Internet Engineering Task Force
Internet-Draft
Intended status: Best Current Practice
Expires: April 30, 2015

J. Durand
CISCO Systems, Inc.
I. Pepelnjak
NIL
G. Doering
SpaceNet

October 27, 2014

BGP operations and security
draft-ietf-opsec-bgp-security-06.txt

Abstract

BGP (Border Gateway Protocol) is the protocol almost exclusively used in the Internet to exchange routing information between network domains. Due to this central nature, it is important to understand the security measures that can and should be deployed to prevent accidental or intentional routing disturbances.

This document describes measures to protect the BGP sessions itself (like TTL, TCP-AO, control plane filtering) and to better control the flow of routing information, using prefix filtering and automatization of prefix filters, max-prefix filtering, AS path filtering, route flap dampening and BGP community scrubbing.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [1].

Copyright Notice

Table of Contents

1. Introduction

BGP (Border Gateway Protocol - RFC 4271 [2]) is the protocol used in the Internet to exchange routing information between network domains. BGP does not directly include mechanisms that control that routes exchanged conform to the various guidelines defined by the Internet community. This document intends to both summarize common existing guidelines and help network administrators apply coherent BGP policies.

2. Scope of the document

The guidelines defined in this document are intended for generic Internet BGP peerings. The nature of the Internet is such that Autonomous Systems can always agree on exceptions to a common framework for relevant local needs, and therefore configure a BGP session in a manner that may differ from the recommendations provided in this document. While this is perfectly acceptable, every configured exception might have an impact on the entire inter-domain routing environment and network administrators SHOULD carefully appraise this impact before implementation.

3. Definitions and Accronyms

- ACL: Access Control List
- ASN: Autonomous System Number
- IRR: Internet Routing Registry
- IXP: Internet eXchange Point
- LIR: Local Internet Registry
- pMTUd: Path MTU Discovery
- RIR: Regional Internet Registry
- Tier 1 transit provider: an IP transit provider which can reach any network on the Internet without purchasing transit services
- uRPF: Unicast Reverse Path Forwarding

4. Protection of the BGP speaker

The BGP speaker needs to be protected from attempts to subvert the BGP session. This protection SHOULD be achieved by an Access Control List (ACL) which would discard all packets directed to TCP port 179 on the local device and sourced from an address not known or permitted to become a BGP neighbor. Experience has shown that natural protection TCP should offer is not always sufficient as it is sometimes run in control-plane software: in the absence of ACLs it is possible to attack a BGP speaker by simply sending a high volume of connection requests to it.

If supported, an ACL specific to the control-plane of the router SHOULD be used (receive-ACL, control-plane policing, etc.), to avoid configuration of data-plane filters for packets transiting through the router (and therefore not reaching the control plane). If the hardware can not do that, interface ACLs can be used to block packets addressed to the local router.

Some routers automatically program such an ACL upon BGP configuration. On other devices this ACL should be configured and maintained manually or using scripts.

In addition to strict filtering, rate-limiting MAY be configured for accepted BGP traffic. Rate-limiting BGP traffic consists in permitting only a certain quantity of bits per second (or packets per second) of BGP traffic to the control plane. This protects the BGP

router control plane in case the amount of BGP traffic overcomes platform capabilities.

Filtering and rate-limiting of control-plane traffic is a wider topic than "just for BGP" (if network administrator brings down a router by overloading one of the other protocols from remote, BGP is harmed as well). For a more detailed recommendation on how to protect the router's control plane, see RFC 6192 [11].

5. Protection of BGP sessions

Current security issues of TCP-based protocols (therefore including BGP) have been documented in RFC 6952 [14]. The following sub-sections list the major points raised in this RFC and give best practices related to TCP session protection for BGP operation.

5.1. Protection of TCP sessions used by BGP

Attacks on TCP sessions used by BGP (aka BGP sessions), for example sending spoofed TCP RST packets, could bring down a BGP peering. Following a successful ARP spoofing attack (or other similar Man-in-the-Middle attack), the attacker might even be able to inject packets into the TCP stream (routing attacks).

BGP sessions can be secured with a variety of mechanisms. MD5 protection of TCP session header, described in RFC 2385 [7], was the first such mechanism. It is now deprecated by TCP Authentication Option (TCP-AO, RFC 5925 [4]) which offers stronger protection. While MD5 is still the most used mechanism due to its availability in vendor's equipment, TCP-AO SHOULD be preferred when implemented.

IPsec could also be used for session protection. At the time this document is published, there is not enough experience on impacts of the use of IPsec for BGP peerings and further analysis is required to define guidelines.

The drawback of TCP session protection is additional configuration and management overhead for authentication information (ex: MD5 password) maintenance. Protection of TCP sessions used by BGP is thus NOT REQUIRED even when peerings are established over shared networks where spoofing can be done (like IXPs), but operators are RECOMMENDED to consider the trade-offs and to apply TCP session protection where appropriate.

Network administrators SHOULD block spoofed packets (packets with a source IP address belonging to their IP address space) at all edges of their network (see RFC 2827 [8] and RFC 3704 [9]). This protects

the TCP session used by iBGP from attackers outside the Autonomous System.

5.2. BGP TTL security (GTSM)

BGP sessions can be made harder to spoof with the Generalized TTL Security Mechanisms (aka TTL security), defined in RFC 5082 [3]. Instead of sending TCP packets with TTL value of 1, the BGP speakers send the TCP packets with TTL value of 255 and the receiver checks that the TTL value equals 255. Since it's impossible to send an IP packet with TTL of 255 to a non-directly-connected IP host, BGP TTL security effectively prevents all spoofing attacks coming from third parties not directly connected to the same subnet as the BGP-speaking routers. Network administrators SHOULD implement TTL security on directly connected BGP peerings.

