BGP Routing Protocol

A Master's Project

Presented to

Department of Telecommunications

In Partial Fulfillment

of the Requirements for the

Master of Science Degree

State University of New York

Polytechnic Institute

By

Sai Kiran Parasa

Aug 2016

BGP Routing Protocol

Declaration

I declare that this project is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

P. Sai kiran

Sai Kiran Parasa
08/13/2016

SUNYIT

DEPARTMENT OF TELECOMMUNICATIONS

Approved and recommended for acceptance as a thesis in partial fulfillment of the requirements for the degree of Master of Science in Telecommunications.

9/6/2016

DATE

Dr. Larry Hash

Project Advisor

Table of Contents:

LIST OF FIGURES:

Abstract:

Border Gateway Protocol is the protocol which makes the Internet work. It is used at the Service provider level which is between different Autonomous Systems (AS). An Autonomous System is a single organization which controls the administrative part of a network. Routing with in an Autonomous System is called as Intra-Autonomous routing and routing between different Autonomous Systems is called as Inter-Autonomous System routing. The routing protocols used within an Autonomous System are called Interior Gateway Protocols (IGP) and the protocols used between the Autonomous Systems are called Exterior Gateway Protocols. Routing Information Protocol (RIP), Open Short Path First (OSPF) and Enhanced Interior Gateway Routing Protocol (EIGRP) are the examples for IGP protocols and Border Gateway Protocol (BGP) is the example for EGP protocols.

Every routing protocol use some metric to calculate the best path to transfer the routing information. BGP rather than using a particular metric, it uses BGP attributes to select the best path. Once it selects the best path, then it starts sending the updates in the network. Every router implementing BGP in the network, configures this best path in its Routing Information Base. Only one best route is selected and forwarded to the whole network. [17] Due to the tremendous increase in the size of the internet and its users, the convergence time during link failure in the protocol is very high.

Introduction:

Border Gateway Protocol is an IETF (Internet Engineering Task Force) standard protocol, which is the most scalable of all routing protocols [6]. BGP is the protocol which makes the Internet work. The Internet is a huge network which connects many devices globally. Around half of the world's population is using internet. Internet is not controlled by any centralized device or host and hence it is considered as decentralized. An Internet Service Provider (ISP) is a company which provides the Internet access to the individuals and other companies. An individual system is connected to the ISP by means of a router. A router is a device which is used to forward the routing information from source to destination over the Internet. These routers use routing protocols to transmit the data by choosing the best path.

The main purpose of the routing protocols is to learn the available routes and build the tables and make the routing decisions. These routers use the route selection mechanism to calculate the feasible paths to send the data. Group of routers with a common policy in a single administration is called an Autonomous System (AS). In other words, a single organization which controls the administrative part of a network is called an Autonomous System (AS). These Autonomous Systems are connected under one tree, called Internet. Autonomous system is a unit of router policy, which is controlled by a network administrator on the behalf of an administrative entity. Here an entity can be a university, or business enterprise or business division or an Internet Service Provider (ISP).

Due to the increase in usage of commercial Internet, the complexity in the network which holds privately-owned Autonomous Systems (ASes) has tremendously increased. Each AS is assigned with a global unique number called Autonomous System Number (ASN). These Autonomous System Numbers are given by an organization called IANA (Internet Assigned Numbers Authority). IANA keeps the Internet to be globally connected. It manages the domain names and provides the number resources for the Internet Protocols and Autonomous Systems. As the size of the Internet has increased tremendously, IANA divided the whole system into 5 RIR's (Regional Internet Registries) and assigned it to provide the Internet Protocol address space and Autonomous System Number within a defined region. ARIN (American Registry for Internet Numbers) is one of the five Regional Internet Registries which administers the US part.

These Autonomous Systems form the internet. The routing with the Autonomous Systems are mainly of two types, Routing within an Autonomous System, which is called Intra-AS routing mechanism and routing between the Autonomous Systems which is called as Inter-AS routing

mechanism. The routing protocols used within the Autonomous System to distribute the routing information are called as Interior Gateway Protocols (IGPs). Some of the Interior Gateway Protocols are; Routing Information Protocol (RIP), Open Short Path First (OSPF) and Enhanced Interior Gateway Protocol (EIGRP). The routing protocols that are used to exchange the routing information between the Autonomous Systems are called as Exterior Gateway Protocols (EGPs). EGP's are the glue which sticks the Autonomous Systems throughout the internet. Border Gateway Protocol (BGP) is an example for Exterior Gateway Protocol. These two types of protocols are shown in the figure 1,

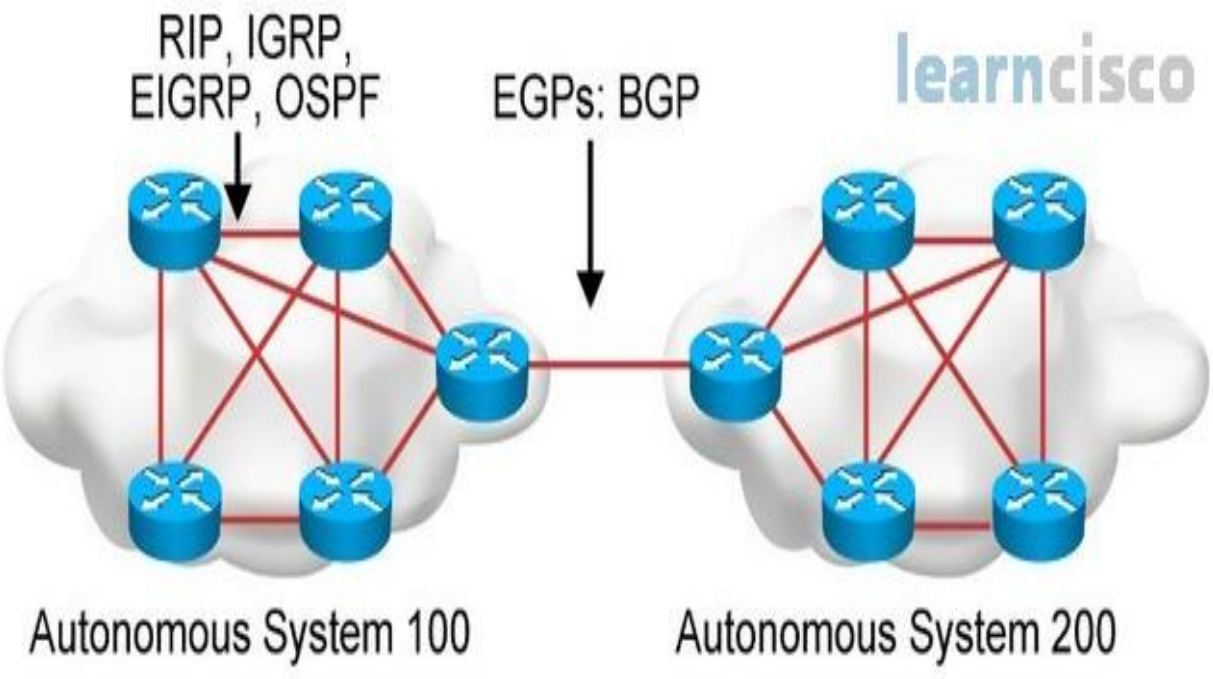

Figure 1: Autonomous Systems with IGP and EGP protocols [26]

http://www.learncisco.net/courses/icnd-1/ip-routing-technologies/dynamic-routing.html

We can see that there are two Autonomous Systems in the network in the figure 1, Autonomous System 100 and Autonomous System 200. The communication between the routers within the Autonomous System is being done using Interior Gateway Protocols like RIP, OSP and EIGRP and the communication between individual Autonomous Systems is done by using Exterior Gateway Protocols like BGP [24].

Each IGP protocol has their own way to propagate to the destination network. They use metrics to find the best path and advertise that path to the other routers in the network. The RIP protocol

uses hop count as a metric to calculate the best path. It chooses the path with least number of hops. It takes distance and direction through which it has to travel to the destination. OSPF uses cost as a metric. It chooses the path with least cost. This cost is calculated by using shortest path first algorithm or Dijkstra's algorithm. EIGRP is a Cisco proprietary protocol which uses bandwidth and delay as the metric to find the best path. It uses DUAL (Diffusing Update Algorithm) for finding the best path. We are going to look at each protocol and their functioning in the upcoming sections.

BGP is mainly used to exchange the routing information between the Autonomous Systems (ASes). The main function of the BGP speaking system (Router implementing BGP) is to exchange the reachability information of the network with the corresponding BGP systems. Each router which implements BGP connect to its neighboring BGP router with some shared policies.

Once the connection is established with the neighboring routers, they exchange their routing tables. The tables consist of all the routing information and the available peers in the network. A single best route per destination is stored in the forwarding table for routing. It uses path vector algorithm to select the best path. It focuses on the path selection to the destination prefix. Unlike the IGP protocols, it does not have any particular metrics to find the best path but instead it uses multiple path attributes to select the best path. We will look into each path attribute in detail in the next chapters.

In the rest of the paper we are going to look into different IGP and EGP protocols. First we are going to see the classification of routing protocols and then look in deep with each protocol. Each IGP protocol is explained with examples. We are going to focus more on the routing in border gateway protocol, the path attributes which it uses to select the best path and how the routing updates happen in the protocol. Next we mention the shortcomings of the existing BGP protocol and suggest some enhancements for the future research.

Audience Definition:

This paper will be understood by an undergraduate final year student and the graduate student who has done their major in Telecommunications and Network Security. Also, this paper can be understood by the people who have basic understanding about functioning of routing protocols and they need not be from the Telecommunication background.

Thesis Statement:

This research paper is about how the Border Gateway Protocol works in the Internet. This paper also gave brief explanation about the IGP protocols and how they work. Each protocol is explained with an example and algorithm that is used to select the best path.

Literature Review:

Y. Rekhter, T. Li, S. Hares, "A Border Gateway Protocol 4 (BGP – 4)", January 2006
https://tools.ietf.org/html/rfc4271

This document gives a clear idea about Border Gateway Protocol (BGP). It discusses about the Autonomous Systems (ASes) and how the communication is happening between them using BGP. The route advertisement mechanism and the message format is clearly explained here. The local Routing Information Base (RIB) consists of all the possible routes but only one route is advertised among all the peers in the network. This best route is selected by the BGP by using path Attributes.

It discussed about the Error handling capability of the protocol and also about the collision detection. It also tells us about the route origination and the route replacement in the case of route failure. This document is the base for my topic as my paper discusses about the BGP enhancements.

As per my view this paper gives a clear introduction about the BGP protocol and how the best path is selected to propagate the information in the network.

D. Walton, A. Retana, E. Chen and J. Scudder, "Advertisement of Multiple Paths in BGP", Network Working Group, Internet Draft, May 23, 2016 (Work in progress).

This draft mainly focuses on the extension to the existing BGP protocol. In traditional BGP only single best path is advertised but this extension allows multiple paths to be advertised for the same destination prefix without replacing the previous ones. Each path in the list is identified by a new attribute called 'Path Identifier' which is added to the address prefix. Previously when a new advertisement for the same destination prefix is made then it is replaced with the previous one. There were no provisions to allow multiple path advertisement for the same address prefix. Each path is uniquely identified by the address prefix and the path identifier. This draft is relevant to my topic as it discusses about the enhancements to the existing BGP protocol.

In my view this draft clearly explains about the advertisement of multiple paths to the same address prefix, which was not possible before. This might open up new doors in the advancement of Internet world.

