Exterior Gateway Protocols: EGP, BGP-4, CIDR

Shivkumar Kalyanaraman
Rensselaer Polytechnic Institute
shivkuma@ecse.rpi.edu
http://www.ecse.rpi.edu/Homepages/shivkuma

Intra-AS and Inter-AS routing

Gateways:

- perform inter-AS routing amongst themselves
- perform intra-AS routing in their AS

Default Routes: limits

- Default routes => partial information
- Routers/hosts w/ default routes rely on other routers to complete the picture.
- In general routing "signposts" should be:
  - Consistent, i.e., if packet is sent off in one direction then another direction should not be more optimal.
  - Complete, i.e., should be able to reach all destinations

Core

- A small set of routers that have consistent & complete information about all destinations.
- Outlying routers can have partial information provided they point default routes to the core
- Partial info allows site administrators to make local routing changes independently.

Peer Backbones

- Initially NSFNET had only one connection to ARPANET (router in Pittsburg) => only one route between the two.
- Addition of multiple interconnections => multiple possible routes => need for dynamic routing
- Single core replaced by a network of peer backbones => more scalable
- Today there are over 30 backbones!
- Routing protocol at cores/peers: GGP -> EGP -> BGP-4
Autonomous Systems (AS)
- AS = set of routers and networks under the same administration
- No theoretical limit to the size of the AS
- All parts within an AS remain connected.
- If two networks rely on core-AS to connect, they don’t belong to a single AS
- AS is identified by a 16-bit AS number
- At least one border router per AS.
- This router also collects reachability information ("external routes") and diffuses it internally and vice versa.

Autonomous Systems (Continued)
- AS types:
  - Stub AS => only single connection to one other AS => it carries only local traffic.
  - Multihomed AS: Connected to multiple AS, but does not allow transit traffic
  - Transit AS: carries transit traffic under policy restrictions
- Traffic types:
  - Local = traffic originating or terminating at AS.
  - Transit = non-local traffic

Exterior Gateway Protocol (EGP)
- A mechanism that allows non-core routers to learn routes from core routers so that they can choose optimal backbone routes
- A mechanism for non-core routers to inform core routers about hidden networks
- Autonomous System (AS) has the responsibility of advertising reachability info to other ASs.
- One+ routers may be designated per AS.
- Important that reachability info propagates to core routers

EGP weaknesses
- EGP does not interpret the distance metrics in routing update messages => cannot be compute shorter of two routes
- As a result it restricts the topology to a tree structure, with the core as the root
- Rapid growth => many networks may be temporarily unreachable
- Only one path to destination => no load sharing

Border Gateway Protocol (BGP)
- Allows multiple cores and arbitrary topologies of AS interconnection.
- Uses a path-vector concept which enables loop prevention in complex topologies
- In AS-level, shortest path may not be preferred for policy, security, cost reasons.
- Different routers have different preferences (policy) => as packet goes thru network it will encounter different policies
- Bellman-Ford/Dijkstra don’t work!
- BGP allows attributes for AS and paths which could include policies (policy-based routing).

BGP (Cont’d)
- When a BGP Speaker A advertises a prefix to its B that it has a path to IP prefix C, B can be certain that A is actively using that AS-path to reach that destination
- BGP uses TCP between 2 peers (reliability)
  - Exchange entire BGP table first (50K+ routes!)
  - Later exchanges only incremental updates
  - Application (BGP)-level keepalive messages
  - Hold-down timer (at least 3 sec) locally config
- Interior and exterior peers: need to exchange reachability information among interior peers before updating intra-AS forwarding table.
CIDR

- Shortage of class Bs => give out a set of class Cs instead of one class B address
  - Problem: every class C n/w needs a routing entry!
  - Solution: Classless Inter-domain Routing (CIDR).
    - Also called “supernetting”
    - Key: allocate addresses such that they can be summarized, i.e., contiguously.
    - Share same higher order bits (i.e., prefix)
    - Routing tables and protocols must be capable of carrying a subnet mask. Notation: 128.13.0/23

CIDR (Continued)

- Eg: allocate class Cs from 194.0.0.0 thru 195.255.255.255 for hosts in Europe (higher order 7 bits the same).
  - Allows one routing entry for Europe
  - Allow other routing entries too. Eg: 194.0.160 + mask of 255.255.240.0
  - When an IP address matches multiple entries (eg 194.0.22.1), choose the one which had the longest mask (“longest-prefix match”)

