TTM4180 - Border Gateway Protocol

Sofía Xiaofan Fernández Marín

March 29, 2022

1. Introduction

The border gateway protocol (BGP) is the method for routing among autonomous systems (ASs). This means that it is an inter-AS routing protocol. This method provides prefix reachability information from neighbouring ASs and determines the best route based on policy and reachability information. (1)

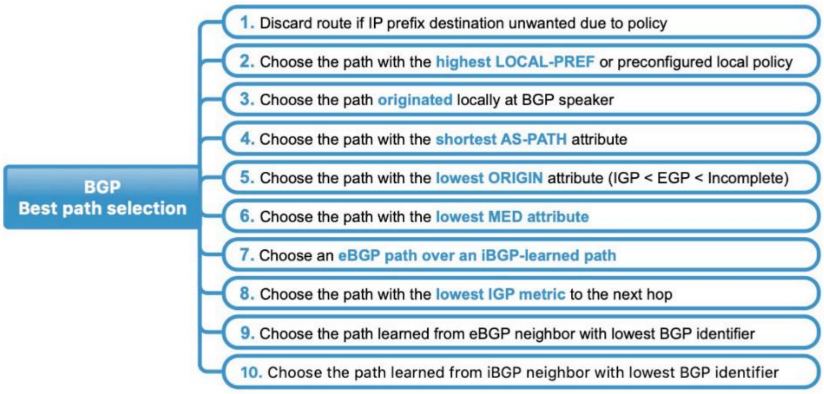
The ASs communicate with each other's through the exterior BGP (eBGP), implementing routing policy and exchanging prefixes between other ASs. These connections use the gateway routers, routers that are on the edge of the AS, making sure that all these routers have the same BGP table. The communications within nodes of the same AS are made through the interior BGP (iBGP), used for choosing which gateway router reaches an address (1, 2). The connections between them are made using the internal and external routers, as well as route reflectors, used for reducing the number of connections of an AS and to avoid full mesh. Another way of avoiding the latter is to create sets of ASs (confederations) for breaking AS into sub-ASs (2).

In this essay, we will discuss how BGP works further. We will also discuss some security incidents along with the history, where people learnt what not to do, and for them not to repeat.

2. Discussion

BGP is the way of routing all the ASs, this is why the destination of BGP is not a host but an IP prefix, it is found with the help of the CIDR, and the IP addresses are assigned by the ICANN (Internet Corporation for Assigned Names and Numbers) (3).

The BGP default route selection chooses the shortest path using a path-vector routing, based on a distance vector-type, a decentralized algorithm, as the information of the network is gradually calculated (3). But BGP also considers an attribute list, describing the path in various ways and letting the administrators implement various routing policies (3). Therefore, this protocol is called policy-based routing, as it might be an AS that wants to control what type of traffic transits in their AS or some traffic originating information, that does not want to go through a specific AS (4). Hence, the first path will be chosen by the policy established by the administrator. The selection of the best BGP route is determined by the following points (4).



1. Discard route if IP prefix destination unwanted due to policy
2. Choose the path with the highest LOCAL-PREF or preconfigured local policy
3. Choose the path originated locally at BGP speaker
4. Choose the path with the shortest AS-PATH attribute
5. Choose the path with the lowest ORIGIN attribute (IGP < EGP < Incomplete)
6. Choose the path with the lowest MED attribute
7. Choose an eBGP path over an iBGP-learned path
8. Choose the path with the lowest IGP metric to the next hop
9. Choose the path learned from eBGP neighbor with lowest BGP identifier
10. Choose the path learned from iBGP neighbor with lowest BGP identifier

Figure 1: BGP Best path selection

As an example, if we had *two identical routes* but with the only difference of the origin, the first one EGP and the second one IGP, the chosen one would be the one with the IGP origin, as it is lower than EGP. The same would have happened if the only distinction were the attribute AS_PATH (first=AS_PATH(1 7 3), second=AS_PATH(6)). The chosen one will be the second, the one with the shortest path.

The before mentioned "AS_PATH" belongs to the object attributes, information about an advertised prefix that helps BGP advertise routes. As prefixes, we are referring to the most significant bits of a subnet IP address (3). Usually, a single subnet has associated adjacent address blocks, an address range with a common prefix that allows to reduce the router forwarding table size, differentiate internal devices of an organization or create new subnets using the same prefix (5).

