IDR and SIDR
Internet-Draft
Intended status: Standards Track
Expires: September 7, 2017

K. Sriram
D. Montgomery
US NIST
B. Dickson

K. Patel
Arrcus
A. Robachevsky
Internet Society

March 6, 2017

Methods for Detection and Mitigation of BGP Route Leaks
draft-ietf-idr-route-leak-detection-mitigation-06

Abstract

RFC 7908 provides a definition of the route leak problem, and also enumerates several types of route leaks. This document first examines which of those route-leak types are detected and mitigated by the existing origin validation (OV) [RFC 6811]. It is recognized that OV offers a limited detection and mitigation capability against route leaks. This document specifies enhancements that significantly extend the route-leak prevention, detection, and mitigation capabilities of BGP. One solution component involves intra-AS messaging from ingress router to egress router using a BGP Community or Attribute. This intra-AS messaging prevents the AS from causing route leaks. Another solution component involves carrying a per-hop route-leak protection (RLP) field in BGP updates. The RLP fields are proposed to be carried in a new optional transitive attribute, called BGP RLP attribute. The RLP attribute helps with detection and mitigation of route leaks at ASes downstream from the leaking AS.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 7, 2017.

Table of Contents

1. Introduction

[RFC7908] provides a definition of the route leak problem, and also enumerates several types of route leaks. This document first examines which of those route-leak types are detected and mitigated by the existing Origin Validation (OV) [RFC6811] method. OV and BGPsec path validation [I-D.ietf-sidr-bgpsec-protocol] together offer mechanisms to protect against re-originations and hijacks of IP prefixes as well as man-in-the-middle (MITM) AS path modifications. Route leaks (see [RFC7908] and references cited at the back) are another type of vulnerability in the global BGP routing system against which OV offers only partial protection. BGPsec (i.e. path validation) provides cryptographic protection for some aspects of BGP update messages, but in its current form BGPsec doesn't offer any protection against route leaks.

For the types of route leaks enumerated in [RFC7908], where the OV method does not offer a solution, this document specifies enhancements that significantly extend the route-leak prevention, detection, and mitigation capabilities of BGP. One solution component involves intra-AS messaging from ingress router to egress router using a BGP Community or Attribute. This intra-AS messaging prevents the AS from causing route leaks. Another solution component involves carrying a per-hop route-leak protection (RLP) field in BGP updates. The RLP fields are proposed to be carried in a new optional transitive attribute, called BGP RLP attribute. The RLP attribute helps with detection and mitigation of route leaks at ASes downstream from the leaking AS.

The solution is meant to be initially implemented as an enhancement of BGP without requiring BGPsec. However, when BGPsec is deployed in the future, the solution can be incorporated in BGPsec, enabling cryptographic protection for the RLP field. That would be one way of implementing the proposed solution in a secure way. It is not claimed that the solution detects all possible types of route leaks but it detects several types, especially considering some significant route-leak occurrences that have been observed in recent years. The document also includes a stopgap method (in Section 5) for detection and mitigation of route leaks for an intermediate phase when OV is deployed but BGP protocol on the wire is unchanged.

2. Related Prior Work

A mechanism embodied in the proposed solution is based on setting an attribute in BGP route announcement to manage the transmission/ receipt of the announcement based on the type of neighbor (e.g. customer, transit provider, etc.). Documented prior work related to said basic idea and mechanism dates back to at least the 1980's. Some examples of prior work are: (1) Information flow rules described in [proceedings-sixth-ietf] (see pp. 195-196); (2) Link Type described in [RFC1105-obsolete] (see pp. 4-5); (3) Hierarchical Recording described in [draft-kunzinger-idrp-ISO10747-01] (see Section 6.3.1.12). The problem of route leaks and possible solution mechanisms based on encoding peering-link type information, e.g. P2C (i.e. Transit-Provider to Customer), C2P (i.e. Customer to Transit-Provider), p2p (i.e. peer to peer) etc., in BGPsec updates and protecting the same under BGPsec path signatures have been discussed in IETF SIDR WG at least since 2011. [draft-dickson-sidr-route-leak-solns] attempted to describe these mechanisms in a BGPsec context. The draft expired in 2012. [draft-dickson-sidr-route-leak-solns] defined neighbor relationships on a per link basis, but in the current document the relationship is encoded per prefix, as routes for prefixes with different peering relationships may be sent over the same link. Also [draft-dickson-sidr-route-leak-solns] proposed a second signature block for the link type encoding, separate from the path signature block in BGPsec. By contrast, in the current document when BGPsec-based solution is considered, cryptographic protection is provided for Route-Leak Protection (RLP) encoding using the same signature block as that for path signatures (see Section 4.3.2).

