                      BGP Operations and Security

Abstract

   The Border Gateway Protocol (BGP) 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
   such as Time to Live (TTL), the TCP Authentication Option (TCP-AO),
   and control-plane filtering. It also describes measures to better
   control the flow of routing information, using prefix filtering and
   automation of prefix filters, max-prefix filtering, Autonomous System
   (AS) path filtering, route flap dampening, and BGP community
   scrubbing.

Status of This Memo

   This memo documents an Internet Best Current Practice.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   BCPs is available in Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc7454.

Table of Contents

1. Introduction

   The Border Gateway Protocol (BGP), specified in 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 whether the 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.

1.1. 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].

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 Acronyms

   o ACL: Access Control List

   o ASN: Autonomous System Number

   o IRR: Internet Routing Registry

   o IXP: Internet Exchange Point

   o LIR: Local Internet Registry

   o PMTUD: Path MTU Discovery

   o RIR: Regional Internet Registry

   o Tier 1 transit provider: an IP transit provider that can reach any
      network on the Internet without purchasing transit services.

   o uRPF: Unicast Reverse Path Forwarding

   In addition to the list above, the following terms are used with a
   specific meaning.

   o Downstream: any network that is downstream; it can be a provider
      or a customer network.

   o Upstream: any network that is upstream.

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) that 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 the 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 cannot 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 surpasses
   platform capabilities.

   Filtering and rate-limiting of control-plane traffic is a wider topic
   than "just for BGP". (If a network administrator brings down a
   router by overloading one of the other protocols remotely, 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
   subsections list the major points raised in this RFC and give the
   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 the TCP session header, described in RFC 2385 [7], was
   the first such mechanism. It has been obsoleted by the 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 vendors' equipment, TCP-AO SHOULD be preferred when
   implemented.

   IPsec could also be used for session protection. At the time of
   publication, there is not enough experience of the impact of using
   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 the maintenance of authentication
   information (for example, MD5 passwords). 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.

   Furthermore, 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 Internal BGP (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 (GTSM 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 an IP host that is not directly connected,
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 a proper value
depending on the distance between BGP speakers (using the principle
described above). Nevertheless, it is not as effective because
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 and advertised on the BGP peerings. Prefixes
exchanged between BGP peers are controlled with inbound and outbound
filters that can match on IP prefixes (as described in this section),
AS paths (as described in Section 9) or any other attributes of a BGP
prefix (for example, BGP communities, as described in Section 11).

6.1. Definition of Prefix Filters

This section lists the most commonly used prefix filters. The
following sections will clarify where these filters should be
applied.

6.1.1. Special-Purpose Prefixes

6.1.1.1. IPv4 Special-Purpose Prefixes

The IANA IPv4 Special-Purpose Address Registry [23] maintains the
list of IPv4 special-purpose prefixes and their routing scope, and it
SHOULD be used for prefix-filter configuration. Prefixes with value
"False" in column "Global" SHOULD be discarded on Internet BGP
peerings.

6.1.1.2. IPv6 Special-Purpose Prefixes

The IANA IPv6 Special-Purpose Address Registry [24] maintains the
list of IPv6 special-purpose prefixes and their routing scope, and it
SHOULD be used for prefix-filter configuration. Only prefixes with
value "False" in column "Global" SHOULD be discarded on Internet BGP
peerings.

6.1.2. Unallocated Prefixes

   IANA allocates prefixes to RIRs that 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
   unallocated prefixes 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 that
   prefixes they receive from BGP peers are in the IANA-allocated IPv4
   address space [25]. 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
   to accept only prefixes derived from those allocated by IANA.
   Administrators can dynamically build this list from the IANA-
   allocated IPv6 space [26]. 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 a
   maximum of one month after any change in the list of IPv6 prefixes
   allocated by IANA.

   If the process in place (whether 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 the latest allocations, and
   required extra work for RIRs that had to "de-bogonize" the newly
   allocated prefixes. (See [18] for information on de-bogonizing.)

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 (ASes) 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, two options can be considered:
   short- and long-term options. They are described in the following
   subsections.

