BGP

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-AS routing*
- *inter-AS routing*
  - every domain must agree to run the same inter-domain routing protocol
- *border router or gateway*
  - router at the border of your AS and a peer, runs
  - must run both intra- and inter-AS routing protocols

Intra-AS Routing

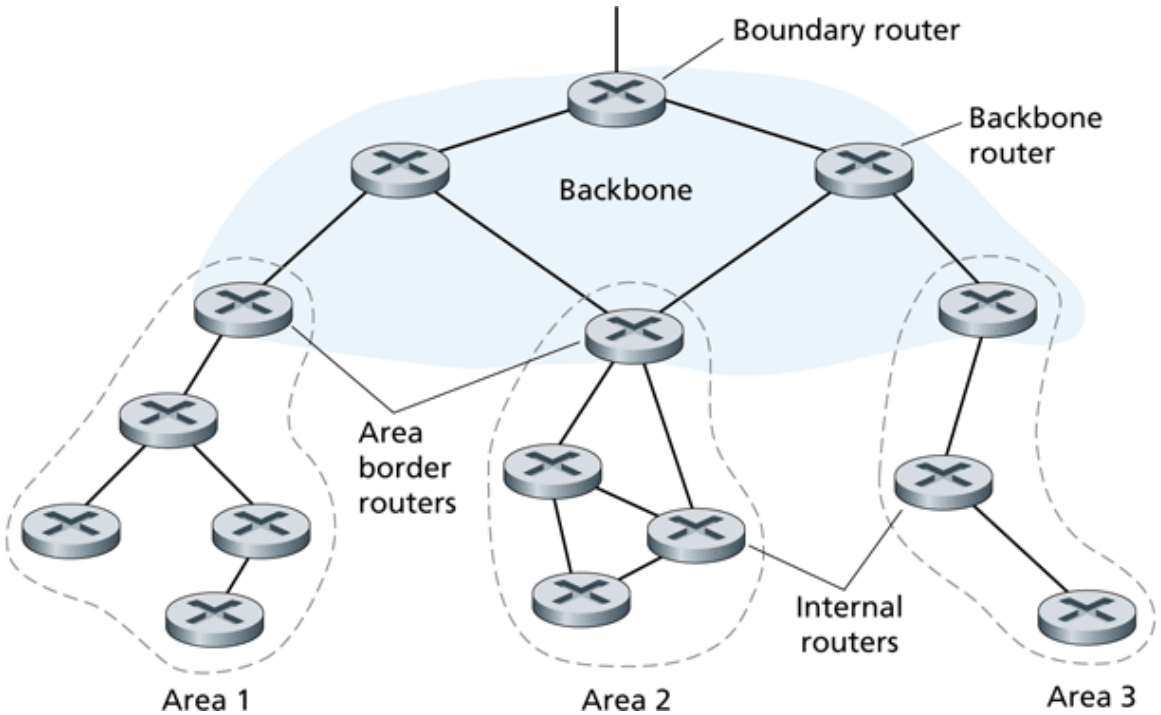
Intra-AS Routing

- most common protocols
  - OSPF: Open Shortest Path First
  - IS-IS: Intermediate System to Intermediate System (very similar to OSPF)
  - EIGRP: Enhanced Interior Gateway Routing Protocol (Cisco)