3. Do Origin Validation and BGPsec Assist in Route-Leak Detection?

Referring to the enumeration of route leaks discussed in [RFC7908], Table 1 summarizes the route-leak detection capability offered by OV and BGPsec for different types of route leaks. (Note: Prefix filtering is not considered here in this table. Please see Section 5.)

A detailed explanation of the contents of Table 1 is as follows. It is readily observed that route leaks of Types 1, 2, 3, and 4 are not detected by OV or BGPsec in its current form. Clearly, Type 5 route leak involves re-origination or hijacking, and hence can be detected by OV. In the case of Type 5 route leak, there would be no existing ROAs to validate a re-originated prefix or more specific, but instead a covering ROA would normally exist with the legitimate AS, and hence the update will be considered Invalid by OV.

Type of Route Leak	Current State of Detection Coverage
Type 1: Hairpin Turn with Full Prefix	Neither OV nor BGPsec (in its current form) detects Type 1.
Type 2: Lateral ISP-ISP-ISP Leak	Neither OV nor BGPsec (in its current form) detects Type 2.
Type 3: Leak of Transit-Provider Prefixes to Peer	Neither OV nor BGPsec (in its current form) detects Type 3.
Type 4: Leak of Peer Prefixes to Transit Provider	Neither OV nor BGPsec (in its current form) detects Type 4.
Type 5: Prefix Re-Origination with Data Path to Legitimate Origin	OV detects Type 5.
Type 6: Accidental Leak of Internal Prefixes and More Specifics	For internal prefixes never meant to be routed on the Internet, OV helps detect their leak; they might either have no covering ROA or have an AS0-ROA to always filter them. In the case of accidental leak of more specifics, OV may offer some detection due to ROA maxLength.

Table 1: Examination of Route-Leak Detection Capability of Origin Validation and Current BGPsec Path Validation

In the case of Type 6 leaks involving internal prefixes that are not meant to be routed in the Internet, they are likely to be detected by OV. That is because such prefixes might either have no covering ROA or have an AS0-ROA to always filter them. In the case of Type 6 leaks that are due to accidental leak of more specifics, they may be detected due to violation of ROA maxLength. BGPsec (i.e. path validation) in its current form does not detect Type 6. However, route leaks of Type 6 are least problematic due to the following reasons. In the case of leak of more specifics, the offending AS is itself the legitimate destination of the leaked more-specific prefixes. Hence, in most cases of this type, the data traffic is neither misrouted not denied service. Also, leaked announcements of Type 6 are short-lived and typically withdrawn quickly following the announcements. Further, the MaxPrefix limit may kick-in in some

receiving routers and that helps limit the propagation of sometimes large number of leaked routes of Type 6.

Realistically, BGPsec may take a much longer time being deployed than OV. Hence solution proposals for route leaks should consider both scenarios: (A) OV only (without BGPsec) and (B) OV plus BGPsec. Assuming an initial scenario A, and based on the above discussion and Table 1, it is evident that the solution method should focus primarily on route leaks of Types 1, 2, 3, and 4.

4. Mechanisms for Prevention, Detection and Mitigation of Route Leaks

There are two considerations for route leaks: (1) Prevention of route leaks from a local AS, and (2) Detection and mitigation of route leaks in ASes that are downstream from the leaking AS.

In Section 4.1, the method of ascertaining peering relationship per prefix is described. Section 4.2 describes intra-AS messaging methods for prevention of route leaks from local AS. Section 4.3 and Section 4.4 describe a simple addition to BGP that facilitates detection and mitigation of route leaks of Types 1, 2, 3, and 4 (see Section 3) at an downstream AS from the leaking AS.

