Border Gateway Protocol

BGP

Ibtihaj Alanazi
Logan Bair
Amirshahram Hematian

Kevin Kuo
Mary Snyder

Abstract— **Border Gateway Protocol (BGP) is a routing protocol that shares reachability and routing information among gateway hosts of autonomous systems. BGP can be classified as a path vector routing protocol. BGP's routing decisions are made utilizing network path information, network policies, and/or administrator set conditions. Routing tables for BGP contain critical information such as a list of reachable routers, the addresses those routers can reach, and a path cost matrix for each router to facilitate choosing the best route to a destination. BGP utilizes Transmission Control Protocol (TCP) to communicate between hosts and only sends updated routing information when a change has been detected. Even then, BGP only sends the part of the routing table that has changed. The most current version of BGP, BGP-4 or BGP4, allows network administrators to configure the path according to organizational policy statements. The AS-Path attribute is an anti-loop mechanism critical to BGP. Any routes that contain the router itself in the path will not be imported.**

I. Introduction

Border Gateway Protocol (BGP) is a routing protocol that is widely used to route traffic across the internet and the most scalable of all routing protocols. BGP exchanges routing information between gateway hosts in a network of autonomous systems (ASs). Routing decisions in BGP are determined using the Best Path Selection, which utilizes known routes/paths and network/routing policies controlled by a common control such as a network administrator.

BGP is not applicable for a small corporate network or networks where the goal is just to provide internet to the simple users (i.e. home networks). Those networks would be fine employing an Interior Gateway Protocol (IGP) such as Routing Information Protocol (RIP) or Open Shortest Path First (OSPF) to exchange routing information, which includes routing tables, in a periodic manner. However, BGP is applicable and often used by network administrators of large organizations who connect two or more Internet Service Providers (ISPs) and/or connect to other network providers. When BGP is used in a network with more than one AS, the protocol is referred to as External BGP (EBGP). If BGP is used inside a single AS, it is referred to as Interior BGP (IBGP).

In this report, we discuss BGP effects on applications, advantages, and disadvantages of BGP over alternatives routing protocols, operation of and reasons behind BGP implementation, BGP standards and the software or hardware requirements for BGP, and definitions of frequently used terms in regards to BGP.

A. Popularity and Use of BGP in Information Technology Applications

BGP is popular among network administrators of networks containing multiple ISPs or single IPSs with a need to connect to other network providers. ISPs use BGP to create routing between different hosts. Many large private Internet Protocol (IP) networks use BGP internally. BGP can be used to join a number of large OSPF networks. BGP is also important in multi homing networks, which creates a connection among various hosts to establish a reliable connection over multiple computer networks. It creates reliable connections that give better redundancy instead of creating multiple access points for single and multiple ISPs.

B. Advantages and Disadvantages of BGP

BGP has several advantages over other protocols including:

- BGP only has one active route for a prefix at a time; however, the IGP may use multiple paths to get to the next-hop.
- BGP ensures routing data is sent to a machine nearer to the end point than itself and the decision process between multiple routes does not cause loops.
- BGP will block sources of bad traffic including spammers, proxies, and/or TCP scanners.
- Provide IP stability, not physically bound to a location/machine

However, there are some disadvantages to BGP as well, including:

- BGP has been found to be vulnerable to attacks and misconfigurations.
- Congestion control/monitoring is not performed with BGP.
- Inability to load balance traffic across multiple connections during high load.

II. Reasons for implementation

The primary function of a BGP speaker is to exchange network reachability information with other BGP systems. This network reachability information includes information on all routes traversed by reachability information within the ASs. This wealth of routing information can be used to construct an AS connectivity graph, which provides a means to identify and remove routing loops while enforcing AS level policy decisions.