GTSM could also be applied to multi-hop BGP peering as well. To achieve this TTL needs to be configured with proper value depending on the distance between BGP speakers (using principle described above). Nevertheless it is not as effective as anyone inside the TTL diameter could spoof the TTL.

Like MD5 protection, TTL security has to be configured on both ends of a BGP session.

6. Prefix filtering

The main aspect of securing BGP resides in controlling the prefixes that are received/advertised on the BGP peerings. Prefixes exchanged between BGP peers are controlled with inbound and outbound filters that can match on IP prefixes (prefix filters, Section 6), AS paths (as-path filters, Section 9) or any other attributes of a BGP prefix (for example, BGP communities, Section 11).

6.1. Definition of prefix filters

This section list the most commonly used prefix filters. Following sections will clarify where these filters should be applied.

6.1.1. Special purpose prefixes

6.1.1.1. IPv4 special purpose prefixes

IANA IPv4 Special-Purpose Address Registry [22] maintains the list of IPv4 special purpose prefixes and their routing scope, and SHOULD be used for prefix filters configuration. Only prefixes with value "False" in column "Global" SHOULD be discarded on Internet BGP peerings.

6.1.1.2. IPv6 special purpose prefixes

IANA IPv6 Special-Purpose Address Registry [23] maintains the list of IPv6 special purpose prefixes and their routing scope, and SHOULD be used for prefix filters configuration. Only prefixes with value "False" in column "Global" SHOULD be discarded on Internet BGP peerings.

6.1.2. Prefixes not allocated

IANA allocates prefixes to RIRs which in turn allocate prefixes to LIRs (Local Internet Registries). It is wise not to accept routing table prefixes that are not allocated by IANA and/or RIRs. This section details the options for building a list of allocated prefixes at every level. It is important to understand that filtering prefixes not allocated requires constant updates as prefixes are continually allocated. Therefore automation of such prefix filters is key for the success of this approach. Network administrators SHOULD NOT consider solutions described in this section if they are not capable of maintaining updated prefix filters: the damage would probably be worse than the intended security policy.

6.1.2.1. IANA allocated prefix filters

IANA has allocated all the IPv4 available space. Therefore there is no reason why network administrators would keep checking prefixes are in the IANA allocated IPv4 address space [24]. No specific filters need to be put in place by administrators who want to make sure that IPv4 prefixes they receive in BGP updates have been allocated by IANA.

For IPv6, given the size of the address space, it can be seen as wise accepting only prefixes derived from those allocated by IANA. Administrators can dynamically build this list from the IANA allocated IPv6 space [25]. As IANA keeps allocating prefixes to RIRs, the aforementioned list should be checked regularly against changes and if they occur, prefix filters should be computed and pushed on network devices. The list could also be pulled directly by routers when they implement such mechanisms. As there is delay between the time a RIR receives a new prefix and the moment it starts allocating portions of it to its LIRs, there is no need for doing this step quickly and frequently. However, network administrators SHOULD ensure that all IPv6 prefix filters are updated within maximum one month after any change in the list of IPv6 prefix allocated by IANA.

If process in place (manual or automatic) cannot guarantee that the list is updated regularly then it's better not to configure any

filters based on allocated networks. The IPv4 experience has shown that many network operators implemented filters for prefixes not allocated by IANA but did not update them on a regular basis. This created problems for latest allocations and required a extra work for RIRs that had to "de-bogonize" the newly allocated prefixes.

6.1.2.2. RIR allocated prefix filters

A more precise check can be performed when one would like to make sure that prefixes they receive are being originated or transited by autonomous systems entitled to do so. It has been observed in the past that an AS could easily advertise someone else's prefix (or more specific prefixes) and create black holes or security threats. To partially mitigate this risk, administrators would need to make sure BGP advertisements correspond to information located in the existing registries. At this stage 2 options can be considered (short and long term options). They are described in the following subsections.

6.1.2.2.1. Prefix filters creation from Internet Routing Registries (IRR)

An Internet Routing Registry (IRR) is a database containing Internet routing information, described using Routing Policy Specification Language objects - RFC 4012 [10]. Network administrators are given privileges to describe routing policies of their own networks in the IRR and information is published, usually publicly. A majority of Regional Internet Registries do also operate an IRR and can control that registered routes conform to prefixes allocated or directly assigned.However, it should be noted that the list of such prefixes is not necessarily a complete list, and as such the list of routes in an IRR is not the same as the set of RIR allocated prefixes.

It is possible to use the IRR information to build, for a given neighbor autonomous system, a list of prefixes originated or transited which one may accept. This can be done relatively easily using scripts and existing tools capable of retrieving this information in the registries. This approach is exactly the same for both IPv4 and IPv6.

The macro-algorithm for the script is described as follows. For the peer that is considered, the distant network administrator has provided the autonomous system and may be able to provide an AS-SET object (aka AS-MACRO). An AS-SET is an object which contains AS numbers or other AS-SETs. An operator may create an AS-SET defining all the AS numbers of its customers. A tier 1 transit provider might create an AS-SET describing the AS-SET of connected operators, which in turn describe the AS numbers of their customers. Using recursion, it is possible to retrieve from an AS-SET the complete list of AS

numbers that the peer is likely to announce. For each of these AS numbers, it is also easy to check in the corresponding IRR for all associated prefixes. With these two mechanisms a script can build for a given peer the list of allowed prefixes and the AS number from which they should be originated. One could decide not use the origin information and only build monolithic prefix filters from fetched data.