4.1. Ascertaining Peering Relationship

There are four possible peering relationships (i.e. roles) an AS can have with a neighbor AS: (1) Provider: transit-provider for all prefixes exchanged, (2) Customer: customer for all prefixes exchanged, (3) Lateral Peer: lateral peer (i.e. non-transit) for all prefixes exchanged, and (4) Complex: different relationships for different sets of prefixes [I-D.ymbk-idr-bgp-open-policy] [Luckie]. On a per-prefix basis, the peering role types simplify to provider, customer, or lateral peer.

Operators rely on some form of out-of-band (OOB) (i.e. external to BGP) communication to exchange information about their peering relationship, AS number, interface IP address, etc. If the relationship is complex, the OOB communication also includes the sets of prefixes for which they have different roles. [I-D.ymbk-idr-bgp-open-policy] introduces a method of confirming the BGP Role during BGP OPEN messaging. It defines a new BGP Role capability, which helps in re-confirming the relationship. BGP Role does not replace the OOB communication since it relies on the OOB communication to set the role type in the BGP OPEN message. However, BGP Role provides a means to double check, and if there is a contradiction detected via the BGP Role messages, then a Role Mismatch Notification is sent [I-D.ymbk-idr-bgp-open-policy].

When the BGP relationship information has been correctly exchanged (i.e. free of contradictions) including the sets of prefixes with different roles (if complex), then this information SHOULD be used to set the role per-prefix with each peer. For example, if the local AS's role is Provider with a neighbor AS, then the per-prefix role is set to 'Provider' for all prefixes sent to the neighbor, and set to 'Customer' for all prefixes received from the neighbor.

4.2. Prevention of Route Leaks at Local AS: Intra-AS Messaging

Note: The intra-AS messaging for route leak prevention can be done using non-transitive BGP Community or Attribute. Both options are described below; one of them will be chosen after IDR working group consensus is established.

4.2.1. Non-Transitive BGP Community for Intra-AS Messaging

The following procedure (or similar) for intra-AS messaging (i.e. between ingress and egress routers) for prevention of route leaks is a fairly common practice used by large ISPs. (Note: This information was gathered from discussions on the NANOG mailing list [Nanog-thread-June2016] as well as through private discussions with operators of large ISP networks.)

Routes are tagged on ingress to an AS with communities for origin, including the type of eBGP peer it was learned from (customer, provider or lateral peer), geographic location, etc. The community attributes are carried across the AS with the routes. Routes that the AS originates directly are tagged with similar origin communities when they are redistributed into BGP from static, IGP, etc. These communities are used along with additional logic in route policies to determine which routes are to be announced to which eBGP peers and which are to be dropped. Route policy is applied to eBGP sessions based on what set of routes they should receive (transit, full routes, internal-only, default-only, etc.). In this process, the ISP's AS also ensures that routes learned from a transit-provider or a lateral peer (i.e. non-transit) at an ingress router are not leaked at an egress router to another transit-provider or lateral peer.

Additionally, in many cases, ISP network operators' outbound policies require explicit matches for expected communities before passing routes. This helps ensure that that if an update has made it into the routing table (i.e. RIB) but has missed its ingress community tagging (due to a missing/misapplied ingress policy), it will not be inadvertently leaked.

The above procedure (or a simplified version of it) is also applicable when an AS consists of a single eBGP router. It is

recommended that all AS operators SHOULD implement the procedure described above (or similar that is appropriate for their network) to prevent route leaks that they have direct control over.

4.2.2. Non-Transitive BGP pRLP Attribute for Intra-AS Messaging

It is possible to use an optional non-transitive BGP Attribute instead of the Community described above for intra-AS messaging for route leak prevention. The following description would be used in case the IDR working group decides on using a BGP Attribute.

A new optional non-transitive BGP Attribute called Preventive Route Leak Protection (pRLP) is used. The attribute type code for the pRLP attribute is to be assigned by IANA. The length of this attribute is 0 as it is used only as a flag.

