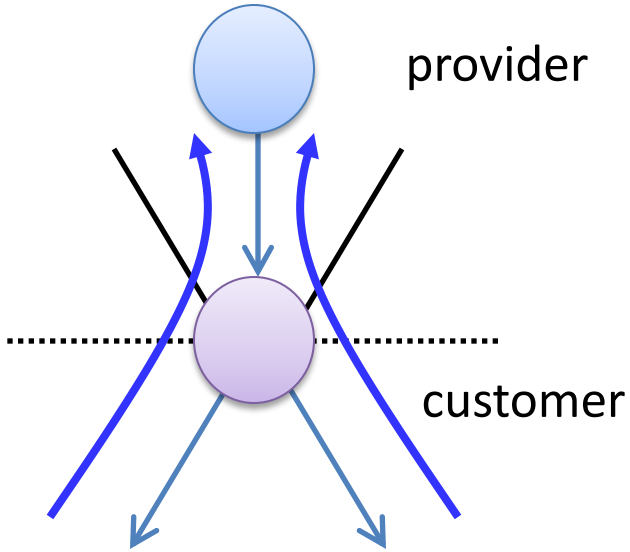
Customer-provider

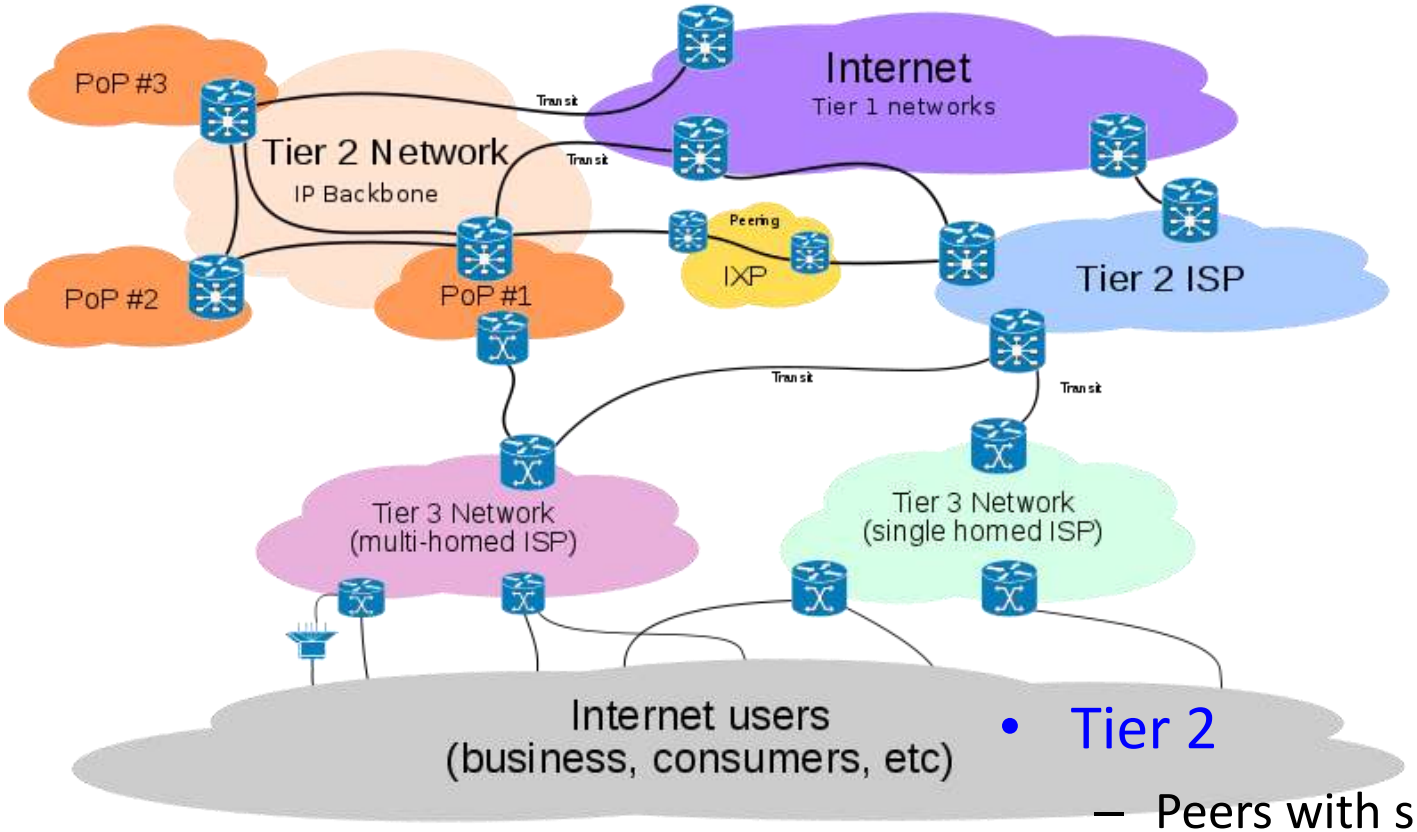
- Customer needs to be reachable from everyone
  - Provider tells all neighbors how to reach the customer
- Customer does not want to provide transit service
  - Customer does not let its providers route through it

Traffic **to** the customer



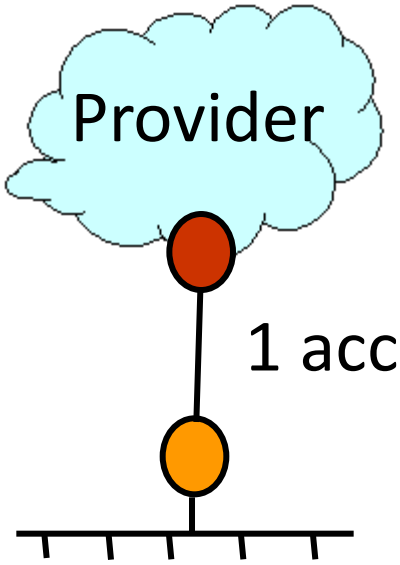
Traffic **from** the customer



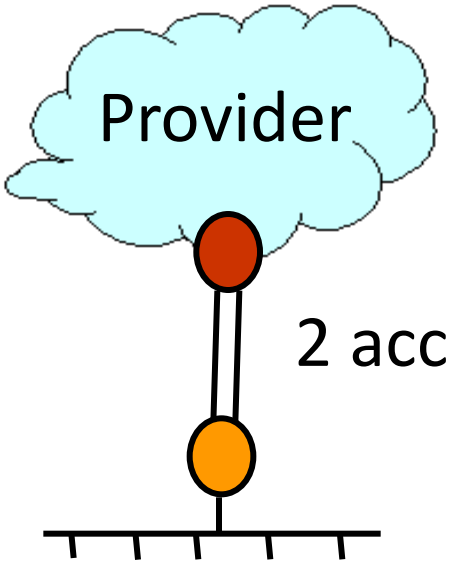


- Tier 1
  - Not a customer of anyone
  - Reach anywhere on Internet without purchasing transit
  - Around ~10, e.g. Centurylink, AT&T, Verizon, Sprint, etc.
- Tier 2
  - Peers with some networks
  - Purchases transit for some destinations
- Tier 3
  - Solely purchase IP transit from other providers
  - Normally single homed