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-AS Routing

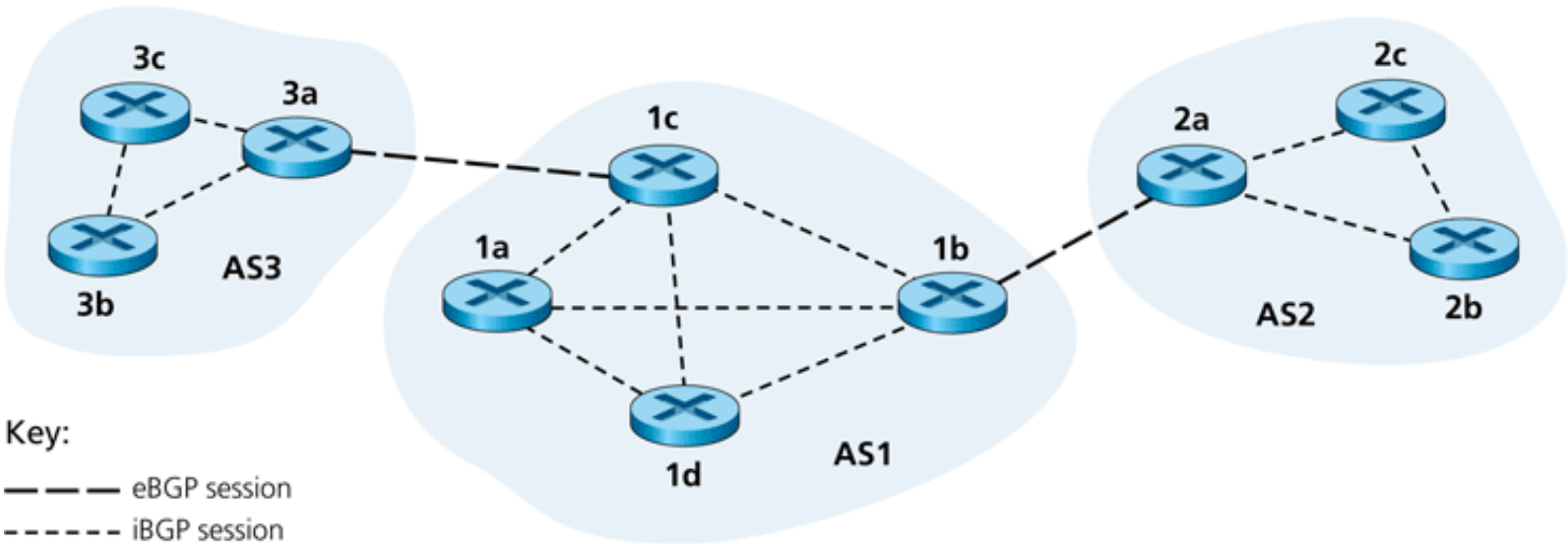
BGP

Inter-AS Routing: BGP

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

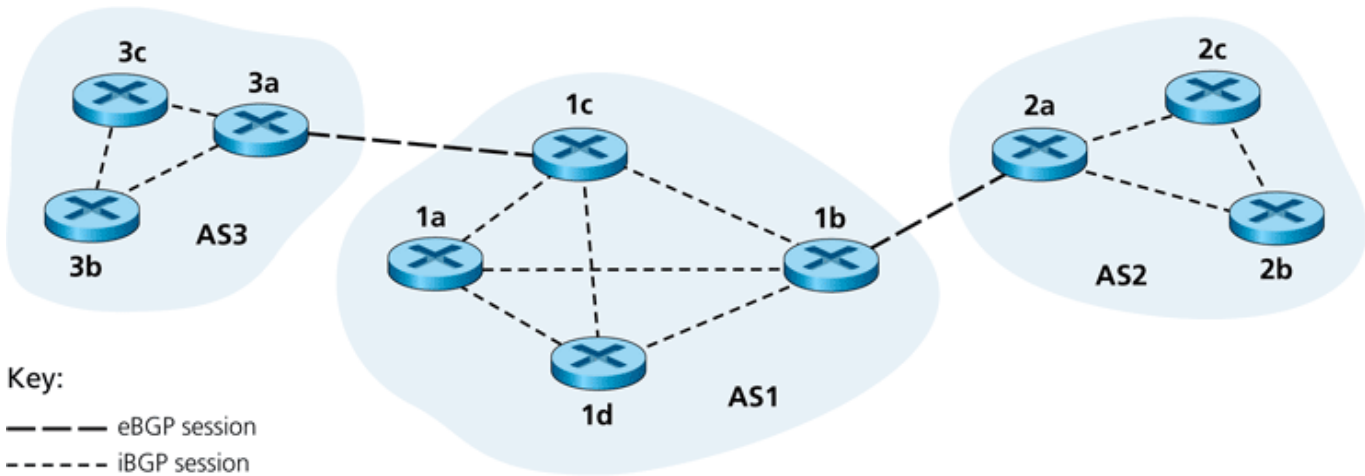
BGP Basics

- BGP peers (routers) establish TCP connections and exchange routing information
- external BGP (eBGP) connections typically correspond to a single link
- internal BGP (iBGP) connections may span several non-BGP routers



