CS 4/55231
Internet Engineering

Kent State University
Dept. of Computer Science

LECT-8A

Inter AS Traffic Engineering

Game Plan: BGP Route Selection

- In the policy engine when there are multiple routes to a destination, the BGP decision process goes through the attributes in the following order. In each step it compares the attributes of both routes and when they are equal it moves down the list.
- All implementations follow the same order- to maintain routing sanity.

Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

INTERNET ENGINEERING

The Three Local "Knobs"

- *Preference* influences which BGP route will be chosen for each destination prefix. Changing preference is done by adding/deleting/modifying route attributes in BGP advertisements. Table 1 shows which attributes can be modified during import to control preference locally, and which can be modified during export to change how much a neighbor prefers the route.
- *Filtering* eliminates certain routes from consideration and also controls who they will be exported to. Filtering may be applied both before preference (inbound filtering) or after preference (outbound filtering). Filtering is done by instructing routers to ignore advertisements with attributes matching certain specified values or ranges.
- *Tagging* allows an operator to associate additional state with a route, which can be used to coordinate decisions made by a group of routers in an AS, or to share context across AS boundaries. The key mechanism is the *community Attribute.*

Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

INTERNET ENGINEERING

How it is Set?

- An ISP implements its policies by applying configuration commands at routers.
- These consist of a set of lists of preference, filtering, and tagging rules, one list for each *session* the router has with a neighboring BGP-speaking router.
- Although the configuration language differs between vendors, a key primitive that is often provided is a *route-map*, a language construct used to modify route attributes and define conditions that determine which routes are exported to peers.
- It consists of two parts: a set of conditions indicating when the map is to be invoked (e.g. the prefix is a specified value, or the AS path matches a specified regular expression), and the action to be taken if the advertisement matches the conditions (e.g. modify a specified attribute, or filter the route).

INTERNET ENGINEERING

Types of Policy

- *Business relationship*
  - policy arising from peering agreement, economic or political relationships an ISP has with its neighbor.
- *Traffic engineering*
  - policy arising from the need to control traffic flow within an ISP and across peering links to avoid congestion and provide good service quality
- *Scalability*
  - policy to reduce control traffic and avoid overloading routers.
- *Security*
  - policy to protect an ISP against malicious or accidental attacks.

INTERNET ENGINEERING

**Business Relationship: Peering Agreement

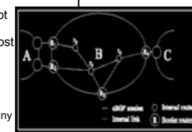
INTERNET ENGINEERING

- Three common relationships ISPs have are:
  - customer-provider, where one ISP pays another to forward its traffic,
  - peer-peer,where two ISPs agree that connecting directly to each other (typically without exchanging payment) would mutually benefit both, perhaps because roughly equal amounts of traffic flow between their networks,
  - backup relationships, where two ISPs set up a link between them that is to be used only in the event that the primary routes become unavailable due to failure.
- Means: ISPs often prefer customer-learned routes over routes learned from peers and providers when both are available. This is often done because sending traffic through customers generates revenue for the ISP while sending traffic through providers costs the ISP money and sending to peers can skew the balance of power in the peering relationship and thereby give incentive to the party receiving more traffic to tear down the relationship or start charging the other party.
- Solution:
  - Assign a non-overlapping range of LocalPref values to each type of peering relationship. For example use LocalPref values in the range 90-99 might be used for customers, 80-89 for peers, 70-79 for providers, and 60-69 for backup links.
  - LocalPref can then be varied within each range to do traffic engineering without violating the constraints associated with the business relationship,
  - A large ISP spanning both North America and Europe may wish to avoid forwarding traffic generated by its customers across an expensive transatlantic link. This can be done by configuring its European routers with a higher LocalPref for routes learned from European ISPs, and giving its North American routers a lower LocalPref for these routes.