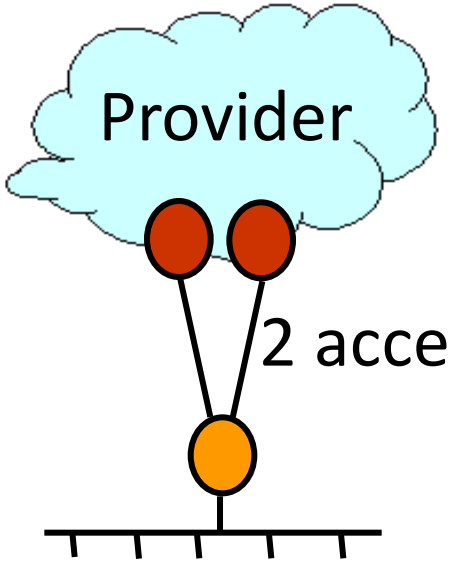
Customer Connecting to a Provider



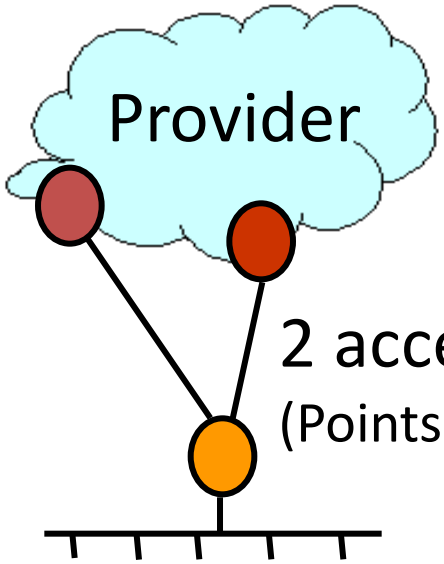
1 access link



2 access links



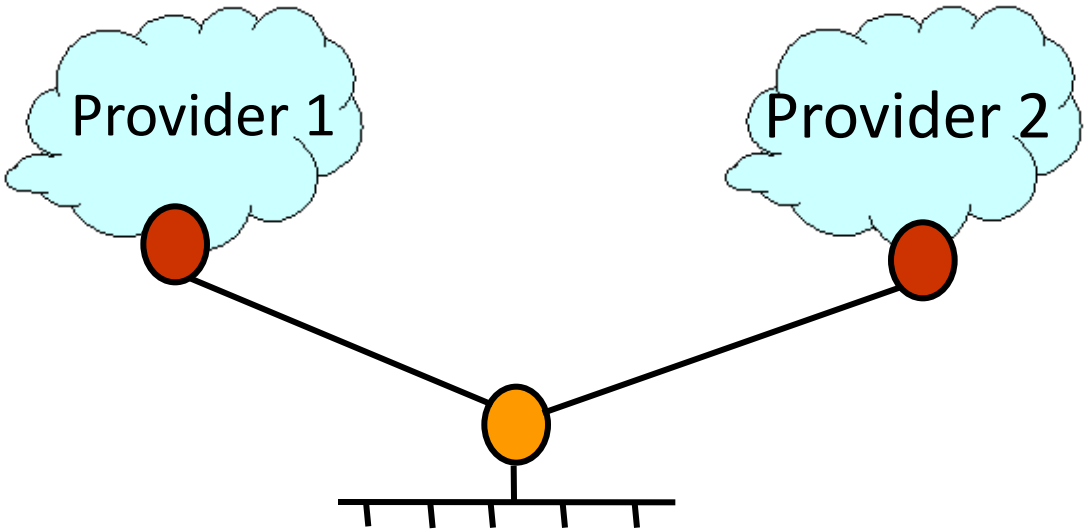
2 access routers



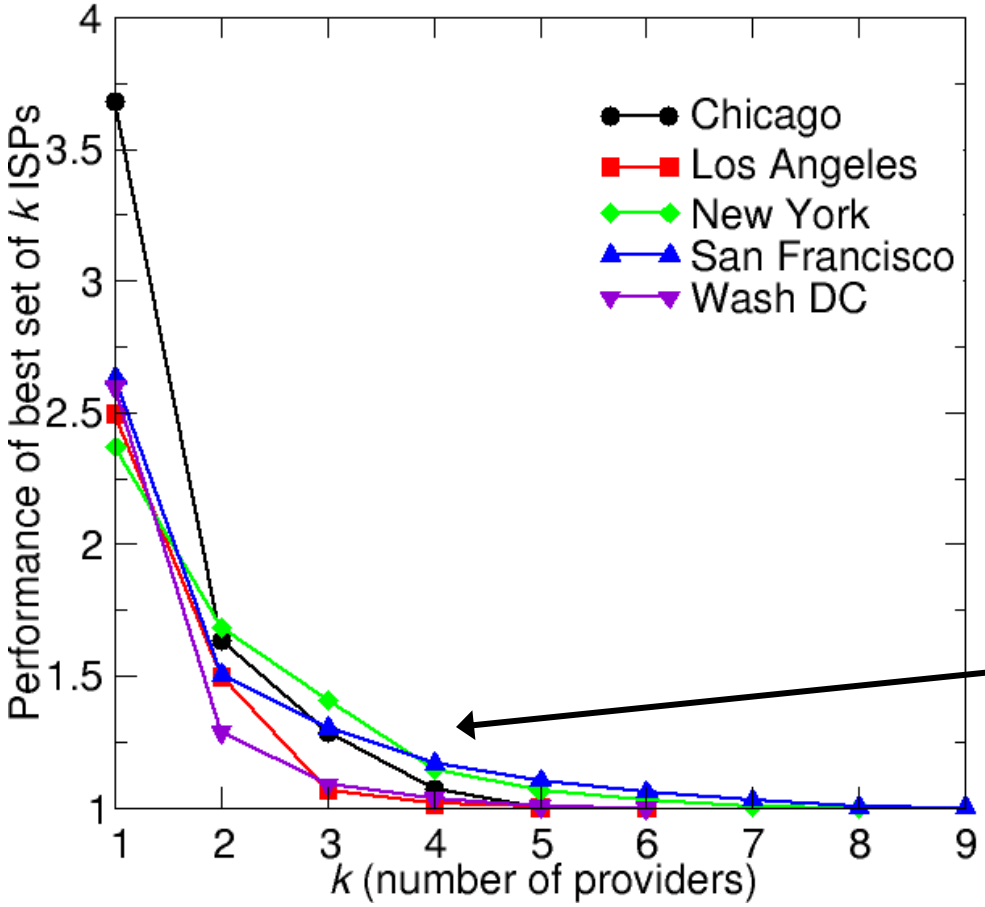
2 access PoPs
(Points of Presence)

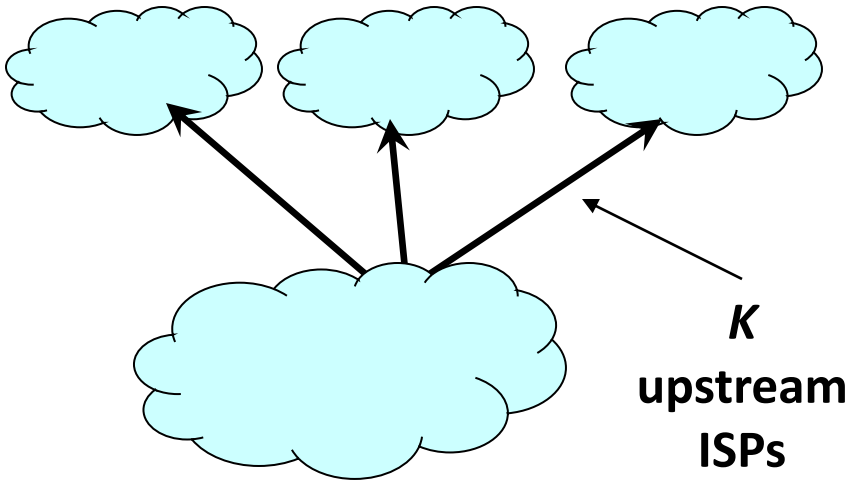
Multi-Homing

- Multi-homing: 2+ providers
  - Extra reliability, survive single ISP failure
  - Financial leverage through competition
  - Better performance by selecting better path



How many links are enough?