P. Brighten Godfrey, Mathew Caesar, Ian Haken, Yaron Singer, Scott Shenker and Ian Stoica, "Stabilizing Route Selection in BGP", IEEE/ACM Transactions on Networking, VOL. 23, NO. 1, February 2015

This paper mainly discussed about the problems in route stability and the trade-offs that occur by modifying the route selection. It also explains about the existing route stability mechanism, RFD (Route Flap Damping) which filters the short-term update routes above some threshold. But this mechanism arises two main problems, slow convergence and degradation of availability. Hence there has been proposed a new approach called SRS (Stable Route Selection). This approach improves the stability by using the flexibility in route selection without sacrificing the availability and also controls the amount of deviation.

The main goal of this paper is to work on improving the BGP's decision process in route selection. It mainly discusses about Stabilizing and route availability of the BGP. This paper also discusses about the points which are feasible for the tradeoff spaces which are novel route selection categories and provable lower bounds. There are a lot of simulations that have been conducted on the BGP route selection by using SRS. The results show that this new selection process improves a significant control-place overhead and reliability in the data-plane with only considerable amount of deviation from the preferred routes. This paper is relevant to my topic as it mainly focuses on the route selection methods and the approach to improve the availability of the routes.

As per my view, this paper heavily focused on the tradeoffs that appear during route switching and the mechanism to overcome the instability.

Aleksandra Cvjetic and Aleksandra Smiljanic, "Improving BGP Protocol to Advertise Multiple Routes for the Same Destination Prefix", IEEE Communications Letters, VOL. 18, NO. 1, January 2014.

This paper mainly discusses about the enhancement in the route selection of the BGP. It clearly explains about the present problems with the routing system and path selection in BGP. A new route selection algorithm called BGP-FRP (BGP with Flexible Routing Policies) was put forward. The BGP-FRP advertises the multiple routes present for a destination prefix, so that the implementation of BGP policies can be done simultaneously.

The BGP-FRP algorithm was implemented in XORP open-source router, in which previously only one route was advertised. In case of change in network topology, it is better to have multiple alternate routes stored in the routers as the routing gets adjusted much faster. A Linux Operating System is used to deploy a virtual network with XORP routers and there are some simulations. After implementing the algorithm, the reliability of the routing is improved as the router stores multiple routes for the given destination.

In my view, this paper explains about how multiple routes are advertised in the network and route selection changes immediately if there is any failure in the present route.

L L Ragha, K V Ghag, "Multiple Route Selector BGP (MSR-BGP)", ICWET'10, February 26–27, 2010, Mumbai, Maharashtra, India.

This paper mainly discusses about the protocol which is an extension to BGP, which is MRS-BGP (Multi Route Selector Border Gateway Protocol). This paper clearly explains about the routing system in the BGP. Even though there are many routes present between origin and destination prefix, the BGP selects only one best route and advertises it to the other ASes (Autonomous Systems). Hence if there is any failure in the route, the BGP looks to construct a new route. Due to this there are two disadvantages, Network bandwidth underutilization and low resistance to link failure.

This paper clearly explains how to overcome these disadvantages by using the MSR-BGP protocol. It also explains how this protocol is less susceptible to looping over other multi route selection algorithms. It also explains about the existing multiple route selection algorithms, their advantages and disadvantages and how the proposed protocol is better than them. It clearly explains how the proposed algorithm perfectly uses the residual bandwidth and avoids the transient loops formed in the network.

According to my view, this paper was able to give a solution to the existing routing problems in BGP by proposing a new protocol.

Classification of Routing Protocols:

A Routing protocol is a set of guidelines that a router has to follow when it communicates or exchanges the routing information. Routing protocols are used to calculate the optimal routes by using the reachability information in the network and update the routing tables. Routers use these routing protocols to know the different possible routes and store them in their tables. Each protocol has its own metrics to calculate the feasible paths. They use metrics to calculate the best paths. As discussed earlier, Routing protocols are mainly divided into two types, Interior Gateway Protocols and Exterior Gateway Protocols. The following classification in figure 2 clearly explains their division,

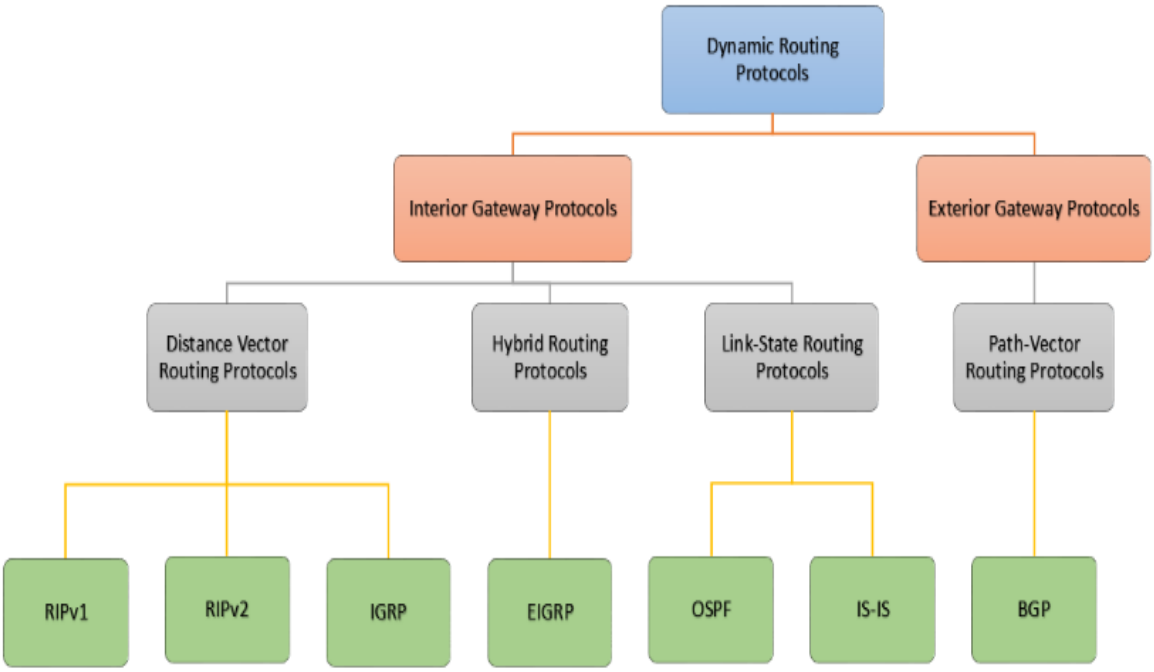

Figure 2: Classification of Routing Protocols [25]

http://www.techtutsonline.com/dynamic-routing/

Interior Gateway Routing Protocols:

The routing protocols used within the Autonomous System to distribute the routing information are called as Interior Gateway Protocols (IGPs). These are responsible for routing updates with in the Autonomous System. Each IGP protocol have their own metrics to choose the best path in the network. These are further classified into two types, distance vector routing protocols and link-state routing protocols. Most of the IGP protocols fall under these two categories.

Routing Information Protocol popularly known as RIP is one of the oldest protocols which is an example for distance vector routing protocol. Open Shortest Path First protocol which is popularly known as OSPF is an IGP protocol which is an example for link state routing protocol. The Enhanced Interior Gateway Routing Protocol (EIGRP) is known as the hybrid routing protocol which has the features of both distance vector and link state routing protocols. We will look into each protocol in detail in the next sections.

Distance Vector Routing Protocols:

Distance vector routing protocols are the first generation protocols. These are fairly simple and not designed for the complex networks. They have limited scope of technology. They can see as far as they are connected to their neighbors or next routers. They have no idea how the topology looks like behind these routers. They have the hop count as their metric. Metrics are used to make the routing decisions by the routers in the network.

The distance vector routing protocol follows the distance vector algorithm, which states that routes are advertised with the vector of distance and direction. The distance implies to the number of hops required to reach the destination i.e. number of routers that it has to pass to get to the destination and the direction refers to the direction of next hop router.

Each router implementing this protocol advertises the network reachability information to the neighboring routers. Neighbors always mean the routers sharing a common data link in the network. These advertisements happen periodically with 10 to 90 seconds interval depending on which type of protocol we are using. The common examples for the distance vector routing protocols are Routing Information Protocol (RIP) and Cisco's Interior Gateway Routing Protocol (IGRP). RIP usually advertises periodically for every 30 seconds.

The distance vector routing protocols usually select the best path which has the lowest cost. Lowest cost here refers to the lowest number of hops. It automatically selects the best path as the path which has the least number of hops. The working of the protocol is further explained in RIP protocol in the next section.

Routing Information Protocol:

Routing Information Protocol (RIP) is one of the oldest distance vector routing protocols. The main function of the routing protocol is to choose the best path to reach the destination prefix. This is done by updating the routing tables in the router by information received from the neighboring routers in the network. Every protocol has a metric to choose the best path among the available paths. RIP uses hop count as the metric to calculate the best path. The maximum number of hops that would allow RIP to reach the destination is 15 hops. If the destination prefix is more than 15 hops then the prefix is marked as unreachable. Hence the RIP is not suitable for the larger networks.

Let us looking at the functioning of the RIP by looking at a small network which has 3 routers A, B and C. E0 and E1 represent the Interfaces through which the routers are configured to the network.

In figure 3, we can see each router has its own routing table with the network information that it is connected to. Router A has directly connected to 1.0.0.0 network and 2.0.0.0 network with the Interface E0 and E1. The third column in the routing table represents the number of hops required to reach that particular network. As the router A is directly connected to 1.0.0.0 and 2.0.0.0 networks, the hop count is going to be 0. Same with Router B and Router C. This is how it looks before running the protocol.

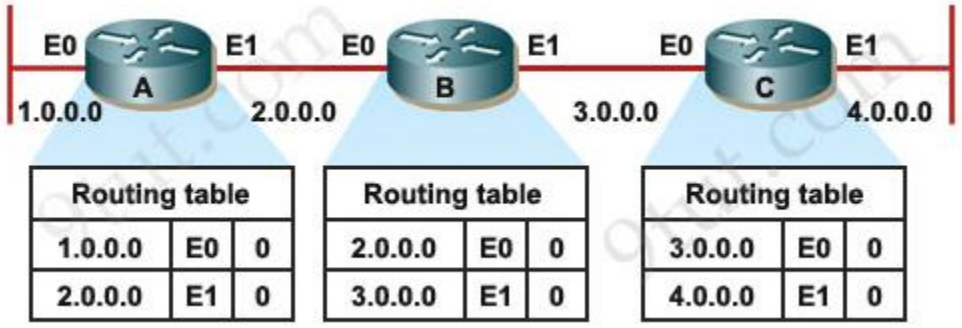

Figure 3: RIP Initial Network [8]

http://www.9tut.com/rip-routing-protocol-tutorial

Now let's assume every router implements RIP protocol in the network. Now router A shares it network information to its neighbor router B. And router B adds the 1.0.0.0 network to its routing table and ignores the 2.0.0.0 network as it is also directly connected to the same network. As we can see the routing table of the router B added the new network prefix with hop count as '1' in the third column. This means it has to cross one router to reach that particular network.