**Business Relationship: Route Propagation Control

INTERNET ENGINEERING

- Issue: Routes learned from providers or peers are usually not exported to other providers or peers, because there is no economic incentive for an ISP to forward traffic it receives from one provider or peer to another.
- Quiz: B does not get paid for transmitting traffic from C to A. B want not to export routes learned from A to C. Note that the problem is that a route was received from a specific peer—this information is ordinarily lost in BGP as the route propagates across an AS.

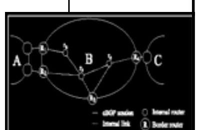


- Solution:
  - Step-1 For every session routers R1 and R2 have with routers in A, B configures an import policy that appends the community attribute X_{peer} to any route learned over these sessions
  - Step-2: After appending the community attribute, B exports the route onwards into its internal iBGP network.
  - Step-3: B configures export policies at R4 that match on this community attribute to determine which routes get exported to C. In particular, every session between R4 and a router in C is configured with an export policy that filters any route with the community attribute X_{peer}.

Traffic Engineering: Outbound Traffic Control

INTERNET ENGINEERING

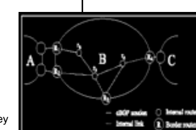
- Operators can influence outbound traffic flow either by configuring import policies that affect which routes get in the set of equally-good border routers, or by modifying IGP link costs.
- Quiz: Operator of ISP B want to send traffic to A via R2 rather than R4.
- One common goal is *early-exit routing* where the ISP forwards traffic to its closest possible exit point, so as to reduce the number of links packets traverse and hence the resulting congestion in its internal network.
- Note: Hot Potato Phenomena: Early-exit routing is known to inflate end-to-end path lengths in the Internet, ISPs often exercise early-exit routing to reduce their costs and network congestion, and because BGP does not support alternatives like determining global shortest paths across multiple ISPs.



**Traffic Engineering: Load Balancing

INTERNET ENGINEERING

- Quiz: suppose B wishes to shift some but not all traffic from its links to A to its link to C, perhaps because the link to A is over utilized or because it is planning to take the link down for maintenance.



- Solution:
  - Decreasing LocalPref for few routes traversing A or increasing LocalPref for routes traversing C.
- Caveat:
  - Achieving a specific level of load balance can be very difficult. The key challenge is to select the proper set of prefixes and change attributes for each appropriately; selecting too large a set will cause too much traffic to shift.
  - Since this is done manually it is subject to mis-configuration. Cannot be done in real time to adjust to changing load, and the outcome from a change can be difficult to predict. There are automated tools to predict the effects of these actions.

**Traffic Engineering: Inbound Traffic Control

INTERNET ENGINEERING

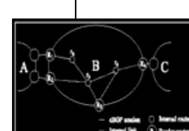
- Goal: An ISP's internal congestion may be exacerbated by its neighbors, because its neighbors might not be aware of the ISP's traffic engineering goals, internal topology, or load on internal links due to privacy reasons. How an ISP can enforce control how much traffic it receives from each of its peering links?
- Quiz: A customer of B is using R1 and ra for internal video conferencing. How it can tell A's routers to use R2 rather than R1?

- Solution:
  - Unfortunately, this is a highly challenging problem, as it requires the local ISP to influence route selection in remote ISPs, which in turn might wish to ignore the local ISP's goals.
  - First, an ISP must convince its neighbor (perhaps through economic incentives and contract) to allow the ISP to control how much traffic it receives on each link from the neighbor.
  - Then establish the real-time control by MED attribute, For the example case, increase the value of the MED attribute R1 advertises to A, causing the link to R2 to become more preferred by A's routers and thereby decreasing R1's load.

**Traffic Engineering: Traffic Shifting from Remote AS

INTERNET ENGINEERING

- Problem: How an ISP can sent path preference information to another ISP which is multi-hop away?
- Quiz: ISP D has peering with both ISP C and A but has no peering with B. Bulk of the traffic is actually generating at D and is sending it via A. Now B wishes to shift some transit traffic originating from D to its link to C. Note, just telling A or B about the preference has little impact. If the traffic is already in A it will use the path from A to B.

