# Routing in the Internet

Daniel Zappala

CS 460 Computer Networking
Brigham Young University

# Scaling Routing for the Internet

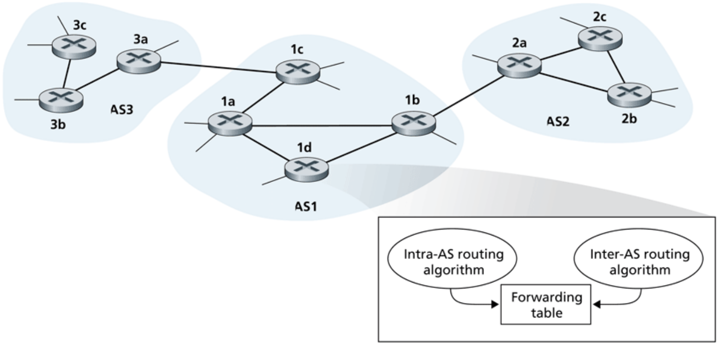
- scale
  - 200 million destinations - can't store all destinations or all prefixes in routing tables
  - link-state: flood link state packets to all hosts in the entire Internet
  - distance-vector: send routing table for all networks to each of your neighbors
- administrative authority
  - the Internet is a network of networks
  - each network administrator wants to control routing in her organization – may even use a different routing algorithm

# Hierarchical Routing

# Hierarchical Routing

- aggregate routers into regions: domains or autonomous systems (AS)
- *intra-domain routing*
    - routing within a domain
    - run a single routing protocol in the domain
- *inter-domain routing*
    - routing between domains
    - every domain must agree to run the same inter-domain routing protocol
- *border router or gateway*
    - router at the border of your domain and a peer, runs
    - must run both intra- and inter-domain routing protocols

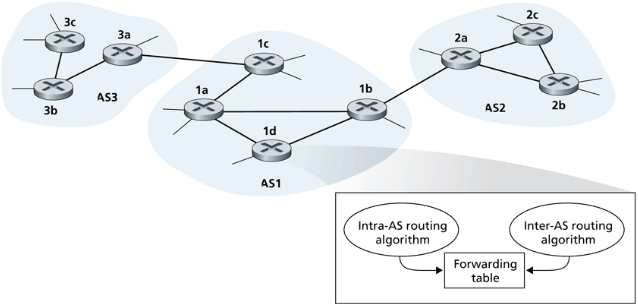
# Domains and Border Routers



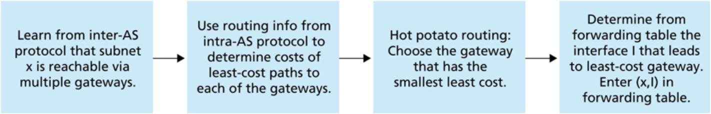

- forwarding table entries on a border router are created by both the intra-domain and inter-domain routing protocols
  - intra-domain sets routes for internal destinations
  - inter-domain sets routes for external destinations

# Hierarchical Routing

- router in AS1 gets a datagram for an external destination
- which border router does it choose?
- inter-domain routing protocol needs to
  - learn destinations reachable through each border router
  - propagate routes to all routers inside the domain
  - some destinations may be reachable by more than one border router – choose the closest one



# Hierarchical Routing Procedure



# Intra-Domain Routing

# Intra-Domain Routing

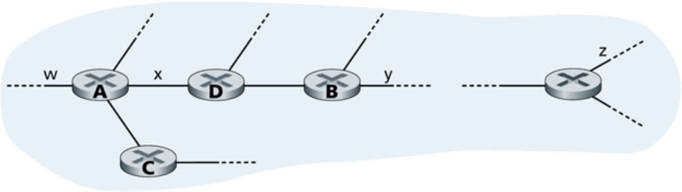
- also known as an interior gateway protocol (IGP)
- most common protocols
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (Cisco)

# RIP

# RIP: Routing Information Protocol

- distance-vector algorithm
- included in BSD-UNIX in 1982, most Unix and Linux distributions since then
- each link cost = 1, infinity = 16 (limits counting to infinity problem)
- exchanges distance vectors with neighbors every 30 seconds (called an advertisement)
- each advertisement contains a list of up to 25 destination networks
- RIP2 - supports subnet masks, adds authentication for advertisements
- RIPng - supports IPv6

# RIP Example




- routing table for D:

| Destination Subnet | Next Router | Number of Hops to Destination |
|---|---|---|
| w | A | 2 |
| y | B | 2 |
| z | B | 7 |
| x | — | 1 |
| .... | .... | .... |

