

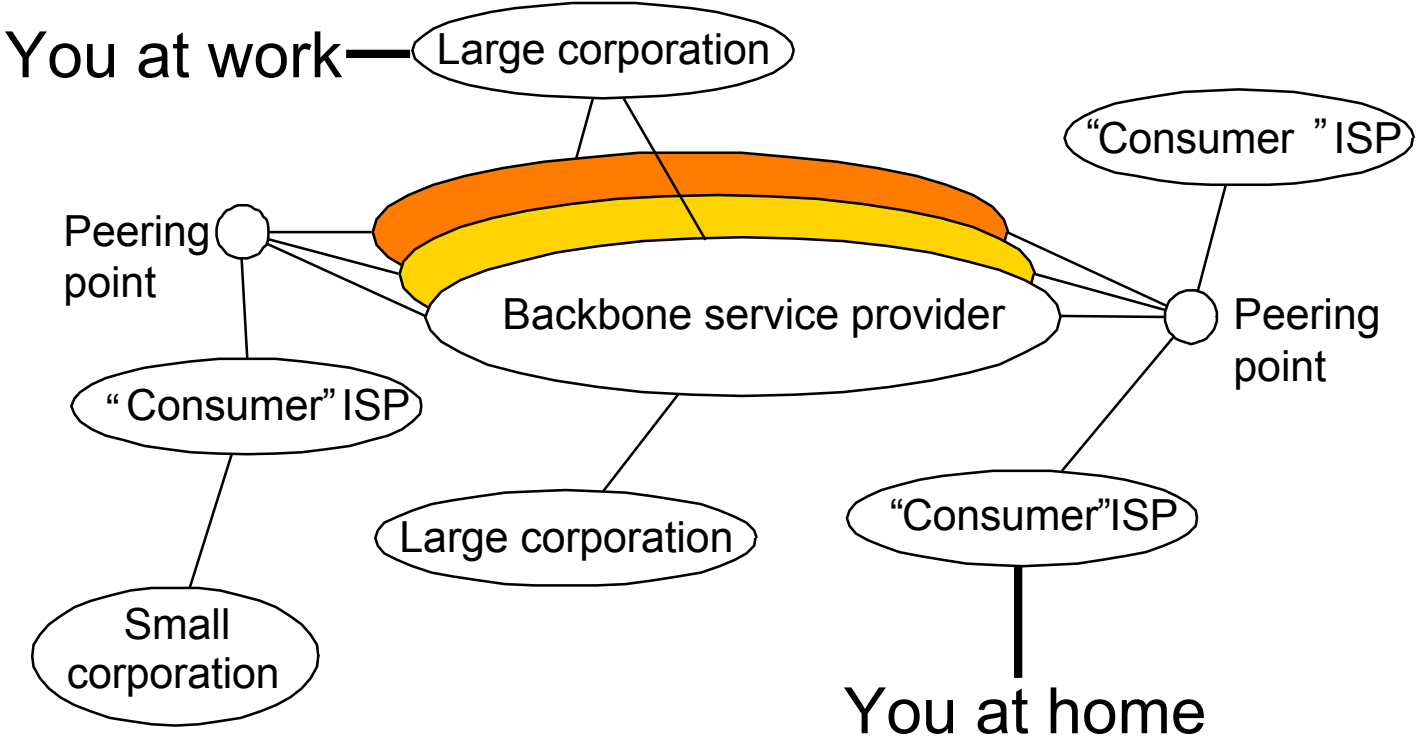
CSC2209
Computer Networks

BGP is Broken

Stefan Saroiu
Computer Science
University of Toronto

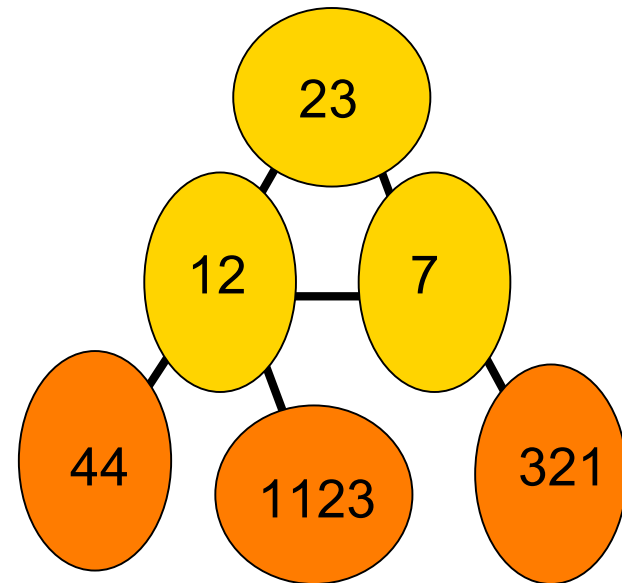
Structure of the Internet

- Inter-domain versus intra-domain routing



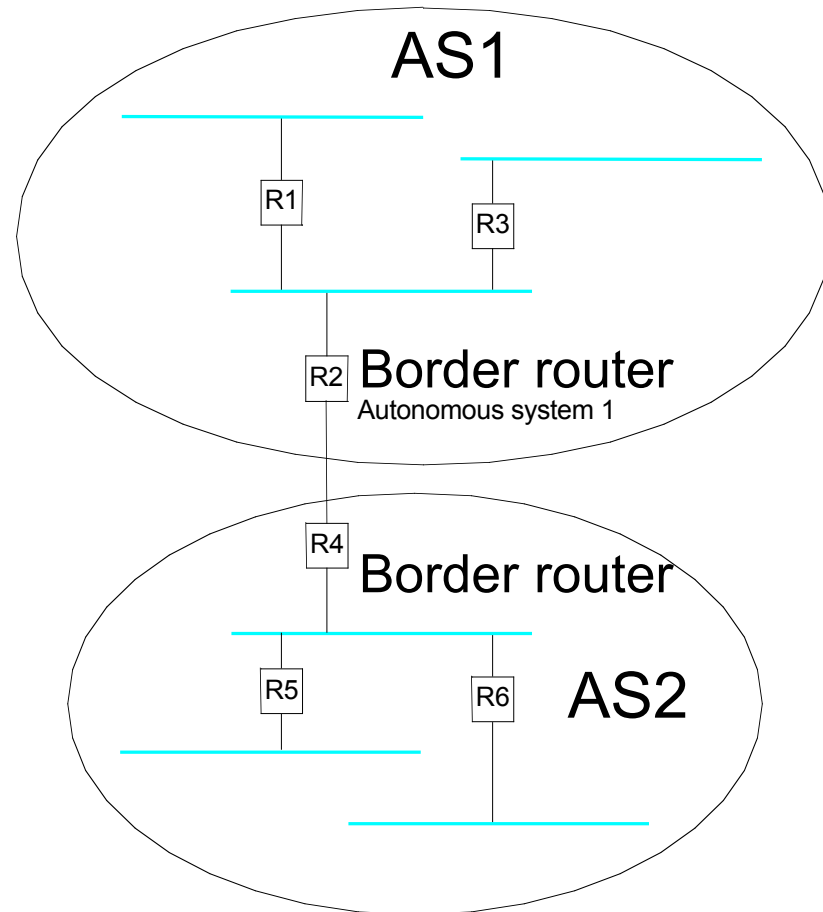
Inter-Domain Routing

- 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