Ingress (receiving) router action: The decision to set or not set the pRLP flag is made by a receiving router upon a route ingress. The flag is set when the route is received from a provider or a lateral peer. The flag is not set when the route is received from a customer. When the relationship is complex, the flag is set based on the per-prefix peering role information discussed in Section 4.1.

Egress (sending) router action: A sending router is allowed to send a route without the pRLP flag to any neighbor (transit-provider, customer, lateral peer). However, if the pRLP flag is present, then the route MUST NOT be sent to a transit-provider or a lateral peer.

An AS that follows the above set of receiver (ingress) and sender (egress) actions, prevents itself from causing route leaks.

4.3. Route-Leak Protection (RLP) Field Encoding by Sending Router

This section, Section 4.4 and Section 4.5 describe methods of detection and mitigation of route leaks in an AS downstream from the leaking AS.

The key principle is that, in the event of a route leak, a receiving router in a transit-provider AS (e.g. referring to Figure 1, ISP2 (AS2) router) should be able to detect from the update message that its customer AS (e.g. AS3 in Figure 1) SHOULD NOT have forwarded the update (towards the transit-provider AS). This means that at least one of the ASes in the AS path of the update has indicated that it sent the update to its customer or lateral (i.e. non-transit) peer, but forbade any subsequent 'Up' forwarding (i.e. from a customer AS to its transit-provider AS). For this purpose, a Route-Leak Protection (RLP) field to be set by a sending router is proposed to be used for each AS hop.

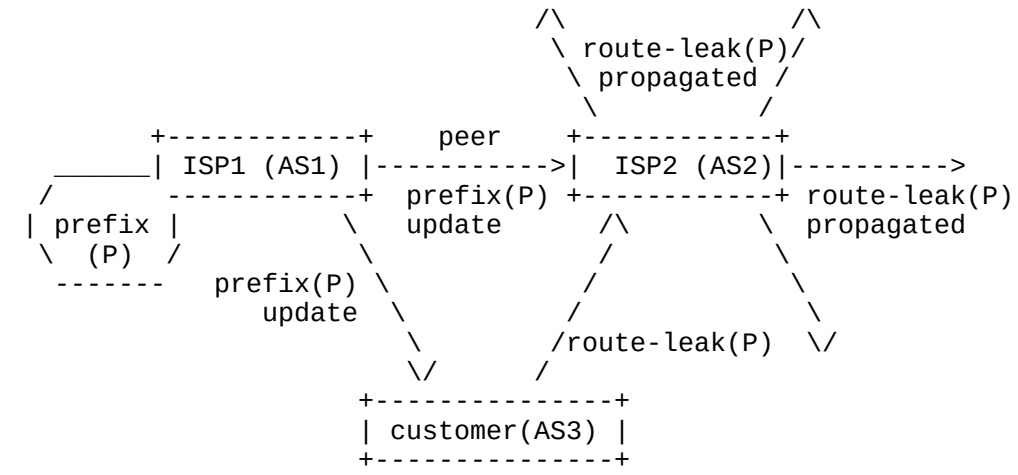

Figure 1: Illustration of the basic notion of a route leak.

For the purpose of route-leak detection and mitigation proposed in this document, the RLP field value SHOULD be set to one of two values as follows:

- 0: This is the default value (i.e. "nothing specified"),

- 1: This is the 'Do not Propagate Up or Lateral' indication; sender indicating that the route SHOULD NOT be forwarded 'Up' towards a transit-provider AS or to a lateral (i.e. non-transit) peer AS.

The RLP indications SHOULD be set on a per prefix basis. This is because some peering relations between neighbors can be complex (see Section 4.1). Further, the RLP indications are set on a per-hop (i.e. per AS) basis.

There are two different scenarios when a sending AS SHOULD set value 1 in the RLP field: (a) when sending the update to a customer AS, and (b) when sending the update to a lateral peer (i.e. non-transit) AS. In essence, in both scenarios, the intent of RLP = 1 is that the neighbor AS and any receiving AS along the subsequent AS path SHOULD NOT forward the update 'Up' towards its (receiving AS's) transit-provider AS or laterally towards its peer (i.e. non-transit) AS.

