Recent Developments in BGP-4

Jarkko Ilomäki
Helsinki University of Technology
Telecommunications Software and Multimedia Laboratory
jarkko.ilomaki@hut.fi

Abstract

The main inter-domain routing protocol, Border Gateway Protocol version 4 (BGP-4) has been a part of the Internet architecture for many years. During these years the protocol has faced new requirements and new problems. Therefore many smaller changes, larger developments, and security additions have been included to enhance the original functionality. This paper describes some of these developments and analyses them. The presented concepts vary from minor functional additions like Capability Advertisement to more radical changes like Autonomous System Confederation and Multiprotocol Extensions. Security additions like the use of MD5 authentication and IPSec are also discussed. The developments presented in this paper have corrected the serious problems of the original protocol and have made it more flexible, manageable and secure. BGP-4 is likely to remain in use for quite some time.

KEYWORDS: Border Gateway Protocol, BGP-4, Inter-domain routing

1 Introduction

The Internet is a network of networks. Although a single network can use any routing algorithm and architecture within itself, there has to be a common set of rules how to exchange data between networks. This problem area can be called inter-domain routing and in this area, the Border Gateway Protocol version 4 stands nearly unchallenged.

The original BGP-4 has been in use for years so most likely some developments and new abilities have been added, or at least have been proposed to it. This paper reviews, what kind of developments and improvements have been made to the original BGP-4 protocol specification. As the concept itself is huge, the study is mostly restricted to Internet Engineering Task Force's (IETF) work on BGP-4 protocol development. The development work of IETF has brought both small additions and larger conceptual changes. Both areas are reviewed in separate sections.

The only exception where more extensive list of sources is used is when discussing the security developments of BGP-4. As the Internet keeps on growing, there are people who would like to harm the current system. BGP-4 is an inviting target due to its popularity and importance. As Murphy states [12], there are simple methods that can be used to disrupt local and overall network behaviour.

This paper discusses specifically to BGP version 4, unless otherwise stated. The discussion is organized as follows. In Section 2 the background and functionality of the original BGP-4 are described in brief. Minor functional developments, larger conceptual changes and security additions are described in Sections 3, 4, and 5.

In Section 6 the future directions of the development work and ideas of BGP-4 are considered. Finally in Section 7 some conclusions about the development work and BGP itself are made.

For every development step presented in this paper, there are two parts: description and analysis. The description part tells what and why the the development does and the analysis part consists of experts' and author's own opinions. The classification is made by the author to structurize the developments and is purely unofficial.

2 BGP-4 Background

2.1 History

Originally the functionality of the Internet was based on core routers which connected different networks together. This structure meant that the core routers had to contain complete information about all the networks. The core router-based approach worked well in the beginning. However, as the Internet started to grow, it soon became clear that the core routers would not be able to handle the load. [4]

The concept of Autonomous System (AS) was developed to divide the Internet into smaller parts. An autonomous system is a single administrative authority, which in a nutshell contain a set of networks and routers. Single autonomous system is identified by its AS number, which are assigned by the Internet Assigned Numbers Authority (IANA). The architecture and routing mechanisms within an autonomous system can be chosen freely. For example, an organization which has connections to multiple Internet providers could be an autonomous system. [4]

The communication between two ASs require a common protocol, which is often called as inter-domain routing protocol. The first inter-domain routing protocol was Exterior Gateway Protocol (EGP) and its specification was defined in RFC 827. [16] The EGP can be seen as the basis for the Border Gateway Protocol (BGP). BGP itself has evolved into four different versions and the fourth version is the one most used today. As the other versions have not been successful, the term BGP most commonly refers to BGP-4.

2.2 Functionality

As stated before, the main function of BGP is to exchange network reachability information between BGP speaking systems. In this section, this functionality is briefly discussed to provide a starting point to the actual development discussion. In the following sections, the changes and additions to the original specification are presented and commented.

As stated before, the communication between autonomous systems is handled with BGP. When an autonomous system has agreed to exchange routing information with some other AS, it first has to select a router that will speak BGP on its behalf. This router is usually selected from the "edge" of the autonomous system's network. Therefore the router is often called as a border gateway or as a border router. When the border routers of different ASs communicate with each other, these routers are referred as BGP peers. These peers "speak" BGP to each other. The Figure 1 illustrates this architecture. BGP does not make any assumptions about intra-autonomous system routing protocols deployed within the autonomous systems.[4, 14]

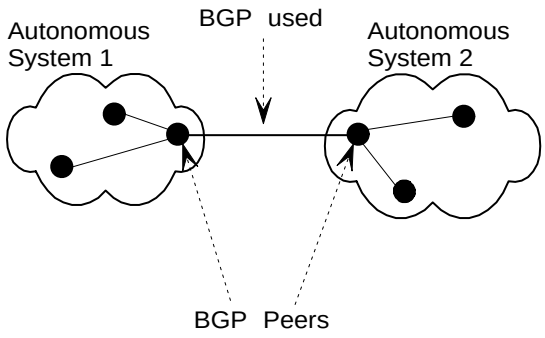

Figure 1: The illustration of BGP usage

BGP peers send small "keep-alive" messages to each other. If a neighbour stops receiving keep-alive messages for a predefined "hold time," it will update its routing table to reflect the loss in available routes. BGP also sends incremental updates when routes become unavailable. Otherwise, full routing tables are exchanged only when two routers first establish or re-establish a peering relationship. If a BGP speaker goes down, all the routes which peers have established with it are removed. [4]

BGP has many properties that make it unusual. First of all, BGP is neither a pure distance-vector protocol nor a pure link state protocol. Instead BGP uses a modified distance vector algorithm referred to as a "Path Vector" algorithm that uses path information to avoid traditional distance vector problems. [4, 11] For example, BGP supplies next hop information for each destination, like distance vector protocols do. However, BGP allows AS to implement policies, which is not common to distance vector protocols.

