



15-441 Computer Networking

Inter-Domain Routing
Border Gateway Protocol -BGP
Peter Steenkiste

Fall 2010
www.cs.cmu.edu/~prs/15-441-F10

Important Lessons From Last Lecture



- Every router needs to be able to forward towards any destination
 - Forwarding table must be complete
 - Can rely on friends to tell you how to get there (DV)
 - Can get an entire map of the network (LS)
- Key challenges
 - What if a router fails or is added? → need to inform everyone
 - Soft-state recovery
 - What if people have inconsistent/different views?
 - Loops, count to infinity

2

Routing Review



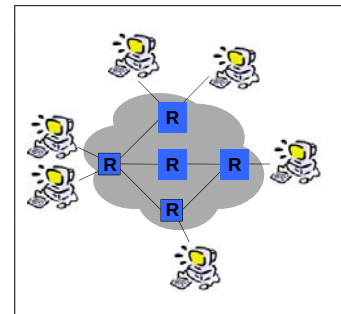
- The Story So Far...
 - IP forwarding requires next-hop information
 - Routing protocols generate the forwarding table
 - Two styles: distance vector, link state
 - Scalability issues:
 - Distance vector protocols suffer from count-to-infinity
 - Link state protocols must flood information through network
- Today's lecture
 - How to make routing protocols support large networks
 - How to make routing protocols support business policies

3

A Logical View of the Internet?



- After looking at RIP/OSPF descriptions
 - End-hosts connected to routers
 - Routers exchange messages to determine connectivity
- NOT TRUE!



5

Routing Hierarchies



- Flat routing doesn't scale
 - Storage → Each node cannot be expected to store routes to every destination (or destination network)
 - Convergence times increase
 - Communication → Total message count increases
- Key observation
 - Need less information with increasing distance to destination
 - Need lower diameters networks
- Solution: hierarchy
 - Example: area hierarchies in OSPF
 - Today: Internet

6

Internet's Area Hierarchy



- What is an Autonomous System (AS)?
 - A set of routers under a single technical administration, using an *interior gateway protocol (IGP)* and common metrics to route packets within the AS and using an *exterior gateway protocol (EGP)* to route packets to other AS's
- Each AS assigned unique ID
- AS's peer at network exchanges

8

AS Numbers (ASNs)



ASNs are 16 bit values 64512 through 65535 are "private"

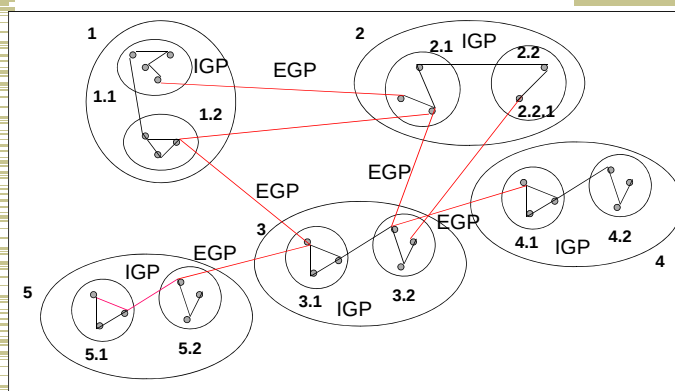
Currently over 15,000 in use

- Genuity: 1
- MIT: 3
- CMU: 9
- UC San Diego: 7377
- AT&T: 7018, 6341, 5074, ...
- UUNET: 701, 702, 284, 12199, ...
- Sprint: 1239, 1240, 6211, 6242, ...
- ...

ASNs represent units of routing policy

9

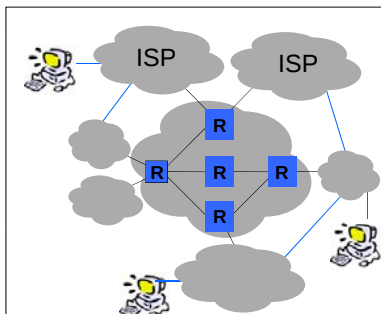
Example



10

A Logical View of the Internet?

