ETH zürich

Networked Systems
ETH Zürich — seit 2015

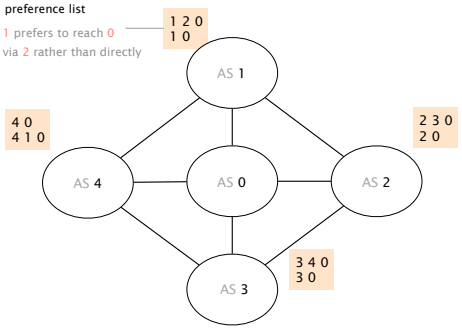
Communication Networks

Prof. Laurent Vanbever

Solution: Exercise 9 – BGP Challenges

BGP Challenges

9.1 Convergence



Does this network ever converge?

Consider this BGP network composed of 5 ASes. Assume that each AS has configured its BGP policies in a way that leads to the preference lists shown in the figure. For example, AS 1 is configured to only accept an announcement for AS 0 if it has path [1,2,0] or [1,0]. In addition, AS 1 prefers the path [1,2,0] over the path [1,0].

Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

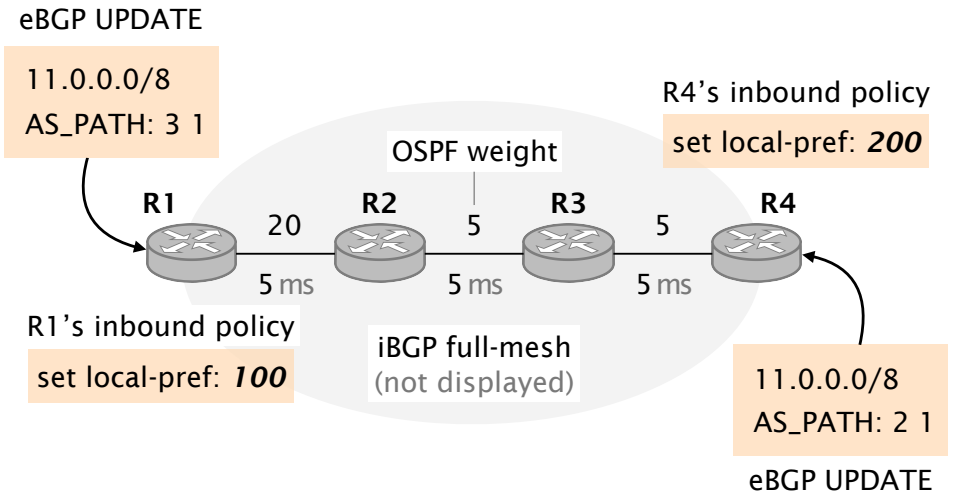
b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution: This BGP network does have a unique, stable solution in which:

- AS 1 selects [1,2,0] (preferred path);
- AS 2 selects [2, 0];
- AS 3 selects [3, 4, 0] (preferred path);
- AS 4 selects [4, 0] (preferred path).

9.2 Left? Right? Both? (Exam Question 2017)

Consider the BGP network composed of 4 routers depicted in Figure below. Two of these routers, $R1$ and $R4$ are egress routers and maintain eBGP sessions with external neighbors. $R1$ is configured to associate a local-preference of 100 to externally-learned routes, while $R4$ is configured to associate a local-preference of 200 to externally-learned routes. $R2$ and $R3$ are internal routers. All four routers are connected in an iBGP full-mesh. OSPF is used as intra-domain routing protocol. The link weights are indicated in the figure, e.g. the $(R1, R2)$ link is configured with a weight of 20. The Figure also indicates the propagation delay for each link (e.g., it takes 5ms for a packet to propagate between $R1$ and $R2$).



A simple BGP network learning external routes via eBGP on $R1$ and $R4$.

a) Considering the above configuration, indicate the next-hop used by each router in the steady state, *i.e.,* once the network has fully converged. Use the keyword "external" to indicate that an edge router is forwarding outside of the domain. Note that we are not looking for the *BGP* next-hop but rather the next-hop a packet would take when being forwarded.

Solution: Since the externally-learned route at $R4$ has a higher local-preference than the one at $R1$ (200 vs. 100), all routers select the route from $R4$. We get the following next-hops:

- R1: $R2$
- R2: $R3$
- R3: $R4$
- R4: <external>

b) One of the network operator decides to lower the local-preference associated by $R4$ to externally-learned routes to 50 (instead of the original 200). Indicate the sequence of BGP messages sent which is triggered following that change along with the timestamps at which they are generated. You can consider that the BGP process on each router is infinitely fast meaning only propagation delay matters. Only indicate when messages are sent, not when messages are received.

Solution: Before the change R1 has two routes for 11/8 available:

(i) 11/8: [2, 1] - LP 200 - R4
(ii) 11/8: [3, 1] - LP 100 - R1

R1 selects the route from R4 as its best route and therefore does not propagate the externally-learned route with lower local-preference.

All other routers have one route available: 11/8: [2, 1] - LP 200 - R4

At first, none of the routers will change their best route, when the local-preference of the externally-learned route at R4 is reduced to 50, as they just have that one route available. However, when R1 learns about the local-preference change, it will select its own, externally-learned route and advertise that route to all the other routers in the network.

Advertisements:

- Timestamp [0 ms] R4 sends the message 11/8 - [2, 1] - LP 50 to R1, R2, R3
- Timestamp [15 ms] R1 sends the message 11/8 - [3, 1] - LP 100 to R2, R3, R4
- Timestamp [30 ms] R4 sends the message 11/8 - withdraw to R1, R2, R3

c) Was a forwarding loop induced due to the configuration change? Briefly explain why or why not. If a loop was created, also indicate its duration (in ms).