Another thing that makes BGP unusual is that it uses a reliable transport protocol, the Transmission Control Protocol (TCP), to exchange routing data [14]. Usually routing protocols use unreliable transport level protocols like User Datagram Protocol (UDP) for data transport [4]. According to the BGP-4 protocol specification [14], this eliminates the need to implement explicit update fragmentation, retransmission, acknowledgement, error handling and sequencing. Any authentication scheme used by the transport protocol may be used in addition to BGP's own authentication mechanisms.

A brief summary of the different messages of BGP-4 are presented in the following list. This information can be used when comparing the developments to the original specification. Especially the behaviour of UPDATE and NOTIFICATION messages is important to understand, because they form a crucial part of the BGP system.

- OPEN - When two BGP peers establish a connection, they send an OPEN message to each other to declare their autonomous system number and establish operating parameters. OPEN message contains also a *hold timer* which tells the maximum number of seconds which may elapse between the receipt of two successive messages.
- UPDATE - BGP peers use UPDATE messages to inform each other when new destinations are available or when some destination is not available any more.
- KEEPALIVE - Two BGP peers exchange KEEPALIVE messages to test the network connection and to verify that both peers still are operational.
- NOTIFICATION - BGP has a support for NOTIFICATION message which is used for some control action or when an error occurs. Errors are permanent - once an error has been detected, BGP sends NOTIFICATION message and then closes the TCP connection.

3 Minor functional changes to BGP-4

In this section the minor functional changes and additions to the original BGP-4 definition are reviewed. This paper represents minor functional changes as changes that add some property or ability to the BGP architecture, but otherwise leave the original specification untouched.

First, a solution to BGP's limiting approach to new additions called "Capabilities Advertisement" is discussed in section 3.1. Next, an improvement "Route Refresh Capability" which allows BGP to refresh its routing data and reduces the overhead of implementing policy changes is discussed in section 3.2.

The length of the autonomous system number and extension to it is discussed in section 3.3. For situations when a BGP speaker must be restarted, the development "Graceful Restart Mechanism" aims to minimize the drawbacks and is presented in section 3.4.

3.1 Capabilities Advertisement with BGP-4

3.1.1 Description

According the current BGP specification [14], when a BGP speaker receives an OPEN message with one or more unrecognized Optional Parameters, the speaker must terminate BGP-4 peering. This kind of behaviour complicates the introduction of new capabilities in BGP-4.

To solve this problem, Chandra and Scudder have developed a new Optional Parameter, called Capabilities. This optional parameter is expected to facilitate the introduction of new capabilities in BGP-4 by providing graceful capability advertisement without requiring that BGP peering be terminated. [2]

The functionality of the capabilities advertisement is not very complex. When a BGP speaker that supports capabilities advertisement sends an OPEN message to its BGP peer, the message may contain the Capabilities parameter. The parameter lists all the capabilities supported by the speaker. If the BGP peer supports the proposed capabilities, the proposals can be taken into use. However, if the BGP peer does not support capabilities advertisement, an NOTIFICATION message is returned to the BGP speaker and the connection is terminated. In this case, the BGP speaker should reattempt to establish connection with its peer without sending the Capabilities Optional Parameter. [2]

The delegation of Capabilities identifiers is centralized. If one has defined an extension to BGP which takes advantage of Capabilities Advertisement, a new unique number must be requested from IANA. [2]

3.1.2 Analysis

Meyer and Patel state [11], that the Capabilities Advertisement provides an easy and flexible way to introduce new features within the protocol. In particular, the BGP capability mechanism allows peers to negotiate various optional features during start-up. This allows the base BGP protocol to contain only essential functionality, while at the same time providing a flexible mechanism for signalling protocol extensions.

In a nutshell, the capabilities advertisement offers both modularity and compatibility. If some BGP router does not support this feature, original specification based communication is still possible [2]. If some new addition to the BGP architecture is designed, a new capability can be taken into use. This capability will then be a signal for more sophisticated features and data processing.