6.1.2.2.1. Prefix Filters Created from Internet Routing Registries
            (IRRs)

   An Internet Routing Registry (IRR) is a database containing Internet
   routing information, described using Routing Policy Specification
   Language objects as described in RFC 4012 [10]. Network
   administrators are given privileges to describe routing policies of
   their own networks in the IRR, and that information is published,
   usually publicly. A majority of Regional Internet Registries do also
   operate an IRR and can control whether registered routes conform to
   the prefixes that are 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 AS, a list of originated or transited prefixes that one may
   accept. This can be done relatively easily using scripts and
   existing tools capable of retrieving this information from the
   registries. This approach is exactly the same for both IPv4 and
   IPv6.

   The macro-algorithm for the script is as follows. For the peer that
   is considered, the distant network administrator has provided the AS
   and may be able to provide an AS-SET object (aka AS-MACRO). An
   AS-SET is an object that 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 look 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, it is difficult to choose 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 [27], 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. Refreshing the filters on a daily basis SHOULD be
considered because routing changes must sometimes be done in an
emergency and registries may be updated at the very last moment.
Note that this approach significantly increases the complexity of the
router configurations, as it can quickly add tens of thousands of
configuration lines for some important peers. To manage this
complexity, network administrators could use, for example, IRRToolSet
[30], a set of tools making it possible to simplify the creation of
automated filter configuration from policies stored in an IRR.

Last but not least, network administrators SHOULD publish and
maintain their resources properly in the 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 of writing this document, 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 RIRs. There are basically two services that SIDR
offers:

o Origin validation, described in RFC 6811 [5], seeks to make sure
   that attributes associated with routes are correct. (The major
   point is the validation of the AS number originating a given
   route.) Origin validation is now operational (Internet

      registries, protocols, implementations on some routers), and in
      theory it can be implemented knowing that the number of signed
      resources is still low at the time of writing this document.

   o Path validation provided by BGPsec [29] seeks to make sure that no
      one announces fake/wrong BGP paths that would attract traffic for
      a given destination; see RFC 7132 [16]. BGPsec is still an
      ongoing work item at the time of writing this document and
      therefore cannot be implemented.

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

If route origin validation is implemented, the reader SHOULD refer to
the rules described in RFC 7115 [15]. In short, each external route
received on a router SHOULD be checked against the Resource Public
Key Infrastructure (RPKI) data set:

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

o If the ROA is found and is INVALID, then the prefix SHOULD be
   discarded.

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

In addition to this, network administrators SHOULD 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"
[31] explains how the RPKI model can impact the Internet if
authorities don't behave as they are supposed to. Further analysis
is certainly required on RPKI, which carries part of BGP security.

6.1.3. Prefixes That Are Too Specific

   Most ISPs will not accept advertisements beyond a certain level of
   specificity (and in return, they do not announce prefixes they
   consider to be too specific). That acceptable specificity is decided
   for each peering between the two 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 notes
   that there are existing practices on the Internet and recommends that
   the reader refer to them. As an example, the RIPE community has
   documented that, at the time of writing of this document, IPv4
   prefixes longer than /24 and IPv6 prefixes longer than /48 are
   generally neither announced nor accepted in the Internet [20] [21].
   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 that may directly suffer if the
   backbone's prefix is suddenly preferred over the Internet.

   In some cases, for example, multihoming 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, when 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 even if 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 prevent other routers in the network from 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 over External BGP (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 -- this will
   usually be automatically achieved by filtering announcements using
   the IRR database.

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 (i.e., 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 the less specific route (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 that are too specific" filters
   described in Section 6.1.3. The easiest way to implement this is for
   the IXP itself to take care of the origination of its prefix and
   advertise it to all IXP members through a BGP peering. Most likely,
   the BGP route servers would be used for this, and the IXP would send
   its entire prefix, which would be equal to or less specific than the
   IXP LAN prefix.

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

6.1.6. The Default Route

6.1.6.1. IPv4

   Typically, the 0.0.0.0/0 prefix is not intended to be accepted or
   advertised except in specific customer/provider configurations;
   general filtering outside of these is RECOMMENDED.

6.1.6.2. IPv6

   Typically, the ::/0 prefix is not intended to be accepted or
   advertised except 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
   Section 6.1.