Not much benefit beyond 4 ISPs

Akella *et al.*, "Performance Benefits of Multihoming", *SIGCOMM 2003*

Interdomain routing

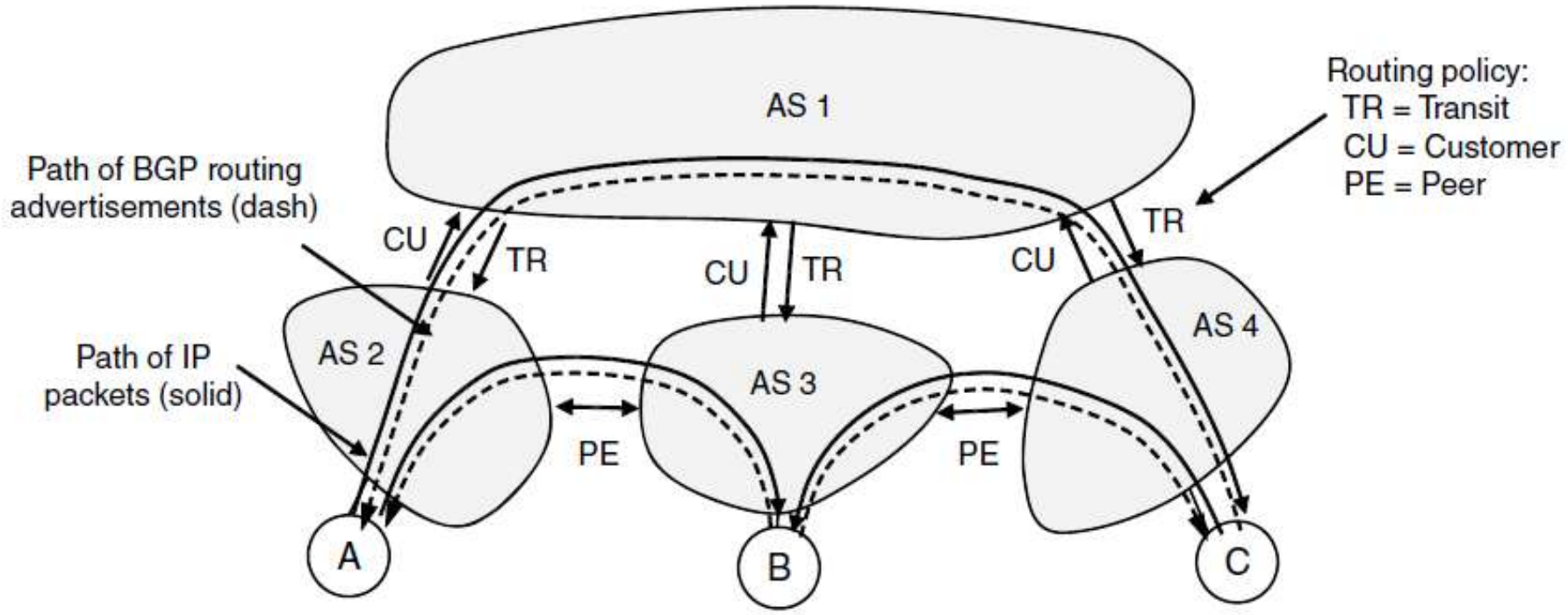
- Exterior Gate Protocol (EGP)
  - Forced a tree-like topology
  - Single backbone and autonomous systems connected as parents/children, not peers
  - Invented in 1982, now obsolete
- Border Gateway Protocol (BGP)
  - Arbitrarily connected ASes
  - Multiple backbone networks

Border Gateway Protocol

- Interdomain routing protocol for the Internet
  - Prefix-based path-vector protocol
  - Policy-based routing using AS paths
  - Evolved over the past 18 years

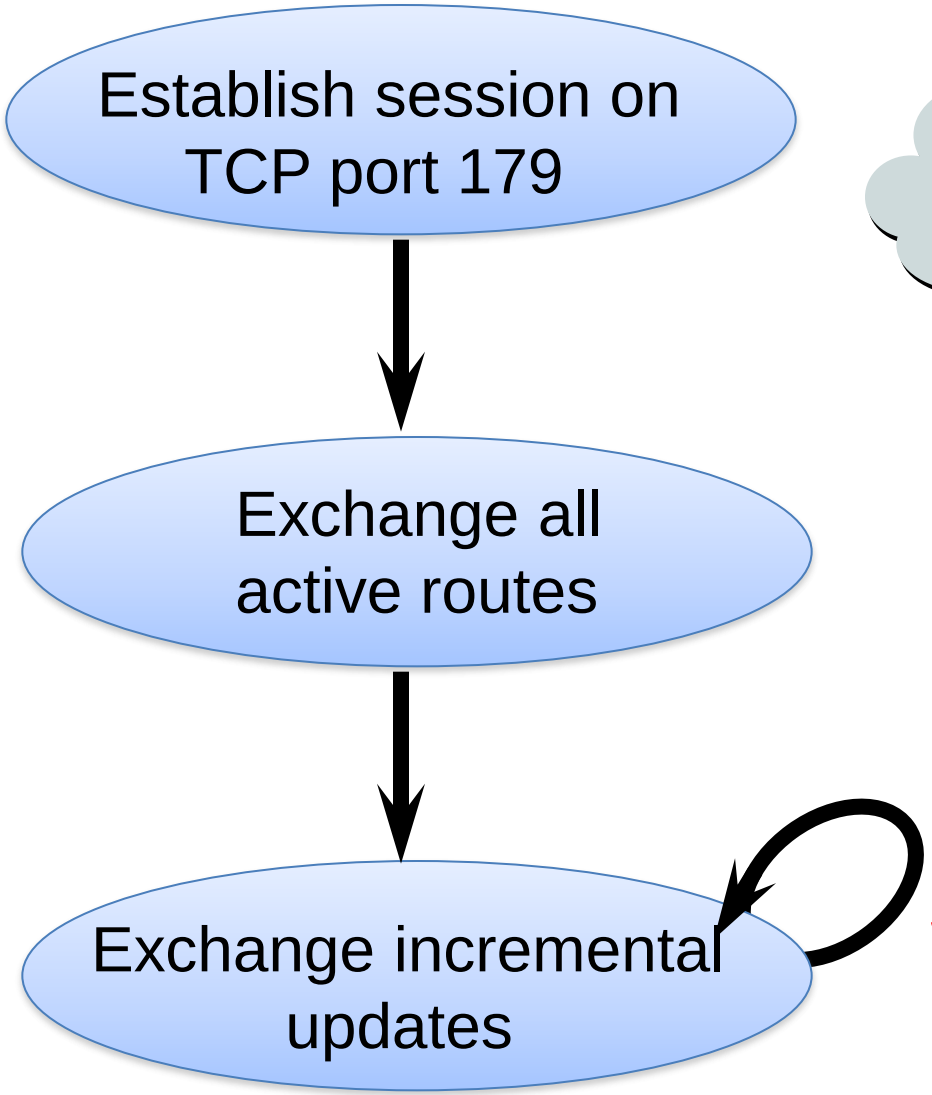
- **1989 : BGP-1 [RFC 1105], replacement for EGP**
- **1990 : BGP-2 [RFC 1163]**
- **1991 : BGP-3 [RFC 1267]**
- **1995 : BGP-4 [RFC 1771], support for CIDR**
- **2006 : BGP-4 [RFC 4271], update**

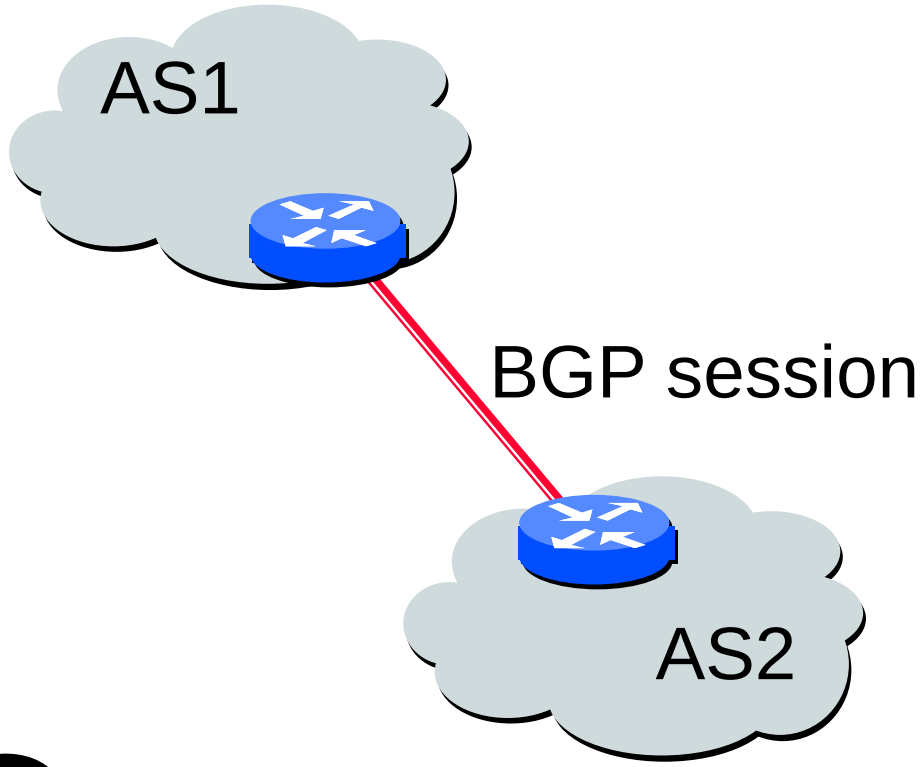
BGP routing



Routing between four Autonomous Systems (ASes)