3.2 Route Refresh Capability

3.2.1 Description

Initially the original BGP specification did not include any mechanism to reset or refresh a BGP peering session without tearing it down and waiting for it to re-establish. It is likely, that the authority of AS wishes to implement some new or revised policy, and do to this, he has to reset the BGP session. Destroying a BGP session carrying a large or the full routing table has severe impact to the AS and its neighbours on the Internet. This impact is caused by the fact that an AS usually has multiple BGP speakers and the reset of one causes the traffic to be rerouted to the "surviving" servers and might overload the network. In a network of one BGP router, the reset of a BGP session will seem like a "disappearance" of reachability information and users in such network will have the impression that the Internet has become unavailable. [14, 3]

One solution to the problem presented above would be to use "soft-reconfiguration". Soft-reconfiguration is an approach in which an unmodified copy of all routes is stored at all times. BGP speaker uses this stored information as a base to which it applies the routing policies. The problem of this solution is that the BGP speaker must store the unmodified copy at all times, even though routing policies do not change frequently. The storage operation consumes both memory and CPU time. [3]

Enke Chen proposes an alternative solution in RFC 2918 that avoids the additional maintenance cost. More specifically, it defines a new BGP capability termed 'Route Refresh Capability', which would allow the dynamic exchange of route refresh request between BGP speakers. Route Refresh implements a messaging system in which a router wishing to refresh or reset its BGP peering with its neighbour simply has to send a notification to its peer. When the neighbour receives the notification, it will send its entire announcement to its peer (obtained from BGP best path table and applicable outbound policy). [3]

To actually refresh the routing information during policy changes, Chen introduces a new BGP message type "ROUTE-REFRESH". This new message is used by a BGP speaker to request an update from its peer. A BGP speaker, which is willing to receive the ROUTE-REFRESH messages from its peer, should advertise the Route Refresh Capability to the peer using BGP Capabilities advertisement extension. [3]

3.2.2 Analysis

Route Refresh Capability introduces a mechanism which allows BGP peers to refresh their BGP peering sessions. The mechanism is both simple and efficient and it is better to use this extension rather than disruptive hard resets. All in all, the Route Refresh Capability is very efficient way to implement new policies without the memory and performance costs of soft-reconfiguration or the disrupted routing and route flapping when resetting a BGP speaker.

For BGP speakers that do not support Route Refresh Capability, the soft-reconfiguration presented in the previous section is the only logical choice to solve the refresh problem. However, soft-reconfiguration consumes more resources and is therefore not a good solution, especially in networks that actively change their routing policies.

3.3 Support for four-octet AS number space

3.3.1 Description

Currently the Autonomous System number is encoded in BGP-4 as a two-octets field. As stated before, the AS number is a unique identifier for every AS. As the Internet grows, the number of ASs is increasing and the two octets reserved for the identifier might not be enough.

An improvement to solve this problem has been proposed by Vohra and Chen. They have defined a way for the BGP to carry the Autonomous System number as a four-octets field. A BGP speaker advertises its four-octet AS number with the Capability Advertisement extension described in Section 3.1. If a BGP speaker has a two-octet AS number, it can be

converted into a four-octet number by setting the high-order two octets of the four-octets field to zero. [20]

3.3.2 Analysis

A question may be asked if this improvement is necessary at the current moment. According the current BGP specification, the autonomous system number takes two octets, which means $2^{16} = 65535$ autonomous systems. Hudson stated in 2001 [6] that the number space consumption would be about 51% each year and warned that the AS number space would be exhausted in 2005. However, during that time 10700 AS numbers were assigned. Smith reports that currently there are less than 17000 autonomous systems registered [18]. The real growth has been about 16,5% during the last three years. In the light of these facts, it seems that the current numbering will be sufficient for many years.

Another problem with the support for four-octet AS number space comes with the interaction with BGP peers that support only two-octet AS numbers. Vohra and Chen state [20] that this kind of intercommunication is possible only if the AS has both four-octet and two-octet AS numbers.

The requirement of two-octet AS number means that today the support for four-octet AS number space is not very useful. However, if BGP is used in the future, the use of the four-octet AS numbers will some day become obligatory. The two-octet numbers will eventually be spent and this extension provides $2^{32} = 4294967296$ unique AS numbers. If no radical changes to the network architectures are invented, four octets will be enough for many decades.

3.4 Graceful Restart Mechanism for BGP-4

3.4.1 Description

According to the specification of BGP, when a connection with on BGP peer is lost, then the BGP peers note that the session is lost. New routes to the destinations advertised by the lost BGP peer must be discovered. This kind of behaviour is not inviting, especially when there is need to reboot the BGP speaking machine.

Usually when BGP on a router restarts, all the BGP peers detect that the session went down, and then came back up. This "down/up" transition results in a "routing flap" and causes BGP route re-computation, generation of BGP routing updates and flap the forwarding tables. It could spread across multiple networks. According to Sangli et al., such routing flaps may create transient forwarding blackholes in which the restarting machine is chosen the best route in the AS, but it necessarily have not received the exterior routing data with BGP. The data sent to machine in the AS will be lost. [17]

Graceful restart is a new capability presented by Sangli et al. A BGP speaker can advertise this ability in the BGP OPEN message as specified in section 3.1. If both peers support this capability, then a graceful restart can be taken into use. If not, then the two peers communicate with each other according the original BGP specification. Under normal circumstances, a restarting server would cause the peer router to clear all routes associated with the restarting router. This does not occur with BGP graceful restart, however. The restarting router sets the restart bit to indicate that it has restarted and sets the forwarding state bit to indicate that it has preserved its forwarding state. When the restart is taking place, the peer of the restarting router hides the restart from the rest of the network by not withdrawing the routes learned via the restarting router. This hiding is acceptable as long as the forwarding state is preserved on the restarting router, allowing traffic to continue to flow through it.[17]