6.2.1. Filters with Internet Peers

6.2.1.1. Inbound Filtering

   There are basically two 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 that are not globally routable (Section 6.1.1)

   o prefixes not allocated by IANA (IPv6 only) (Section 6.1.2.1)

   o routes that are 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, etc.). 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 the 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 that's used as the
source of information by the script is not fully trusted:

o prefixes that are not globally routable (Section 6.1.1)

o routes that are 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)

6.2.1.2. Outbound Filtering

The configuration should ensure 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 BGP communities, AS paths, or both. Also, it may be desirable
to add the following filters before any policy to avoid unwanted
route announcements due to bad configuration:

o Prefixes that are not globally routable (Section 6.1.1)

o Routes that are too specific (Section 6.1.3)

o IXP LAN prefixes (Section 6.1.5)

o The default route (Section 6.1.6)

If 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
   customer 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 when the customer is a network connecting other
   customers (for example, a Tier 1 transit provider connecting service
   providers). An exception is when the customer network applies strict
   inbound/outbound prefix filtering, and there are too many prefixes
   announced by that network 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
   the customer wants to receive. In the simplest possible scenario,
   the customer may want to receive only the default route; this can be
   done easily by applying a filter with the default route only.

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

   o prefixes that are not globally routable (Section 6.1.1)

   o routes that are too specific (Section 6.1.3)

   o the default route (Section 6.1.6)

   In some cases, the customer may desire to receive the default route
   in addition to the full BGP table. This can be done by the provider
   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

   If 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. Sometimes, the default

   route (in addition to the full BGP table) can be desired from an
   upstream provider. If 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). However,
   different policies could be applied if a particular upstream should
   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. If a default route is requested,
   a simple inbound filter can be applied to accept only the default
   route (Section 6.1.6). If the leaf network is not capable of listing
   the prefixes because there are too many (for example, if it requires
   the full Internet routing table), then it should configure the
   following filters to avoid receiving bad announcements from its
   upstream:

   o prefixes not routable (Section 6.1.1)

   o routes that are too specific (Section 6.1.3)

   o prefixes belonging to local AS (Section 6.1.4)

   o the default route (Section 6.1.6) depending on whether or not the
      route is requested

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
   prefix 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 that impact a single network. Studies have
   shown that implementations of BGP route flap dampening could cause

more harm than benefit; therefore, in the past, the RIPE community
has recommended against using BGP route flap dampening [19]. Later,
studies were conducted to propose new route flap dampening thresholds
in order to make the solution "usable"; see RFC 7196 [6] and [22] (in
which RIPE reviewed its recommendations). This document RECOMMENDS
following IETF and RIPE recommendations and using BGP route flap
dampening with the 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. The following rules are generally RECOMMENDED:

   o 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.

   o From upstreams that 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 shut down 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.

   o Network administrators SHOULD accept from customers only 2-byte or
      4-byte AS paths containing ASNs belonging to (or authorized to
      transit through) the customer. If network administrators cannot
      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 SHOULD try to discourage
      excessive prepending in such paths. This loose policy could be

combined with filters for specific 2-byte or 4-byte AS paths that
must not be accepted if advertised by the customer, such as
upstream transit providers or peer ASNs.

o Network administrators SHOULD NOT accept prefixes with private AS
   numbers in the AS path unless the prefixes are from customers. An
   exception could occur when an upstream is offering some particular
   service like black-hole origination based on a private AS number:
   in that case, prefixes SHOULD be accepted. Customers should be
   informed by their upstream in order to put in place ad hoc policy
   to use such services.

o Network administrators SHOULD NOT accept prefixes when the first
   AS number in the AS path is not the one of the peer's 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 deactivated, 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.

o Network administrators SHOULD NOT advertise prefixes with a
   nonempty AS path unless they intend to provide transit for these
   prefixes.

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

o 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 they SHOULD
   be stripped when received from BGP peers that are not party to
   such private arrangements.

o Network administrators SHOULD NOT override BGP's default behavior,
   i.e., they should not accept their own AS number in the AS path.
   When considering an exception, the impact (which may be severe on
   routing) should be studied carefully.

AS path filtering should be further analyzed when ASN renumbering is
done. Such an 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 the ASN of a router without reconfiguring
all EBGP peers at the same time (as that operation would require
synchronization with all peers attached to that router). During this
renumbering operation, the 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 third-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 the
   capacity nor the 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 into sending packets into a black
   hole (unreachable next hop) or to an unsuspecting third 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 of 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 third-party next hops.

   This policy also SHOULD be adjusted if the best practice of Remote
   Triggered Black Holing (aka RTBH as described in 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 a 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-hole information to their peers
   using an agreed-upon 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:

   o 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

   o Networks administrators SHOULD NOT remove other communities
      applied on received routes (communities not removed after
      application of the 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. Security Considerations

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

   This document does not aim to describe existing BGP implementations,
   their potential vulnerabilities, or ways they handle errors. It does
   not detail how protection could be enforced against attack techniques
   using crafted packets.