BGP Operations





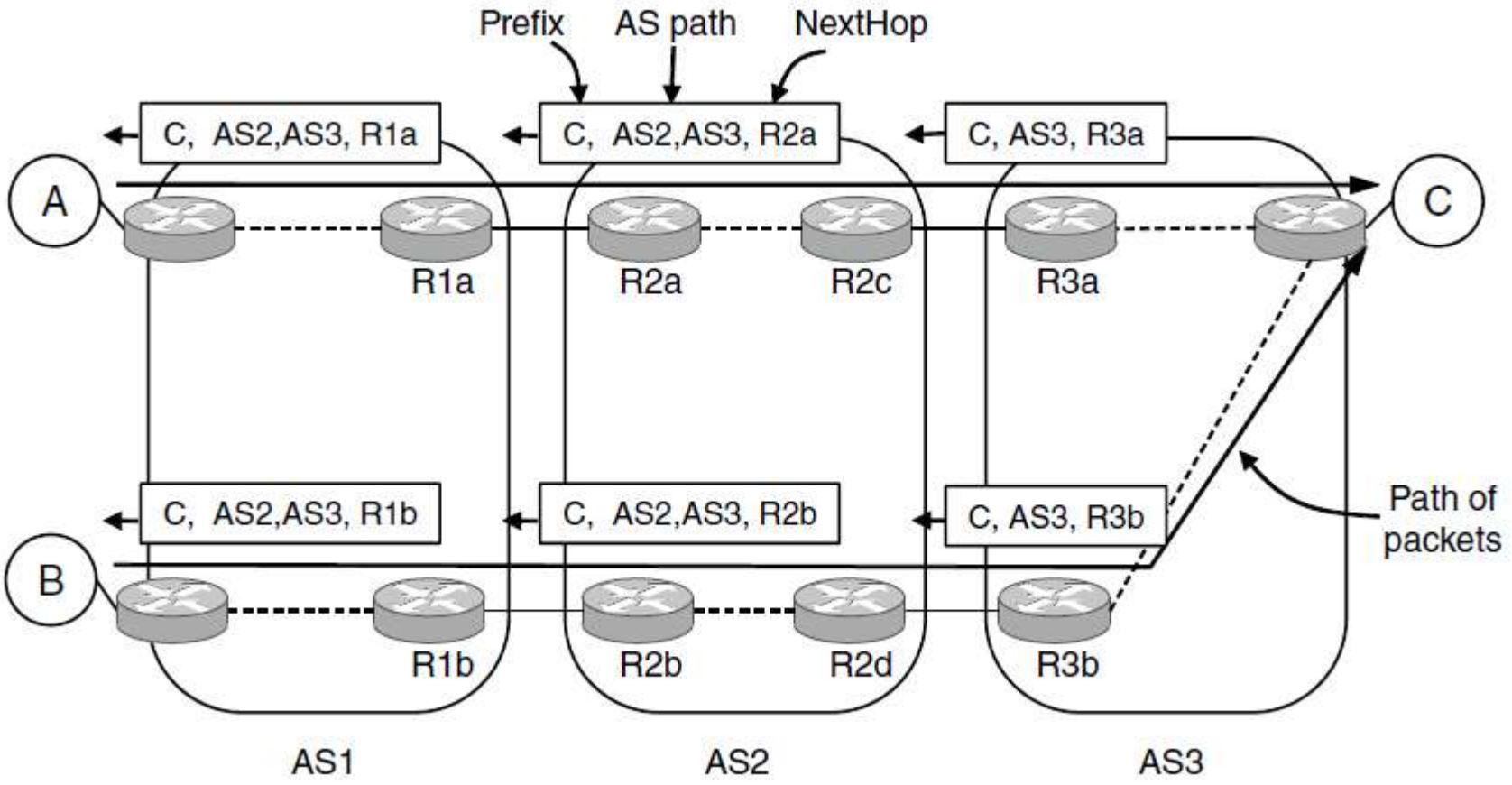
While connection is ALIVE:

Exchange route UPDATE msgs

Incremental Protocol

- Routers form mesh over TCP
- A node learns multiple paths to destination
  - Stores all routes in routing table
  - Applies policy to select single active route
  - May advertise route to neighbors
- Incremental updates
  - Announcement
    - Upon selecting new active route, add node id to path
    - Optionally advertise to each neighbor
  - Withdrawal
    - If active route is no longer available, send message to neighbors

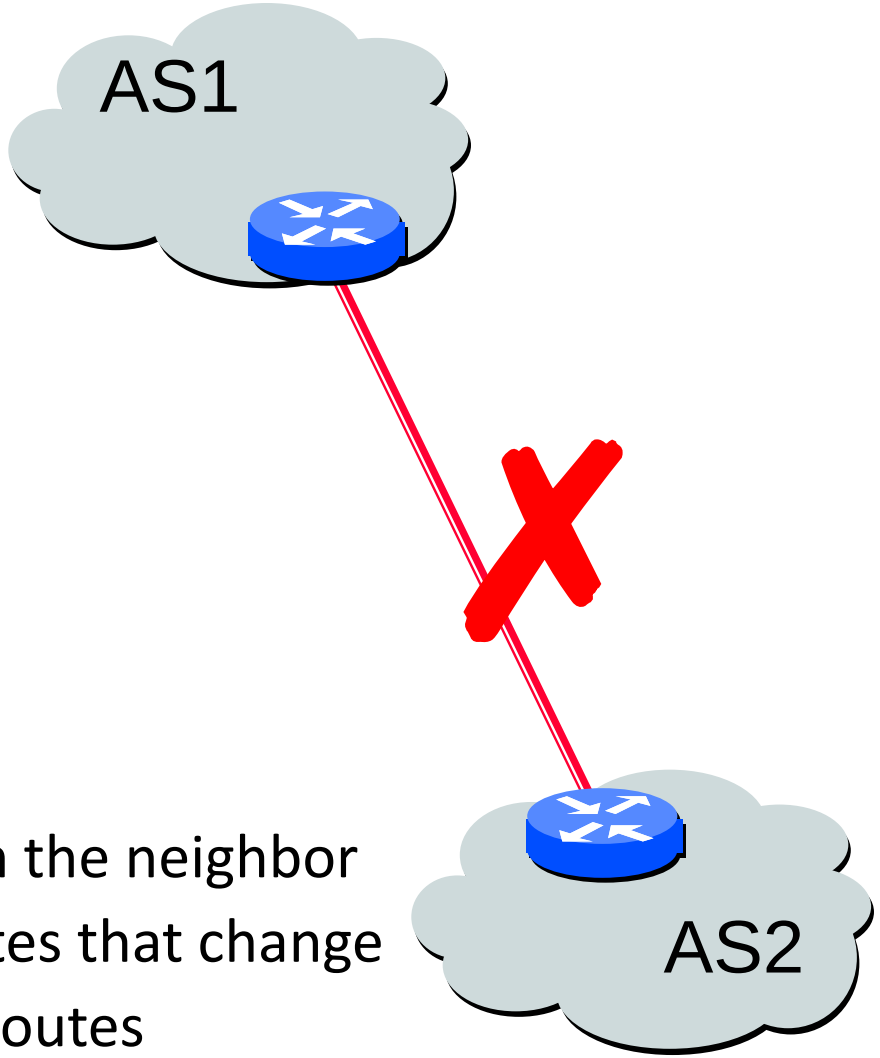
BGP advertisements



Propagation of BGP route advertisements.
Advertisements contain: AS path + next-hop router.