As a part of the specification Sangli et al. define also an end-of-RIB (routing information base) marker, which is an empty update message. A BGP speaker sends this end-of-RIB marker to a peer when it has completed sending all of its initial routing updates. Within the graceful restart, the end-of-RIB marker is used by the restarting router to know when it has received all of the update information from its peers. Only after it has learned all of its peers' updates does it perform route selection and send out its own update. This ensures that the first updates sent out by the router after it has restarted reflect the current network state as completely as possible. [17]

3.4.2 Analysis

The graceful restart is a welcome solution for the problem of restarting BGP speaking machines. In some networks, it might take many minutes to even transfer the whole routing table from the BGP peers. With the graceful restart capability, the server can preserve its original routing data, commit a restart and then take the earlier data back into use. The BGP peers know when restart has been done and inform the restarting server of changes done during the restart.

4 Conceptual changes in BGP-4

Earlier section discussed minor functional changes to the original BGP-4 protocol specification. These changes included new optional parameter to allow more easier usage of new additions, enhanced route refresh mechanism, support for larger amount of autonomous systems and a non-disruptive restart mechanism. All these changes did not severely alter the original specification.

However, there are some developments to the BGP which radically change the original specification. These developments are called "Conceptual" by the author to amplify their nature. In this section, two of these developments are discussed. First, the advantages of breaking one autonomous system into separate subsystems are presented in Section 4.1. Second, in Section 4.2 the nature of original BGP as a routing protocol of IP version 4 traffic is extended to allow BGP carry routing information for many other network level protocols.

4.1 Autonomous System Confederations for BGP-4

4.1.1 Description

A serious problem with original specification of BGP concerns the scalability of BGP in a large autonomous sys-

tem. By the original definition, all BGP speakers of an autonomous system must be fully meshed [19]. Full mesh in this context means that every BGP speaker must send its routing data to every other BGP speaker in the AS. An illustration of the situation is presented in Figure 2. For simplicity, only the structure of one AS without other ASs is shown.

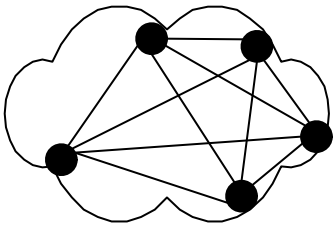

Figure 2: Illustration of fully meshed autonomous system

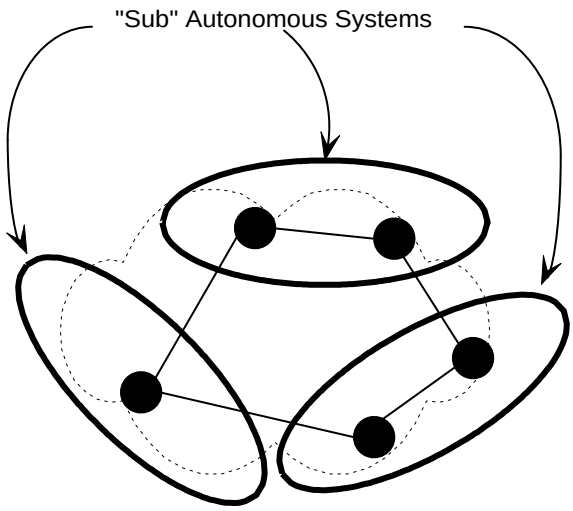

Figure 3: Illustration of AS confederation

The reason for this mechanism is to guarantee that all BGP routers of an autonomous systems propagate valid data. However, as Traina et al. point out [19], the result is that in a AS which has n BGP routers, $n*(n-1)/2$ unique connection between the routers must be established. In the beginning of the AS ideology this was fine, but today many networks have grown quite large and most likely have multiple BGP speakers.

Basic solution to the problem would be to divide the autonomous into smaller, independent autonomous systems. The division could be done without the addition of confederation ideology, but in that case the complexity of routing policies would increase within the area of the original autonomous system. In addition, the plain division of one AS to smaller autonomous systems increases the maintenance effort needed when some changes to the original topology take place.

To solve this scaling problem without the need to break the original AS into smaller ASs, numerous solutions have been presented. One of these solutions, called *Autonomous System Confederations for BGP-4*, was designed by Traina, McPherson and Scudder. The structure of this solution is represented in RFC 3065 [19]. In this paper, a brief introduction to the solution will be presented next.

The main idea in Autonomous System Confederations is to divide the autonomous system into smaller "sub"-autonomous systems [19]. Figure 3 is a confederation example of the AS presented earlier in Figure 2. These new "sub"-autonomous systems form a *confederation* which represents itself to external BGP peers as nothing has changed. However, there are substantial changes in the way that the BGP speakers inside the confederation exchange route information.

In the original specification [14], attributes named as AS_PATH are used when two BGP peers exchange route information. The AS_PATH contains a list of ASs through which a UPDATE message has traversed through. The information in the list can be used for example to help when making routing decisions or preventing routing loops.