Inter-Domain Routing

- Border routers summarize and advertise internal routes to external neighbors and vice-versa
- Border routers apply policy
- Internal routers can use notion of default routes
- Core is “default-free”; routers must have a route to all networks in the world

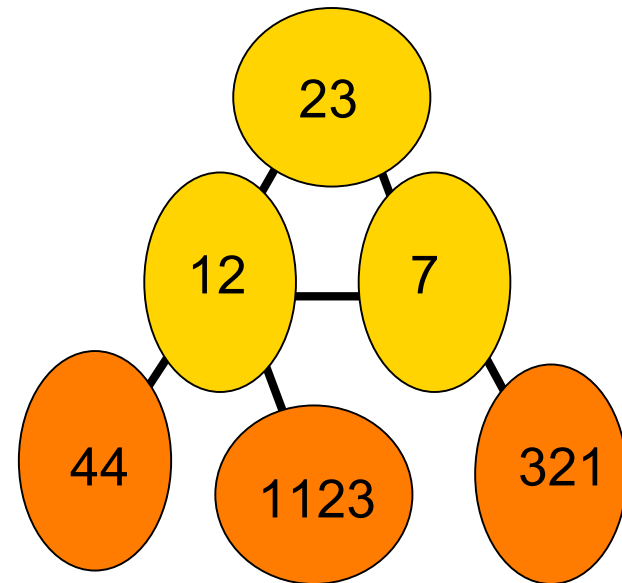


Border Gateway Protocol (BGP-4)