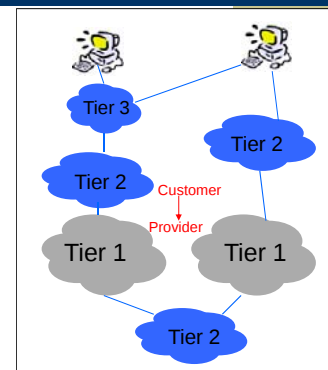
- RIP/OSPF not very scalable → area hierarchies
- NOT TRUE EITHER!
- ISP's aren't equal
 - Size
 - Connectivity



11

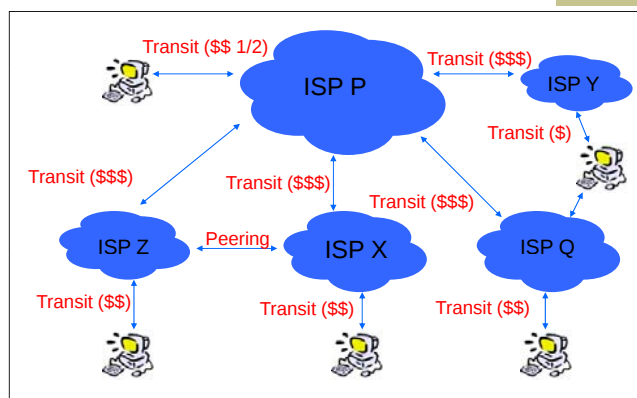
A Logical View of the Internet

- Tier 1 ISP
 - “Default-free” with global reachability info
- Tier 2 ISP
 - Regional or country-wide
- Tier 3 ISP
 - Local



12

A More Interesting Example



13

Policy Impact

- “Valley-free” routing
 - Number links as (+1, 0, -1) for provider, peer and customer
 - In any path should only see sequence of +1, followed by at most one 0, followed by sequence of -1
- WHY?
 - Consider the economics of the situation
 - Why does an ISP forward packets?

14

History



- Mid-80s: EGP
 - Reachability protocol (no shortest path)
 - Did not accommodate cycles (tree topology)
 - Evolved when all networks connected to NSF backbone
- Result: BGP introduced as routing protocol
 - Latest version = BGP 4
 - BGP-4 supports CIDR
 - Primary objective: connectivity not performance

16

Choices



- Link state or distance vector?
 - No universal metric – policy decisions
- Problems with distance-vector:
 - Bellman-Ford algorithm may not converge
- Problems with link state:
 - Metric used by routers not the same – loops
 - LS database too large – entire Internet
 - May expose policies to other AS's

17

Solution: Distance Vector with Path



- Each routing update carries the entire path
- Loops are detected as follows:
 - When AS gets route, check if AS already in path
 - If yes, reject route
 - If no, add self and (possibly) advertise route further
- Advantage:
 - Metrics are local - AS chooses path, protocol ensures no loops

18

Interconnecting BGP Peers



- BGP uses TCP to connect peers
- Advantages:
 - Simplifies BGP
 - No need for periodic refresh - routes are valid until withdrawn, or the connection is lost
 - Incremental updates
- Disadvantages
 - Congestion control on a routing protocol?
 - Poor interaction during high load

19

Hop-by-hop Model



- BGP advertises to neighbors only those routes that it uses
 - Consistent with the hop-by-hop Internet paradigm
 - e.g., AS1 cannot tell AS2 to route to other AS's in a manner different than what AS2 has chosen (need source routing for that)
- BGP enforces policies by
 1. choosing paths from multiple alternatives and
 2. controlling advertisement to other AS's

20

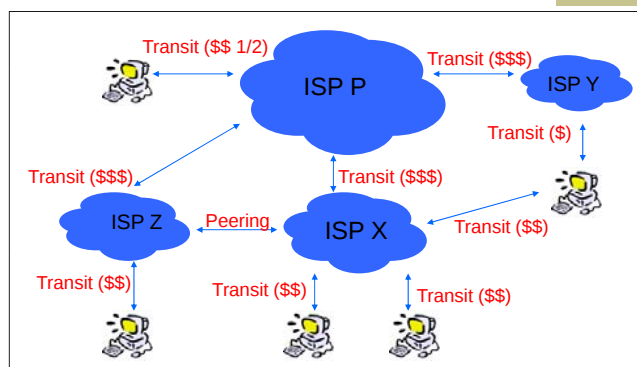
Examples of BGP Policies



- A multi-homed AS refuses to act as transit
 - Limit path advertisement
- A multi-homed AS can become transit for some AS's
 - Only advertise paths to some AS's
- An AS can favor or disfavor certain AS's for traffic transit from itself
 - By choosing those paths among the options

21

Some Examples



22

BGP Messages



- Open
 - Announces AS ID
 - Determines hold timer – interval between keep_alive or update messages, zero interval implies no keep_alive
- Keep_alive
 - Sent periodically (but before hold timer expires) to peers to ensure connectivity.
 - Sent in place of an UPDATE message
- Notification
 - Used for error notification
 - TCP connection is closed *immediately* after notification

23

BGP UPDATE Message



- List of withdrawn routes
- Network layer reachability information
 - List of reachable prefixes
- Path attributes
 - Origin
 - Path
 - Metrics
- All prefixes advertised in message have same path attributes