Autonomous system confederation adds a new, similar attribute to AS_PATH called AS_CONFED to the specification. In addition, a new number called *Confederation Member AS Number* is assigned to the sub-systems. The AS_CONFED is used within the confederation as an identifier between different sub-ASs. The route exchange between some confederation's BGP speakers follows the same logic than the original BGP specification. In the confederation, the Member AS Number is used to identify different sub-ASs and when some BGP speaker propagates information to external BGP peers, it uses the confederation AS identifier. This way the original AS seems to be intact from outsider's point of view. [19]

4.1.2 Analysis

In a nutshell, the Autonomous System Confederation can be seen as a autonomous system composed of autonomous systems. The original AS will remain intact to other ASs, because the new sub-ASs represent themselves to other autonomous systems' BGP speakers as a single (original) AS. The division of original complex system structure into smaller, more manageable parts first of all lessens the number of internal sessions between the BGP speakers.

The requirement for the use of this solution is that all BGP speakers in the AS must support the presented extension. A BGP speaker that does not support the new attributes needed by the new mechanism will regard incoming messages as faulty. However, when an AS divides into a confederation of ASs, it is most likely that all BGP speakers within the original AS support the needed properties.

One observation must be made concerning the maintenance and administration of a confederation. Traina et al. state, that although the solution they have presented reduces the management complexity, it is still is advisable to maintain single administrative authority [19]. Otherwise the extension might cause problems between the sub-systems other external BGP-peers. All in all, the BGP confederations extension has been taken into wide use and many network and software vendors support it [19].

4.2 Multiprotocol Extensions for BGP-4

4.2.1 Description

The original specification states that BGP-4 is able to carry routing information only for Internet Protocol version 4 (IPv4) [14]. This is serious limitation if BGP-4 is wanted to be used in a network which uses some other Network Layer protocol.

To remove this limitation, RFC 2858 defines extension to BGP-4 to enable it to carry routing information for multiple protocols. In addition to the IPv4 mentioned earlier, the supported protocols include for example Internet Protocol version 6 (IPv6) and Novell's Internetwork Packet Exchange (IPX). In addition, the extension mechanism allows the propagated network reachability information refer to either unicast or multicast or both unicast and multicast traffic. The BGP with multiprotocol extension is called MBGP or MP-BGP for convenient. Sometimes the BGP-4 with multiprotocol extension is referred as BGP-4+. [1]

The extensions represented by Bates et al. are backward compatible: a BGP router that supports the extensions can communicate with a router that does not support the extensions. There is one limitation: the BGP speaker using Multiprotocol extensions must have an IPv4 address. [1]

The multiprotocol extension bases its functionality on the fact that the original BGP specification carries only three pieces of information that are IPv4-specific. These pieces of information are attributes which are used to tell the next hop target (NEXT_HOP), the last BGP speaker that performed route aggregation (AGGREGATOR), and Network Layer Reachability Information (NLRI). [14]

Bates et al. suggest [1], that to add multiprotocol extension to BGP, the attributes described earlier must be replaced with new protocol specific information. Therefore, the following changes are proposed in the RFC 2858 [1]:

- The AGGREGATOR attribute is filled with the IPv4 address of the BGP speaker. This can be done, because the extension assumes that the BGP speaker must have IPv4 address.
- Two new attributes, Multiprotocol Reachable NLRI (MP_REACH_NLRI) and Multiprotocol Unreachable NLRI (MP_UNREACH_NLRI) are added to remove the limitations of the original NEXT_HOP and NLRI attributes.

The first new attribute (MP_REACH_NLRI) is used to carry the set of reachable destinations together with the next hop information. This attribute makes it possible to associate a particular network layer protocol with the next-hop information. The second one (MP_UNREACH_NLRI) is used to carry the set of unreachable destinations. MP_UNREACH_NLRI can be used for example to withdraw multiple unavailable routes.

Both of the presented new attributes are optional. This way a BGP speaker that does not support the multiprotocol capabilities will ignore the information and will not pass it to other BGP speakers. [1]

According to the specification, a BGP speaker that uses Multiprotocol Extensions is strongly advised to use the Capability Advertisement procedures described in the Section 3.1 to determine whether the speaker could use Multiprotocol Extensions with a particular peer. Otherwise the new optional parameters will distrupt connection with a BGP peer that does not support Multiprotocol Extensions. [1]

4.2.2 Analysis

Multiprotocol extensions for BGP alter the original concept of exchanging only IPv4 routing information. Now routing information for many different protocols can be propagated. This ability opens new doors to many new applications. For example, in addition to the IPX and IPv6 routing data exchange, RFC 3107 [13] defines a way for carrying Multiprotocol Label Switching (MPLS) label data with BGP and multiprotocol extensions.

The extensions do come with a cost, however. Although Bates et al. have designed the mechanism to be relatively simple, a BGP speaker utilizing the extensions must do extra work. The exchanged data must be validated and errors must be handled. In this case, the error handling is very strict. For example, if a BGP speaker receives from a neighbour an UPDATE message that contains an incorrect MP_REACH_NLRI or MP_UNREACH_NLRI attribute, the speaker must delete all the BGP routes received from that neighbour [1].