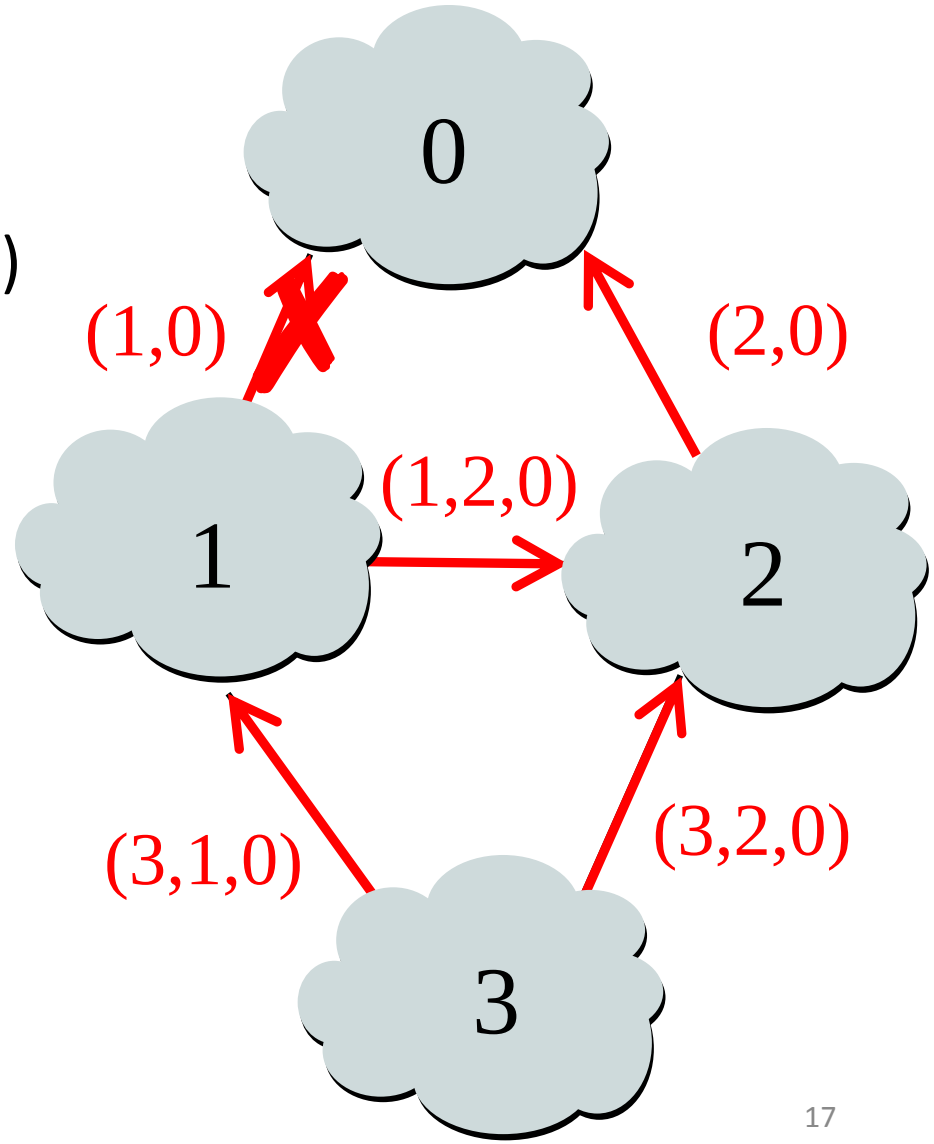
BGP Session Failure

- BGP runs over TCP
  - BGP only sends updates when changes occur
  - TCP doesn't detect lost connectivity on its own
- Detecting a failure
  - Keep-alive: 60 seconds
  - Hold timer: 180 seconds
- Reacting to a failure
  - Discard all routes learned from the neighbor
  - Send new updates for any routes that change
  - Overhead increases with # of routes



Routing Change: Path Exploration

- AS 1
  - Delete the route (1,0)
  - Switch to next route (1,2,0)
  - Send route (1,2,0) to AS 3
- AS 3
  - Sees (1,2,0) replace (1,0)
  - Compares to route (2,0)
  - Switches to using AS 2

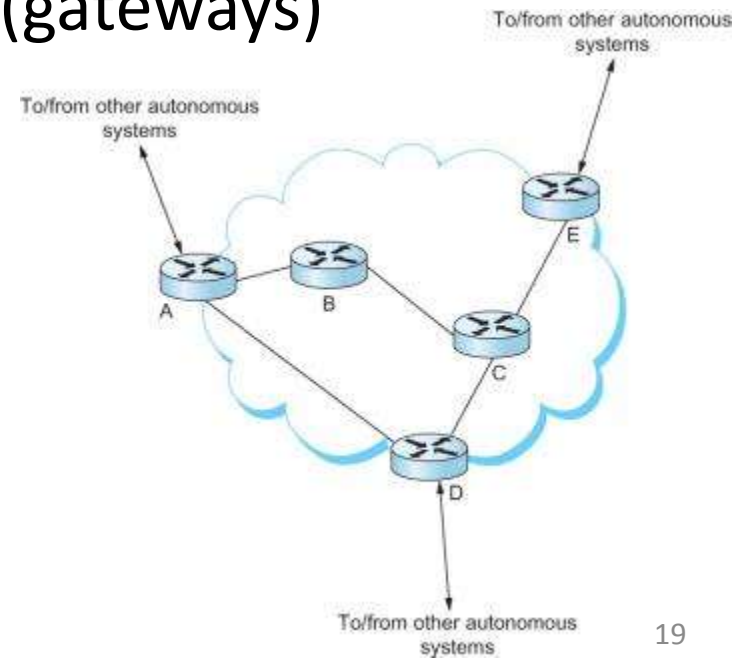


BGP converges slow

- Path vector avoids count-to-infinity
  - But ASes still must explore many alternative paths
  - Find highest-ranked path still available
- In practice:
  - Most popular destinations have stable BGP route
  - Instability lies in a few unpopular destinations
- Low convergence delay is a goal
  - Can be tends of seconds/minutes
  - Important for interactive applications

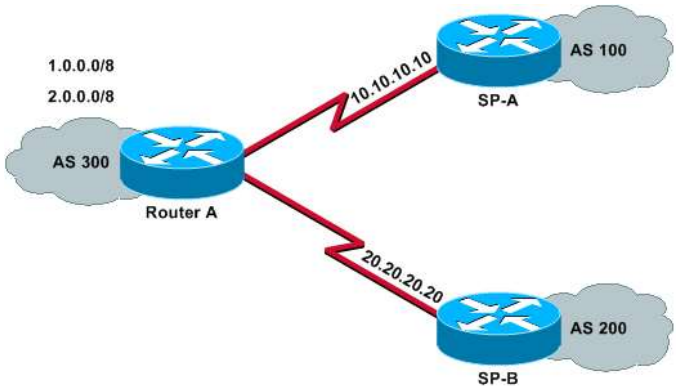
Running BGP in an AS

- Each AS has:
  - At least one BGP speaker advertising:
    - local networks
    - other reachable networks (if transit AS)
  - One or more border routers (gateways)
    - Where packets enter/exit AS



Configuring BGP

- BGP speaker in an AS:
  - Manually config to talk to routers in other ASes



AS 300 is multi-homed, connected to two different ISPs.