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

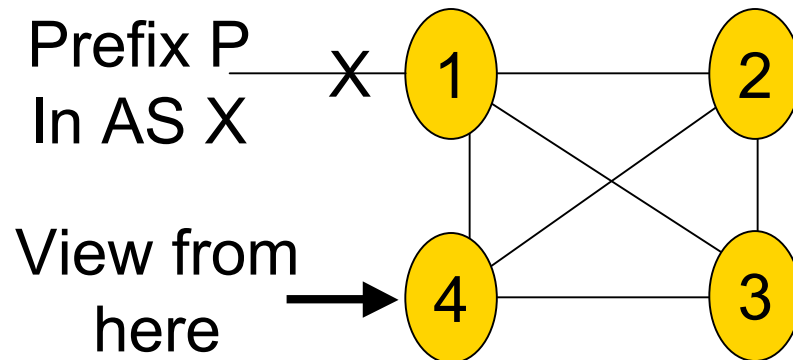
Path Vectors

- 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



An Ironic Twist on Convergence

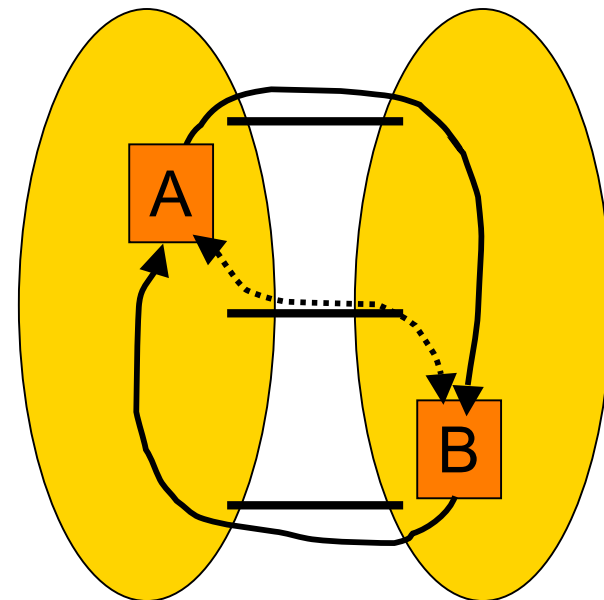
- BGP convergence can undergo a process analogous to count-to-infinity!



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- So 4 uses 4 2 1 X, which is soon withdrawn, then 4 3 2 1 X, ...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

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



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?

Q:

- Should a router keep re-sending its UPDATES to other routers?

Q:

- Should a router keep re-sending its UPDATES to other routers?
 - No, because TCP guarantees reliable delivery
 - How are failures detected then?

Q:

- Should a router keep re-sending its UPDATES to other routers?
 - No, because TCP guarantees reliable delivery
 - How are failures detected then?
 - BGP implements a KEEPALIVE mechanism on top of TCP

eBGP Sessions

- Between routers in different AS's
- Each eBGP router of an AS learns about some subset of all prefixes in the AS
 - Advertises them outside of the AS
- Ensures:
 - Loop-free forwarding
 - Complete visibility:
 - How do two eBGP routers of the same AS exchange information?

iBGP Sessions

- Between iBGP routers of same AS
 - Full mesh
 - How are these messages routed in an AS?

iBGP Sessions

- Between iBGP routers of same AS
 - Full mesh
 - How are these messages routed in an AS?
 - Through the intra-domain routing protocol
 - Why not use the intra-domain routing protocol in the first place

iBGP Sessions

- Between iBGP routers of same AS
 - Full mesh
 - How are these messages routed in an AS?
 - Through the intra-domain routing protocol
 - Why not use the intra-domain routing protocol in the first place
 - BGP's announcements use rich set of attributes to advertise paths
 - Intra-domain protocols are very "chatty"

Does iBGP scale well?

Does iBGP scale well?

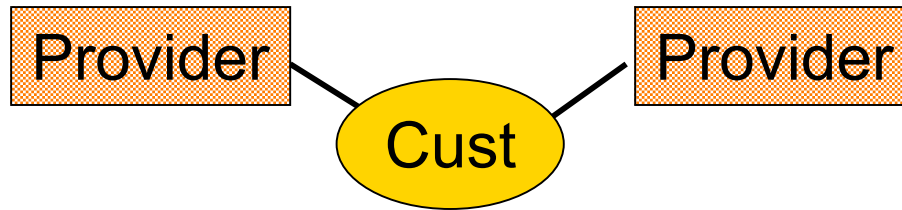
- No:
 - $O(n^2)$ sessions
 - Not a problem for small AS's
 - But there are AS's with hundreds of iBGP routers
 - Tier-1 ISPs
- How do we make iBGP scale then?

How to Scale iBGP

- Introduce hierarchy:
 - Route reflectors
 - Each reflector has a set of BGP routers as clients
 - If route update received from another iBGP
 - Send update to clients
 - If route update received from clients or if route update from eBGP
 - Send update to other iBGP

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 the customer to influence this