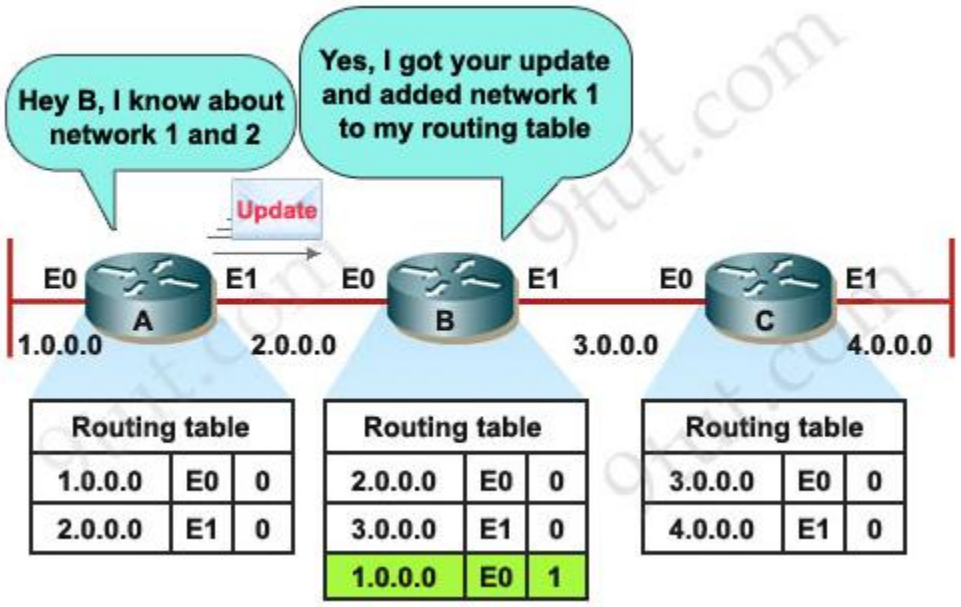

Figure 4: RIP: Router A sharing reachability information [8]

http://www.9tut.com/rip-routing-protocol-tutorial

In the same way, router B shares its routing table with router A. And router A adds the 3.0.0.0 network into its routing table with hop count as 1. We can see this in figure 5,

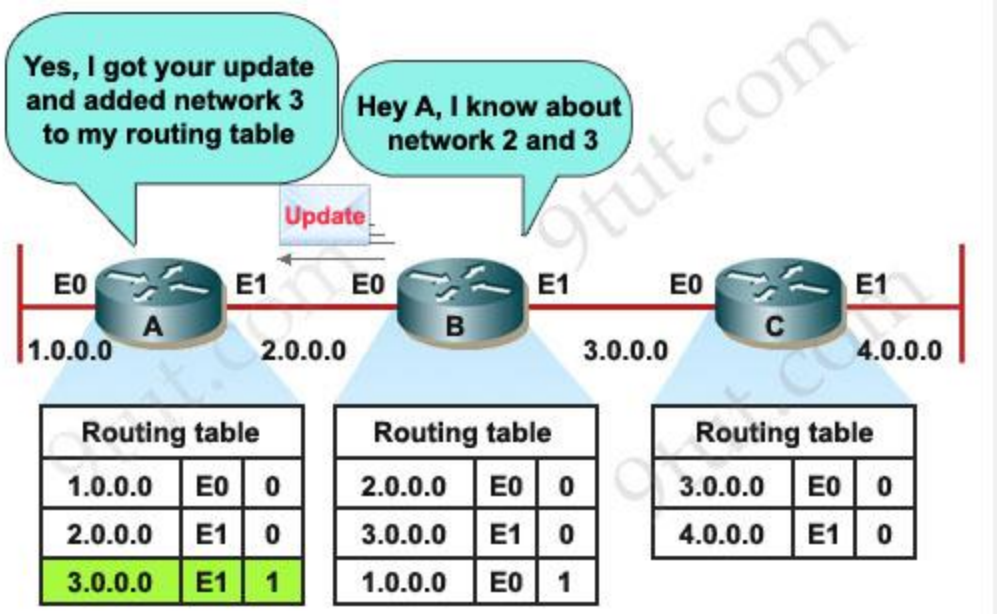

Figure 5: Router A table updated from Router B [8]

http://www.9tut.com/rip-routing-protocol-tutorial

In figure 6, we can see router B shares its Routing table with router C. And router C updates its routing table with router B's information and adds the two new networks, 1.0.0.0 and 2.0.0.0 to its routing table. As we can see the hop count to reach the 1.0.0.0 network is 2 as it has to cross two routers, router A and router B in order to reach that network. In order to reach 2.0.0.0 network, it just needs to cross router B, hence its hop count is 1.

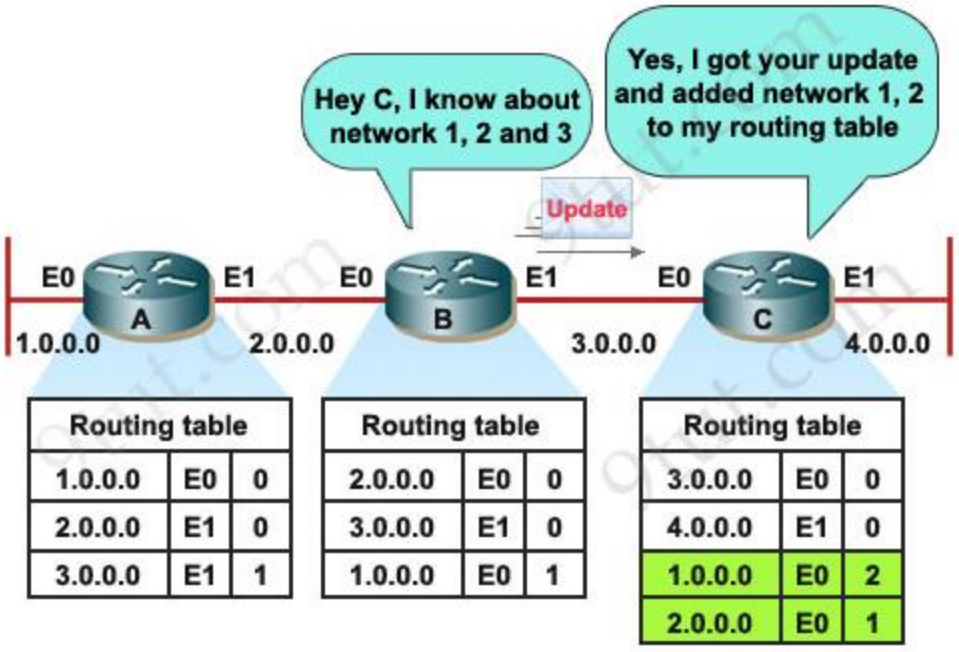

Figure 6: Router C updated with neighbor Router B [8]

http://www.9tut.com/rip-routing-protocol-tutorial

The figure 7 shows the full updated routing tables of the routers in the network. We can see router B got to know about the 4.0.0.0 network from router C and added that network in its routing table. In the same way in next periodic update, the router A gets updated with the new network 4.0.0.0 from router B. It keeps the hop count as 2 as it has to cross router B and router C to reach that network.

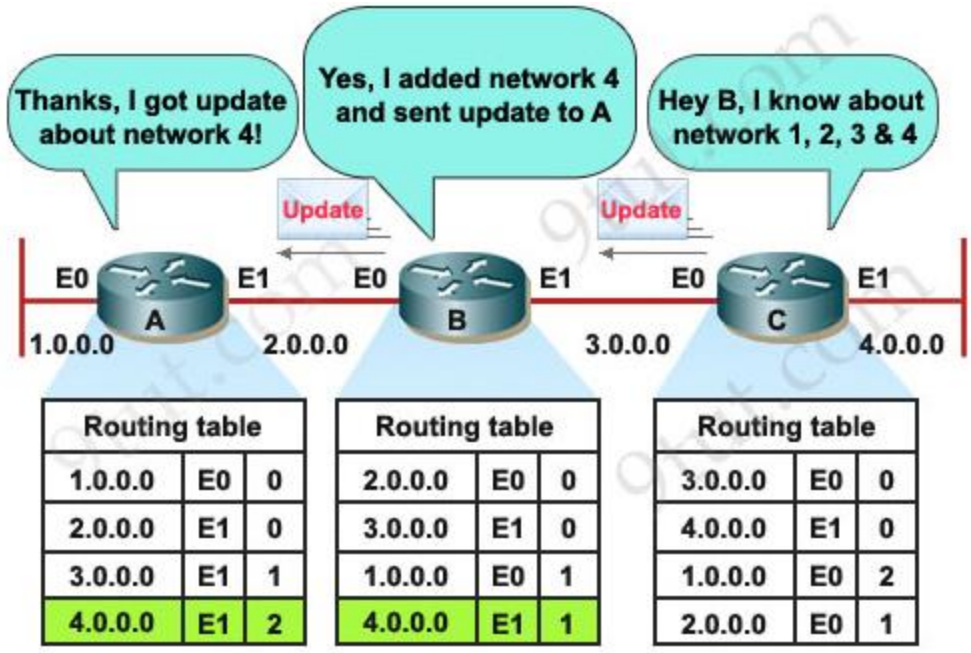

Figure 7: Complete Network setup in Routing tables [8]

http://www.9tut.com/rip-routing-protocol-tutorial

Hence in this way all the routing tables in the network gets updated. The RIP sends the periodic updates for every 30 seconds. Suddenly if a link gets failed in the network, it will wait till the next periodic update to update the other routers in the network. Hence the convergence time taken will be more in the network.

Link State Routing Protocols:

Link state routing protocols are like road map and whereas the distance vector routing protocols are like road signs. They have full scope of the topology. Each router in the network implementing Link state routing protocol initiates its reachability information and passes on to the neighboring routers. In this way it has the whole topology information of the network. The ultimate goal is to have the identical information in every router about the network.

Link state routing protocols are also called as the distributed database protocol or the short path first which runs on the Dijkstra's shortest path algorithm. Open Short Path First (OSPF) is the suitable example for the Link state routing protocol. The routers in the network does not send routing updates periodically like distance vector protocol, instead it only sends updates if there is any link failure or if there is any change in the topology. This update from the router is sent to each and every router that is connected in the network not only to the directly connected router. It reduces the CPU overhead by great amount as there won't be any periodic updates.

First of all, all the routers in the network implementing same protocol should get to know about each other. This is done by sending the 'Hello' message by a router to the neighboring router by prepending its router id into the message. A router will wait for the reply for certain period before declaring the link as 'dead'. When there is a neighbor discovery, then these hello message are used to monitor the health of the link. These hello packets are exchanged for every 10 seconds. A router is flagged unreachable if it does not receive the response back in 40 seconds. All the network updates are flooded in the network rather than updating periodically.

Link state flooding:

Once the network is established, the routers are allowed to send the advertisements which Link State Advertisements (LSA) to report any event. Her the event can be a link failure or any change in the topology. Thus advertisement should be broadcasted to every neighbor in the network. Each LSA is assigned with the sequence number which is a unique identifier to that particular update. Every LSA is stored in the link state database of the router and forwarded to the neighboring router. If a router receives the same LSA which it already knows then it simply ignores.

This is how the updates happen in the link state routing protocol. As the router does not wait for any periodical route updates, it simply broadcasts the update when it receives the LSA. This makes the Link state routing protocol converge faster when compared the distance vector routing protocol. And also the LSA is attached with a unique identifier the routers have full knowledge on the change in the network. Even though a router receives the same LSA from different router it simply ignores and hence there won't be any loops in the network.

The application of the link state protocol is explained in OSPF protocol in the below section.

Open Short Path First (OSPF):

Open Short Path First (OSPF) protocol is the widely used Interior Gateway Protocol (IGP). It is a public routing protocol which means it is not owned by any company like EIGRP (Enhance Interior Gateway Routing Protocol) which is a Cisco owned routing protocol. OSPF is one of the best Link state routing protocol. [9] As discussed earlier in link state routing protocol, router has the knowledge of topology and updates are not sent periodically. The updates are sent only when there is a link failure at any router or if there is any topology change.