24

Path Selection Criteria



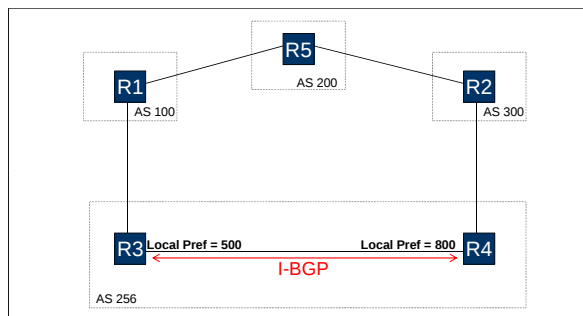
- Attributes + external (policy) information
- Examples:
 - Hop count
 - Policy considerations
 - Preference for AS
 - Presence or absence of certain AS
 - Path origin
 - Link dynamics

25

LOCAL PREF



- Local (within an AS) mechanism to provide relative priority among BGP routers (e.g. R3 over R4)



26

LOCAL PREF – Common Uses

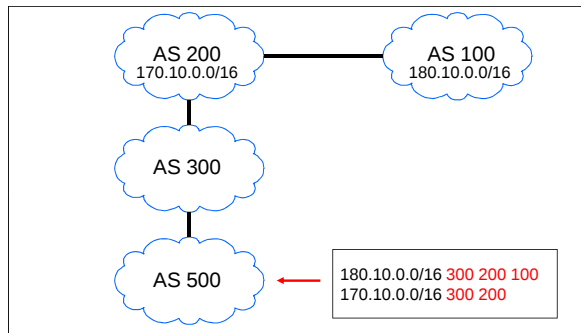


- Routers have a default LOCAL PREF
 - Can be changed for specific ASes
- Peering vs. transit
 - Prefer to use peering connection, why?
- In general, customer > peer > provider
 - Use LOCAL PREF to ensure this

27

AS_PATH

- List of traversed AS's



28

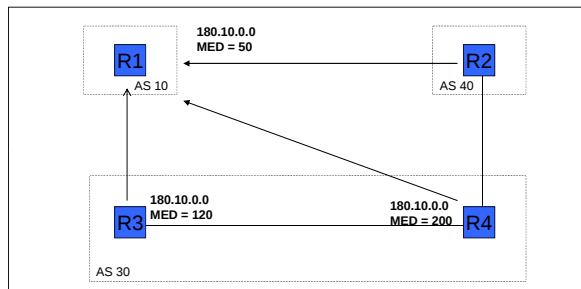
Multi-Exit Discriminator (MED)

- Hint to external neighbors about the preferred path into an AS
 - Non-transitive attribute
 - Different AS choose different scales
- Used when two AS's connect to each other in more than one place

29

MED

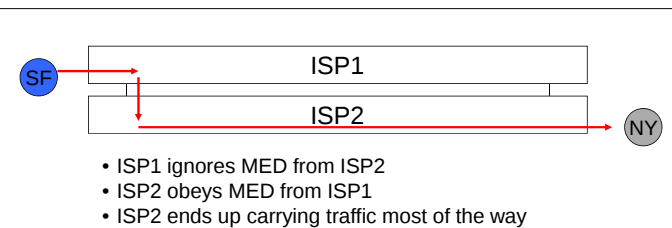
- Hint to R1 to use R3 over R4 link
- Cannot compare AS40's values to AS30's



30

MED

- MED is typically used in provider/subscriber scenarios
- It can lead to unfairness if used between ISP because it may force one ISP to carry more traffic:



31

Decision Process



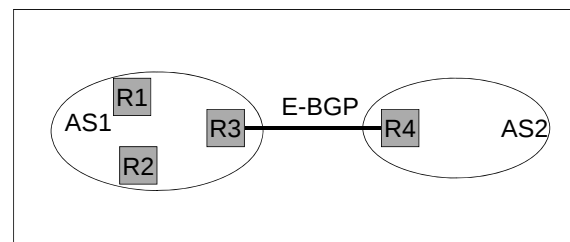
- Processing order of attributes:
 - Select route with highest LOCAL-PREF
 - Select route with shortest AS-PATH
 - Apply MED (if routes learned from same neighbor)

32

Internal vs. External BGP



- BGP can be used by R3 and R4 to learn routes
- How do R1 and R2 learn routes?
- Border gateways also need to run an internal routing protocol
 - Establish connectivity between routes inside AS



38

Important Concepts



- Wide area Internet structure and routing driven by economic considerations
 - Customer, providers and peers
- BGP designed to:
 - Provide hierarchy that allows scalability
 - Allow enforcement of policies related to structure
- Mechanisms
 - Path vector – scalable, hides structure from neighbors, detects loops quickly

44