A. Seqence of Steps for BGP - Best Path Selection Algorithm

BGP will assign the first valid path from the list of valid paths as the current best path. BGP will compare this current best path with the next path in the list, until there are no more valid paths in the list. The following are the rules used to select the best path [2]:

1. Prefer the route with highest WEIGHT [Cisco-specific and local to the router]
2. Prefer the route with highest LOCAL_PERF [Local within an AS, default is 100]
3. Prefer locally originated routes. Self-sourced routes are preferred over aggregate-address sourced routes
4. Prefer the route with the shortest AS_PATH
5. Prefer the route with lowest origin code (i.e. IGP preferred over EGP, which is preferred over an incomplete route)
6. Prefer the route with the lowest multi-exit discriminator (MED)
7. Prefer EBGP learnt routes over IBGP learnt routes. If best route is selected, skip to step 9
8. Prefer the route with the smallest IGP metric to the BGP next-hop
9. Determine if multiple routes require installation into the routing table for BGP Multipath. If best route is not selected, continue to step 10
10. When both routes are external, prefer the oldest route (received first)
11. Prefer the route from the BGP router with the lowest router ID
12. When both router IDs are the same for multiple routes, prefer the route with the smallest cluster list
13. Prefer the route with the lowest neighbor address

B. Advantages of Best Path Selection Algorithm

BGP avoids the pit fall of distance-vector protocols, by utilizing the AS-PATH attribute for each advertised route. This allows BGP to efficiently and quickly not only detect, but also avoid any routes that may contain loops. The best path selection process can also be customized through use of the BGP Cost Community attribute. This is an additional step after the required steps in the algorithm that allows the cost communities of the routes to be compared to determine the degree to which a particular path may be preferred.

III. Implementation Of Standard

In this section we discuss [1] Message formats, Route updates, Path attributes, Error handling, Version negotiation, Timers and Security in Border Gateway Protocol (BGP).

A. Message Format

The maximum size of a BGP message is 4096 bytes. All BGP messages are transmitted over TCP connections. Message processing will start whenever all bytes of a BGP message are received. The smallest message that may be transmitted consists of a BGP header without any data portion (19 bytes). All implementations of BGP are required to support the maximum message size. All multi-byte fields are in big-endian byte order.

Each message has a fixed-size header. There may or may not be a data portion following the header, depending on the message type. The header contains Marker, Length, and Type fields that are 16, 2, and 1 bytes long, respectively. The Marker field is included for compatibility that should be set to all ones. The Length field is an unsigned integer that indicates the total length of the message, including the header in bytes. Thus, it allows one to locate the (Marker field of the) next message in the TCP stream. The value of the Length field must always be at least 19 and no greater than 4096 bytes, and may be further constrained, depending on the message type. "Padding" of extra data after the message is not allowed. Therefore, the Length field must have the smallest value required, given the rest of the message. The Type field is an unsigned integer that indicates the type code of the message. There are four types of BGP messages: Open, Update, Notification, and Keep Alive messages.

B. Routing Updates

Dissimilar to Routing Information Protocol (RIP), a distance-vector routing protocol, which utilizes the hop count as a routing metric, BGP, does not broadcast its entire routing table. At boot, your peer will hand over its entire table. After that, everything depends on updates received. Route updates are stored in a Routing Information Base (RIB). A routing table will just store one route for every end point, but the RIB normally contains numerous paths to a destination. It is depended on the router to choose which routes will make it into the routing table, and therefore which paths will actually be used. In case that a route is withdrawn, an alternate route to the same place can be taken from the RIB. The RIB is only used to keep track of routes that could possibly be utilized. If a route withdrawal is received and it only existed in the RIB, it is silently deleted from the RIB. No update is sent to peers. RIB entries never time out. They continue to exist until it is assumed that the route is no longer valid.

C. Path Attributes

In many cases, there will be multiple routes to the same destination. BGP therefore uses path attributes to decide how to route traffic to specific networks. The easiest of these to understand is Shortest AS_Path. What this means is the path that traverses the least number of AS "wins." Another important attribute is Multi_Exit_Disc (Multi-exit discriminator, or MED). This makes it possible to tell a remote AS that if there are multiple exit points on to your network, a specific exit point is preferred. The Origin attribute specifies the origin of a routing update. If BGP has multiple routes, then origin is one of the factors in determining the preferred route.

D. Error Handling

When an error occurs, a Notification message with the indicated Error Code, Error Subcode, and Data fields is sent, and the BGP connection is closed. If no Error Subcode is specified, then a zero must be used. The phrase "the BGP connection is closed" means that the transport protocol connection has been closed and that all resources for that BGP connection have been de-allocated. Routing table entries associated with the remote peer are marked as invalid. The fact that the routes have become invalid is passed to other BGP peers before the routes are deleted from the system. Unless specified explicitly, the Data field of the Notification message that is sent to indicate an error is empty.