UPDATE message in BGP

- Primary message between two BGP speakers.
- Used to advertise/withdraw IP prefixes (NLRI)
- Path attributes field: unique to BGP
  - Apply to all prefixes specified in NLRI field
  - Optional vs Well-known; Transitive vs Non-transitive

Path Attributes: ORIGIN

- ORIGIN:
  - Describes how a prefix came to BGP at the origin AS
  - Prefixes are learned from a source and “injected” into BGP:
    - Directly connected interfaces, manually configured static routes, dynamic IGP or EGP
  - Values:
    - IGP (EGP): Prefix learnt from IGP (EGP)
    - INCOMPLETE: Static routes
Path Attributes: AS-PATH
- List of **ASs** thru which the prefix announcement has passed. AS on path adds ASN to AS-PATH
- Eg: 138.39.0.0/16 originates at AS1 and is advertised to AS3 via AS2.
- Eg: AS-SEQUENCE: “100 200”
- Used for loop detection and path selection

**AS1** (100) 138.39.0.0/16 **AS2** (200) **AS3** (15)

Path Attributes: NEXT-HOP
- Next-hop: node to which packets must be sent for the IP prefixes. May not be same as peer.
- **UPDATE for 180.20.0.0, NEXT-HOP= 170.10.20.3**

Attributes: MULTI-EXIT Discriminator
- Also called METRIC or MED Attribute
- AS1: multihomed customer. AS2 includes MED to AS1
- AS1 chooses which link (NEXTHOP) to use

Path Attribute: LOCAL-PREF
- Locally configured indication about which path is preferred to exit the AS in order to reach a certain network. Default value = 100.

I-BGP
- So far we have talked about E-BGP, i.e. interaction between R3 and R4
- How do R1, R2, R5 (termination points of internal default routes) learn of external routes?
- Need a way to internally distribute routes

I-BGP
- Why is IGP (OSPF, ISIS) not used?
  - In large ASs full route table is very large
  - Rate of change of routes is frequent
  - Tremendous amount of control traffic
- I-BGP:
  - Within an AS
  - Same protocol/state machines as EBGP
  - But different rules about advertising prefixes
  - Prefix learned from an I-BGP neighbor cannot be advertised to another I-BGP neighbor to avoid looping ⇒ need full IBGP mesh!
- **AS-PATH cannot be used internally. Why?**
IBGP vs EBGP
- I-BGP sessions between every pair of routers within an AS: full mesh.
- Independent of physical connectivity.

Physical link
IBGP session
A
B

D

IBGP

AS1

Other Attributes
- AGGREGATOR
  - If a BGP speaker aggregates on some of the prefixes heard from other neighbors, it may attach the AGGREGATOR attribute specifying the router which performed aggregation
- COMMUNITY STRING
  - The community attribute is a transitive, optional attribute in the range 0 to 4,294,967,295.
  - Way to group destinations (NLRI) or ASs and apply policy routing decisions (accept, prefer, redistribute, etc.) on them.

BGP Route Selection Process
Series of tie-breaker decisions...
- If NEXTHOP is inaccessible do not consider the route.
- Prefer largest LOCAL-PREF
- If same LOCAL-PREF prefer the shortest AS-PATH.
- If all paths are external prefer the lowest ORIGIN code (IGP<EGP<INCOMPLETE).
- If ORIGIN codes are the same prefer the lowest MED.
- If MED is same, prefer min-cost NEXTHOP
- If routes learned from EBGP or IBGP, prefer paths learnt from EBGP
- Final tie-break: Prefer the route with I-BGP ID (IP address)

IBGP Scaling: Route Reflection
- Add hierarchy to I-BGP
- Route reflector: A router whose BGP implementation supports the re-advertisement of routes between I-BGP neighbors
- Route reflector client: A router which depends on route reflector to re-advertise its routes to entire AS and learn routes from the route reflector

Route Reflection

AS Confederations
- Divide and conquer: Divides a large AS into sub-ASs
Summary

- Cores, peers, autonomous systems, EGP
- BGP avoids EGP-induced tree structure and allows policy-based routing, and scaling.
- BGP details: CIDR, Path Attributes, IBGP, scaling, route selection.