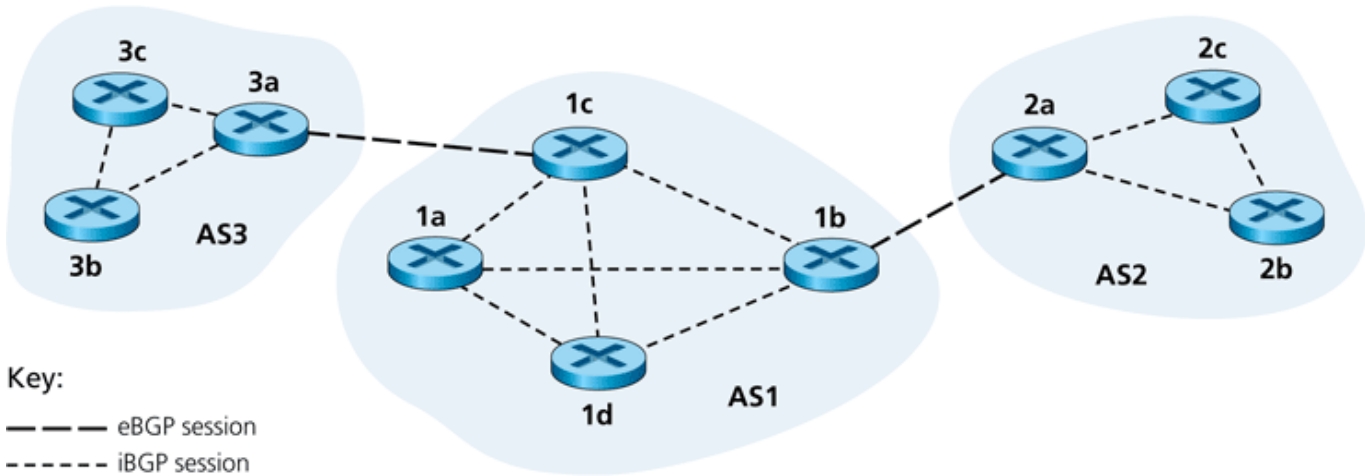
BGP Advertisements

- advertise prefixes that are *reachable* by your AS
- example
  - 3a uses eBGP to advertise network x to 1c
  - 1c uses iBGP to advertise this to other routers in AS1
  - 1b uses eBGP to advertise network x to 2a
  - accumulated path is AS1 AS2 (x)



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



Hot Potato Routing

- choose the path with the least cost to the NEXT-HOP router
  - two routes: AS1 AS2 x, AS2 x
  - find cost to the NEXT-HOP for AS1 and AS2
  - choose the least cost
- does not choose shortest AS path!
- gets the packet out of your AS as quickly as possible

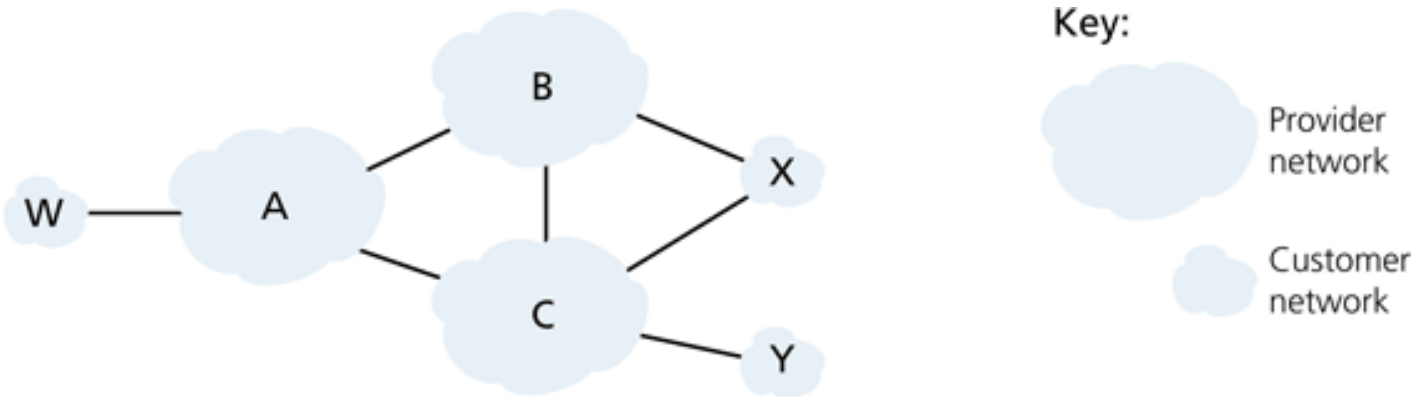
BGP Route Selection in Practice

- 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 : follow these rules in order)
  - highest local preference (e.g. prefer directly-connected routes, or routes over Internet2)
  - shortest AS-PATH
  - hot potato
  - use BGP identifiers (e.g. IP address)