# RIP Example

- advertisement from A:

| Destination Subnet | Next Router | Number of Hops to Destination |
|---|---|---|
| z | C | 4 |
| w | — | 1 |
| x | — | 1 |
| .... | .... | .... |

- routing table for D:

| Destination Subnet | Next Router | Number of Hops to Destination |
|---|---|---|
| w | A | 2 |
| y | B | 2 |
| z | A | 5 |
| .... | .... | .... |

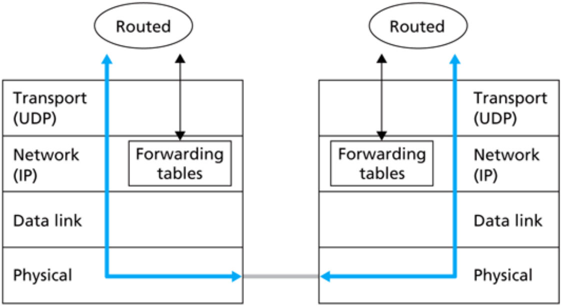
- D changes its route for z to use A instead of B (cost of 5 instead of 7)

# RIP Link Failure and Recovery

- if no advertisement heard after 180 sec, neighbor/link declared dead
    - routes using neighbor invalidated
    - new advertisements sent to neighbors
    - neighbors in turn send out new advertisements (if tables changed)
    - link failure info quickly propagates to entire network
    - poison reverse used help with count-to-infinity

# RIP Table Processing

- RIP run as application-level process called routed (route daemon)
- advertisements sent in UDP packets, periodically repeated

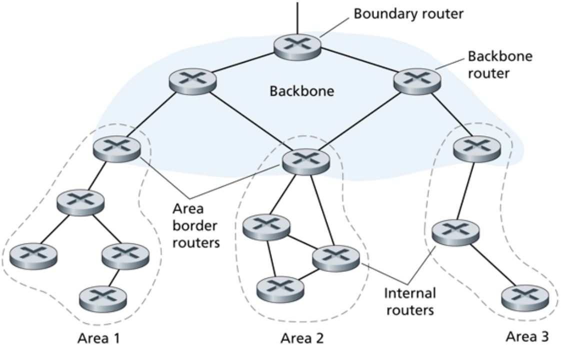


# OSPF

# OSPF: Open Shortest Path First

- open: publicly available
- uses link-state algorithm
  - link-state advertisements (LSAs) contain one entry per neighbor router
  - LSAs sent to each router in the domain
  - LSAs sent as OSPF messages directly over IP (no TCP and no UDP)
- security: all messages authenticated
- multi-path: multiple same-cost paths allowed
- TOS: multiple cost metrics per link (e.g. satellite can be low cost for bandwidth, high cost for latency)
- multicast support: MOSPF uses OSPF link-state database
- hierarchical: divide a domain into multiple areas

# OSPF Hierarchy




- area routers learn topology and routes for area
- area border routers summarize distances for networks in their area, advertise to other area routers on backbone
- backbone routers run OSPF on the backbone
- boundary routers connect to Internet

# Inter-Domain Routing

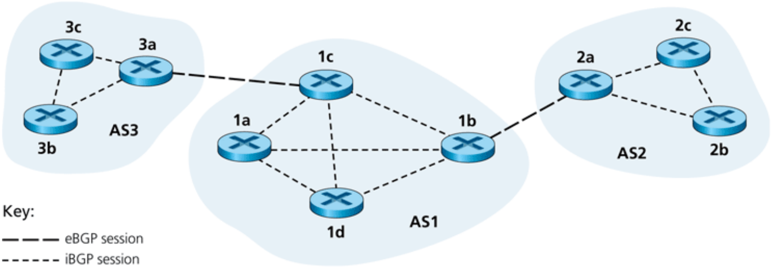
# BGP

# Inter-Domain Routing: BGP

- Border Gateway Protocol (BGP) - the standard for Internet inter-domain routing
- BGP allows domains to
  - advertise routes for internal networks to the rest of the Internet
  - obtain routes for external networks from other domains
  - use policy to select routes (not just shortest path)

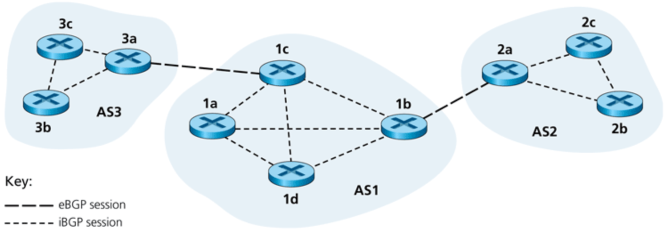
# BGP Basics

- BGP peers (routers) establish TCP connections and exchange routing information (may span several non-BGP routers)
- when AS1 advertises a prefix (network) to AS2, AS1 is promising it will forward any datagrams sent to that prefix
- prefixes can be aggregated along any bit boundary