E. Version Negotiation

BGP speakers may negotiate the version of the protocol by making multiple attempts at opening a BGP connection, starting with the highest version number each BGP speaker supports. If an open attempt fails with an Error Code, Open Message Error, and an Error Subcode, Unsupported Version Number, then the BGP speaker has available the version number it tried, the version number its peer tried, the version number passed by its peer in the Notification message, and the version numbers it supports. If the two peers support one or more common versions, this will allow them to rapidly determine the highest common version. In order to support BGP version negotiation, future versions of BGP must retain the format of the Open and Notification messages.

F. Timers

BGP employs five timers: ConnectRetryTimer, HoldTimer, KeepaliveTimer, MinASOriginationIntervalTimer, and MinRouteAdvertisementIntervalTimer. Two optional timers may be supported: DelayOpenTimer, IdleHoldTimer by BGP. Section 8 describes their use. The full operation of these optional timers is outside the scope of this document. ConnectRetryTime is a mandatory FSM attribute that stores the initial value for the ConnectRetryTimer. The suggested default value for the ConnectRetryTime is 120 seconds. HoldTime is a mandatory FSM attribute that stores the initial value for the HoldTimer. The suggested default value for the HoldTime is 90 seconds. During some portions of the state machine (see Section 8), the HoldTimer is set to a large value. The suggested default for this large value is 4 minutes. The KeepaliveTime is a mandatory FSM attribute that stores the initial value for the KeepaliveTimer. The suggested default value for the KeepaliveTime is 1/3 of the HoldTime. The suggested default value for the MinASOriginationIntervalTimer is 15 seconds. The suggested default value for the MinRouteAdvertisementIntervalTimer on EBGP connections is 30 seconds. The suggested default value for the MinRouteAdvertisementIntervalTimer on IBGP connections is 5 seconds. An implementation of BGP must allow the HoldTimer to be configurable on a per-peer basis, and may allow the other timers to be configurable. To minimize the likelihood that the distribution of BGP messages by a given BGP speaker will contain peaks, jitter should be applied to the timers associated with MinASOriginationIntervalTimer, KeepaliveTimer, MinRouteAdvertisementIntervalTimer, and ConnectRetryTimer. A given BGP speaker may apply the same jitter to each of these quantities, regardless of the destinations to which the updates are being sent; that is, jitter need not be configured on a per-peer basis. The suggested default amount of jitter shall be determined by multiplying the base value of the appropriate timer by a random factor, which is uniformly distributed in the range from 0.75 to 1.0. A new random value should be picked each time the timer is set. The range of the jitter's random value may be configurable.

G. Security

BGP makes use of TCP for reliable transport of its traffic between peer routers. To provide connection-oriented integrity and data origin authentication on a point-to-point basis, BGP specifies use of the mechanism defined in RFC 2385 [3]. These services are intended to detect and reject active wiretapping attacks against the inter-router TCP connections. Absent the use of mechanisms that affect these security services, attackers can disrupt these TCP connections and/or masquerade as a legitimate peer router. Because the mechanism defined in the RFC does not provide peer-entity authentication, these connections may be subject to some forms of replay attacks that will not be detected at the TCP layer. Such attacks might result in delivery (from TCP) of "broken" or "spoofed" BGP messages. The mechanism defined in RFC 2385 augments the normal TCP checksum with a 16-byte message authentication code (MAC) that is computed over the same data as the TCP checksum. This MAC is based on a one-way hash function (MD5) and use of a secret key. The key is shared between peer routers and is used to generate MAC values that are not readily computed by an attacker who does not have access to the key. A compliant implementation must support this mechanism, and must allow a network administrator to activate it on a per-peer basis. RFC 2385 does not specify a means of managing (e.g., generating, distributing, and replacing) the keys used to compute the MAC. RFC 3562 [4] (an informational document) provides some guidance in this area, and provides rationale to support this guidance. It notes that a distinct key should be used for communication with each protected peer. If the same key is used for multiple peers, the offered security services may be degraded, e.g., due to an increased risk of compromise at one router that adversely affects other routers. The keys used for MAC computation should be changed periodically, to minimize the impact of a key compromise or successful cryptanalytic attack. RFC 3562 suggests a crypto period (the interval during which a key is employed) of, at most, 90 days. More frequent key changes reduce the likelihood that replay attacks (as described above) will be feasible. However, absent a standard mechanism for effecting such changes in a coordinated fashion between peers, one cannot assume that BGP-4 implementations complying with this RFC will support frequent key changes. Obviously, each key should also be chosen to be difficult for an attacker to guess. The techniques specified in RFC 1750 [5] for random number generation provide a guide for generation of values that could be used as keys. RFC 2385 calls for implementations to support keys "composed of a string of printable ASCII of 80 bytes or less." RFC 3562 suggests keys used in this context be 12 to 24 bytes of random (pseudo-random) bits. This is fairly consistent with suggestions for analogous MAC algorithms, which typically employ keys in the range of 16 to 20 bytes. To provide enough random bits at the low end of this range, RFC 3562 also observes that a typical ACSII text string would have to be close to the upper bound for the key length specified in RFC 2385.