In an Autonomous System running with OSPF protocol, the whole region is divided into multiple areas. This is done to decrease the traffic bursts in the network. Area 0 is called as the backbone are which will be in touch with every other area in the network. Every router in the other areas are local to that particular area and have the routing tables to that particular network.

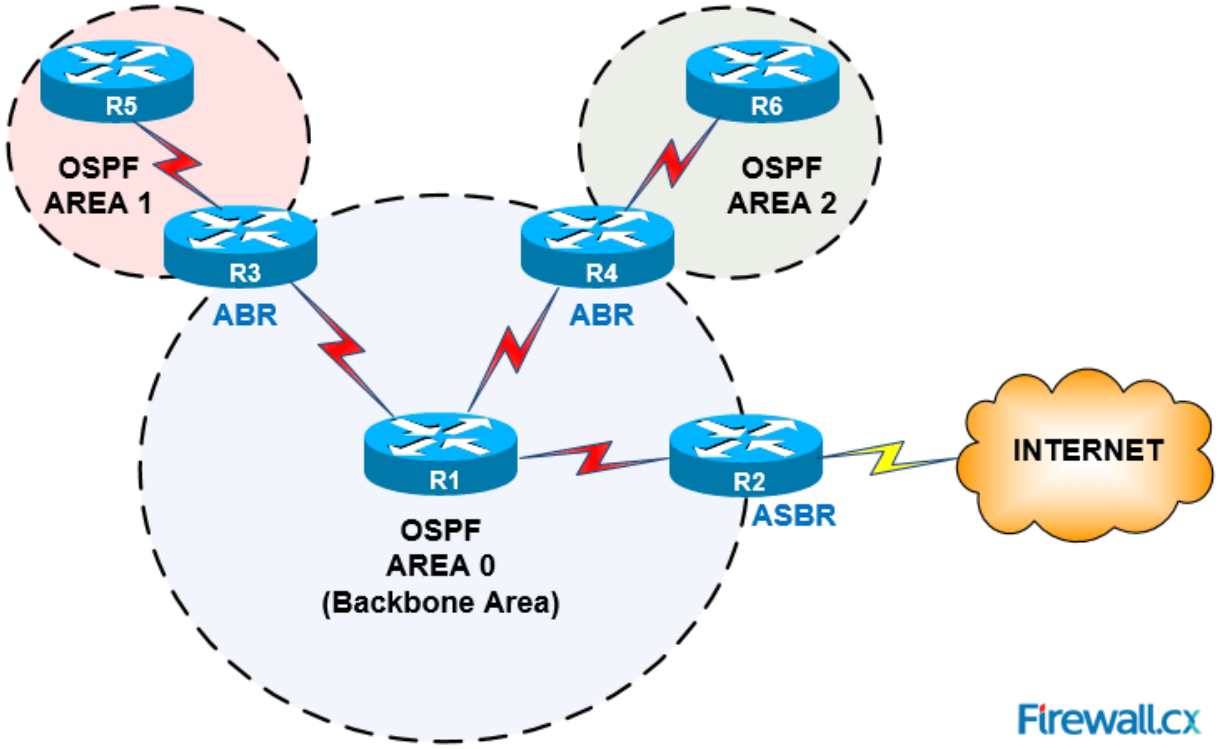

Figure 8: OSPF Areas in an Autonomous System [12]

http://www.firewall.cx/networking-topics/routing/ospf-routing-protocol/1110-ospf-operation-basic-advanced-concepts-ospf-areas-roles-theory-overview.html

As seen in figure 8, every Area has an ABR which is Area Border Router. These routers summarize the IP addresses for each area. There is also ASBR (Autonomous System Boundary Router). This router is connected to the other routing systems like BGP from another routing System which acts like a gateway router.

There are 5 types of link state packets,

Hello: It is used to maintain the neighbor relation with the adjacent routers.

Database Description: It contains all the list of links and this used by the routers receiving in updating the database.

Link State Request: It is used by receiving routers to request any information about a particular router.

Link State Update: It is a reply to the above step which announces new information.

Link State Acknowledgement: It is an acknowledgement sent to LSU.

We can see how routers become neighbors in the following figure,

Figure 9 (a): Router 1 peering Router 2 [9]

Figure 9 (b): Router 2 peering Router 1 [9]

http://www.9tut.com/ospf-routing-protocol-tutorial/2

From the figure 9 (a), we can see there are two routers in the network trying to become neighbors. First, Router R1 sends a Hello message to R2. R2 receives the message and if it has the same network policies as R1, it will send the Hello message from its end. Once they agree to become neighbors they exchange their Link state database with each other. Hence all the routers in the network will have the same Link state database. If there is any change in the network the

information will be sent to the whole network. The convergence time taken by OSPF is very less when compared to the RIP protocol and hence it is faster.

Enhanced Interior Gateway Routing Protocol (EIGRP):

Enhanced Interior Gateway Routing Protocol which is popularly known as EIGRP is the Cisco proprietary protocol. It is called as the hybrid routing protocol as it has both the features of distance-vector and link state routing protocols. It uses Diffusing Update Algorithm (DUAL) to select the best path. It guarantees the loop free paths. EIGRP uses the bandwidth and delay as metrics to choose the best path.

EIGRP sends the hello packets to the adjacent routers to become neighbors. Each hello packet consists of Autonomous System Number (ASN), Subset number and metric components. Each router contains a hold-down timer which is 3 times the hello packet interval. The router is marked as unreachable if the router does not get any response.

Route calculation:

EIGRP routers primarily contains 3 types of tables,

1. Neighbor table: It contains the list of neighbors
2. Topology table: It contains the topology information
3. Routing table: It contains the best path to propagate

We will look into each table information in detail with an example in the next section.

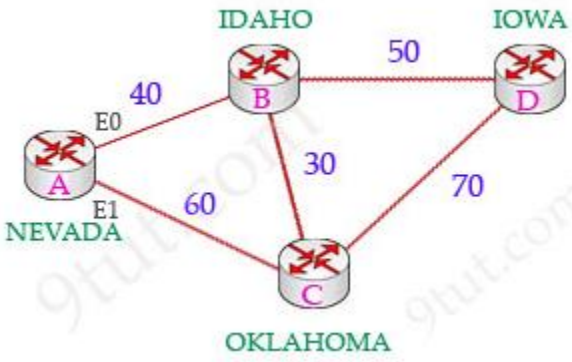

Figure 10: EIGRP routing demonstration [10]

http://www.9tut.com/eigrp-routing-protocol-tutorial/2

Let us consider a network as shown in the above figure 10. We have to travel from Nevada to Iowa. There are two routes how we can propagate to Iowa from Nevada,

Nevada → Idaho → Iowa

Nevada → Oklahoma → Iowa

Nevada has two neighboring routers Idaho and Oklahoma through which it has to travel to Iowa. The distance between Nevada and Iowa is called the Feasible Distance. Feasible distance is the total distance required to reach the destination prefix and this information is stored in the routing table. From figure 10, the feasible distance through Idaho is 40+50 = 90 and feasible distance through Oklahoma is 60+70 = 130.

The distance between Idaho and Iowa is called Reported Distance. Reported distance is the distance from the neighboring router to the destination prefix. In figure 10 the reported distance from Idaho is 50 and reported distance through Oklahoma is 70.

By default, EIGRP selects the route with lowest feasible route. And this path is called as Successor. Successor is the path with best metric to reach the destination prefix which is installed in the routing table for forwarding.

The second path which is Nevada → Oklahoma → Iowa will become the Feasible Successor. Feasible Successor is the backup path which can be used immediately when the best path fails. This path is stored in the topology table. For a path to be a feasible successor, neighbors reported distance should be less than the successor's feasible distance [11].

Below is how the neighbor table in the above topology look like. It will contain the next-hop router and the interface. The neighboring routers for Router A in Nevada are Router B and Router C.

EIGRP Neighbor Table	
Next-hop Router	Interface
Router B	E0
Router C	E1

Figure 11: EIGRP Neighbor Table [10]

http://www.9tut.com/eigrp-routing-protocol-tutorial/2

Below is the topology table which contains all the link distance. The topology table consists of each route's feasible distance and reported or advertised distance.

Topology Table			
Network	Feasible Distance	Advertised Distance	EIGRP Neighbor
IOWA	90	50	IDAHO
IOWA	130	70	OKLAHOMA

Figure 12: EIGRP Topology Table [10]

http://www.9tut.com/eigrp-routing-protocol-tutorial/2

The last table is the Routing table which is configured with the best route in the network to reach the destination. Whenever this route fails EIGRP gets updated with the backup path from the topology table and puts that route into the Routing table.

Routing Table			
Network	Metric (Feasible Distance)	Outbound Interface	Next hop (EIGRP Neighbor)
IOWA	90	50	IDAHO

Figure 13: EIGRP Routing Table [10]

http://www.9tut.com/eigrp-routing-protocol-tutorial/2

Exterior Gateway Routing Protocol:

The Exterior Gateway Protocol handles the routing outside an Autonomous System. The most common EGP protocol used in the Internet is BGP protocol, which is the most efficient Internet routing protocol. It is mainly used to exchange the routing information between the Autonomous Systems (ASes). The routers implementing the EGP peers with the routers in the neighboring Autonomous System, establish a connection and then forwards the routing table. The EGP protocol is complex than as it happens at the Internet Service Provider Level. They are considered to be slow due to their size of the network. The working of this protocol is explained in the border gateway protocol in the next section.

Border Gateway Protocol (BGP):

Border Gateway Protocol is the Internet routing protocol. The transmission is done between different Autonomous Systems (AS) and within Autonomous Systems. It is the widely used Exterior Gateway Routing Protocol. Each AS contains multiple routers and each AS is paired up with the other ASes, up on a signed agreement and shared policies. There exist multiple routes between the source AS and destination AS, but the BGP protocol selects only one best route to propagate the traffic by using the best path algorithm.

There are two different modes in which the BGP operates eBGP (External BGP) and iBGP (Internal BGP). The eBGP exchanges the routing information between different ISPs (Internet Service Providers) and the iBGP exchanges the routing information in the same ISP between the routers. Both these configurations are important as they share the commercial policies with the other ISPs. BGP has always been an interesting topic for the research community mainly due to several concerns related to its convergence, its churn, its limitations in terms of traffic engineering, documented anomalies, and other issues, but the recent large scale outages in the Internet acted as a catalyst for reviving the research focus towards its routing.

BGP does not periodically flood the network with routing information, but it sends messages in an incremental manner. It is a peer-to-peer protocol which distributes the routing information and keeps the routing table up-to-date. The messages or routing updates should be sent only when a new route is discovered or older route is withdrawn. Just like IGP protocols use distance vector and Link state routing algorithm, BGP use Path-vector algorithm for route discovery. Routing Decisions are based on,

1. Path
2. Network Policies
3. Rules.

The BGP has certain policies with the incoming and outgoing AS routes to the Internet. In BGP only single best route is selected to propagate to the destination prefix. The BGP route selection algorithm consists of several steps in which a router compares BGP route attributes and selects the one with following preferences [2];

1. Highest local preference,
2. Lowest AS path,
3. Lowest origin value,
4. Lowest Multi-Exit Discriminator (MED),
5. eBGP route over the iBGP route,
6. Lowest internal cost of the BGP next hop,
7. Lowest BGP ID, and
8. Lowest peer IP address.