When sending an update 'Up' to a transit-provider AS, the RLP encoding SHOULD be set to the default value of 0. When a sending AS sets the RLP encoding to 0, it is indicating to the receiving AS that the update can be propagated in any direction (i.e. towards transit-provider, customer, or lateral peer).

The two-state specification in the RLP field (as described above) works for detection and mitigation of route leaks of Types 1, 2, 3, and 4 which are the focus here (see Section 4.4 and Section 4.5).

An AS MUST NOT rewrite/reset the values set by any preceding ASes in their respective RLP fields.

The proposed RLP encoding SHOULD be carried in BGP-4 [RFC4271] updates in a new BGP optional transitive attribute (see Section 4.3.1). In BGPsec, it SHOULD be carried in the Flags field (see Section 4.3.2).

4.3.1. BGP RLP Attribute

The BGP RLP attribute is a new BGP optional transitive attribute. The attribute type code for the RLP attribute is to be assigned by IANA. The length field of this attribute is 2 octets. The value field of the RLP attribute is defined as a set of one or more pairs of ASN (4 octets) and RLP (one octet) fields as described below (Figure 2).

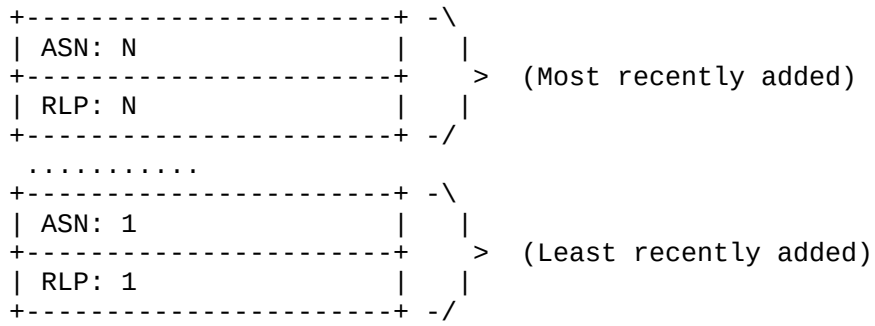

Figure 2: BGP RLP Attribute format.

The RLP Attribute value is a sequence of these two components (see Figure 2):

ASN: Four octets encoding the public registered AS number of a BGP speaker.

RLP Field: One octet encoding the RLP Field bits. The value of the RLP Field octet can be 0 (decimal) or 1 (decimal) as described above in Section 4.3.1. Its usage will be further discussed in subsequent sections.

If all ASes in the AS_PATH of a route are upgraded to participate in RLP, then the ASNs in the RLP TLV in Figure 2 will correspond one-to-one with sequence of ASes in the AS_PATH (excluding prepends). If some ASes do not participate, then one or more {ASN, RLP} tuples may be missing in the RLP attribute relative to the AS_PATH.

4.3.2. Carrying RLP Field Values in the BGPsec Flags

In BGPsec enabled routers, the RLP encoding SHOULD be accommodated in the existing Flags field in BGPsec updates. The Flags field is part of the Secure_Path Segment in BGPsec updates [I-D.ietf-sidr-bgpsec-protocol]. It is one octet long, and one Flags field is available for each AS hop, and currently only the first bit is used in BGPsec. So there are 7 bits that are currently unused in the Flags field. One of these bits can be designated for the RLP field value (see Section 4.3.1). This bit can be set to 0 when the RLP Field value is 0 and set to 1 when the RLP Field value is 1. Since the BGPsec protocol specification requires a sending AS to include the Flags field in the data that are signed over, the RLP field for each hop (assuming it would be part of the Flags field as described) will be protected under the sending AS's signature.

4.4. Recommended Actions at a Receiving Router for Detection of Route Leaks

The following receiver algorithm is RECOMMENDED for detecting route leaks:

A receiving router MUST mark an update as a 'Route Leak' if ALL of the following conditions hold true:

1. The update is received from a customer or lateral peer AS.

2. The update has the RLP Field set to 1 (i.e. 'Do not Propagate Up or Lateral') indication for one or more hops (excluding the most recent) in the AS path.