5 Developments in BGP-4 Security

In the beginning of the Internet, hostile attacks were highly unlikely and there was no reason to design and implement protection against them. This historical issue must be taken into account when the security of BGP is discussed.

BGP is an old protocol which has been in use for decades. It was not designed with protection against deliberate or accidental errors. There are no internal mechanisms in the BGP protocol to protect against attacks that modify, delete, forge, or replay data [12]. BGP could for example be used to blackhole traffic. If an attacker has access to the forwarding path of the target system, he can quietly discard the traffic while continuing to function as a BGP speaker. BGP also relies on lower level protocols like TCP and IP. Any attack on these lower level protocols, like IP-spoofing or session stealing, is an attack against the BGP.

According to Murphy, the faulty, misconfigured or deliberately malicious sources could disrupt overall Internet behaviour. This could be done by injecting faulty routing information into the BGP distributed routing database (by modifying, forging, or replaying BGP packets) [12]. Kent et al continue, that BGP has a number of vulnerabilities that can be exploited to cause problems such as misdelivery or nondelivery of user traffic, misuse of network resources, network congestion and packet delays, and violation of local routing policies [9].

Currently the security of BGP is often handled in routers with Access Control Lists (ACL), which restrict communication to allowed routers [7]. ACLs are a good way to add security, if the rules of the list stay relatively simple. However, ACLs do not provide data encryption or integrity validation.

Following sections present some solutions to the security issues associated with BGP-4. First an implementation of the

BGP's original authentication routine is reviewed in Section 5.1. Next, a solution for BGP's security problems by using the IPSec security architecture is discussed in Section 5.2. Finally, a secure, scalable, deployable architecture for an authorization and authentication system that addresses most of the security problems associated with BGP is presented in Section 5.3.

5.1 MD5 Authentication

5.1.1 Description

The current BGP specification [14] requires that an implementation must support MD5 authentication. However, the specification does not require the usage of authentication which has lead to a situation in which autonomous systems are not authenticated at all. The lack of authentication in BGP means that neither BGP peers nor the validity of BGP messages can be validated. Murphy points out that without a peer authentication a man-in-the-middle attack is fairly easy to commit [12].

The specification for the MD5 authentication with BGP is presented by Heffernan in RFC 2385 [5]. In short, Heffernan defines a new TCP option for carrying an MD5 digest (described in RFC 1321 [15]) in a TCP segment. This digest acts like a signature for that segment, and it is created with information which only the peers of the connection know. Since BGP uses TCP as its transport protocol, using the MD5 digest significantly reduces the danger from certain security attacks on BGP.

The MD5 digest is calculated from the TCP pseudo-header, TCP header, TCP segment data and an independently-specified key or password, known only by the two BGP peers. Heffernan points, that the key should be connection-specific. The sender calculates the digest using its own key and sends it in the TCP segment. The receiver must validate the digest by calculating its own digest from the same data using its own key and compare the two digests. If the comparison fails, the segment is dropped as faulty. [5]

When the MD5 authentication is used, an attacked entity would have to guess or obtain the TCP sequence numbers as well as know the password used in the MD5 algorithm to spoof a BGP connection. This password never appears in the connection stream, and the actual form of the password can be decided by the application. [5]

5.1.2 Analysis

MD5 provides a relatively easy way to add authentication to BGP traffic. Although the MD5 algorithm has some known theoretical weaknesses [5], it still is "good enough" for the task. However, as Heffernan himself points out, the calculation of MD5 digest takes time depending on the segment size [5]. This might have negative effects on performance and might be a reason not to use MD5 authentication with BGP.

There is also the difficulty of the shared secret key used by the BGP peers. This key has to be shared in a secure way or the basis of the authentication is lost. Heffernan notes, that although the keys can be changed on-the-fly, retransmissions can become problematical in some TCP implementations with changing passwords. [5]

MD5 is a promising solution despite the performance and key exchange issues. Although there are some things that might make the MD5 authentication unfitting for BGP, Heffernan states [5] that the solution he has defined provides a currently practiced security mechanism for BGP. In addition, Kanclirz [7] notes, that the deployment of the MD5 authentication is quite simple after the key exchange problem has been solved.

5.2 IPSec with BGP-4

5.2.1 Description

One way to solve most of the security problems with BGP would be the use of IP security (IPSec). IPSec is defined in RFC [8]. IPSec is a security mechanism dedicated to the IP layer so the structure and functionality of the original BGP does not need changes.

IPSec provides both authentication and data encryption. This way the BGP peers could first authenticate each other and then transfer the routing information in an encrypted form. If only plain authentication is needed, then the Authencation Header (AH) property to add authentication routines to normal IP-based traffic. If both privacy as well as authentication would be wanted with BGP, then the Encapsulated Security Payload (ESP) could be used. [8].

Another way to take advantage of IPSec would be the usage of IPSec tunnelling [7]. In IPSec tunnelling the original IP packet containing the BGP data would be capsulated in a new, encrypted packet. These packets could then be sent through the Internet securely to the corresponding BGP peer. The receiver would then decrypt the received package and extract the original IP packet containing the BGP data. [8].