As prefixes, AS numbers and AS-SETs may not all be under the same RIR authority, a difficulty resides choosing for each object the appropriate IRR to poll. Some IRRs have been created and are not restricted to a given region or authoritative RIR. They allow RIRs to publish information contained in their IRR in a common place. They also make it possible for any subscriber (probably under contract) to publish information too. When doing requests inside such an IRR, it is possible to specify the source of information in order to have the most reliable data. One could check a popular IRR containing many sources (such as RADB [26], the Routing Assets Database) and only select as sources some desired RIRs and trusted major ISPs (Internet Service Providers).

As objects in IRRs may frequently vary over time, it is important that prefix filters computed using this mechanism are refreshed regularly. A daily basis could even be considered as some routing changes must be done sometimes in a certain emergency and registries may be updated at the very last moment. It has to be noted that this approach significantly increases the complexity of the router configurations as it can quickly add tens of thousands configuration lines for some important peers. To manage this complexity, network adminstrators could for example use IRRToolSet [29], a set of tools making it possible to simplify the creation of automated filters configuration from policies stored in IRR.

Last but not least, authors recommend that network administrators publish and maintain their resources properly in IRR database maintained by their RIR, when available.

6.1.2.2.2. SIDR - Secure Inter Domain Routing

An infrastructure called SIDR (Secure Inter-Domain Routing), described in RFC 6480 [12] has been designed to secure Internet advertisements. At the time this document is written, many documents have been published and a framework with a complete set of protocols is proposed so that advertisements can be checked against signed routing objects in RIR routing registries. There are basically two services that SIDR offers:

- o Origin validation, described in RFC 6811 [5], seeks at making sure that attributes associated with a routes are correct (the major point being the validation of the AS number originating this route). Origin validation is now operational (Internet registries, protocols, implementations on some routers...) and in theory it can be implemented knowing that the proportion of signed resources is still low at the time this document is written.

- o Path validation provided by BGPsec [27] seeks at making sure that no ones announce fake/wrong BGP paths that would attract trafic for a given destination, see RFC 7132 [16]. BGPsec is still an on-going work item at the time this document is written and therefore cannot be implemented.

Implementing SIDR mechanisms is expected to solve many of BGP routing security problems in the long term but it may take time for deployments to be made and objects to become signed. It also has to be pointed that SIDR infrastructure is complementing (not replacing) the security best practices listed in this document. Authors therefore recommend to implement any SIDR proposed mechanism (example: route origin validation) on top of the other existing mechanisms even if they could sometimes appear targeting the same goal.

If route origin validation is implemented, authors recommend to refer to rules described in RFC 7115 [15]. In short, each external route received on a router SHOULD be checked against the RPKI data set:

- o If a corresponding ROA (Route Origin Authorization) is found and is valid then the prefix SHOULD be accepted.

- o It the ROA is found and is INVALID then the prefix SHOULD be discarded.

- o If an ROA is not found then the prefix SHOULD be accepted but corresponding route SHOULD be given a low preference.

Authors also recommend that network operators sign their routing objects so their routes can be validated by other networks running origin validation.

One should understand that the RPKI model brings new interesting challenges. The paper On the Risk of Misbehaving RPKI Authorities [30] explains how RPKI model can impact the Internet if authorities don't behave as they are supposed to do. Further analysis is certainly required on RPKI, which carries part of BGP security.

6.1.3. Prefixes too specific

Most ISPs will not accept advertisements beyond a certain level of specificity (and in return do not announce prefixes they consider as too specific). That acceptable specificity is decided for each peering between the 2 BGP peers. Some ISP communities have tried to document acceptable specificity. This document does not make any judgement on what the best approach is, it just recalls that there are existing practices on the Internet and recommends the reader to refer to what those are. As an example the RIPE community has documented that as of the time of writing of this document, IPv4 prefixes longer than /24 and IPv6 prefixes longer than /48 are generally not announced/accepted in the Internet [19] [20]. These values may change in the future.

6.1.4. Filtering prefixes belonging to the local AS and downstreams

A network SHOULD filter its own prefixes on peerings with all its peers (inbound direction). This prevents local traffic (from a local source to a local destination) from leaking over an external peering in case someone else is announcing the prefix over the Internet. This also protects the infrastructure which may directly suffer in case backbone's prefix is suddenly preferred over the Internet.

In some cases, for example in multi-homing scenarios, such filters SHOULD NOT be applied as this would break the desired redundancy.

To an extent, such filters can also be configured on a network for the prefixes of its downstreams in order to protect them too. Such filters must be defined with caution as they can break existing redundancy mechanisms. For example in case an operator has a multihomed customer, it should keep accepting the customer prefix from its peers and upstreams. This will make it possible for the customer to keep accessing its operator network (and other customers) via the Internet in case the BGP peering between the customer and the operator is down.

6.1.5. IXP LAN prefixes

6.1.5.1. Network security

When a network is present on an IXP and peers with other IXP members over a common subnet (IXP LAN prefix), it SHOULD NOT accept more specific prefixes for the IXP LAN prefix from any of its external BGP peers. Accepting these routes may create a black hole for connectivity to the IXP LAN.

If the IXP LAN prefix is accepted as an "exact match", care needs to be taken to avoid other routers in the network sending IXP traffic towards the externally-learned IXP LAN prefix (recursive route lookup pointing into the wrong direction). This can be achieved by preferring IGP routes before eBGP, or by using "BGP next-hop-self" on all routes learned on that IXP.