Even though there are multiple routes present between the same source and destination, only single route is selected for transmission and the remaining routes are not at all utilized. The BGP Source router has all the information about the multiple paths that are available for transmission. It not only transmits but also advertises only the single best route to all its peers. There are few problems associated with the selection of only one route,

Let us take a scenario where in a network there exits multiple routes, but the traditional BGP selects only one best route. We can see this in the following diagram,

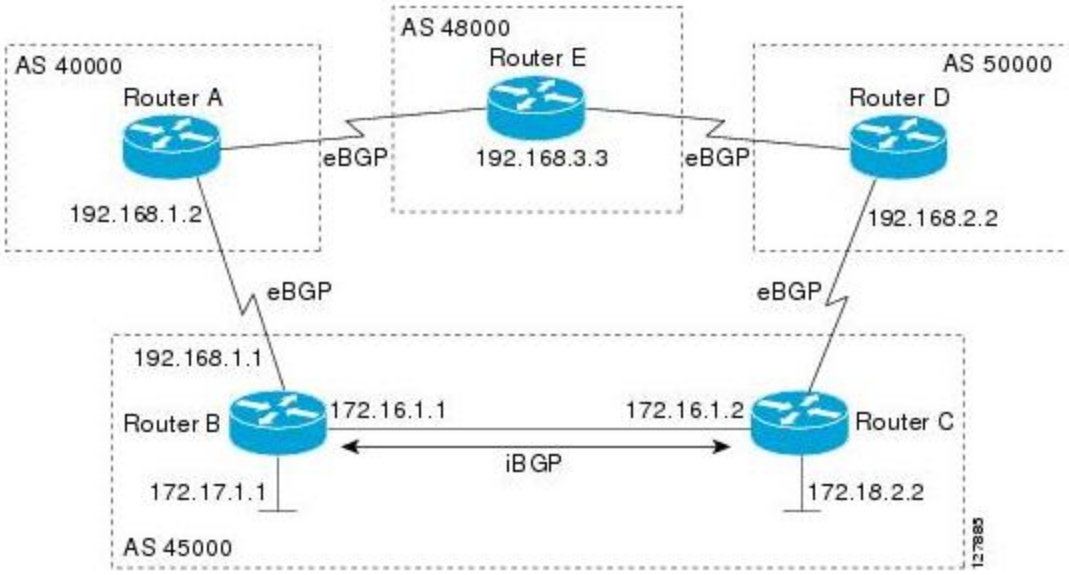

Figure 14: Advertisement of routes in BGP protocol [20]
http://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_bgp/configuration/xe-3s/irg-xe-3s-book/irg-external-sp.html

From figure 14, let us assume that Router A is the destination and Router C is the origin. Router B and Router C belongs to the same Autonomous System, AS 45000 and hence the communication between these two routers would be iBGP. The remaining Autonomous Systems are considered as the peers, since they have a combined agreement and shared policies. There are two routes available in the above network,

1. Router C >>> Router D >>> Router E >>> Router A and
2. Router C >>> Router B >>> Router A.

Based on the route selection criteria in BGP mentioned above, route 2 is selected and also advertised throughout the network. Even though there is another possible way how the traffic can propagate, it is simply ignored. If a link in the route fails, then the route withdrawal advertisement is given to all the routers. Again the route selection process starts and looks for the next best route and then propagates the traffic through the next best route. If this route also fails, then it has to follow the same procedure from the scratch. This failure message is sent via UPDATE message in the network. Let's see what type of messages are propagated in the BGP protocol.

BGP Message Formats

In a network if routers are running BGP they exchange messages to share their routing information and to report any incidents in the network. Incidents can be route change or network failure or configuration failure. BGP has different message formats for each of such incidents. There are basically 4 main types of messages,

1. Open: When a BGP is configured in the router, it sends this message to establish the peering with the neighbor.
2. Update: After peering this update message is sent to share the routing information in between the peers.
3. Notification: This message is sent to notify if there is any problem in the network. Typically to close a connection.
4. Keepalive: This message is exchanged to track the status of the session.

There is another optional message type called Route-Refresh. This capability should be enabled in the BGP implementing router. This message is used to refresh the routes in case of policy changes or any other changes.

Now let's look at each message formats functionality in detail.

BGP OPEN Message:

This is the first message that is exchanged between the BGP peers after the connection gets established. Each BGP peer exchanges this message to introduce itself and its parameters to neighbors. They match their operational parameters and maintain some agreement to make its peering successful. The OPEN message has following parameters as shown in figure 15,

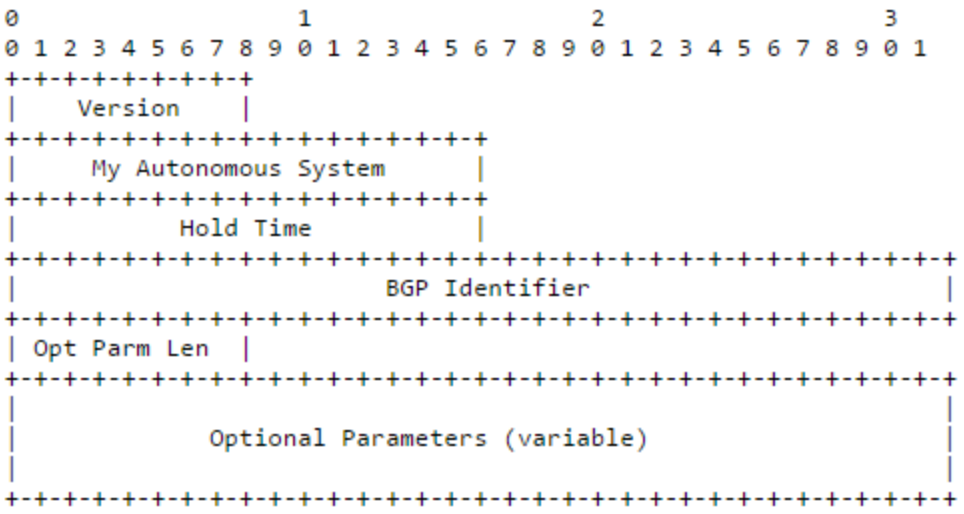

Figure 15: BGP Open Message Format [17]
https://tools.ietf.org/html/rfc4271

Version:

It specifies which version of the BGP is being implemented. The default version of the BGP is 4. The BGP versions should match between the peers. If the versions are different then the higher version BGP speaker can lower its version capabilities to meet the peering policy. Else it can close the connection.

My Autonomous System:

This field tells the neighbor in which Autonomous System (AS) the sender BGP peer is residing. Each AS is assigned with the unique Autonomous System Number (ASN) by IANA. The ASN ranges from 0 – 65535. Hence 'My Autonomous System' field in the OPEN message gives the autonomous system number to the neighbor. Each router in the network announces its ASN to its neighbors by open message.

Hold Time:

The hold time field in the open message denotes the maximum time required to elapse the receipt of successive UPDATE or KEEPALIVE messages. Each BGP speaker must configure their hold time after peering. By default, it is set to zero.

BGP Identifier:

The BGP Identifier is the field which denotes the IP address of that particular router.

Optional Parameters & Length:

Optional parameters can be set upon the agreement. The length of this field is defaulted to zero. This field becomes active if the neighbors agree to share some additional parameters.

UPDATE Message:

The UPDATE messages are the messages which carry routing information among the BGP peers [17]. An UPDATE message has certain information by using which we can construct a graph, which describes the relationship between various Autonomous Systems. This UPDATE message may be used to detect the routing loops and some other defects and then remove them from inter-AS routing.

An UPDATE message has all the information regarding advertising a feasible route to a peer or to withdraw multiple routes from the service [17]. An UPDATE message can both, advertise a feasible path and also withdraw multiple unreliable routes from the service simultaneously. We will look more into the UPDATE message in the Route Advertisement section as it plays a key role in advertising the routes.

KEEPALIVE Message:

BGP uses keepalive message to stop the hold timer to get expired. This message is exchanged between the peers to maintain reachability. The time interval between these messages would be 1/3rd of the hold time interval. [17] If the hold timer is set to zero then there should be no keepalive message sent.

NOTIFICATION Message:

This message is sent when there is any error condition in the network. When this message is received the BGP connection gets closed. We can see its format in the below figure,

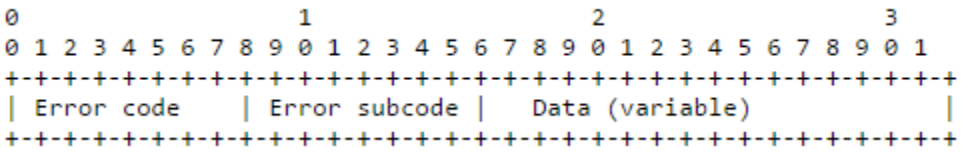

Figure 16: Notification Message Format [17]
https://tools.ietf.org/html/rfc4271

It contains three fields;

1. Error Code
2. Error Subcode

3. Data

Looking at the error code and subcode we can identify where it is going wrong. There are around six types of error codes and each has its their subcodes. Below are the classification of codes and subcodes which tell us about the triggering event;

Error Code 1: Message Header Error

Error Subcode:

1.1. Connection Not Synchronized
1.2. Bad Message Length
1.3. Bad Message Type

Error Code 2: OPEN Message Error

Error Subcode:

2.1. Unsupported Version Number
2.2. Bad Peer AS
2.3. Bad BGP Identifier
2.4. Unsupported Version Number
2.5. Authentication Failure
2.6. Unacceptable Hold Time

Error Code 3: UPDATE Message Error

Error Subcode:

3.1.Malformed Attribute List
3.2.Unrecognized Well-known Attribute
3.3. Attribute Flags Error
3.4. Attribute Length Error
3.5. Attribute Length Error
3.6. Invalid Origin Attribute
3.7. AS Routing Loop
3.8. Invalid Next-hop Attribute
3.9. Optional Attribute Error

We have looked at the BGP message formats. In the same way BGP has its neighbor states through which every BGP session has to pass through. These states tell us what exactly is happening with that particular session. Now let us look at each state in detail by its mechanism called BGP Finite State Mechanism.

BGP Finite State Machine:

There are totally six different BGP states in which we can find a BGP session. They are;

1. Idle State
2. Connect State
3. Active State
4. OpenSent State
5. OpenConfirm State
6. Established State

The following figure 17 shows the flow in BGP Finite State Machine. Whenever there is a BGP session then the first state which appears would be idle state. We will look into each state and their flow in the following section,

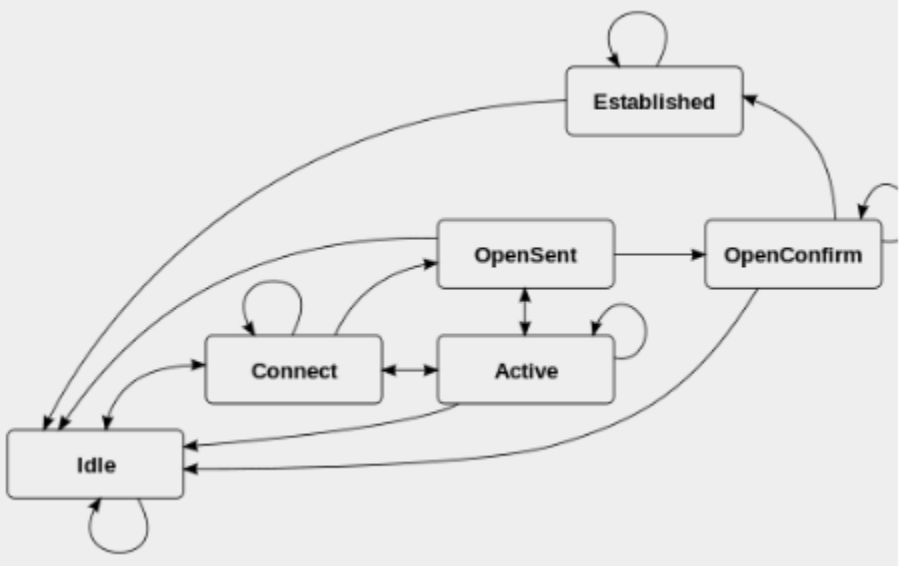

Figure 17: BGP Finite State Machine [27]

http://gponsolution.com/bgp-routing-protocol-overview.html

Idle State:

This the first stage that starts when BGP is configured on to the device. Idle state can also be achieved when an error occurs in any other states. The peers will lose its connection and the

router comes to idle state. BGP is always started with an idle state and proceed to the Connect state for a TCP connection.