The attributes accompany each prefix of a BGP update and allow the ASs to implement their routing policies and they can be classified into two categories: *well-known*, where all the attributes of this category should be recognized by all BGP. They can be split into another two sub-categories [1.1] *mandatory*, every route update has to have them, and [1.2] *discretionary*, which can be sent or not. The other category is *optional*, where the attributes do not need to be supported by all BGP, which also can be spitted into two sub-categories [2.1] *transitive*, which must be forward to the BGP peers, and [2.2] *non-transitive*, only significant to the receiving AS. (3)

The most common attributes used for inter-domain routing systems are (3, 6):

- ORIGIN: Belongs to the 1.1 category. It specifies how the IP prefix has been obtained, from an iBGP [i], eBGP [e] or unknown [?].
- AS_PATH: Also, the 1.1 category . Describes the BGP path vector, and the route visited.
- NEXT_HOP: IP address of the next-hop router, used by the border router. Category 1.1 too.
- MULTI_EXIT_DISC: Belongs to the 2.2 category. Multiple Exit Discriminator (MED) is used by the source AS to tell the target AS the preference of a traffic output.
- LOCAL_PREF: Belongs to the 1.2 category. Inform other routers in the same AS of the preference of an advertised route.
- ATOMIC_AGGR: Informs other BGP routers that the local system selected a less specific route without selecting a more specific route that is included in it. From category 1.2.
- AGGREGATOR: Category 2.1. Tells the last AS number that formed the aggregated route and the IP address of the BGP router within the AS.

As seen before, these attributes are used for choosing the best path selection, but they also are useful when it is needed to do an BGP UPDATE, advertising changes in routes (2). These updates belong to the messages BGP interchanges after two nodes have done a connection. BGP exchanges messages using TCP, these messages have a common header where it is indicated which type of message it is. The existing messages are *open*, starts a BGP session; *update*, exchange reachability information and advertise or withdraw changes in routes; *keepalive,* confirm availability if there are no updates; *notification,* report error and can terminate connections or *route-refresh*, that request updates. (2)

Each update message is able to advertise changes, as when a connection is established, this message type is sent and it is composed of a *withdrawal route path* (addresses no longer reachable), *path attribute* (list of attributes BGP can interpret) and *network layer reachability information* (updated view of attributes of the "best" route object). With this method, it is reduced the number of messages over the network, as each update message forms a unit and has no fragmentation because in each message there is a whole route. (3)

Another scenario where attributes are used is to implement policies. The ISP (internet service provider) uses the attributes for filtering routes it does not want in a certain AS, avoiding neighbours or changing the value of the attributes for influencing the path selection seen in Figure 1. This will be implemented in the router via configuration commands from the ISP. (7)

The attributes can also be used when there is the need of controlling the next hope because of business agreements or relationships via assigning LocalPrefs values or by controlling the route export. The common relationships are *customer-provider*, *peer-to-peer,* and *backup.* (8)

But using attributes can work against, since they can be tampered for indicating a false origin, path, or an AS origin IP prefix (3, 5). Another vulnerability of BGP is that it has no way of determining the authenticity of the source of the BGP messages nor the if an AS can announce an IP prefix (or if that prefix is valid). (3)

This policy attacks can be done in various ways, the *prefix hijack* is one of the most common, where one AS impersonates another, by taking the IP prefix of the AS1 and announcing that it also has that IP. In that way, the non-legitimate AS also gets the messages that are supposed to be just for the AS1 (5, 9). This is why more than 20% of the global prefixes are not public, as if someone could get them, they could create a prefix hijacking (9). Associated with this problem, there exist the *route leaks*, where an AS propagates routes to some others ASs that are not supposed to know those IP prefixes, usually due to misconfigurations of the policies (5). For instance, on the 26th of August 2017, Google leaked BGP prefixes providers' routes, so it became a transit AS instead of just exchanging traffic between two networks (10).

It caused a large-scale internet disruption mainly in Japan, slowing or blocking access to websites and online services as well as making prefixes getting de-aggregated as they also exposed some internal traffic engineering (10). As an example,[1] the AS45629 (Jastel out of Thailand) announced 171.5.0.0/17, which became reachable with Google [15169], as Jastel [45629] peers with it. The next AS in the path was Verizon [707] which started providing transit for Jastel via Google. After that, Verizon announced that advertisement to some of their customers. (10)
It lasted 10 minutes but around 135.000 prefixes were leaked via Google in Verizon path causing many networks to reroute traffic (10, 11). Some services that noted the incident were Merukari, Mobile Suica or Rakuten Securities (12).
After this event, Verizon improved their security and subsequent to this incident, now we take more into account the Mutually Agreed Norms for Routing Security (MANRS) for better protection against traffic anomalies, for, at least, communicating to the peers what announcements they should expect (11).