5.2.2 Analysis

IPSec is a very good solution to many of the current Internet's security questions. It provides both authentication and data encryption which are both valuable qualities.

However, the IPSec does have some drawbacks. First the question of key management must be answered. In which way the keys are exchanged and how does the system react to key changes? A second and fairly important issue is the time IPSec operations take. Border routers exchanging routing information are even now very busy and the addition of IPSec-based security would increase the load even further [7]. Third problem associated with the IPSec is the complexity of operations and maintenance which lead to a high possibility of misconfiguration [7].

In general, the IPSec would be a fine solution and it offers many promising features. However, IPSec is most likely too heavy structure to be used with BGP. The problems associated with the key management and performance would have to be solved first before any large-scale utilization of IPSec is likely. Kanclirz even raises a question, should the BGP traffic be encrypted at all [7]. Plain authentication with MD5 as presented in Section 5.1 without any other encryption might be a better and lighter way to increase security.

5.3 Secure BGP (S-BGP)

5.3.1 Description

The MD5 Authentication and IPSec security mechanism presented before do not have a huge affect on existing BGP implementations. They both add authentication and IPSec adds encryption and tunnelling possibility. In this section the basic structure of a secure, scalable, deployable architecture called Secure BGP (S-BGP) is presented. S-BGP is designed to be an authorization and authentication system that addresses most of the security problems associated with BGP [9].

In a nutshell, S-BGP is a collection of cryptographic solutions to the well-known security weaknesses described for example by Murphy [12]. S-BGP relies its functionality on the idea that BGP messages are digitally signed. This way the source of the messages can be authenticated, and the source's authorization to send updates for specific routes is verified. [9]

Digital signatures in S-BGP are based on IP addressed and for that reason Lynne et al. recommend to use Public Key Infrastructure (PKI). Actually S-BGP will require two different PKIs. The first PKI is used to attach one or more address blocks to some organization which has right to the specified addresses. In comparison to the normal PKI ideology, the S-BGP certificate for the first PKI is not a singular subject's identity but a block of addresses specified in the certificate. The second PKI is used to authenticate autonomous systems and individual BGP speakers and their mutual relationship. [9, 10]

A new path attribute for the BGP called *Attestation* is used to describe that the certificate issuer has authorized a subject to advertise a path to a specified addresses. In addition, to ensure the message authenticity and integrity and to prevent replay-attacks, S-BGP uses IPSec ESP. ESP is used with no encryption and Internet Key Exchange (IKE) is used for key exchanged. [9, 10]

5.3.2 Analysis

S-BGP is a very complex architecture. In practice, S-BGP adds three databases to a router:

- an originating AS database, which is derived from the address attestations,
- a public key database which is derived from the end-entity certificates of the S-BGP speakers and
- a database expressing for describing the AS's S-BGP related policy.

Although the designers of the S-BGP architecture state, that S-BGP has been designed to minimize bandwidth utilization, CPU power and memory consumption, it still has an impact on all of those [10]. Processing power is needed to encrypt and decrypt the data and much memory must be allocated to store the necessary data. On the other hand, the administration of a system utilizing S-BGP fully could be very difficult.

It is likely that the S-BGP is too complex to be used properly. Even if the system would solve all the security problems of BGP, the use of resources is likely to be too high to be acceptable.

6 Future Directions

The previous chapters presented some of the developments made to the original BGP. In this chapter, the future of BGP is discussed. The ideas presented in this chapter are author's own ideas based on the research done for this paper.

The question of IPv6 is likely to surface from time to time. BGP now supports the exchange of IPv6 routing data via the Multiprotocol Extensions 4.2 and this ability might play a crucial part if and when networks move to use IPv6 instead of IPv4. Still, it is fair to question if the whole transition even takes place.

Security of BGP is most likely to become an important issue. As the world becomes more perilous and more networked every day, it is likely that BGP will attract adversaries against it. New security architectures like the presented S-BGP are likely to be seen, although they might be too heavy for proper use.

Since BGP is the core routing protocol on the Internet, the future development must be tightly controlled. Any changes to it must be performed with great care and with careful planning. On the other hand, the user community of BGP mainly consists of autonomous systems which obtain their routers from a handful of vendors. Therefore the changes are possible.

In future it is likely that BGP will be developed to handle Multicast, Multihoming and Mobility. These three M's are currently the "hot potato" of the Internet. Multicast is seen as a boon to the Internet architecture as a single message could be transmitted to many without the need to send it multiple times (like unicast) or sending it to all (broadcast). With Multihoming, multiple connections can be used by an entity for example for backup purposes or load-balancing. Mobile devices have become extremely popular during the last few years and will most likely influence BGP in some way.

7 Conclusion

BGP has been a part of the Internet architecture for many years. Although the protocol itself has proven to be very useful to the whole Internet architecture, the world today has changed a lot since the dawn of the Internet age. The rapid development of Internet technologies, performance and security issues have forced the BGP to evolve. In this paper, some of these additions and upgrades were presented.