Connect State:

In this state, BGP is waiting for the TCP connection to establish between the peers. If the TCP connection is successful, an OPEN message is sent to the neighbor and progresses to the OpenSent state. If the connection doesn't go well then it will go to Active state for reconnection. If there is any problem in the link, then it will go back to the idle state.

Active State:

This state means it is actively looking for the connection to be established. We shouldn't go by the name as it is not yet connected. Any unsuccessful connection will result into Active state. When timer expires in this state it will go back to the Connect state for reconnecting. If the connection is successful without any problems, then it proceeds to next step.

OpenSent State:

This is appeared when the neighbor receives the Open message and waiting for its reply. If it receives the neighbor's Open message, then it goes to the OpenConfirm state. If there are any errors in the neighbor's Open message, then a Notification message is sent to the neighbor and it goes back to the idle state. If there is any connectivity problem, then it goes to the Active state for reconnection.

OpenConfirm State:

In this state, it is simply waiting for a Keepalive message or Notification message. If keepalive message is sent, then it goes to the established state else if something goes wrong then notification message is sent and goes back to the idle state.

Established State:

This state is achieved when BGP session is fully established. It means the BGP peers are able to successfully exchange their updates. If there is any link failure or any error then the same process as above, notification message will be sent which contains error code and it is sent back to idle state.

Route Advertisement in BGP:

The route advertisement in the BGP is carried by using an UPDATE message. It plays a vital role in carrying the routing information to the peers in the network. The UPDATE message in the network carries the Network Layer Reachability Information and all the routing information. It can contain both, the route to be advertised in the network and the unfeasible routes to be withdrawn. Let us look at each module in the UPDATE message which is present in the BGP protocol. In the BGP, an UPDATE message can contain only one single route to a particular destination. If a peer receives an UPDATE message from its neighboring peer, then it implicitly replaces the new route with the existing one. At a time only one route can be advertised.

An UPDATE message has all the information regarding advertising a feasible route to a peer or to withdraw multiple routes from the service [17]. An UPDATE message can both, advertise a feasible path and also withdraw multiple unreliable routes from the service simultaneously. The UPDATE message contains the fixed-size BGP header and the other fields as shown in figure 19,

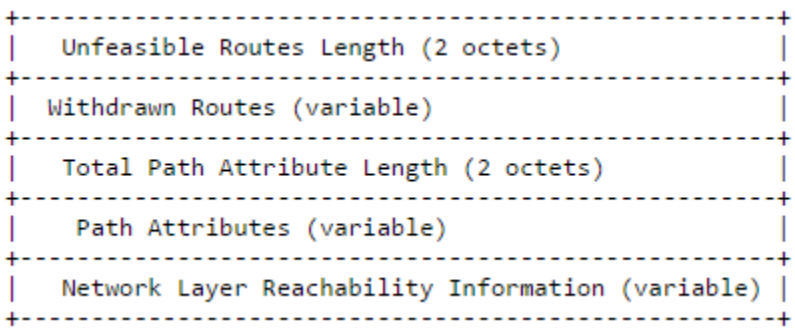

Figure 18: Fields in UPDATE message [17]
https://tools.ietf.org/html/rfc4271

Unfeasible Routes Length:

The first field in the UPDATE message is Unfeasible Route Length. This field indicates the withdrawn routes and total length in octets. The value of this field is determined in the Network Layer Reachability Information (NLRI) field. If there are not any WITHDRAWN ROUTES present in the UPDATE message, then the value in the NLRI field is assigned to '0' and there won't be any WITHDRAWN ROUTES field in the UPDATE message.

Withdrawn Routes:

This a variable length field which contains a list of number of IP addresses that are to be withdrawn from the service. Every IP address prefix again contains two fields which are <length, prefix>,

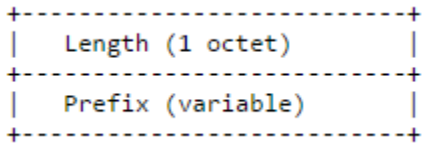

Figure 19: IP address prefix fields [17]
https://tools.ietf.org/html/rfc4271

a) Length:
   The length of the IP address prefix length field is denoted in bits. A length of zero indicates a prefix which matches all the IP addresses [17].
b) Prefix:
   The prefix field contains IP address prefixes which needs to be discarded [17].

Total Path Attribute Length:

The value of the Total Path Attribute Length field is related to NLRI field. If a value '0' is present for the Total Path Attribute Length field, then there won't be any NLRI field present in the UPDATE message. This means that there is no new route that needs to be added to the memory buffer.

Path Attributes:

This is a mandatory field in an UPDATE message. Each Path Attribute consists of three main parts <attribute type, attribute length, attribute value> which is of variable length.

The attribute type field in the Path Attribute consists of two parameters which are Attribute Flags and Attribute Type Code. It is shown in the figure 21,

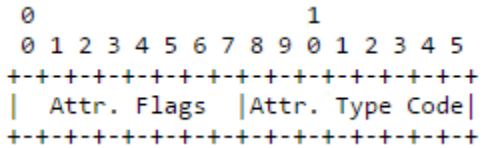

Figure 20: Fields in Attribute Type in Path Attribute [17]
https://tools.ietf.org/html/rfc4271

The first higher-order bit, bit 0 in the Attribute Flag field denotes an Optional bit. If the value is set to 1 then it marked as optional or if it is set to 0 then it is marked as well-known.

The second higher-order bit, bit 1 in the Attribute Flag field denotes a Transitive bit. If the value is set to 1 then the Optional Attribute is said to be Transitive and if the value is set to 0, then the Optional Attribute is said to be Non-Transitive.

The third higher-order bit, bit 2 in the Attribute Flags octet is denoted as a Partial bit. If the value is set to 1 then the Optional Transitive Attribute is said to be Partial and of the value is set to 0 then it is considered as Complete. The Optional Non-Transitive attributes and well-known attributes partial bit value should be 0 [17].

The fourth higher-order bit, bit 3 in the Attribute Flags octet is denoted as Extended Length bit. If the length of the value of the attribute exceeds 255 octets, then only the extended length attribute is used [17].

The lower-order four bits in Attribute Flag should be assigned a value 0 and must be ignored when they are received.

Network Layer Reachability Information (NLRI):

NLRI field consists of list of all IP address prefixes in the network. It is mainly exchanged between the BGP routers by using UPDATE message. It contains all the path information. It contains two fields which are length and prefix. The below diagram shows their representation,

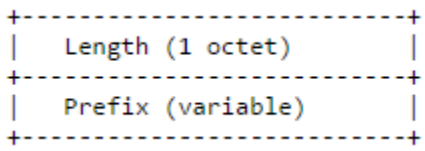

Figure 21: Fields in Network Layer Reachability Information Attribute [17]
https://tools.ietf.org/html/rfc4271

By using this NLRI information only one path can be advertised according to the above architecture. Even though there are multiple paths present in the network, the BGP policies compliance to advertise only the best path to all its peers. If a route is advertised to the neighboring peer with the same NLRI information as the existing route, then the existing route is replaced by the new one. There will be only one route present at a time. Hence here we are making an enhancement to the protocol. There is a research going on to advertise multiple paths to the same destination, but it is in still progress.

BGP Path Selection Mechanism:

The Border Gateway Protocol (BGP) is a protocol that exchanges the routing information among the routers for the traffic flow through the available optimal paths. A BGP router connects to its peers and come to a shared agreement and exchange their routing information. BGP uses path vector algorithm to choose its best path. Unlike the IGP protocols BGP does not look at one particular metric but instead it has some path attributes to look into and decide the best path. It will look at the whole path and direction (next-hop) through which it is travelling

The BGP routers receive multiple routes to the same destination prefix. But the data transfer is being done by only one route which is called the best Path. There is an algorithm called best path selection Algorithm based on which this best route is selected. This algorithm decides the best path that has to be installed in the IP routing table, which is used to forward the traffic. In the present day Internet, the BGP routing tables include more than 40,000 routes which comes from different ISP providers. BGP has to compare all the routing tables and has to choose the best route on the router.

In order to select the best path among the available paths, it has a two-step process. First step is to find and reject the unfeasible paths or paths which does not comply with the local policy. Let us look at the following scenarios which shows what type of paths get discarded,

1. The paths without any access to the NEXT_HOP.
2. The paths marked with 'Not Synchronized'.
3. Paths shown by EBGP neighbor, when the local AS originates the same path.
4. The paths marked as 'Received-Only'.
5. If the neighboring AS number is not listed first in the AS-SEQUENCE, when 'enforce-first-as' command is enabled.

Second step is to find the best path. For the best path selection, BGP uses path attributes to compare between the available paths. The selected best path will be stored in routing information base – out (adj-rib-out) forwarding table. This route will be advertised to the neighboring peers. Let us look at the path attributes in detail to understand how a best route is selected.

Path Attributes are classified into four categories;

1. Well-known mandatory
2. Well-known discretionary
3. Optional transitive
4. Optional non-transitive

Every UPDATE message should contain the well-known attributes if it contains NLRI information. These well-known attributes include origin, as_path, next_hop and well-known discretionary are local_pref and atomic_aggregate. multi_exit_disc and router_id are optional non-transitive attributes. Aggregator and community are optional transitive path attributes. There are few other optional attributes like weight which have high priority but are used only at their proprietary devices. The optional attributes are set by the administrator depending upon the peering policy. Let us look at each path attribute in detail.

Preferences in Route Selection:

The list of preference list is given below. BGP selects the best path based on the following criteria and injects that into the routing table.

1. **Weight:**
   Weight is the primary criteria based on which the router selects the route. It is a Cisco proprietary attribute which is installed in Cisco devices. It is not a mandatory attribute. But if it exists then it has the high priority than other attributes. The BGP selects a route which has the highest weight i.e., if there are multiple routes available for the route propagation, BGP selects a route with the highest weight assigned to the router. The weight parameter is set through route maps, neighbor command or via AS-Path Access list.