IP Anycast

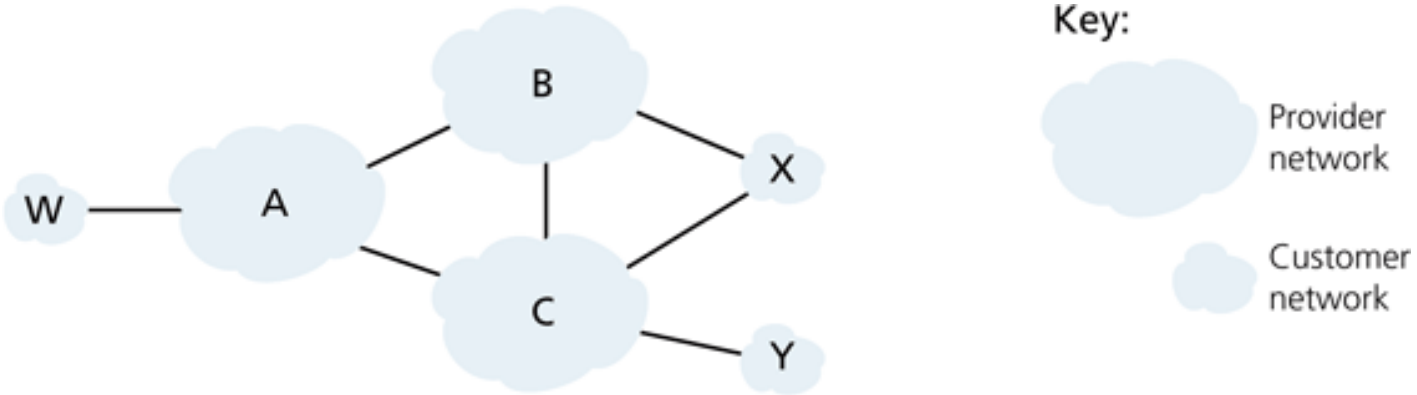
- motivation
  - many applications want to replicate data across many servers in diverse geographic locations
  - each user should get data from the closest server
- example: DNS
- simple with BGP: advertise a route to a prefix from multiple locations

BGP Policy



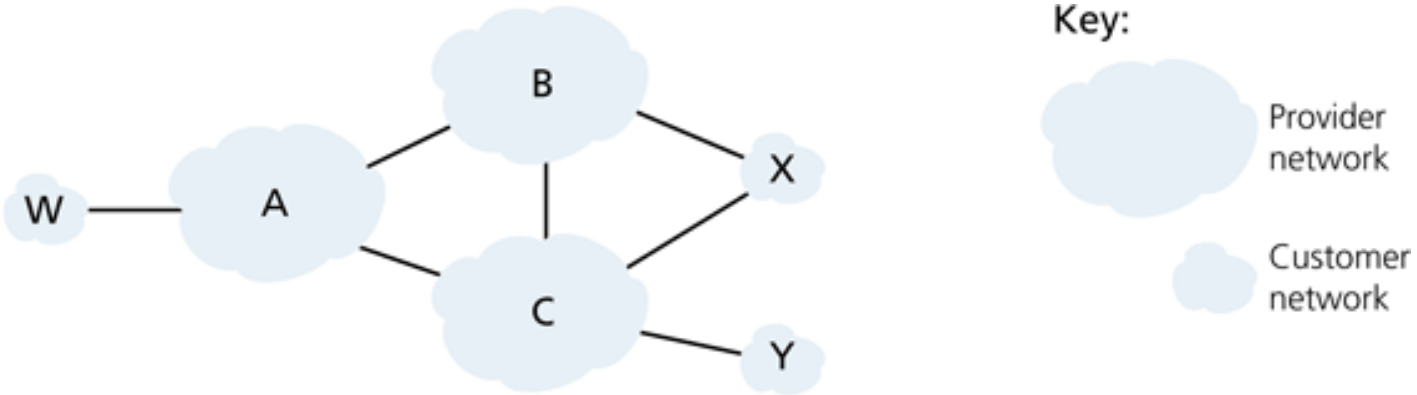
- provider networks: A, B, C
- customer networks: X, W, Y
- X is dual-homed: attached to two networks
- policy
  - X does not want to route from B to C via itself
  - X will not advertise to B a route for C

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