Another possible attack is a *Denial-of-service attack (DoS)*, where the attacker causes heavy congestion on routers or links used by BGP messages. When BGP sessions are recovered, there are delays as they need to exchange dill routing tables. The attackers can also create withdraws and re-advertisement of target routes of the AS for causing link flapping (9). For avoiding this, it is often used *blackholing*. This is a way of mitigating the attacks, dropping all the traffic towards the *blackholed* prefix, and rewriting the next hop to null. In this way, the DoS attacks are redirected to this prefix and dropped from the network, stopping the flow of unwanted data. (13)
But for securing BGP, instead of having to mitigate attacks, there are other approaches, explained further (9). *Secure BGP (S-BGP)*, giving a public key infrastructure (PKI) so the BGP speakers and the IPs prefixes of the ASs are certified; *Secure origin BGP (soBGP)*, similar to S-BGP but without hierarchical PKI, achieved by having three different certifications, [1] EntityCert: a public key associated with an AS number and private key with an AS, [2] AuthCert: associates IP prefix with an AS, [3] ASPolicyCert: feasibility testing; *Resource Public key infrastructure (RPKI)*, where it verifies that an AS is authorized to announce a specific IP prefix for preventing IP prefix hijacking and mi-origination (9). RPKI uses *trust anchors (TAs)*, independent regional internet registries (RIRs) that maintain the infrastructure and offer RPKI as a service; *Route Origination Authorizations (ROA)*, for authorizing ASNs to originate IP prefixes; and *validators*, where the speaker obtains the ROA information (9, 14).

3. Conclusion

As a summary of the essay, we have seen how the border gateway protocol works and its role on the global internet. We have also discussed why BGP is called policy-based routing, the path selection of BGP and the attributes this protocol uses and how. We have also seen how to secure BGP and the challenges it has faced as well as an incident and what we have learned about that and how to avoid such a problem in the future.

[1] Example using path [286 701 15169 45629] and reading from right to left.

Reference List

1. Kurose J.F. and Ross K.W. Chapter 5.4 - Routing Among the ISPs: BGP. In: Horton M., editor. Computer Networking: A Top-Down Approach. 7th ed. Hoboken, New Jersey: Pearson 2017. p. 395-406.
2. Huston G. An Introduction to BGP – the Protocol Column [Internet]. Australia: The ISP Column; 2006 May [cited 2022 Apr 8].
   Available from: https://www.potaroo.net/ispcol/2006-05/bgp.html
3. Medhi D. and Ramasamy K. Chapter 9 - BGP. In: Kaufmann M., editor. Network Routing – Algorithm, Protocols and Architectures. 2nd ed. Boston: Elsevier 2018. p. 286-331.
4. Moldeklev K. Lecture 07, slide 35, March 2022. Figure: The default BGP route selection process is to prefer a path with the longest prefix match, then using attributes to choose route; p. 35.
5. Medhi D. and Ramasamy K. Chapter 10 - Routing in the global internet. In: Kaufmann M., editor. Network Routing – Algorithm, Protocols and Architectures. 2nd ed. Boston: Elsevier 2018. p. 334-375.
6. Goralski W. Chapter 16 - Border Gateway Protocol. In: Kaufmann M., editor. The Illustrated Network. 2nd ed. Boston: Elsevier 2017. p. 409-430.
7. Feamster, N., and Balakrishnan, H. (2005, May). Detecting BGP configuration faults with static analysis. In Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2 (pp. 43-56).
8. Caesar M and Rexford J. BGP routing policies in ISP networks. IEEE network. 2005; 19(6):5-11.
9. Mitseva A, Panchenko A and Engel T. The state of affairs in BGP security: A survey of attacks and defenses. Computer Communication. 2018; 124:45-60.
10. CISCO. BGP leak causing Internet outages in Japan and beyond [Internet]. California: Cisco Systems; 2017 Aug 26 [cited 2022 Apr 12].
    Available from: https://bgpmon.net/bgp-leak-causing-internet-outages-in-japan-and-beyond/
11. Internet Society. Google leaked prefixes – and knocked Japan off the Internet [Internet]. USA: Internet Society. 2017 Aug 28 [cited 2022 Apr 19].
    Available from: https://www.internetsociety.org/blog/2017/08/google-leaked-prefixes-knocked-japan-off-internet/
12. Morris Cornell-Morgan. Friday's Widespread Internet Outage in Japan [Internet]. 2017 Aug 26 [cited 2022 Apr 19].
    Available from: https://morimori.tokyo/2017/08/fridays-widespread-internet-outage-in-japan/
13. Giotsas V, Smaragdakis G, Dietzel C, Richter P, Feldmann A, Berger A. Inferring BGP Blackholing Activity in the Internet. In: Proceedings of the 2017 Internet Measurement Conference, editor. London, UK. 2017. p. 1-14.
14. Kristoff J, Bush R, Kanich C, Michaelson G, Phokeer A, Schmidt TC, Wählisch M. On measuring RPKI relying parties. In: Proceedings of the ACM Internet Measurement, editor. Virtual event, USA. 2020. p. 184-491.