Lecture 8: Inter-Domain Routing
Border Gateway Protocol - BGP
Justine Sherry
Peter Steenkiste
Fall 2017
www.cs.cmu.edu/~prs/15-441-F17

Outline
- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

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
  - Only transit domains really need it
- ASes peer with other ASes at network exchanges
  - “Gateway routers” forward packets across ASes
AS Numbers (ASNs)

ASNs are 16 bit values 64512 through 65535 are “private”
- 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

A Logical View of the Internet?

- Logical consequence of hierarchy: repeat the intra-domain solutions at inter-net level
- Based on IP and OSPF style routing protocols
- NOT TRUE!
- Lots of problems with this picture

A Logical View of the Internet

- Ases are commercial entities
  - Must make money!
- They play different roles in the Internet
  - Tier 1 ISP: global, internet wide connectivity
  - Tier 2 ISP: regional or country-wide
  - Tier 3 ISP: local
- This is an emergent property:
  - Businesses specialize
  - Business relationships

A More Interesting Example

- ISP X
  - Transit ($$)
  - Transit ($$$)
  - Peering
- ISP Z
  - Transit ($$)
  - Transit ($$$)
  - Peering
- ISP Q
  - Transit ($$$)
  - Transit ($$)
  - Transit ($$
- ISP Y
  - Transit ($$$)
  - Transit ($$)
  - Transit ($$)
- ISP P
  - Transit ($$ 1/2)
  - Transit ($$$)
  - Transit ($$)
Policy and Economics Rules

- WHY?
  - Consider the economics of the Internet
  - Why does an ISP forward packets?
- Emergent property: “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
  - -1 → 0 → +1 corresponds to a valley and means an ISP is forwarding packets for free
    - Worse: it is paying its providers for forwarding

Outline

- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

History

- Mid-80s: EGP
  - Reachability protocol (no shortest path)
  - Did not accommodate cycles (assumes tree topology)
  - Evolved when all networks connected to NSF backbone
- Commercialization led to richer topologies – Result: BGP introduced as routing protocol
  - Latest version is BGP-4 - supports CIDR
  - Primary objective:
    - Connectivity not performance
    - Respect business relationships
    - Allow for local policies in each AS

Choices

- Link state or distance vector?
  - Constraint: No universal metric – local policy decisions
- Problems with distance-vector:
  - Bellman-Ford algorithm may converge slowly
  - Problems with “count to infinity”
- Problems with link state:
  - If routers do not use the same metric – loops
  - LS database too large – entire Internet
  - May expose policies to other AS’s
Solution: Distance Vector with Path

- Each routing update carries the entire path
- Loops are detected as follows:
  - When AS gets route, check if its AS number is already in the 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

Policy-based Routing: AS 1 to X

1. Receive reachability destination for destination X
   - AS1 selects its path to X based on local policies
2. AS1 advertise its path to X selectively
   - Use local policies to decide who to advertise it to
   - Colors are flipped for AS 2

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 with other traffic during high load

Hop-by-hop Model

- BGP only advertises routes that it uses to its neighbors
- Consistent with the hop-by-hop Internet paradigm
  - e.g., AS1 cannot forward AS2’s packets to other AS’s in a manner different than what AS2 has chosen
  - Worse: can lead to forwarding loops
- BGP enforces policies by
  1. choosing paths from multiple alternatives and
  2. controlling advertisement to other AS’s
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

Some Examples

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

BGP UPDATE Message

- List of withdrawn routes
- Network layer reachability information
  - List of reachable prefixes
- Path attributes
  - Origin
  - Path
  - Metrics: used by policies for path selection
- All prefixes advertised in message have the same path attributes
AS_PATH

- List of traversed AS’s

LOCAL PREF

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

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

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
**MED**

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

### 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:
  - ISP1 ignores MED from ISP2
  - ISP2 obeys MED from ISP1
  - ISP2 ends up carrying traffic most of the way

**Path Selection Criteria**

- Attributes + external (policy) information
- Rough ordering for path selection
  - Highest LOCAL-PREF
    - Captures business relationships and other factors
  - Shortest AS-PATH
  - Lowest origin type
  - Lowest MED (if routes learned from same neighbor)
  - eBGP over iBGP-learned
  - Lowest internal routing cost to border router
  - Tie breaker, e.g., lowest router ID

**Routing and Forwarding in the Internet: Prefixes**

- Prefix “grows” along path
BGP and Prefixes

- BGP advertisements specify prefix reachability
  - Prefix \( \approx \) network ID in a CIDR world
- BGP can also merge advertisements:
  - Example: 4 “/20” advertisements that share the top 18 bits in their prefix can become a single “/18” adv., if the reachability information is the same
- Can also leverage the longest prefix rule to merge entries:
  - Example: if only three of the prefix share reachability information, you can create a “/18” and a “/20” prefix

Outline

- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

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 routers inside AS
  - I-BGP: uses same messages as E-BGP

I-BGP Route Advertisements

- I-BGP uses different rules about re-advertising prefixes:
  - Prefix learned from E-BGP can be advertised to I-BGP neighbor and vice-versa, but
  - Prefix learned from I-BGP neighbors cannot be advertised to other I-BGP neighbors → direct connections (TCP) for I-BGP routers
  - Reason: AS PATH is the same AS and thus danger of looping.
How Do ISPs Peer?

- Public peering: use network to connect large number of ISPs in Internet eXchange Point (IXP)
  - Managed by IXP operator
  - Layer 2 private network
  - Efficient: can have 100s of ISPs
  - Has led to increase in peering
- Private peering: directly connect ISP border routers
  - Set up as private connection
  - Typically done in an Internet eXchange Point (IXP)

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