13. References

13.1. Normative References

   [1]   Bradner, S., "Key words for use in RFCs to Indicate Requirement
          Levels", RFC 2119, March 1997,
          <http://www.rfc-editor.org/info/rfc2119>.

   [2]   Rekhter,, Y., Li,, T., and S. Hares,, "A Border Gateway
          Protocol 4 (BGP-4)", RFC 4271, January 2006,
          <http://www.rfc-editor.org/info/rfc4271>.

   [3]   Gill, V., Heasley, J., Meyer, D., Savola,, P., and C.
          Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC
          5082, October 2007, <http://www.rfc-editor.org/info/rfc5082>.

   [4]   Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication
          Option", RFC 5925, June 2010,
          <http://www.rfc-editor.org/info/rfc5925>.

   [5]   Mohapatra, P., Scudder, J., Ward, D., Bush, R., and R.
          Austein, "BGP Prefix Origin Validation", RFC 6811, January
          2013, <http://www.rfc-editor.org/info/rfc6811>.

   [6]   Pelsser, C., Bush, R., Patel, K., Mohapatra, P., and O.
          Maennel, "Making Route Flap Damping Usable", RFC 7196, May
          2014, <http://www.rfc-editor.org/info/rfc7196>.

13.2. Informative References

   [7]   Heffernan, A., "Protection of BGP Sessions via the TCP MD5
          Signature Option", RFC 2385, August 1998,
          <http://www.rfc-editor.org/info/rfc2385>.

   [8]   Ferguson, P. and D. Senie, "Network Ingress Filtering:
          Defeating Denial of Service Attacks which employ IP Source
          Address Spoofing", RFC 2827, May 2000,
          <http://www.rfc-editor.org/info/rfc2827>.

   [9]   Baker, F. and P. Savola, "Ingress Filtering for Multihomed
          Networks", RFC 3704, March 2004,
          <http://www.rfc-editor.org/info/rfc3704>.

   [10]   Blunk, L., Damas, J., Parent, F., and A. Robachevsky, "Routing
          Policy Specification Language next generation (RPSLng)", RFC
          4012, March 2005, <http://www.rfc-editor.org/info/rfc4012>.

   [11]   Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router
          Control Plane", RFC 6192, March 2011,
          <http://www.rfc-editor.org/info/rfc6192>.

   [12]   Lepinski, M. and S. Kent, "An Infrastructure to Support Secure
          Internet Routing", RFC 6480, February 2012,
          <http://www.rfc-editor.org/info/rfc6480>.

   [13]   Hilliard, N. and D. Freedman, "A Discard Prefix for IPv6", RFC
          6666, August 2012, <http://www.rfc-editor.org/info/rfc6666>.

   [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, <http://www.rfc-editor.org/info/rfc6952>.

   [15] Bush, R., "Origin Validation Operation Based on the Resource
         Public Key Infrastructure (RPKI)", RFC 7115, January 2014,
         <http://www.rfc-editor.org/info/rfc7115>.

   [16] Kent, S. and A. Chi, "Threat Model for BGP Path Security", RFC
         7132, February 2014, <http://www.rfc-editor.org/info/rfc7132>.

   [17] Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker,
         "Internet Exchange Route Server", Work in Progress,
         draft-ietf-idr-ix-bgp-route-server-06, December 2014.

   [18] Karrenberg, D., "RIPE-351 - De-Bogonising New Address Blocks",
         October 2005.

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

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

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

   [22] 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.

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

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

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

   [26] IANA, "Internet Protocol Version 6 Address Space",
         <http://www.iana.org/assignments/ipv6-address-space>.

   [27] Merit Network Inc., "Merit RADb", <http://www.radb.net>.

   [28] George, W. and S. Amante, "Autonomous System (AS) Migration
         Features and Their Effects on the BGP AS_PATH Attribute", Work
         in Progress, draft-ga-idr-as-migration-03, January 2014.

   [29] Bellovin, S., Bush, R., and D. Ward, "Security Requirements for
         BGP Path Validation", RFC 7353, August 2014,
         <http://www.rfc-editor.org/info/rfc7353>.

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

   [31] 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
   France

   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