If the IXP LAN prefix is accepted at all, it SHOULD only be accepted from the ASes that the IXP authorizes to announce it - which will usually be automatically achieved by filtering announcements by IRR DB.

6.1.5.2. pMTUd and the loose uRPF problem

In order to have pMTUd working in the presence of loose uRPF, it is necessary that all the networks that may source traffic that could flow through the IXP (ie. IXP members and their downstreams) have a route for the IXP LAN prefix. This is necessary as "packet too big" ICMP messages sent by IXP members' routers may be sourced using an address of the IXP LAN prefix. In the presence of loose uRPF, this ICMP packet is dropped if there is no route for the IXP LAN prefix or a less specific route covering IXP LAN prefix.

In that case, any IXP member SHOULD make sure it has a route for the IXP LAN prefix or a less specific prefix on all its routers and that it announces the IXP LAN prefix or less specific (up to a default route) to its downstreams. The announcements done for this purpose SHOULD pass IRR-generated filters described in Section 6.1.2.2.1 as well as "prefixes too specific" filters described in Section 6.1.3. The easiest way to implement this is that the IXP itself takes care of the origination of its prefix and advertises it to all IXP members through a BGP peering. Most likely the BGP route servers would be used for this. The IXP would most likely send its entire prefix which would be equal or less specific than the IXP LAN prefix.

Appendix Appendix A gives an example of guidelines regarding IXP LAN prefix.

6.1.6. The default route

6.1.6.1. IPv4

The 0.0.0.0/0 prefix is likely not intended to be accepted nor advertised other than in specific customer / provider configurations, general filtering outside of these is RECOMMENDED.

6.1.6.2. IPv6

The ::/0 prefix is likely not intended to be accepted nor advertised other than in specific customer / provider configurations, general filtering outside of these is RECOMMENDED.

6.2. Prefix filtering recommendations in full routing networks

For networks that have the full Internet BGP table, some policies should be applied on each BGP peer for received and advertised routes. It is RECOMMENDED that each autonomous system configures rules for advertised and received routes at all its borders as this will protect the network and its peer even in case of misconfiguration. The most commonly used filtering policy is proposed in this section and uses prefix filters defined in previous section Section 6.1.

6.2.1. Filters with Internet peers

6.2.1.1. Inbound filtering

There are basically 2 options, the loose one where no check will be done against RIR allocations and the strict one where it will be verified that announcements strictly conform to what is declared in routing registries.

6.2.1.1.1. Inbound filtering loose option

In this case, the following prefixes received from a BGP peer will be filtered:

- o Prefixes not globally routable (Section 6.1.1)
- o Prefixes not allocated by IANA (IPv6 only) (Section 6.1.2.1)
- o Routes too specific (Section 6.1.3)
- o Prefixes belonging to the local AS (Section 6.1.4)
- o IXP LAN prefixes (Section 6.1.5)
- o The default route (Section 6.1.6)

6.2.1.1.2. Inbound filtering strict option

In this case, filters are applied to make sure advertisements strictly conform to what is declared in routing registries (Section 6.1.2.2). Warning is given as registries are not always

accurate (prefixes missing, wrong information...) This varies across the registries and regions of the Internet. Before applying a strict policy the reader SHOULD check the impact on the filter and make sure solution is not worse than the problem.

Also in case of script failure each administrator may decide if all routes are accepted or rejected depending on routing policy. While accepting the routes during that time frame could break the BGP routing security, rejecting them might re-route too much traffic on transit peers, and could cause more harm than what a loose policy would have done.

In addition to this, network administrators could apply the following filters beforehand in case the routing registry used as source of information by the script is not fully trusted:

- Prefixes not globally routable (Section 6.1.1)
- Routes too specific (Section 6.1.3)
- Prefixes belonging to the local AS (Section 6.1.4)
- IXP LAN prefixes (Section 6.1.5)
- The default route (Section 6.1.6)

6.2.1.2. Outbound filtering

Configuration should be put in place to make sure that only appropriate prefixes are sent. These can be, for example, prefixes belonging to both the network in question and its downstreams. This can be achieved by using a combination of BGP communities, AS-paths or both. It can also be desirable that following filters are positioned before to avoid unwanted route announcement due to bad configuration:

- Prefixes not globally routable (Section 6.1.1)
- Routes too specific (Section 6.1.3)
- IXP LAN prefixes (Section 6.1.5)
- The default route (Section 6.1.6)

In case it is possible to list the prefixes to be advertised, then just configuring the list of allowed prefixes and denying the rest is sufficient.

6.2.2. Filters with customers

6.2.2.1. Inbound filtering

The inbound policy with end customers is pretty straightforward: only customers prefixes SHOULD be accepted, all others SHOULD be discarded. The list of accepted prefixes can be manually specified, after having verified that they are valid. This validation can be done with the appropriate IP address management authorities.

The same rules apply in case the customer is also a network connecting other customers (for example a tier 1 transit provider connecting service providers). An exception can be envisaged in case it is known that the customer network applies strict inbound/outbound prefix filtering, and the number of prefixes announced by that network is too large to list them in the router configuration. In that case filters as in Section 6.2.1.1 can be applied.

6.2.2.2. Outbound filtering

The outbound policy with customers may vary according to the routes customer wants to receive. In the simplest possible scenario, the customer may only want to receive only the default route, which can be done easily by applying a filter with the default route only.