The reason for stating "excluding the most recent" in the above algorithm is as follows. An ISP should look at RLP values set by ASes preceding the immediate sending AS in order to ascertain a leak. The receiving router already knows that the most recent hop in the update is from its customer or lateral-peer AS to itself, and it does not need to rely on the RLP field value set by that AS (i.e the immediate neighbor AS in the AS path) for detection of route leaks.

If the RLP encoding is secured by BGPsec (see Section 4.3) and hence protected against tampering by intermediate ASes, then there would be

added certainty in the route-leak detection algorithm described above (see discussions in Section 6.1 and Section 6.2).

4.5. Possible Actions at a Receiving Router for Mitigation

After applying the above detection algorithm, a receiving router may use any policy-based algorithm of its own choosing to mitigate any detected route leaks. An example receiver algorithm for mitigating a route leak is as follows:

- If an update from a customer or lateral peer AS is marked as a 'Route Leak' (see Section 4.4), then the receiving router SHOULD prefer an alternate unmarked route.

- If no alternate unmarked route is available, then a route marked as a 'Route Leak' MAY be accepted.

A basic principle here is that if an AS receives and marks a customer route as 'Route Leak', then the AS should override the "prefer customer route" policy, and instead prefer an alternate 'clean' route learned from another customer, a lateral peer, or a transit provider. This can be implemented by adjusting the local preference for the routes in consideration.

5. Stopgap Solution when Only Origin Validation is Deployed

A stopgap method is described here for detection and mitigation of route leaks for the intermediate phase when OV is deployed but BGP protocol on the wire is unchanged. The stopgap solution can be in the form of construction of a prefix filter list from ROAs. A suggested procedure for constructing such a list comprises of the following steps:

- ISP makes a list of all the ASes (Cust_AS_List) that are in its customer cone (ISP's own AS is also included in the list). (Some of the ASes in Cust_AS_List may be multi-homed to another ISP and that is OK.)

- ISP downloads from the RPKI repositories a complete list (Cust_ROA_List) of valid ROAs that contain any of the ASes in Cust_AS_List.

- ISP creates a list of all the prefixes (Cust_Prfx_List) that are contained in any of the ROAs in Cust_ROA_List.

- Cust_Prfx_List is the allowed list of prefixes that is permitted by the ISP's AS, and will be forwarded by the ISP to upstream ISPs, customers, and peers.

- o A route for a prefix that is not in Cust_Prfx_List but announced by one of ISP's customers is 'marked' as a potential route leak. Further, the ISP's router SHOULD prefer an alternate route that is Valid (i.e. valid according to origin validation) and 'clean' (i.e. not marked) over the 'marked' route. The alternate route may be from a peer, transit provider, or different customer.

Special considerations with regard to the above procedure may be needed for DDoS mitigation service providers. They typically originate or announce a DDoS victim's prefix to their own ISP on a short notice during a DDoS emergency. Some provisions would need to be made for such cases, and they can be determined with the help of inputs from DDoS mitigation service providers.

For developing a list of all the ASes (Cust_AS_List) that are in the customer cone of an ISP, the AS path based Outbound Route Filter (ORF) technique [draft-ietf-idr-aspath-orf] can be helpful (see discussion in Section 6.4).

Another technique based on AS_PATH filters is described in [Snijders]. This method is applicable to very large ISPs (i.e. big networks) that have lateral peering. For a pair of such very large ISPs, say A and B, the method depends on ISP A communicating out-of-band (e.g. by email) with ISP B about whether or not it (ISP A) has any transit providers. This out-of-band knowledge enables ISP B to apply suitable AS_PATH filtering criteria for routes involving the presence of ISP A in the path and prevent certain kinds of route leaks (see [Snijders] for details).

6. Design Rationale and Discussion

This section provides design justifications for the methodology specified in Section 4, and also answers some questions that are anticipated or have been raised in the IETF IDR and SIDR working group meetings.