Router A

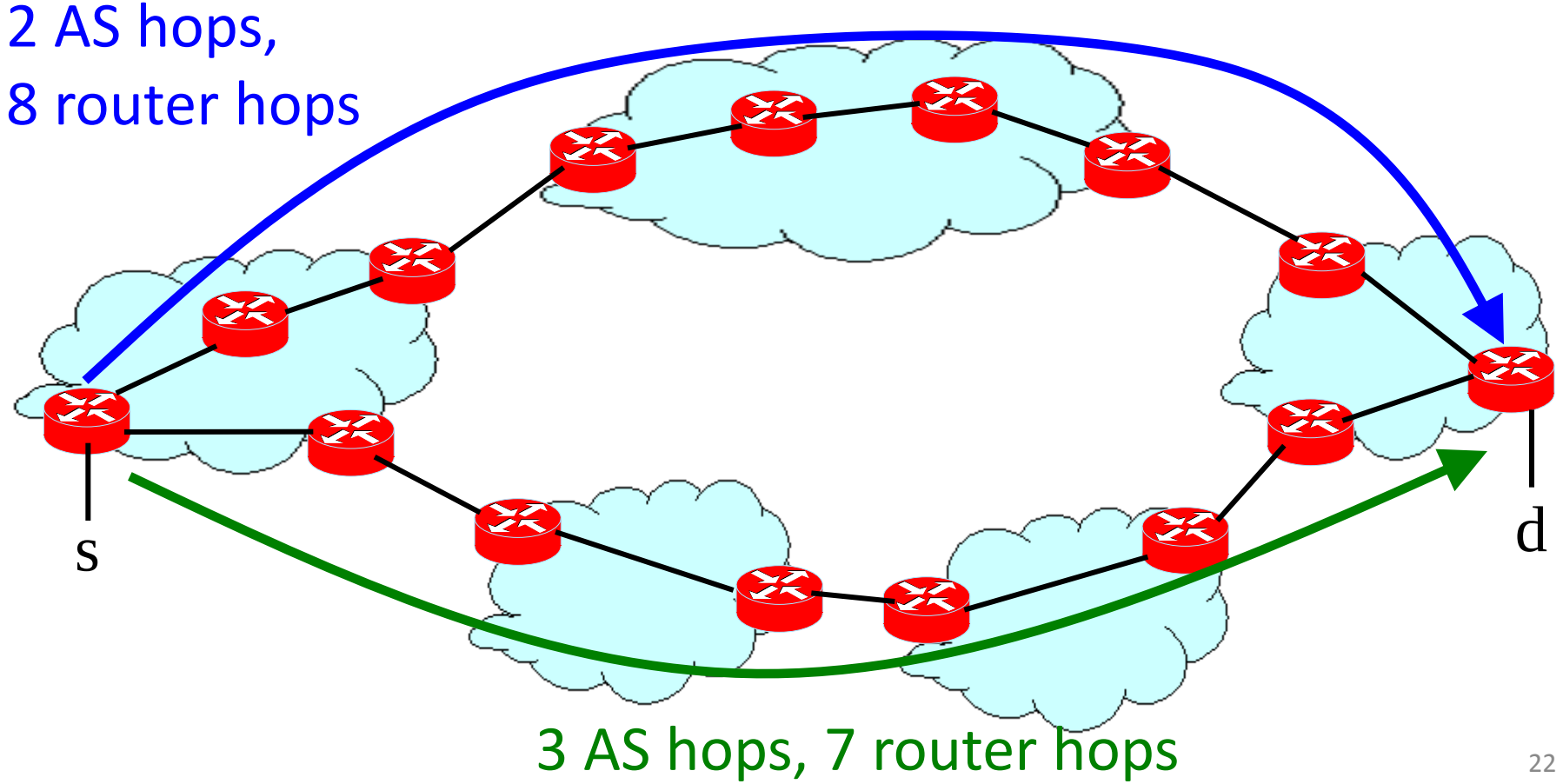
```
Current configuration:

router bgp 300
 network 1.0.0.0
 network 2.0.0.0

 neighbor 10.10.10.10 remote-as 100
 neighbor 10.10.10.10 route-map localonly out

!--- Outgoing policy route-map that filters routes to service provider A (SP-A).

 neighbor 20.20.20.20 remote-as 200
 neighbor 20.20.20.20 route-map localonly out

!--- Outgoing policy route-map that filters routes to service provider B (SP-B).

end
```

BGP decision process

- Policy decision by AS, various possibilities:
  - Route via peered network instead of transit
  - Shorter AS path better
    - Debatable since we don't know how many hops in AS
  - Lowest cost for your AS
    - Get it off your network sooner
  - Provide best quality of service for your customer

AS Path Length != Router Hops

- AS path may be longer than shortest AS path
- Router path may be longer than shortest path

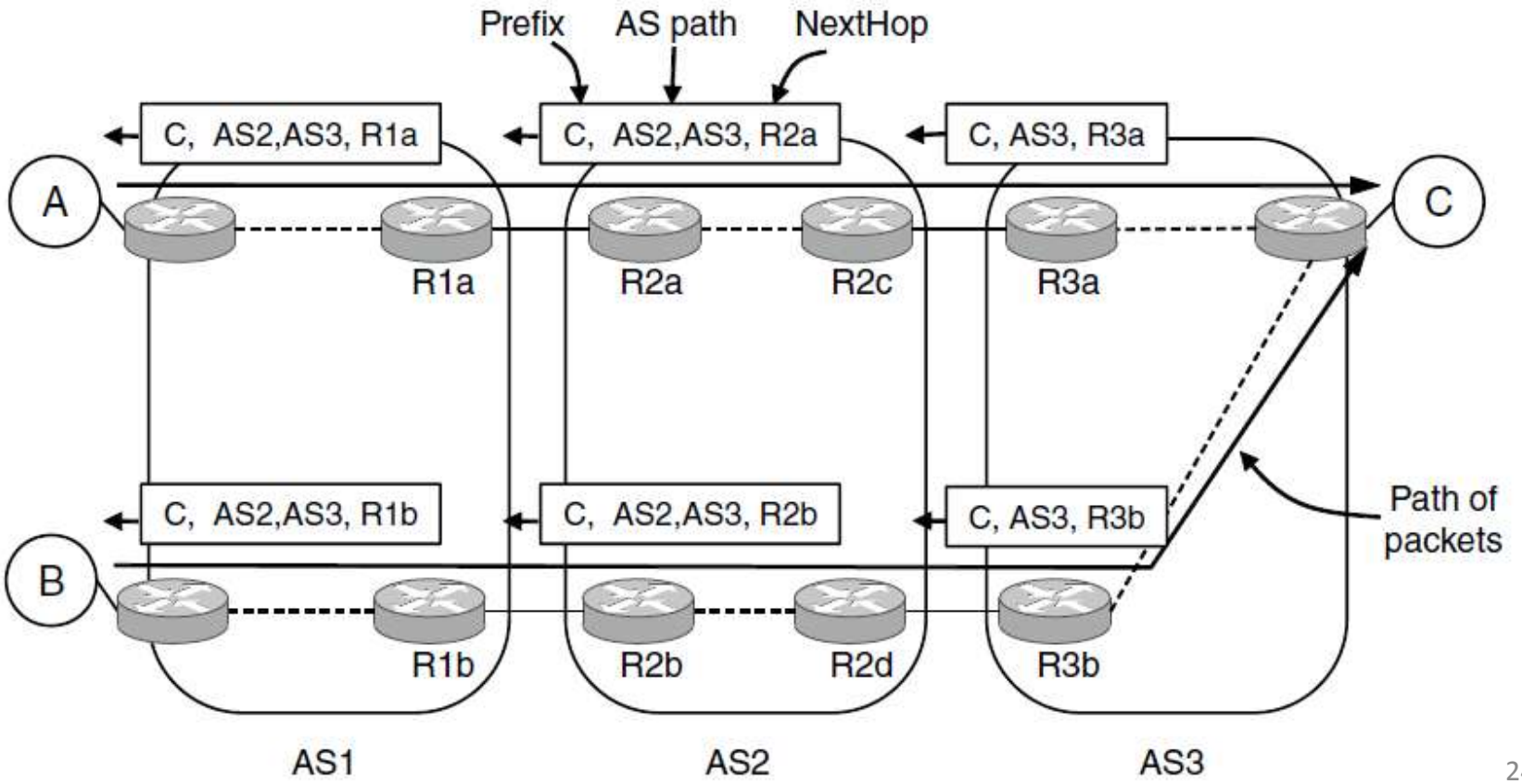


Routing packet inside your AS

- Hot-potato (early exit) routing
  - Each router selects closest exit point from AS
  - Minimize your costs in shipping around data
  - Based on intra-domain routing (e.g. OSPF)
- Cold-potato (late exit) routing
  - Keep packet in your AS as long as possible
  - Maximize control and quality of service