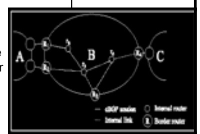


- Solution:
  - Unfortunately BGP was not designed with a mechanism to control route selection in ASes multiple hops away.
  - However, a workaround commonly used is for an AS to prepend multiple copies of its AS number to the AS path in order to artificially inflate the AS-path length. B can do this by prepending additional copies of its AS number onto the AS paths in BGP advertisements it sends to A. This increases the AS-path length in these advertisements, which causes routes advertised by C to other ISPs to become more desirable in comparison.

**Traffic Engineering: Remote Control

INTERNET ENGINEERING

- Problem: In certain cases, an ISP may want to give some remote control to a friendly ISP to let it manage its router's configuration.
- Quiz: Suppose C also directly peers with A (not shown in the figure). Suppose B suddenly wishes to have all inbound traffic from C for A to be routed not through it. If C has a LocalPref to prefer the direct route to B, no change in MED or AS prepending will force C to use alternate routes through A to B. B could request C to manually change its router configurations, but this can be time consuming for human operators if B changes its policy often (e.g. for traffic engineering purposes). How to let B control is automatically?



- Solution
  - C can allow B to control C's routing policy with respect to B's routes by configuring its routers to map certain community attributes to certain LocalPref values [2]. If desired, C can limit the degree of B's control to prevent certain policies of its own from being subverted. For example, C can configure its routers to map community value X1 to a LocalPref of 60, and X2 to a LocalPref of 75, allowing B to disable the route, but not allowing B to have it chosen over routes C wants to prefer more (by setting a higher LocalPref.

Scalability: Excessive Route Advertisement

INTERNET ENGINEERING

- Problem: Some mis-configurations and faults in neighboring ISPs can lead them to generate excessive rates of updates. Sending updates too frequently can trigger route instability, leading to poor service quality, or can overload a router's processing capability or memory capacity, which can cause outages and router failure. ISP B wants to protect itself.
- **Response:**
  - Step-1: Limit routing table size (by filtering and using the community attribute): Routing table overflow can cause the router to crash. This can be a particularly important issue for smaller ISPs which may have less expensive routers with less memory capacity.
  - Step-2: Protect from excessive advertisements from neighbors by: (a) Filtering long prefixes (e.g., longer than /24) to encourage use of aggregation. (b) As a safety check, routers often maintain a fixed per-session prefix limit that limits the number of prefixes a neighbor can advertise. (c) An ISP with a small number of routes may not need the entire routing table, and may instead configure a default route through which most destinations can be reached.

**Scalability: Flap Damping

INTERNET ENGINEERING

- Problem:
  - An ISP is receiving some routes which are frequently changing states. Routing instability is undesirable, as it can increase CPU load on routers. Also, frequent shifting of traffic to different paths can introduce jitter and packet loss in applications like Voice-over-IP and interfere with TCP's round-trip-time calculations.
- Response:
  - The key mechanism is *flap damping*. It limits propagation of unstable routes. It works by maintaining a penalty value associated with the route that is incremented whenever an update is received. When the penalty value surpasses a configurable threshold, the route is *suppressed* for some time, i.e., it is made unavailable to the decision process.
  - An ISP can lower the penalty threshold to improve route stability at the cost of worsening availability. ISPs may wish to less aggressively dampen or disable damping for certain prefixes, for example routes to the root Domain Name System servers, or routes from customers with high availability requirements.
  - Also, ISPs sometimes more aggressively dampen longer prefixes than shorter prefixes, with the motivation that damping a shorter prefix can have a large effect on reachability.