In case the customer wants to receive the full routing (in case it is multihomed or if wants to have a view of the Internet table), the following filters can be simply applied on the BGP peering:

- Prefixes not globally routable (Section 6.1.1)

- Routes too specific (Section 6.1.3)

- The default route (Section 6.1.6)

There can be a difference for the default route that can be announced to the customer in addition to the full BGP table. This can be done simply by removing the filter for the default route. As the default route may not be present in the routing table, network administrators may decide to originate it only for peerings where it has to be advertised.

6.2.3. Filters with upstream providers

6.2.3.1. Inbound filtering

In case the full routing table is desired from the upstream, the prefix filtering to apply is the same as the one for peers Section 6.2.1.1 with the exception of the default route. The default route can be desired from an upstream provider in addition to the full BGP table. In case the upstream provider is supposed to announce only the default route, a simple filter will be applied to accept only the default prefix and nothing else.

6.2.3.2. Outbound filtering

The filters to be applied would most likely not differ much from the ones applied for Internet peers (Section 6.2.1.2). But different policies could be applied in case it is desired that a particular upstream does not provide transit to all the prefixes.

6.3. Prefix filtering recommendations for leaf networks

6.3.1. Inbound filtering

The leaf network will deploy the filters corresponding to the routes it is requesting from its upstream. In case a default route is requested, a simple inbound filter can be applied to accept only the default route (Section 6.1.6). In case the leaf network is not capable of listing the prefixes because the amount is too large (for example if it requires the full Internet routing table) then it should configure filters to avoid receiving bad announcements from its upstream:

- Prefixes not routable (Section 6.1.1)
- Routes too specific (Section 6.1.3)
- Prefixes belonging to local AS (Section 6.1.4)
- The default route (Section 6.1.6) depending if the route is requested or not

6.3.2. Outbound filtering

A leaf network will most likely have a very straightforward policy: it will only announce its local routes. It can also configure the following prefixes filters described in Section 6.2.1.2 to avoid announcing invalid routes to its upstream provider.

7. BGP route flap dampening

The BGP route flap dampening mechanism makes it possible to give penalties to routes each time they change in the BGP routing table. Initially this mechanism was created to protect the entire Internet from multiple events impacting a single network. Studies have shown that implementations of BGP route flap dampening could cause more harm than they solve problems and therefore RIPE community has in the past recommended not using BGP route flap dampening [18]. Studies have then been conducted to propose new route flap dampening thresholds in order to make the solution "usable", see RFC 7196 [6] and RIPE has reviewed its recommendations in [21]. Authors of this document propose to follow IETF and RIPE recommendations and only use BGP route flap dampening with adjusted configured thresholds.

8. Maximum prefixes on a peering

It is RECOMMENDED to configure a limit on the number of routes to be accepted from a peer. Following rules are generally RECOMMENDED:

- From peers, it is RECOMMENDED to have a limit lower than the number of routes in the Internet. This will shut down the BGP peering if the peer suddenly advertises the full table. Network administrators can also configure different limits for each peer, according to the number of routes they are supposed to advertise plus some headroom to permit growth.

- From upstreams which provide full routing, it is RECOMMENDED to have a limit higher than the number of routes in the Internet. A limit is still useful in order to protect the network (and in particular the routers' memory) if too many routes are sent by the upstream. The limit should be chosen according to the number of routes that can actually be handled by routers.

It is important to regularly review the limits that are configured as the Internet can quickly change over time. Some vendors propose mechanisms to have two thresholds: while the higher number specified will shutdown the peering, the first threshold will only trigger a log and can be used to passively adjust limits based on observations made on the network.

9. AS-path filtering

This section lists the RECOMMENDED practices when processing BGP AS-paths:

- Network administrators SHOULD accept from customers only AS(4)-Paths containing ASNs belonging to (or authorized to transit

through) the customer. If network administrators can not build and generate filtering expressions to implement this, they SHOULD consider accepting only path lengths relevant to the type of customer they have (as in, if these customers are a leaf or have customers of their own), and try to discourage excessive prepending in such paths. This loose policy could be combined with filters for specific AS(4)-Paths that must not be accepted if advertised by the customer, such as upstream transit provider or peer ASNs.

- Network administrators SHOULD NOT accept prefixes with private AS numbers in the AS-path except from customers. Exception: an upstream offering some particular service like black-hole origination based on a private AS number. Customers should be informed by their upstream in order to put in place ad-hoc policy to use such services.

- Network administrators SHOULD NOT accept prefixes when the first AS number in the AS-path is not the one of the peer unless the peering is done toward a BGP route-server [17] (for example on an IXP) with transparent AS path handling. In that case this verification needs to be de-activated as the first AS number will be the one of an IXP member whereas the peer AS number will be the one of the BGP route-server.

- Network administrators SHOULD NOT advertise prefixes with non-empty AS-path unless they intend to be transit for these prefixes.

- Network administrators SHOULD NOT advertise prefixes with upstream AS numbers in the AS-path to their peering AS unless they intend to be transit for these prefixes.

- Private AS numbers are conventionally used in contexts that are "private" and SHOULD NOT be used in advertisements to BGP peers that are not party to such private arrangements, and should be stripped when received from BGP peers that are not party to such private arrangements.

- Network administrators SHOULD NOT override BGP's default behavior accepting their own AS number in the AS-path. In case an exception to this is required, impacts should be studied carefully as this can create severe impact on routing.