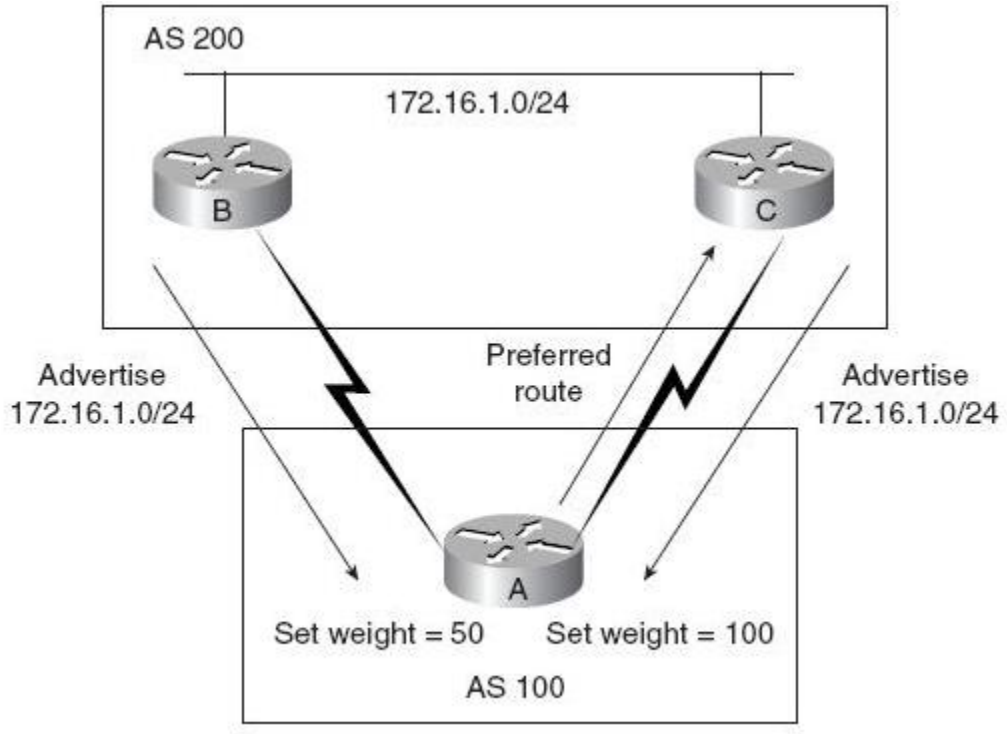

Figure 22: Weight Attribute [13]
http://routemyworld.com/wp-content/uploads/2008/12/bgpweightattribute.jpg

In figure 15, Router A receives a route advertisement for the network 172.16.1.0/24 from both routers B and C. But each router has their own Set weight, like router B has weight of 50 and router C has weight of 100. Both the paths advertised, are for the same network 172.16.1.0. Hence as per the guidelines, the BGP selects the route with highest weight and installs it into the routing table.

2. **Local Preference:**

The next criteria which BGP looks into is Local Preference. If the two routes have the same weight, then the BGP takes a route with highest local preference. If a local preference is not assigned any value then the default value a network assigns is 100.

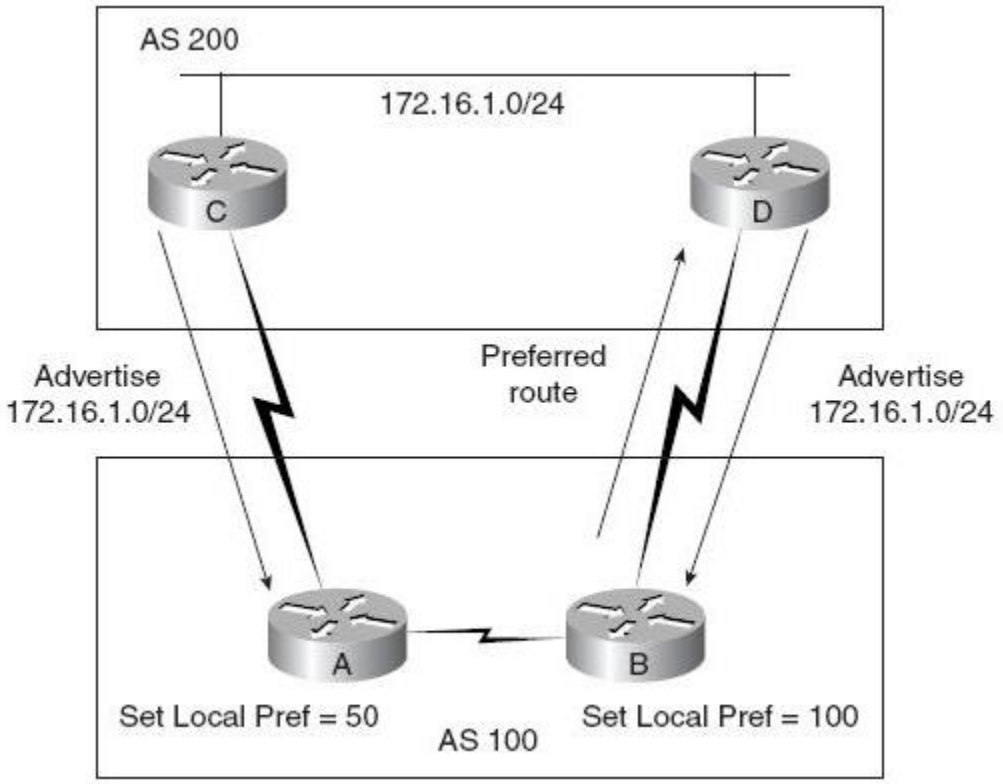

Figure 23: Local Preference Attribute [14]
http://routemyworld.com/wp-content/uploads/2008/12/bgplocal_prefattribute.jpg

In figure 16, we can there are two Autonomous Systems AS 200 and AS 100. AS 200 advertises the route for the network 172.16.1.0/24 to AS 100 through two different routers, Routers C and D. Router A in AS 100 receives the route through C and sets its

LOCAL_PREF value as 50 and Router B receives a route from D for the same network and sets its LOCAL_PREF value as 100. The Local Preference values are set according to the policies between the two autonomous systems. As per the algorithm, the BGP selects the route with highest LOCAL_PREF value. Hence the route which is advertised by Router D in AS 200 is preferred over Router C's route from the same Autonomous System (AS 200).

3. **AS Path Attribute:**

The Autonomous System prepends its AS number into the IP routing table, whenever the route passes through it. These AS numbers are assigned to the Autonomous System by IANA (Internet Assigned Numbers Authority). The AS Path attribute is nothing but it records all the AS numbers through which the route has been traversed. Hence the BGP prefers the route with less number of AS traversal. We can see this in below example.

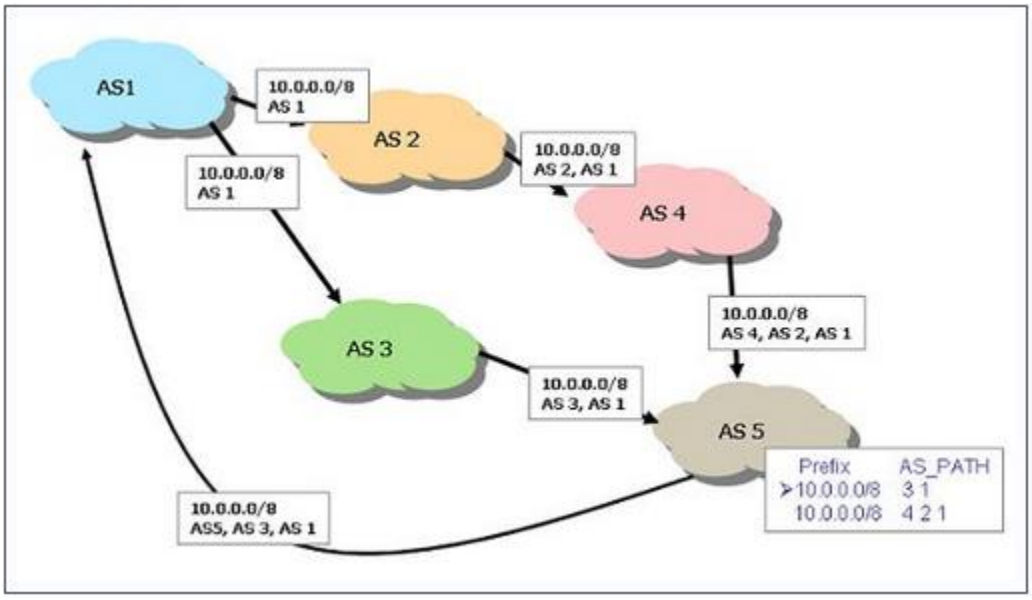

Figure 24: The AS-Path Attribute [15]
http://routemyworld.com/category/routing-protocols/bgp/

From figure 17, we can see AS1 is the origin of the network 10.0.0.0/8 and advertises it to its neighbors AS2 and AS3. Hence in the path vector AS1 adds its own AS_PATH (<AS1>). In the same way AS2 and AS3 advertises the same network to its neighboring Autonomous Systems by prepending their AS_PATH values. The path vector from AS2 will look like <AS2, AS1> and from AS3 it looks like <AS3, AS1>.

Now AS4 receives a route from AS2 and it advertises to its neighbor AS5. In the same way AS5 also receives a route advertisement from AS3. Hence the path vectors from both the Autonomous Systems might look like (<AS3, AS1>) and (<AS4, AS2, AS1>). AS5 will certainly select the shortest path to reach the network 10.0.0.0/8. Hence it selects the route coming from AS3 with path vector <AS3, AS1> and adds its AS number (AS5) to the same table.

4. **Origin value:**
   The Origin value is nothing but the origination of the network. It can be IBGP or EBGP or Incomplete. If a network is originated with in an AS then it is preferred over the network coming from another AS, which is nothing but EBGP. And EBGP is preferred over incomplete origin value. Incomplete here, refers to network originated from unknown source. Hence the BGP prefers the route with lowest Origin Value.

5. **Lowest Multi-Exit Discriminator (MED):**
   The Multi-Exit Discriminator attribute is opposite to the local preference attribute, as it deals with the traffic leaving from AS and MED deals with traffic entering into the AS. The value of the MED is set to zero by default. This attribute is considered in only Inter domain routing i.e., between different ASes but not with in the same AS. The MED attribute is considered only for the Exterior BGP updates. It is exchanged only between directly connected ASes.

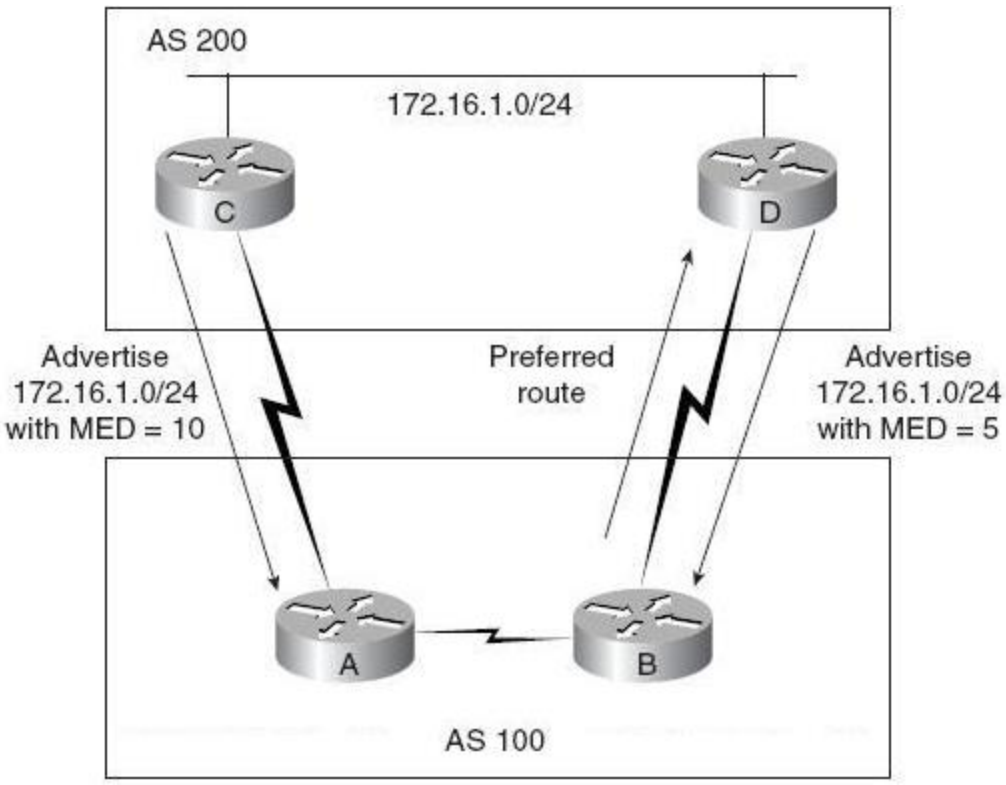

Figure 25: Multi-Exit Discriminator Attribute [16]
http://routemyworld.com/wp-content/uploads/2008/12/bgp-medattribute1.jpg

In figure 18, we can see that AS 200 is advertising the network 172.16.1.0/24 to the Autonomous System AS 100. AS 100 receives the advertisement through two routes, one from router C and another from router D. Router sends the network information to router A in AS 100 and router D sends the information to router B in AS 100. As there will be communication between the routers in the same Autonomous System, the routers in the AS 100 will exchange the route information which they got it from AS 200. The BGP will prefer the route with lowest MED value and hence it selects the route coming from router D in AS 200 to router B in AS 100.