IV. Hardware and software requirements for BGP

Routers usually used in small offices and/or home networks may not be capable of running BGP. Some routers may contain table size limits that are too low to support BGP for the network size. The amount of memory required for a BGP speaker depends on many factors including the volume of BGP information being exchanged with the other BGP speakers, the BGP attributes configured/used on the BGP speaker, the BGP speaker's way of storing BGP information, as well as possible VPN configurations.

Commercial grade routers found in large enterprise networks and used by Internet Service Providers (ISPs) typically have operating systems capable of supporting BGP. Common manufacturers of BGP capable routers include Cisco, Alcatel Lucent, and Juniper Networks. There are some open source packages capable of running BGP including GNU Zebra, Ouagga, Open BGPD, BIRD, XORP, and Vyatta.

V. Vulnerabilities of BGP

BGP has been found to be subject to the following attacks:

1. Confidentiality violations (eavesdropping)
2. Replay
3. Message Insertion - since BGP uses TCP, this is much more difficult after the TCP connection has been completely established
4. Message Deletion – again this is much more difficult, but not impossible, once the TCP connection has been completely established
5. Message Modification – any modification that maintains the syntax and length of the TCP payload may not be detected
6. Man-in-the-Middle – BGP does not perform any authentication for BGP speakers/peers
7. Denial of Service (DoS) – the most critical DoS attack would be on the BGP routing protocol itself

These various attacks can be summarized into three fundamental vulnerabilities:

1. No authentication of or protection against the integrity or newness of BGP speaker/peer communication
2. No validation of the authority of an AS to send NLRI information
3. No authentication of the path attributes sent by an AS

Definition of Frequentlty used terms

A. *Autonomous System (AS)*

An autonomous system is a collection of routers, a connected group of one or more blocks of IP addresses assigned to a particular organization and provide a single routing policy. AS are controlled by a network administrator.

B. *BGP Identifier*

The BGP Identifier is the identifier for a BGP speaker. The format is a 4-octet unsigned, non-zero integer, normally the IPv4 host address, and should be unique to the AS the BGP belongs, if possible.

C. *BGP Multipath*

When enabled, BGP Multipath allows multiple BGP paths for the same destination to be included in the IP routing table.

D. *BGP Speaker*

A router configured to handle BGP messages is called a BGP speaker. BGP speakers call each other peers or neighbors.

E. *CIDR (Classless Inter-Domain Routing) notation*

CIDR notation is constructed from the IP address (in IPv4 or IPv6 standard) and the prefix size (as a decimal).

F. *Exterior Gateway Protocol (EGP)*

EGP is a routing protocol that is used to transfer or share data between neighbor gateways hosts in a network of autonomous systems. The information is shared among the neighbors through periodic messages and commands polling for reachability and requests to update responses.

G. *External BGP (EBGP)*

External Border Gateway Protocol (EBGP) is a BGP extension used for data transfer and communication between BGP enabled systems in different autonomous systems.

H. *Interior Gateway Protocol (IGP)*

An Interior Gateway Protocol (IGP) is a network protocol used for distributing routing information among gateways within an AS.

I. *Internal BGP (IBGP)*

Internal BGP is a network protocol used to provide routing information between routers within the same AS.