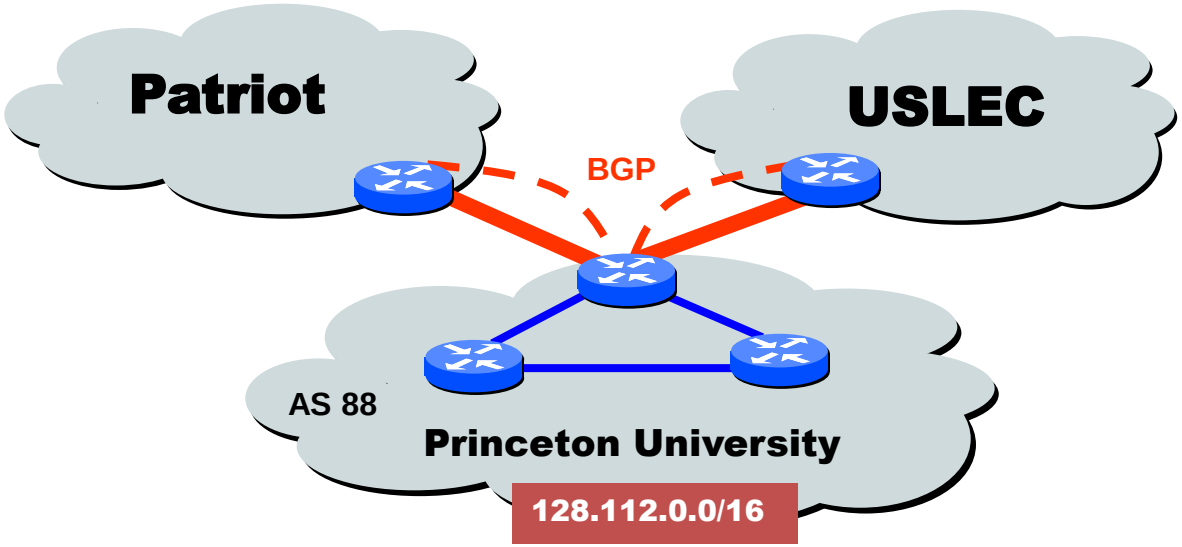
Paths not always symmetric

- Asymmetry of paths
  - Path A->B may not be same as B->A!



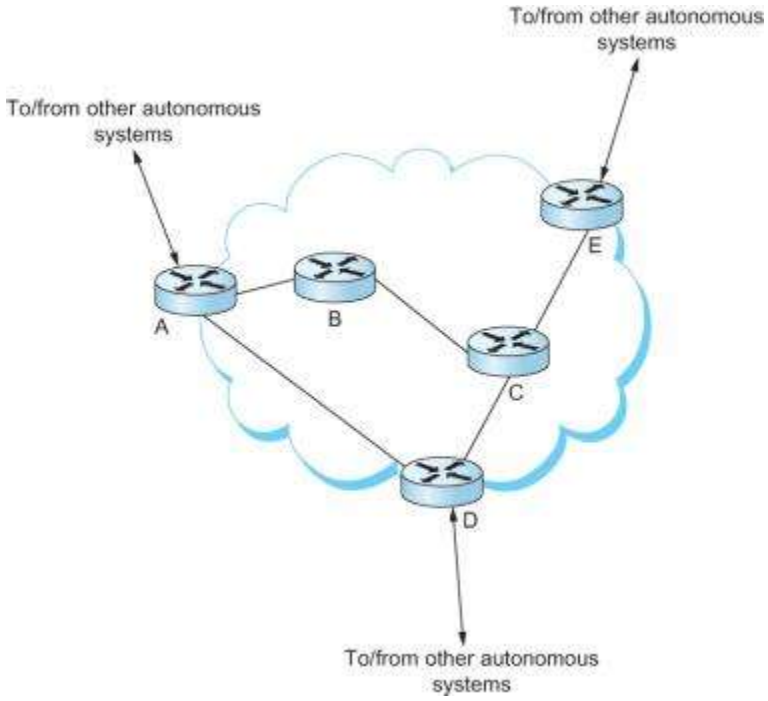
Integration of routing

- Combine interdomain & intradomain routing
  - Stub network
    - Border BGP router injects default route into intradomain protocol
    - Anything not destined for AS, goes to border router



Integration of routing

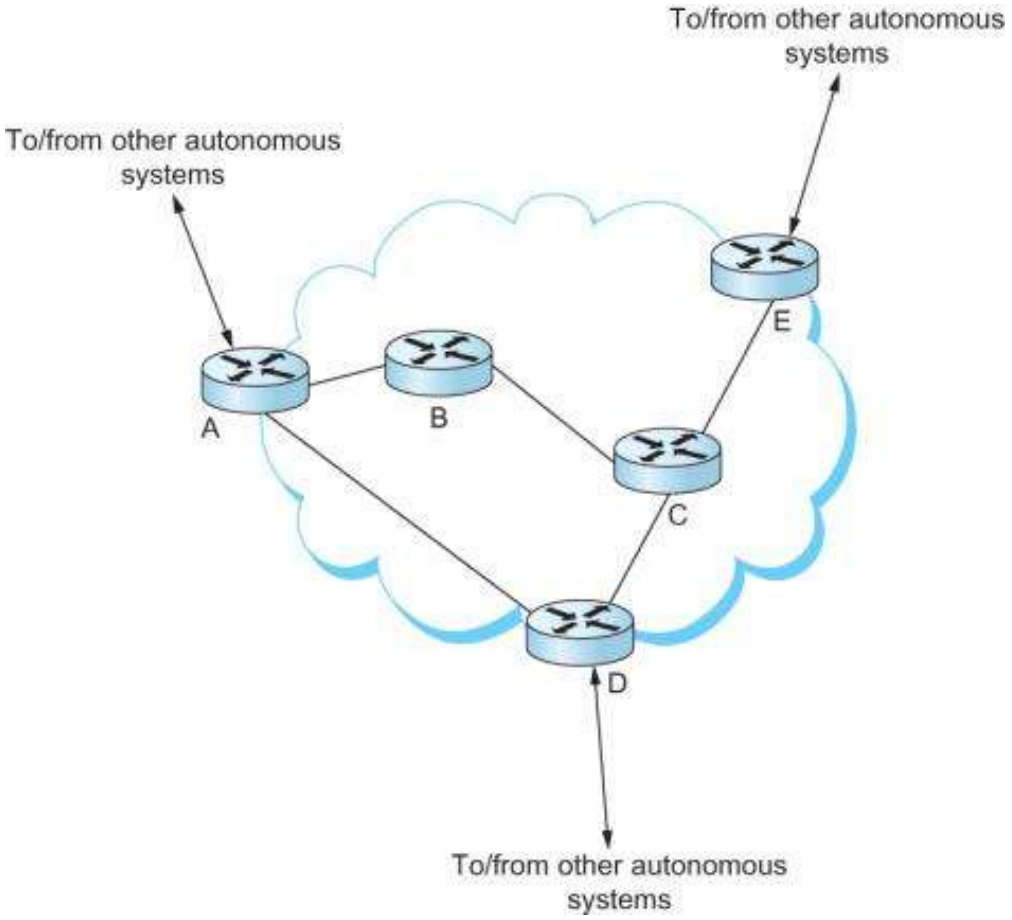
- Combine interdomain & intradomain routing
  - Border router injects routes learned from other AS into intradomain protocol
  - Other routers in AS can then route to prefix



Integration of routing

- Backbone networks
  - Too many routes to inject into normal link-state intradomain protocol
- Interior BGP (iBGP)
  - BGP running inside an AS
  - Best border router to use for a prefix
  - Run conventional protocol such as OSPF or RIP (generically called an IGP) to route inside the AS

Integration of routing



Prefix	BGP Next Hop
18.0/16	E
12.5.5/24	A
128.34/16	D
128.69./16	A

BGP table for the AS

Router	IGP Path
A	A
C	C
D	C
E	C

IGP table for router B

Prefix	IGP Path
18.0/16	C
12.5.5/24	A
128.34/16	C
128.69./16	A

Combined table for router B

Routing security

- BGP: glue that binds the modern Internet
- How secure is it?
  - Not very
  - Relies on trust and best practices between ASes
  - Fat finger mistakes can happen
  - Malicious attacks can happen

1. Enhance law enforcement's capabilities for preventing and prosecuting cyber-space attacks;
2. Create a process for national vulnerability assessments to better understand the potential consequences of threats and vulnerabilities;
3. Secure the mechanisms of the Internet by improving protocols and routing;
4. Foster the use of trusted digital control systems/supervisory control and data acquisition systems;
5. Reduce and remediate software vulnera-bilities;
6. Understand infrastructure interdepen-dencies and improve the physical security of cyber systems and telecommunications;
7. Prioritize federal cybersecurity research and development agendas; and
8. Assess and secure emerging systems.

