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.

<table>
<thead>
<tr>
<th>Step</th>
<th>Attribute</th>
<th>Controlled by local or neighbor AS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Highest LocalPref</td>
<td>Local</td>
</tr>
<tr>
<td>2</td>
<td>Lower AS path length</td>
<td>Neighbor</td>
</tr>
<tr>
<td>3</td>
<td>Lower origin AS</td>
<td>Neighbor</td>
</tr>
<tr>
<td>4</td>
<td>Lower MED</td>
<td>Neighbor</td>
</tr>
<tr>
<td>5</td>
<td>Other imported over BGP-learned</td>
<td>Neighbor</td>
</tr>
<tr>
<td>6</td>
<td>Local best to foreign router</td>
<td>Local</td>
</tr>
<tr>
<td>7</td>
<td>Local best to best router</td>
<td>Neighbor</td>
</tr>
</tbody>
</table>

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.

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).

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.
**Business Relationship: Peering Agreement**

- Three common relationships ISPs have are:
  - customer-provider, where one ISP pays another to forward its traffic,
  - peer-peer, where two ISPs agree to connect directly to each other (typically without
    charging), and
  - customer-peer, where one ISP pays another to forward its traffic.

- Means: ISPs often prefer customer-learned routes over routes learned from peers and
  thereby give incentives to the party receiving more traffic to tear down the relationship
  or start charging the other party.

- Nature: ISPs often prelearn learned routes over routes learned from peers and
  incentivize the party receiving more traffic to tear down the relationship
  or start charging the other party.

- 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.

- Quiz: Operator of ISP B wants to send traffic to A via R2 rather than R4.

- Note: Hot Potato Phenomena: Early-exit routing is known to infuse
  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: Outbound Traffic Control**

- 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 BGP path costs.

- Quiz: Operator of ISP B wants 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.

- Quiz: A customer of B is using R1 and R4 for internal video
  conferencing. How it can tell A’s routers to use R2 rather than R1?

**Traffic Engineering: Load Balancing**

- 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.

**Traffic Engineering: Inbound Traffic Control**

- 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 over
  how much traffic it receives from each of its peering links?

- Quiz: A customer of B is using R1 and R4 for internal video
  conferencing. How it can tell A’s routers to use R2 rather than R1?

**Traffic Engineering: Routing Shifting from Remote AS**

- Problem: How an ISP can send path preference information to
  another ISP which is multipath away?

- Quiz: ISP D has peering with both ISP C and A but has no peering with
  B. Bulk of the traffic is actually generated at D and is sending it
  via A. Now D 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 for an AS to control
  multipath is 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
  and thereby causing BGP routes to be more attractive in comparison.
**Traffic Engineering: Remote Control**

- **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 B. 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. C could request B to manually change its router configurations, which B can do if the route is policy safe (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). Either C or B can limit the degree of B's control to prevent certain policies of its own from being overwritten. 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 preventing B from changing other routes. C wants to prevent more (by setting a higher LocalPref).

**Scalability: Excessive Route Advertisement**

- **Problem:** Some mis-configurations and faults in neighboring ISPs can overload a router's processing capability or memory capacity, which can overload a router's processing capability or memory capacity, which can cause routes or router failures. An ISP wants to prevent its neighboring AS from violating their peering agreement.

- **Measure:** 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.

**Security: Reengineered BGP Advertisement**

- **Creating 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 routes 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 next hop attribute wrong 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 examine some attributes with the expectation that, for example, the import prefix policy in New York could set the next hop attribute to the neighbor's IP address, or that in San Francisco, it could set the next hop attribute to the neighbor's IP address plus one.

**Security: Malicious Advertisement**

- **Attack:** ISPs may be vulnerable to false advertisement from other ISPs. An malicious ISP can deliberately inject faulty routes to degrade quality of service or its neighbor. An malicious ISP can influence route selection to favor its own AS over another, such as routes to certain countries or regions.

- **Response:** Apply Import Filtering.

- **Step-1:** Filter incoming 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.

- **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.

- **Step-3:** 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**

- **Problem:** 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.

- **Response:** The key mechanism is flap damping. It limits propagation of unstable routes. It works by assigning 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.

- **Step-1:** Filter incoming 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.

- **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.

- **Step-3:** 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.

**Security: Hide Critical Infrastructure**

- **Threats:** An ISP should protect its critical infrastructure from detection and misuse by preventing attacks and building defenses.

- **Response:** An ISP should implement a robust security system, including firewalls, intrusion detection systems, and regular vulnerability assessments. An ISP should ensure that all hosts and devices are protected by up-to-date and appropriate security measures.

- **Measure:** An ISP should perform periodic security audits and vulnerability assessments. An ISP should implement a secure configuration management process to prevent unauthorized access to critical infrastructure.
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.

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.

Next Topic: AS Peering Infrastructure