J. *Multi-Exit Discriminator (MED)*

Multi-Exit Discriminator is a metric attribute advertised by an AS to suggest a preferred route into the AS for external autonomous systems.

K. *Multiprotocol BGP (MBGP)*

Multiprotocol BGP is an enhancement or extension to BGP that allows multiprotocol routing. While BGP supports only IPv4 unicast addresses, MBGP supports IPv4 and IPv6 addresses as well as unicast and multicast variants of each.

L. *NLRI (Network Layer Reachable Information)*

Network layer reachable information is exchange between BGP routers consisting of a network mask in CIDR notation specifying the number of network bits (length) and the network address for the subnet (prefix).

M. *Open Shortest Path First (OSPF)*

Open Shortest Path First is a routing protocol used to find the shortest and/or most efficient route between routers operating within a single AS.

N. *Route*

A route is a pairing of a destination and the attributes of the path to that destination (gateway). There are two types of routes:

1) *Feasible Routes*

Routes in which message communication is possible and contain no loops are called feasible routes.

2) *Unfeasible Routes*

Routes in which message communication is not possible or that contain loops are called unfeasible routes.

O. *Routing Information Base (RIB)*

A Routing Information Base (RIB) contains all route updates for destinations to which the router may forward information. The RIB is at the heart of BGP's system of routing management and consists of three databases or tables used by each BGP speaker:

1) *Adj-RIBs-In (Adjacent Routing Information Base, Incoming):*

Routes received from the BGP speakers are contained in a Adj-RIBs-In table. Each BGP speaker has a separate Adj-RIBs-In routing table.

2) *Adj-RIBs-Out (Adjacent Routing Information Base, Outgoing):*

Routes advertised to BGP speakers are contained in a Adj-RIBs-Out table. Each BGP speaker has a separate Adj-RIBs-Out routing table.

3) *LOC-RIB (Local Routing Information Base):*

The LOC-RIB is separate from the main routing table of the router and holds the best routes from the Adj-RIB-In and Adj-RIB-Out tables found using the BGP route selection algorithm.

P. *Routing Information Protocol (RIP)*

Routing Information Protocol uses a hop count to find the best path/route from source to destination. Routing loops are prevented by imposing a limit on the number of hops allowed for a route.

Q. *AS_PATH*

The AS_PATH attribute is an ordered list of AS numbers the route passed through to arrive at each destination.

References

[1] RFC 4271 – A Border Gateway Protocol 4 (BGP-4). January 2006.

[2] Cisco Systems, BGP Best Path Selection Algroithm. Document ID: 13753, May 2012 http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html

[3] RFC 2385 – Protection of BGP Sessions via the TCP MD5 Signature Option. August 1998 https://www.ietf.org/rfc/rfc2385.txt

[4] RFC 3562 – Key Management Considerations for the TCP MD5 Signature Option. July 2003 https://www.ietf.org/rfc/rfc3562.txt

[5] RFC 1750 – Randomness Recommendations for Security. December 1994 https://www.ietf.org/rfc/rfc1750.txt

[6] Cisco Systems, Border Gateway Protocol. Last modified 13 September 2013 at 11:02. http://docwiki.cisco.com/wiki/Border_Gateway_Protocol

[7] Cisco Systems, BGP Case Sutdies. Document ID: 16634, March 2014 http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/26634-bgp-toc.html

[8] Cisco Systems, Achieve Optimal Routing and Reduce BGP Memory Consumption. Document ID: 12512, February 2008 http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/12512-41.html

[9] D.G. Anderson, N. Feamster, S. Bauer and H. Balakrishnan, Topology inference from BGP routing dynamics, Internet Measurement Workshop 2002, 2002.

[10] G. Huston, Analyzing the Internet's BGP routing table, The Internet Protocol Journal, 4(1), March 2001.

[11] G. Huston, BGP routing table statistics, http://www.telstra.net/ops/bgp

[12] RFC 5291 – Outbound Route Filtering Capability for BGP-4, August 2008.

[13] RFC 3221 – Commentary on Inter-Domain Routing in the Internet, December 2001.

[14] B. Christian, T. Tauber, BGP Security Requirements, Internet-Draft: IETF (2006).