6.1. Is route-leak solution without cryptographic protection a serious attack vector?

It has been asked if a route-leak solution without BGPsec, i.e. when RLP Fields are not protected, can turn into a serious new attack vector. The answer seems to be: not really! Even the NLRI and AS_PATH in BGP updates are attack vectors, and RPKI/OV/BGPsec seek to fix that. Consider the following. Say, if 99% of route leaks are accidental and 1% are malicious, and if route-leak solution without BGPsec eliminates the 99%, then perhaps it is worth it (step in the right direction). When BGPsec comes into deployment, the route-leak protection (RLP) bits can be mapped into BGPsec (using the Flags

field) and then necessary security will be in place as well (within each BGPsec island as and when they emerge).

Further, let us consider the worst-case damage that can be caused by maliciously manipulating the RLP Field values in an implementation without cryptographic protection (i.e. sans BGPsec). Manipulation of the RLP bits can result in one of two types of attacks: (a) Upgrade attack and (b) Downgrade attack. Descriptions and discussions about these attacks follow. In what follows, P2C stands for transit provider to customer (Down); C2P stands for customer to transit provider (Up), and p2p stands for peer to peer (lateral or non-transit relationship).

(a) Upgrade attack: An AS that wants to intentionally leak a route would alter the RLP encodings for the preceding hops from 1 (i.e. 'Do not Propagate Up or Lateral') to 0 (default) wherever applicable. This poses no problem for a route that keeps propagating in the 'Down' (P2C) direction. However, for a route that propagates 'Up' (C2P) or 'Lateral' (p2p), the worst that can happen is that a route leak goes undetected. That is, a receiving router would not be able to detect the leak for the route in question by the RLP mechanism described here. However, the receiving router may still detect and mitigate it in some cases by applying other means such as prefix filters [RFC7454]. If some malicious leaks go undetected (when RLP is deployed without BGPsec) that is possibly a small price to pay for the ability to detect the bulk of route leaks that are accidental.

(b) Downgrade attack: RLP encoding is set to 1 (i.e. 'Do not Propagate Up or Lateral') when it should be set to 0 (default). This would result in a route being mis-detected and marked as a route leak. By default RLP encoding is set to 0, and that helps reduce errors of this kind (i.e. accidental downgrade incidents). Every AS or ISP wants reachability for prefixes it originates and for its customer prefixes. So an AS or ISP is not likely to change an RLP value 0 to 1 intentionally. If a route leak is detected (due to intentional or accidental downgrade) by a receiving router, it would prefer an alternate 'clean' route from a transit provider or peer over a 'marked' route from a customer. It may end up with a suboptimal path. In order to have reachability, the receiving router would accept a 'marked' route if there is no alternative that is 'clean'. So RLP downgrade attacks (intentional or accidental) would be quite rare, and the consequences do not appear to be grave.

6.2. Combining results of route-leak detection, OV and BGPsec validation for path selection decision

Combining the results of route-leak detection, OV, and BGPsec validation for path selection decision is up to local policy in a receiving router. As an example, a router may always give precedence to outcomes of OV and BGPsec validation over that of route-leak detection. That is, if an update fails OV or BGPsec validation, then the update is not considered a candidate for path selection. Instead, an alternate update is chosen that passed OV and BGPsec validation and additionally was not marked as route leak.

If only OV is deployed (and not BGPsec), then there are six possible combinations between OV and route-leak detection outcomes. Because there are three possible outcomes for OV (NotFound, Valid, and Invalid) and two possible outcomes for route-leak detection (marked as leak and not marked). If OV and BGPsec are both deployed, then there are twelve possible combinations between OV, BGPsec validation, and route-leak detection outcomes. As stated earlier, since BGPsec protects the RLP encoding, there would be added certainty in route-leak detection outcome if an update is BGPsec valid (see Section 6.1).