IP prefix delegation

Butler *et al.*: A Survey of BGP Security Issues and Solutions

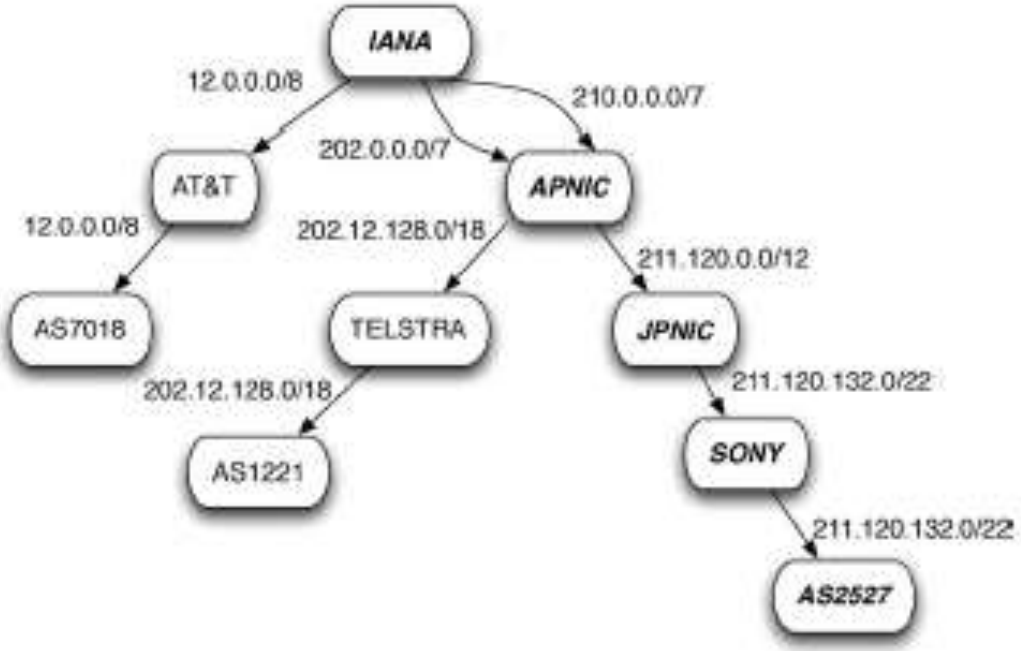


Fig. 1. ***An example of address delegation from the root (IANA) to regional and national registries.***

Routing security

- Prefix hijacking
  - Advertise you handle a prefix of another AS
  - e.g. Pakistan Telecom vs. YouTube, Feb 24th 2008
    - Government didn't like video, orders ISPs to block:

Corrigendum- Most Urgent

GOVERNMENT OF PAKISTAN
PAKISTAN TELECOMMUNICATION AUTHORITY
ZONAL OFFICE PESHAWAR
Plot-11, Sector A-3, Phase-V, Hayatabad, Peshawar.
Ph: 091-9217279- 5829177 Fax: 091-9217254
www.pta.gov.pk

NWFP-33-16 (BW)/06/PTA February ,2008

Subject: **Blocking of Offensive Website**

Reference: This office letter of even number dated 22.02.2008.

I am directed to request all ISPs to immediately block access to the following website

URL: http://www.youtube.com/watch?v=o3s8jtvvg00

IPs: 208.65.153.238, 208.65.153.253, 208.65.153.251

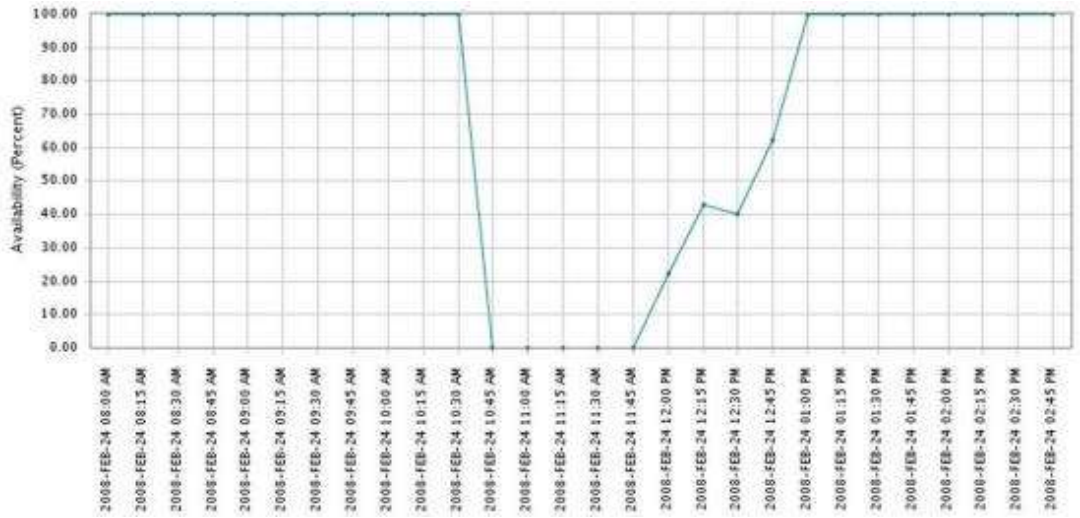
Compliance report should reach this office through return fax or at email peshawar@pta.gov.pk today please.

Prefix hijacking

- 18:48 Pakistan Telecom (AS 17557) starts advertising 208.65.153.0/24
- Its provider PCCW (AS 3491) propagates change, spreads worldwide
- YouTube only advertising 208.65.152.0/22, less specific so all YouTube traffic starts routing to Pakistan Telecom black hole
- 20:07 YouTube starts advertising 208.65.153.0/24
- 20:18 YouTube starts advertising 208.65.153.128/25, 208.65.153.0/25
- 21:01 PCCW withdraws prefixes from Pakistan Telecom

18:47
http://www.youtube.com/watch?v=l69Vi5IDc0g

18:48
http://www.youtube.com/watch?v=IzLPKuAOe50



Worldwide availability of YouTube (Keynote Systems)

Prefix hijacking

- Apr 1997: AS 7007 incident
  - Router at MAI Network services accidently leaks entire routing table
  - Leaks with /24 prefix, make it a more specific route to most of the Internet
- Dec 2004: TTNet pretends to be entire Internet
- Jan 2006: Con-Edison hijacks chunk of Internet
- Apr 2010: Chinese ISP hijacks Internet

Hijacking hard to debug

- Victim AS may not see a problem
  - Can continue to route inside its AS
- Hijack may not cause loss of connectivity
  - Hijacker may just be snooping and still deliver traffic
  - May cause performance degradation
- Loss of connectivity may be isolated
  - Only certain parts of Internet affected

Secure routing

- Origin authentication
  - Secure database mapping IP prefixes to owner ASes
- Protecting advertisements
  - Avoid inserting, deleting thing into path
  - Protecting TCP conversations between routers
- Secure BGP
  - Accurate registries, public key infrastructure, encryption, needs to be fast
  - Deployment difficult

Summary

- Business relationships between ASes
  - Customer-provider, paying for transit
  - Peer-peer, settlement-free
  - Tier 1, 2, 3
- Border Gateway Protocol (BGP)
  - Global Internet routing
  - Path-vector protocol
  - Allows ASes to enforce business policies
  - Security issues