Solution: No, there was no forwarding loop. The route change happens from left (R1) to right (R4). This is due to the fact that R1 also uses the path through R4 and therefore does not advertise its alternative route until it becomes the best route.

d) It turns out that the network operator changed her mind. This time, she configures $R4$ to associate a local-preference of 100 to externally-learned routes (i.e. the same local-preference value as on $R1$). Indicate the next-hop used by each router in the steady state (once the network has fully converged). Again use the keyword “external” to indicate that an egress router is forwarding outside of the domain.

Solution: Since both externally-learned routes are now equally preferred, the routers consider the next criteria in the decision process. Finally, they will select the route with the lower IGP metric to the BGP next-hop. We get the following next-hops:

- R1: <external>
- R2: R3
- R3: R4
- R4: <external>

e) Soon after the network has fully converged due to the configuration change of $R4$, a failure happens disconnecting $R4$ from all its external neighbors. The connection between $R4$ and $R3$ is still working fine though. Indicate the sequence of BGP messages sent following that failure along with the timestamps at which they are generated. Only indicate when messages are sent, not when messages are received.

Solution: Before the failure all routers have two routes for 11/8 available:

(i) 11/8: [2, 1] - LP 100 - R4
(ii) 11/8: [3, 1] - LP 100 - R1

R4 detects the failure and switches to the route from R1 while withdrawing its own route.

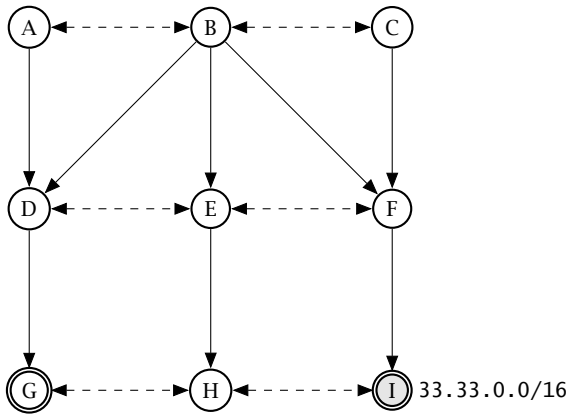
Advertisements:

- Timestamp [0 ms] R4 sends the message 11/8 - [2, 1] - withdraw to R1, R2, R3

f) Was a forwarding loop induced due to the failure? Briefly explain why or why not. If a loop was created, also indicate its duration (in ms).

Solution: Yes, there exist forwarding loops for 10 ms in total. The routers receive the withdrawal with a 5 ms time difference from right (R4) to left (R1). This leads to a forwarding loop of 5 ms for the two pairs of routers (R3 and R4, R2 and R3). In both cases, the router closer to R4 changes first its best path, while the router further away is still using the withdrawn path.

9.3 BGP Security (Exam Question 2020)



An Internet topology of 9 ASes in which AS I announces a prefix and AS G tries to hijack it.

Consider the Internet topology consisting of 9 Autonomous Systems (ASes) in the Figure above. Single-headed plain arrows point from providers to their customers (AS A is the provider of AS D) while double-headed dashed arrows connect peers (AS A and AS B are peers). Each AS is made up of a single BGP router and applies the default selection and exportation BGP policies based on their customers, peers and providers.

In this task, the routers break ties using the AS number of the neighbor: in case multiple routes are equally good, the router selects the route of the neighbor with the lowest AS number (in alphabetical order; e.g., a route from AS A is preferred over AS B in case of a tie).

AS I is the origin of prefix `33.33.0.0/16` and advertises it to its neighbors. Independently of what the external advertisements are, AS I ***always*** prefers its internal route to reach any IP destination in `33.33.0.0/16`.

a) AS G wants to hijack the traffic going to AS I for `33.33.0.0/16`. It starts advertising the exact same prefix with itself, AS G, as origin. From which ASes is it able to hijack the traffic?

Solution: We can hijack traffic from the following ASes: A, B, D, E, H.

b) The ASes notice the hijack and, as a counter-measure, deploy Resource Public Key Infrastructure (RPKI) Internet-wide. After that, from which ASes is the attacker able to hijack the traffic by still advertising the exact same prefix with itself as origin?

Solution: AS G is not able to attract any traffic anymore as all the ASes can tell that the route is invalid.

c) RPKI has a flaw. What is the problem of RPKI? How can AS G hijack the prefix `33.33.0.0/16` despite RPKI? From which ASes is AS G able to hijack the traffic?

Solution: RPKI only checks whether the origin AS indeed owns the advertised prefix, but it has no way of validating that the announcement actually originated at the correct AS and has not just been added to the beginning of the AS path.

AS G can simply add a fake origin to its announcements and advertise `33.33.0.0/16` with an AS path of G, I.

By doing so, it can only attract the traffic from ASes A, D. Since AS G prepends the true origin to the AS path, the AS path length of its announcements increases by one. Hence, some ASes prefer the true announcement over the hijack because of that (e.g., AS B).

d) In response, the ASes switch to BGPSec (Secure BGP). Explain what security it provides and how AS E can detect that the announcement from AS G has a forged AS path.

Solution: BGPSec provides origin and path authentication through cryptographic signatures. It allows to verify the entire path.

In BGPSec, every AS signs the message with its private key and includes the AS number of the next AS on the path. When a router receives an announcement, it can check each signature on the path using the public keys of the ASes. It then quickly sees whether the path is valid or not. In this case, AS E could detect that the origin (i.e., the very first hop on the AS path) is not correctly signed.