6. **eBGP route over the iBGP route:**
   If all the above conditions are same between the routes then the BGP will look into this attribute. If a route is coming from another AS (EBGP), then it is preferred over the route coming from the same AS (IBGP). It checks whether the route is originated in the same AS or in the different AS. If it is not originated in the same AS then it prefers EBGP (route coming from another AS).

7. **BGP next hop attribute:**
   The BGP next hop attribute comes into play when all the above attributes are same between the routes. The network routing table consists of the next hop attribute which

specifies the next available router information. The BGP selects the next hop based on the cost (router resources). It selects the route which has the low cost next hop.

8. **Lowest BGP ID:**
   When all the above attributes cannot specify the best path then the BGP looks into the BGP ID attribute. Every BGP implementing Router is assigned with an ID and it is called BGP ID. According to the BGP route selection algorithm, the BGP selects a router which has the lowest BGP ID to traverse the route.

9. **Lowest peer IP address:**
   Every router has an IP address that is associated with it. The BGP prefers the route comes from a lowest neighboring IP address. This IP address is used in the BGP neighbor configuration. The IP address corresponds to a router which used in the TCP connection.

Challenges in Existing BGP:

Bandwidth Underutilization:

Even though there are multiple routes between the source and destination prefix, only one route will be advertised for the routing purpose and this route is called as the 'best route'. The best route is selected based on various attributes which were discussed earlier. Everything is fine until the best route fails. If there is any link failure or disconnection in the route, the network again builds the new path for the transmission. The BGP has to construct the new path from the scratch even though it has multiple paths available. It is because the multiple paths were not advertised to the peers. The BGP protocol is not able to capitalize with the available bandwidth. It can be shown in the below example,

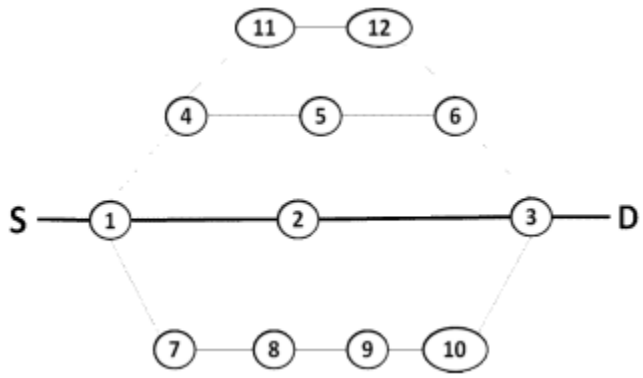

Figure 26: Default BGP path and Multiple BGP paths [3]

http://delivery.acm.org.ezproxy.sunyit.edu/10.1145/1750000/1741974/p304-ragha.pdf?ip=149.15.108.152&id=1741974&acc=ACTIVE%20SERVICE&key=7777116298C9657D%2E5234849A804BC5E3%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=698264581&CFTOKEN=44306304&__acm__=1438321824_1bc0c5d73a3e3db7d60e8a323f8ceee1

From figure 23, we can see there is a source host S and a destination host D and few intermediate routers. There are multiple routes between the source host S and destination host D,

S → R1 → R2 → R3 → D,
S → R1 → R7 → R8 → R9 → R10 → R3 → D,
S → R1 → R4 → R5 → R6 → R3 → D,
S → R1 → R4 → R11 → R12 → R6 → R3 → D.

Out of these available paths, the BGP selects the best path as the one with lowest hop count which is shown in dark line in the figure 23 (S → R1 → R2 → R3 → D). The same route is broadcasted all over the network leaving other routes behind. Hence the other peers don't know about the remaining routes as only best path is advertised. Hence when the best path fails, it has to again construct a new path even though there exist multiple paths. Hence the residual bandwidth available for these paths is not used properly [3].

Delay in Convergence:

If there is any link failure in the selected best path, the path failure is notified to all the participating peers in the network. Again the path selection process starts from the scratch. After all the routes available from the network, BGP applies the best path algorithm and chooses its best path and advertises the new best path again to the peers. During this interval the transmission of the data has to be stopped till the new link is formed.

Even though there are multiple paths in the network, the transmission is interrupted whenever the best path fails. There is no guarantee that the next best path selected will function without any failure. If the second best path also fails, then again it has to construct the new best path. As we know link failures are quite common in a network due to the continuous topology changes and the complexity in the network structure. The convergence time taken by the BGP protocol to continue the transmission after the best path failure is very high. Hence there will be transmission delay in the network.

Future Recommendations:

BGPsec:

The Internet Engineering Task Force (IETF) has put their focus on the enhancement of BGP which is called BGPsec which introduces the new security features in routing. It is an extension to the existing BGP which introduces the Resource Public Key Infrastructure (RPKI). It is mainly used for maintaining integrity between the Autonomous Systems with cryptographic certificates. This enhancement is put forth to overcome the BGP Hijacking attack. Due to the size of the protocol, it will be easy for the attacker to introduce a false IP-prefix and reroute the data in the favor of attacker. The IETF group are testing with different scenarios where they can identify this false advertisement by introducing this new feature.

Multi path BGP:

Another enhancement that IETF focus group are planning to enhance the existing BGP is introducing Multi path BGP. In this enhancement they are introducing a new capability in the UPDATE message called ADD_PATH capability which enables the router to advertise multiple equal cost paths to the neighboring peers. [2] These paths are given a unique identifier called Path Identifier which is introduced in the Network Layer Reachability Information field of the UPDATE message. Hence when an UPDATE message is received by the neighboring router with multiple paths to the same destination prefix, then it can store and process all the paths by giving priority by using the path identifier. This is in still testing phase.

Conclusion:

The Border Gateway Protocol is the most efficient Exterior Gateway protocol. It deals with the routing between the Autonomous Systems at the Service provider level. The routing with in the Autonomous System is done by Interior Gateway Protocols like RIP, OSPF and EIGRP. All these IGP protocols have some metrics to calculate the best path. Unlike the IGP protocols which have metrics to choose the best path, BGP has path attributes to choose the best path. It is the widely used Internet routing protocol. The network administrators handling the networks can fine tune the path attributes according to the network. As it deals with the huge network like Internet, it is little slow. There is a lot of research conducted by the Network Research Community and other big companies like Cisco to enhance the protocol like Multi path BGP, providing QOS enhancements in the route propagation, etc. It would be an interesting topic in the future.

References:

1. P. Brighten Godfrey, Mathew Caesar, Ian Haken, Yaron Singer, Scott Shenker and Ian Stoica "Stabilizing Route Selection in BGP", IEEE/ACM Transactions on Networking, VOL. 23, NO. 1, February 2015.
2. Aleksandra Cvjetic and Aleksandra Smiljanic, "Improving BGP Protocol to Advertise Multiple Routes for the Same Destination Prefix", IEEE Communications Letters, VOL. 18, NO. 1, January 2014.
3. L L Ragha, K V Ghag, "Multiple Route Selector BGP (MSR-BGP)", ICWET'10, February 26–27, 2010, Mumbai, Maharashtra, India.
4. BGP Best Path Selection Algorithm, http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html.
5. BGP Case Studies, http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/26634-bgp-toc.html#weight.
6. Cisco – "Border Gateway Protocol", http://www.cisco.com/c/en/us/products/ios-nx-os-software/border-gateway-protocol-bgp/index.html
7. D. Walton Cumulus Networks, A. Retana, E. Chen Cisco Systems, INC., J. Scudder Juniper Networks, "Advertisement of multiple paths in BGP", Network Working Group, Internet-Draft: draft-ietf-ide-add-paths-10, October, 2014.
8. CCNA Training, "RIP Routing Protocol", http://www.9tut.com/rip-routing-protocol-tutorial
9. CCNA Training, "OSPF Routing Protocol", http://www.9tut.com/ospf-routing-protocol-tutorial/2
10. CCNA Training, "EIGRP Routing Protocol", http://www.9tut.com/eigrp-routing-protocol-tutorial/2
11. Study CCNA, "EIGRP", http://study-ccna.com/eigrp-overview/
12. Firewall.cx, "OSPF Routing Protocol", http://www.firewall.cx/networking-topics/routing/ospf-routing-protocol/1110-ospf-operation-basic-advanced-concepts-ospf-areas-roles-theory-overview.html
13. Route My World, "BGP Weight Attribute", http://routemyworld.com/wp-content/uploads/2008/12/bgpweightattribute.jpg
14. Route My World, "BGP Local_Pref Atrribute", http://routemyworld.com/wp-content/uploads/2008/12/bgplocal_prefattribute.jpg
15. Route My World, "Routing Protocols", http://routemyworld.com/category/routing-protocols/bgp/

16. Route My World, “BGP MED Attribute”, http://routemyworld.com/wp-content/uploads/2008/12/bgp-medattribute1.jpg
17. Y. Rekhter T.J. Watson Research Center, IBM Corp., T. Li Cisco Systems, “A Border Gateway Protocol 4 (BGP-4)”, Network Working Group, Request for Comments: 4271, Obseletes: 1771, Category: Standards Track, January 2006
18. J. Uttaro AT&T, P. Francois IMDEA Networks, K. Patel Cisco Systems, P. Mohapatra Cumulus Networks, J. Haas Juniper Networks, A. Simpson, R. Fragassi Alcatel-Lucent, “Best practices for advertisement of multiple paths in ibgp”, Internet Draft: draft-ietf-idr-add-paths-guidelines-07.txt, Dec 3, 2014.
19. Routing Bits, Filling the Gaps, “Output 101: BGP AFI/ SAFI , http://routing-bits.com/2009/11/26/output-101-bgp-afisafi/
20. IP Routing: BGP configuration Guide, Cisco IOS XE Release 3S, http://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_bgp/configuration/xe-3s/irg-xe-3s-book/irg-external-sp.html
21. C. Hedrick, Rutgers University (1988), “Routing Information Protocol”, Network Working Group, Request for Comments: 1058, June 1988.
22. J. Moy, Proteon, Inc., (1991), “OSPF Version 2”, Network Working Group, Request for Comments: 1247
23. D. Savage, D. Slice, J. Ng, S. Moore, R. White Cisco Systems (2013), “Enhanced Interior Gateway Routing Protocol”, IETF Internet Draft.
24. Microsoft, Technet, “Autonomous Systems”, https://technet.microsoft.com/en-us/library/Cc957842.aspx
25. Tech Tuts Online, “Dynamic Routing”, http://www.techtutsonline.com/dynamic-routing/
26. Learn Cisconet RSS, “Dynamic Routing and Routing Protocols”, http://www.learncisco.net/courses/icnd-1/ip-routing-technologies/dynamic-routing.html
27. GPON Solution, “BGP Routing Protocol Overview”, http://gponsolution.com/bgp-routing-protocol-overview.html