AS-path filtering should be further analyzed when ASN renumbering is done. Such operation is common and mechanisms exist to allow smooth ASN migration [28]. The usual migration technique, local to a router, consists in modifying the AS-path so it is presented to a peer with the previous ASN, as if no renumbering was done. This

makes it possible to change ASN of a router without reconfiguring all eBGP peers at the same time (as this operation would require synchronization with all peers attached to that router). During this renumbering operation, rules described above may be adjusted.

10. Next-Hop Filtering

If peering on a shared network, like an IXP, BGP can advertise prefixes with a 3rd-party next-hop, thus directing packets not to the peer announcing the prefix but somewhere else.

This is a desirable property for BGP route-server setups [17], where the route-server will relay routing information, but has neither capacity nor desire to receive the actual data packets. So the BGP route-server will announce prefixes with a next-hop setting pointing to the router that originally announced the prefix to the route-server.

In direct peerings between ISPs, this is undesirable, as one of the peers could trick the other one to send packets into a black hole (unreachable next-hop) or to an unsuspecting 3rd party who would then have to carry the traffic. Especially for black-holing, the root cause of the problem is hard to see without inspecting BGP prefixes at the receiving router at the IXP.

Therefore, an inbound route policy SHOULD be applied on IXP peerings in order to set the next-hop for accepted prefixes to the BGP peer IP address (belonging to the IXP LAN) that sent the prefix (which is what "next-hop-self" would enforce on the sending side).

This policy SHOULD NOT be used on route-server peerings, or on peerings where network administrators intentionally permit the other side to send 3rd-party next-hops.

This policy also SHOULD be adjusted if Remote Triggered Black Holing best practice (aka RTBH - RFC 6666 [13]) is implemented. In that case network administrators would apply a well-known BGP next-hop for routes they want to filter (if an Internet threat is observed from/to this route for example). This well known next-hop will be statically routed to a null interface. In combination with unicast RPF check, this will discard traffic from and toward this prefix. Peers can exchange information about black-holes using for example particular BGP communities. Network administrators could propagate black-holes information to their peers using agreed BGP community: when receiving a route with that community a configured policy could change the next-hop in order to create the black hole.

11. BGP community scrubbing

Optionally we can consider the following rules on BGP AS-paths:

- Network administrators SHOULD scrub inbound communities with their number in the high-order bits, and allow only those communities that customers/peers can use as a signaling mechanism

- Networks administrators SHOULD NOT remove other communities applied on received routes (communities not removed after application of previous statement). In particular they SHOULD keep original communities when they apply a community. Customers might need them to communicate with upstream providers. In particular network administrators SHOULD NOT (generally) remove the no-export community as it is usually announced by their peer for a certain purpose.

12. Change logs

!!! NOTE TO THE RFC EDITOR: THIS SECTION WAS ADDED TO TRACK CHANGES AND FACILITATE WORKING GROUP COLLABORATION. IT MUST BE DELETED BEFORE PUBLICATION !!!

12.1. Diffs between draft-jdurand-bgp-security-01 and draft-jdurand-bgp-security-00

Following changes have been made since previous document draft-jdurand-bgp-security-00:

- "This documents" typo corrected in the former abstract

- Add normative reference for RFC5082 in former section 3.2

- "Non routable" changed in title of former section 4.1.1

- Correction of typo for IPv4 loopback prefix in former section 4.1.1.1

- Added shared transition space 100.64.0.0/10 in former section 4.1.1.1

- Clarification that 2002::/16 6to4 prefix can cross network boundaries in former section 4.1.1.2

- Rationale of 2000::/3 explained in former section 4.1.1.2

- o Added 3FFE::/16 prefix forgotten initially in the simplified list of prefixes that must not be routed by definition in former section 4.1.1.2

- o Warn that filters for prefixes not allocated by IANA MUST only be done if regular refresh is guaranteed, with some words about the IPv4 experience, in former section 4.1.2.1

- o Replace RIR database with IRR. A definition of IRR is added in former section 4.1.2.2

- o Remove any reference to anti-spoofing in former section 4.1.4

- o Clarification for IXP LAN prefix and pMTUd problem in former section 4.1.5

- o "Autonomous filters" typo (instead of Autonomous systems) corrected in the former section 4.2

- o Removal of an example for manual address validation in former section 4.2.2.1

- o RFC5735 obsoletes RFC3300

- o Ingress/Egress replaced by Inbound/Outbound in all the document

12.2. Diffs between draft-jdurand-bgp-security-02 and draft-jdurand-bgp-security-01

Following changes have been made since previous document draft-jdurand-bgp-security-01:

- o 2 documentation prefixes were forgotten due to errata in RFC5735. But all prefixes were removed from that document which now point to other references for sake of not creating a new "registry" that would become outdated sooner or later

- o Change MD5 section with global TCP security session and introducing TCP-AO in former section 3.1. Added reference to BCP38

- o Added new section 3 about BGP router protection with forwarding plane ACL

- o Change text about prefix acceptable specificity in former section 4.1.3 to explain this doc does not try to make recommendations

- Refer as much as possible to existing registries to avoid creating a new one in former section 4.1.1.1 and 4.1.1.2

- Abstract reworded

- 6to4 exception described (only more specifics MUST be filtered)

- More specific -> more specifics

- should -> MUST for the prefixes an ISP needs to filter from its customers in former section 4.2.2.1

- Added "plus some headroom to permit growth" in former section 7

- Added new section on Next-Hop filtering

12.3. Diffs between draft-ietf-opsec-bgp-security-00 and draft-jdurand-bgp-security-02

Following changes have been made since previous document draft-jdurand-bgp-security-02:

- Added a subsection for RTBH in next-hop section with reference to RFC6666

- Changed last sentence of introduction

- Many edits throughout the document

- Added definition of tier 1 transit provider

- Removed definition of a BGP peering

- Removed description of routing policies for IPv6 prefixes in IANA special registry as this now contains a routing scope field

- Added reference to RFC6598 and changed the IPv4 prefixes to be filtered by definition section

- IXP added in accronym/definition section and only term used throughout the doc now

12.4. Diffs between draft-ietf-opsec-bgp-security-01 and draft-ietf-opsec-bgp-security-00

Following changes have been made since previous document draft-ietf-opsec-bgp-security-00:

- Obsolete RFC2385 moved from normative to informative reference
- Clarification of preference of TCP-AO over MD5 in former section 4.1
- Mentioning KARP efforts in TCP session protection section in former section 4 and adding 3 RFC as informative references: 6518, 6862 and 6952
- Removing reference to SIDR working-group
- Better dissociating origin validation and path validation to clarify what's potentially available for deployment
- Adding that SIDR mechanisms should be implemented in addition to the other ones mentioned throughout this document
- Added a paragraph in former section 8 about ASN renumbering
- Change of security considerations section
- Added the newly created IANA IPv4 Special Purpose Address Registry instead of references to RFCs listing these addresses

12.5. Diffs between draft-ietf-opsec-bgp-security-02 and draft-ietf-opsec-bgp-security-01

Following changes have been made since previous document draft-ietf-opsec-bgp-security-01:

- Added a reference to draft-ietf-sidr-origin-ops
- Added a reference to RFC6811 and RFC6907
- Changes "Most of RIR's" to "A majority of RIR's" on IRR availability
- Various edits
- Added NIST BGP security recommendations document
- Added that it's possible to get info from ISPs from RADB
- Correction of the url for IPv4 special use prefixes repository
- Clarification of the fact only prefixes with Global Scope set to False MUST be discarded

- o IANA list could be pulled directly by routers (not just pushed on routers).

- o Warning added when prefixes are checked against IRR

- o Recommend network operators to sign their routing objects

- o Recommend network operators to publish their routing objects in IRR of their IRR when available

- o Dissociate rules for local AS and downstreams in former section 5.1.4

12.6. Diffs between draft-ietf-opsec-bgp-security-03 and draft-ietf-opsec-bgp-security-02

Following changes have been made since previous document draft-ietf-opsec-bgp-security-02:

- o Added a note on TCP-AO to be preferred over MD5

- o Mention that loose AS filtering with customers can be combined with precise filters for important ASNs (example those of transits) that are must not be received on theses peers in former section 8.

- o MD5 removed from abstract

- o recommended -> RECOMMENDED where appropriate

- o Reference to BCP38 and BCP84 in former section 4.1

- o Added a note to RFC Editor to remove change section before publication

- o Removal of "future work" section

- o Added rate-limiting in addition to filtering in former section 3

- o Reference to IRRToolSet in former section 5.1.2.3

- o Removed "foreword" section

12.7. Diffs between draft-ietf-opsec-bgp-security-04 and draft-ietf-opsec-bgp-security-03

Following changes have been made since previous document draft-ietf-opsec-bgp-security-03:

- RFC6890 updates RFC5735
- RFC6890 updates RFC5156
- Removed reference RFC2234 and RFC 4234
- Moved route-server draft into informative reference section

12.8. Diffs between draft-ietf-opsec-bgp-security-05 and draft-ietf-opsec-bgp-security-04

Following changes have been made since previous document draft-ietf-opsec-bgp-security-04:

- RFC7196 updates draft-ietf-idr-rfd-usable
- RFC7115 updates draft-ietf-sidr-origin-ops
- draft-ietf-idr-ix-bgp-route-server-05 updates ietf-idr-ix-bgp-route-server-00

12.9. Diffs between draft-ietf-opsec-bgp-security-06 and draft-ietf-opsec-bgp-security-05

Following changes have been made since previous document draft-ietf-opsec-bgp-security-05:

- Wording improvements
- Introduction improved
- References are expanded (not just reference numbers are displayed but also the title of the document
- First occurence of accronyms expanded
- GTSM for multi-hop peerings
- Remove eBGP as protected by BCP38
- Add a caveat for IPsec for session protection

- o Changed MUST for SHOULD everywhere

- o Small changes in communities section

- o Removed simplified IPv6 prefix list

- o Removed note in section 9 about 32 bits ASN

- o IXP LAN prefix example in appendix

- o Make sure all references are in the text. Most of them were removed as they were initially here for previous version when IANA registries with routing scopes did not exist

13. Acknowledgements

The authors would like to thank the following people for their comments and support: Marc Blanchet, Ron Bonica, Randy Bush, David Freedman, Wesley George, Daniel Ginsburg, David Groves, Mike Hugues, Joel Jaeggli, Tim Kleefass, Warren Kumari, Jacques Latour, Lionel Morand, Jerome Nicolle, Hagen Paul Pfeifer, Thomas Pinaud, Carlos Pignataro, Jean Rebiffe, Donald Smith, Kotikalapudi Sriram, Matjaz Straus, Tony Tauber, Gunter Van de Velde, Sebastian Wiesinger, Matsuzaki Yoshinobu.

Authors would like to thank once again Gunter Van de Velde for presenting the draft at several IETF meetings in various working groups, indeed helping dissemination of this document and gathering of precious feedback.

14. IANA Considerations

This memo includes no request to IANA.