**Security: Malicious Advertisement

INTERNET ENGINEERING

- Attack
  - ISP's a vulnerable to false advertisement from other ISPs. An malicious ISP can deliberately inject faulty routes to degrade quality of service in rival ISP. Can come from Faulty BGP of neighbor as well.
- Response
  - Apply Import Filtering
  - Perform common sanity checks. Look for special-use private address, address blocks that have not yet been allocated are obviously invalid. Advertisements from customers for prefixes they do not own should not be propagated.
  - Look for problematic AS. A Tier-1 ISP should not accept any routes from its customers that contain another Tier-1 ISP in the AS path. Also, advertisements containing private AS numbers in the AS path may be considered invalid. Configure your filters based on the contents of public repositories.

Security: Reengineered BGP Advertisement

INTERNET ENGINEERING

- **Cheating with Routing Policies:**
  - An ISP may want to prevent its neighboring AS from violating their peering agreement. Otherwise, the ISP could be duped into carrying traffic a longer distance across its backbone on the neighbor's behalf.
- Example: An ISP peers with a neighbor in both New York and San Francisco. By advertising a prefix with a MED of 0 in New York and a MED of 1 in San Francisco, the peer could trick the ISP into having all of its routers direct traffic for this destination through the New York peering point, even if the San Francisco peering point is closer. The peer could achieve the same goal by configuring its San Francisco router to advertise the route with the nexthop attribute wrongly set to the IP address of the New York router.
- Response
  - To defend against violations of peering agreements, the ISP can configure the import policy to overwrite some attributes with the expected values. For example, the import policy could set all MED values to 0, unless the ISP has agreed in advance to honor the neighbor's MEDs. Similarly, the import policy could set the next-hop attribute to the IP address of the remote end of the BGP session, and remove any unexpected community values.

Security: Hide Critical Infrastructure

INTERNET ENGINEERING

- Threat:
  - An ISP should prevent external entities from accessing certain critical internal resources.
- Measure:
  - Determine the IP addresses of critical assets and apply export filtering and include export filters not to divulge them. ISP should protect its own backbone infrastructure by not divulging the IP addresses used to number the router interfaces. The ISP should protect hosts running network-management software.

BGP Denial-of-Service Attack (1)

- **Attack:**
  - Denial-of-service attacks can degrade service by overloading the routers with extra BGP update messages or consuming excessive amounts of link bandwidth. For example, the ISP's routers could run out of memory if a neighbor sends route advertisements for a large number of destination prefixes.
- ISP Response
  - The ISP can configure each BGP session with a maximum acceptable number of prefixes, tearing down the session when the limit is exceeded.
  - The import policy could filter prefixes with large mask lengths (e.g., longer than /24). As another example, a neighbor sending an excessive number of BGP update messages can easily deplete the CPU resources on the ISP's routers. Upon detecting the excessive BGP updates, the operators could modify the import policy to discard advertisements for the offending prefixes or disable the BGP session.
  - Upon identifying the neighbor or prefix responsible for the excessive BGP updates, the ISP can more aggressively dampen (Section 5) or even completely filter updates it receives from these sources.

INTERNET ENGINEERING

Denial-of-Service Attack (2)

- **Denial-of-service or Spam attack on ISP or Its Customer**
  - An ISP (or its customers) may be subject to a denial-of-service attack where excessive data traffic is sent to victim hosts.
- Response:
  - An ISP can block the offending traffic by installing a *blackhole route* that drops traffic destined to the victim addresses. Blackhole routes may be statically configured, or operators may run a special BGP session that advertises the prefixes of the victims. Routers receiving prefixes on this session then assign the next-hop to be an address associated with the "null" route or to a monitoring system that can perform further analysis of the traffic.
  - Using a similar technique, the ISP can advertise the address blocks of known spammers to blackhole traffic sent to these addresses. These blackhole routes prevent the spammers from establishing bidirectional communication (i.e., a TCP connection, which depends on receiving a SYNACK packet) with the ISP's mail servers.

INTERNET ENGINEERING

Next Topic:

AS Peering Infrastructure