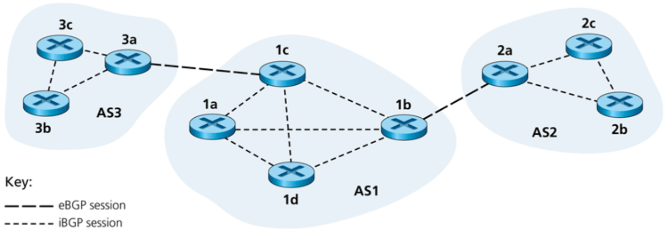
# BGP Reachability

- advertise prefixes that are *reachable* by your domain
- example
  - 3a uses BGP to send reachability info to 1c for internal networks
  - 1c uses OSPF/IGRP to distribute reachability to other routers in AS1
  - 1b uses BGP to advertise these networks to 2a
  - any router that learns about a new/updated prefix creates/updates forwarding table entry



# BGP Attributes

- attributes may be attached to prefixes = *route*
- important attributes
  - AS-PATH: an ordered list of ASs in the route
  - NEXT-HOP: IP address of the router which should be used as the BGP next hop to the destination
- example
  - when 3a advertises a route to 1c, it uses its own IP address as the NEXT-HOP and the AS-PATH is AS3.
  - when 1b advertises the same route to 2a, it changes the NEXT-HOP to 1b's address, and the AS-PATH is AS3-AS1



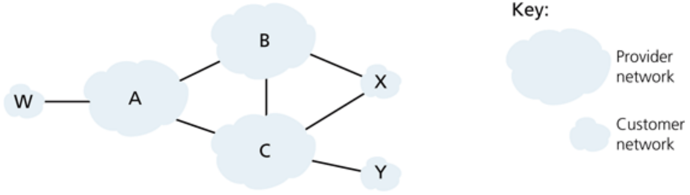
# BGP Route Selection

- BGP routers use policies to determine which routes they will accept and advertise to their peers
- policy eliminates routes for which you don't want to carry traffic)
- route selection among multiple routes for same prefix : complicated rules)
  - largest weight
  - highest local preference (e.g. prefer directly-connected routes, or routes over Internet2)
  - shortest AS-PATH
  - cheapest internal route to BGP NEXT-HOP

# BGP Messages

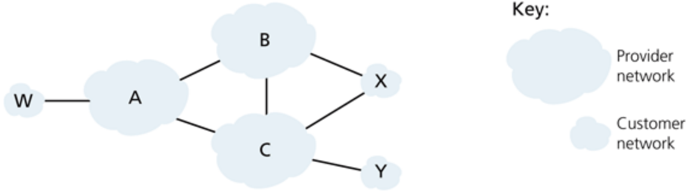
- exchanged using TCP
- message types
  - OPEN: open TCP connection to peer and authenticate sender
  - UPDATE: advertise new paths, withdraw old paths
  - KEEPALIVE: keep connection alive in absence of updates, ACKs OPEN request
  - NOTIFICATION: reports errors, can also close connection

# BGP Policy



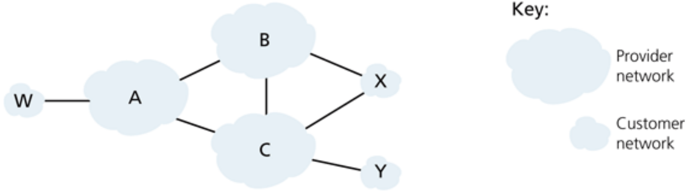

- provider networks: A, B, C
- customer networks: X, W, Y
- X is dual-homed: attached to two networks
- policy
    - X does not want to route from B to C via itself
    - X will not advertise to B a route for C

# BGP Policy




- A advertises path AW to B
- B advertises path BAW to X
- should B advertise path BAW to C?

# BGP Policy




- A advertises path AW to B
- B advertises path BAW to X
- should B advertise path BAW to C?
  - No! B gets no benefit from routing CBAW since neither C nor W are customers of B
  - B wants to force C to route via A
  - B wants to route only to/from its own customers

# Separation of Concerns

- policy
  - inter-domain: want control over how traffic is routed, who routes to domain, needs policy
  - intra-domain: single administrator, so no policy decisions needed
- scale
  - hierarchical routing saves table size, reduces update traffic
- performance
  - intra-domain focuses on performance
  - inter-domain focuses on global reachability, policy