15. Security Considerations

This document is entirely about BGP operational security. It depicts best practices that one should adopt to secure its BGP infrastructure: protecting BGP router and BGP sessions, adopting consistent BGP prefix and AS-path filters and configure other options to secure the BGP network.

On the other hand this document doesn't aim at depicting existing BGP implementations and their potential vulnerabilities and ways they handle errors. It does not detail how protection could be enforced against attack techniques using crafted packets.

16. References

16.1. Normative References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <http://xml.resource.org/public/rfc/html/rfc2119.html>.

[2] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[3] Gill, V., Heasley, J., Meyer, D., Savola, P., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC 5082, October 2007.

[4] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

[5] Mohapatra, P., Scudder, J., Ward, D., Bush, R., and R. Austein, "BGP Prefix Origin Validation", RFC 6811, January 2013.

[6] Pelsser, C., Bush, R., Patel, K., Mohapatra, P., and O. Maennel, "Making Route Flap Damping Usable", RFC 7196, May 2014.

16.2. Informative References

[7] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385, August 1998.

[8] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

[9] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, March 2004.

[10] Blunk, L., Damas, J., Parent, F., and A. Robachevsky, "Routing Policy Specification Language next generation (RPSLng)", RFC 4012, March 2005.

[11] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, March 2011.

[12] Lepinski, M. and S. Kent, "An Infrastructure to Support Secure Internet Routing", RFC 6480, February 2012.

[13] Hilliard, N. and D. Freedman, "A Discard Prefix for IPv6", RFC 6666, August 2012.

[14] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, May 2013.

[15] Bush, R., "Origin Validation Operation Based on the Resource Public Key Infrastructure (RPKI)", BCP 185, RFC 7115, January 2014.

[16] Kent, S. and A. Chi, "Threat Model for BGP Path Security", RFC 7132, February 2014.

[17] "Internet Exchange Route Server", <http://tools.ietf.org/id/draft-ietf-idr-ix-bgp-route-server-05.txt>.

[18] Smith, P. and C. Panigl, "RIPE-378 - RIPE Routing Working Group Recommendations On Route-flap Damping", May 2006.

[19] Smith, P., Evans, R., and M. Hughes, "RIPE-399 - RIPE Routing Working Group Recommendations on Route Aggregation", December 2006.

[20] Smith, P. and R. Evans, "RIPE-532 - RIPE Routing Working Group Recommendations on IPv6 Route Aggregation", November 2011.

[21] Smith, P., Bush, R., Kuhne, M., Pelsser, C., Maennel, O., Patel, K., Mohapatra, P., and R. Evans, "RIPE-580 - RIPE Routing Working Group Recommendations On Route-flap Damping", January 2013.

[22] "IANA IPv4 Special Purpose Address Registry", <http://www.iana.org/assignments/iana-ipv4-special-registry/iana-ipv4-special-registry.xhtml>.

[23] "IANA IPv6 Special Purpose Address Registry", <http://www.iana.org/assignments/iana-ipv6-special-registry/iana-ipv6-special-registry.xml>.

[24] "IANA IPv4 Address Space Registry", <http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml>.

[25] "IANA IPv6 Address Space Registry", <http://www.iana.org/assignments/ipv6-unicast-address-assignments/ipv6-unicast-address-assignments.xml>.

[26] "Routing Assets Database", <http://www.radb.net>.

[27] "Security Requirements for BGP Path Validation", <http://datatracker.ietf.org/doc/draft-ietf-sidr-bgpsec-reqs/>.

[28] "Autonomous System (AS) Migration Features and Their Effects on the BGP AS_PATH Attribute", <http://datatracker.ietf.org/doc/draft-ga-idr-as-migration/>.

[29] "IRRToolSet project page", <http://irrtoolset.isc.org>.

[30] Cooper, D., Heilman, E., Brogle, K., Reyzin, L., and S. Goldberg, "On the Risk of Misbehaving RPKI Authorities", <http://www.cs.bu.edu/~goldbe/papers/hotRPKI.pdf>.

Appendix A. IXP LAN prefix filtering - example

An IXP in the RIPE region is allocated an IPv4 /22 prefix by RIPE NCC (X.Y.0.0/22 in this example) and uses a /23 of this /22 for the IXP LAN (let say X.Y.0.0/23). This IXP LAN prefix is the one used by IXP members to configure eBGP peerings. The IXP could also be allocated an AS number (AS64496 in our example).

Any IXP member SHOULD make sure it filters prefixes more specific than X.Y.0.0/23 from all its eBGP peers. If it received X.Y.0.0/24 or X.Y.1.0/24 this could seriously impact its routing.

The IXP SHOULD originate X.Y.0.0/22 and advertise it to its members through an eBGP peering (most likely from its BGP route servers, configured with AS64496).

The IXP members SHOULD accept the IXP prefix only if it passes the IRR generated filters (see Section 6.1.2.2.1)

IXP members SHOULD then advertise X.Y.0.0/22 prefix to their downstreams. This announce would pass IRR based filters as it is originated by the IXP.

Authors' Addresses

Jerome Durand
CISCO Systems, Inc.
11 rue Camille Desmoulins
Issy-les-Moulineaux 92782 CEDEX
FR

Email: jerduran@cisco.com

Ivan Pepelnjak
NIL Data Communications
Tivolska 48
Ljubljana 1000
Slovenia

Email: ip@ipspace.net

Gert Doering
SpaceNet AG
Joseph-Dollinger-Bogen 14
Muenchen D-80807
Germany

Email: gert@space.net