One peculiar thing is that the development steps are mostly backward compatible, but more sophisticated features rely heavily on previous improvements. A word "feature-chain" could be used to describe the situation: although the extensions are mostly compatible with the original specification, the use of some advanced ability require some other extension and so on. For example, RFC 3107 [13] defines a way for carrying Multiprotocol Label Switching (MPLS) label data with BGP-4, but it assumes that multi-

procotol extensions are taken into use. In turn, multiprotocol extensions strongly rely on the "Capabilities Advertisement" ability.

From the initial setting, in which plain IPv4 routing data was exchanged between a pair of ASs, BGP has developed greatly. Now multiprotocol routing information can be exchanged and a single AS can be a confederation of sub-ASs. New capabilities can be added with ease and BGP routers rebooted without the problems of losing BGP sessions. Routing data can be refreshed easily between BGP peers as refresh mechanisms are taken into use. The initial limit to number of ASs has been raised tremendously. Additionally, as new security additions have been introduced, ASs can be authorized and authenticated and the routing data can be encrypted.

All the developments presented in this paper confirm that BGP is very flexible. One might note that there are two things that make BGP so successful. First, the original specification contains only the things that are essential to the functionality. This probably has been a blessing to the Internet since the number of misconfigurations and the complexity of the inter-domain routing might be many times bigger compared to the current situation. Second, although the protocol initially contained only the necessary properties, it has very flexible mechanisms which allow relatively easy extensions to the protocol.

The future of the BGP-4 seems to be very bright. Much work has been done to add new features to the original specification and this work has shown that even new additions can be added. BGP-4 will stand unchallenged for many years to come.

References

[1] Tony Bates, Yakov Rekhter, Ravi Chandra, Dave Katz. *Multiprotocol Extensions for BGP-4*. RFC 2858, IETF Network Working Group, June 2000

[2] Ravi Chandra, John Scudder. *Capabilities Advertisement with BGP-4*. RFC 3392, IETF Network Working Group, November 2002

[3] Enke Chen. *Route Refresh Capability for BGP-4*. RFC 2918, IETF Network Working Group, September 2000

[4] Douglas Comer. *Internetworking with TCP/IP Principles, Protocols, and Architectures Vol 1*. 4th edition, Prentice Hall, 2000.

[5] Andy Heffernan. *Protection of BGP Sessions via the TCP MD5 Signature Option*. RFC 2385, IETF Network Working Group, August 1998

[6] Geoff Huston. *Commentary on Inter-Domain Routing in the Internet*. RFC 3221, IETF Internet Architecture Board, December 2001

[7] Jan Kanclirz Jr. *Make Secure - BGP*. http://www.makesecure.com/makesecure-bgp.pdf, September 2003

[8] Stephen Kent, R. Atkinson. *Security Architecture for the Internet Protocol*. RFC 2401, IETF Network Working Group, November 1998

[9] Stephen Kent, Charles Lynn, Kares Seo. *Secure border gateway protocol (secure-BGP)*. *IEEE Journal on Selected Areas in Communications* 18(4):582-592; 2000.

[10] Charles Lynn, Joanne Mikkelson, Karen Seo. *Secure BGP (S-BGP)*. Internet Draft, IETF Network Working Group, June 2003 http://www.ir.bbn.com/projects/sbgp/draft-clynn-s-bgp-protocol-01.txt

[11] David Meyer, Keyur Patel. *BGP-4 Protocol Analysis*. Internet Draft, IETF Network Working Group, June 2003 http://www.ietf.org/internet-drafts/draft-ietf-idr-bgp-analysis-04.txt

[12] Sandra Murphy. *BGP Security Vulnerabilities Analysis*. Internet Draft, IETF Network Working Group, June 2003 http://www.ietf.org/internet-drafts/draft-ietf-idr-bgp-vuln-00.txt

[13] Yakov Rekhter, Eric Rosen *Carrying Label Information in BGP-4*. RFC 3107, IETF Network Working Group, May 2001.

[14] Yakov Rekhter, Tony Li. *A Border Gateway Protocol 4 (BGP-4)*. RFC 1771, IETF Network Working Group, March 1995.

[15] Ronald Rivest. *The MD5 Message-Digest Algorithm*. RFC 1321, IETF Network Working Group, April 1992.

[16] Eric Rosen. *Exterior Gateway Protocol (EGP)*. RFC 827, IETF, October 1982.

[17] Srihari Sangli, Yakov Rekhter, Rex Fernando, John Scudder, Enke Chen. *Graceful Restart Mechanism for BGP*. Internet Draft, IETF Network Working Group, October 2003 http://www.ietf.org/internet-drafts/draft-ietf-idr-restart-08.txt

[18] Philip Smith. *CIDR Report for 11 February 2004*. URL http://www.cidr-report.org/as6447

[19] Paul Traina, D. McPherson, John Scudder. *Autonomous System Confederations for BGP*. RFC 3065, IETF Network Working Group, February 2001

[20] Quaizar Vohra, Enke Chen. *BGP support for four-octet AS number space*. Internet Draft, IETF Network Working Group, September 2003 http://www.ietf.org/internet-drafts/draft-ietf-idr-as4bytes-07.txt