6.3. Are there cases when valley-free violations can be considered legitimate?

There are studies in the literature [Anwar] [Giotsas] [Wijchers] observing and analyzing the behavior of routes announced in BGP updates using data gathered from the Internet. In particular, the studies have focused on how often there appear to be valley-free (e.g. Gao-Rexford [Gao] model) violations, and if they can be explained [Anwar]. One important consideration for explanation of violations is per-prefix routing policies, i.e. routes for prefixes with different peering relationships may be sent over the same link. One encouraging result reported in [Anwar] is that when per-prefix routing policies are taken into consideration in the data analysis, more than 80% of the observed routing decisions fit the valley-free model (see Section 4.3 and SPA-1 data in Figure 2). [Anwar] also observes, "it is well known that this model [the basic Gao-Rexford model and some variations of it] fails to capture many aspects of the interdomain routing system. These aspects include AS relationships that vary based on the geographic region or destination prefix, and traffic engineering via hot-potato routing or load balancing." So there may be potential for explaining the remaining (20% or less) violations of valley-free as well.

One major design factor in the methodology described in this document is that the Route-Leak Protection (RLP) encoding is per prefix. So

the proposed solution is consistent with ISPs' per-prefix routing policies. Large global and other major ISPs will be the likely early adopters, and they are expected to have expertise in setting policies (including per prefix policies, if applicable), and make proper use of the RLP indications on a per prefix basis. When the large ISPs participate in this solution deployment, it is envisioned that they would form a ring of protection against route leaks, and co-operatively avoid many of the common types of route leaks that are observed. Route leaks may still happen occasionally within the customer cones (if some customer ASes are not participating or not diligently implementing RLP), but said leaks would be much less likely to propagate from one large participating ISP to another.

6.4. Comparison with other methods (routing security BCPs)

It is reasonable to ask if techniques considered in BCPs such as[RFC7454] (BGP Operations and Security) and [NIST-800-54] may be adequate to address route leaks. The prefix filtering recommendations in the BCPs may be complementary but not adequate. The difficulty is in ISPs' ability to construct prefix filters that represent their customer cones (CC) accurately, especially when there are many levels in the hierarchy within the CC. In the RLP-encoding based solution described here, AS operators signal for each route propagated, if it SHOULD NOT be subsequently propagated to a transit provider or peer.

AS path based Outbound Route Filter (ORF) described in [draft-ietf-idr-aspath-orf] is also an interesting complementary technique. It can be used as an automated collaborative messaging system (implemented in BGP) for ISPs to try to develop a complete view of the ASes and AS paths in their CCs. Once an ISP has that view, then AS path filters can be possibly used to detect route leaks. One limitation of this technique is that it cannot duly take into account the fact that routes for prefixes with different peering relationships may be sent over the same link between ASes. Also, the success of AS path based ORF depends on whether ASes at all levels of the hierarchy in a CC participate and provide accurate information (in the ORF messages) about the AS paths they expect to have in their BGP updates.

6.5. Per-Hop RLP Field or Single RLP Flag per Update?

The route-leak detection and mitigation mechanism described in this document is based on setting RLP Fields on a per-hop basis. There is another possible mechanism based on a single RLP flag per update.

Method A - Per-Hop RLP Field: The sender (eBGP router) on each hop in the AS path sets its RLP Field = 1 if sending the update to a

customer or lateral peer (see Section 4.3) and Section 4.3.1). No AS (if operating correctly) would rewrite the RLP Field set by any preceding AS.

Method B - Single RLP Flag per Update: As it propagates, the update would have at most one RLP flag. Once an eBGP router (in the update path) determines that it is sending an update towards a customer or lateral peer AS, it sets the RLP flag. The flag value equals the AS number of the eBGP router that is setting it. Once the flag is set, subsequent ASes in the path must propagate the flag as is.

To compare Methods A and B, consider the example illustrated in Figure 3. Consider a partial deployment scenario in which AS1, AS2, AS3 and AS5 participate in RLP, and AS4 does not. AS1 (2 levels deep in AS3's customer cone) has imperfect RLP operation. Each complying AS's route leak mitigation policy is to prefer an update not marked as route leak (see Section 4.5). If there is no alternative, then a transit-provider may propagate a marked update from a customer. In this example, multi-homed AS4 leaks a route received for prefix Q from transit-provider AS3 to transit-provider AS5.

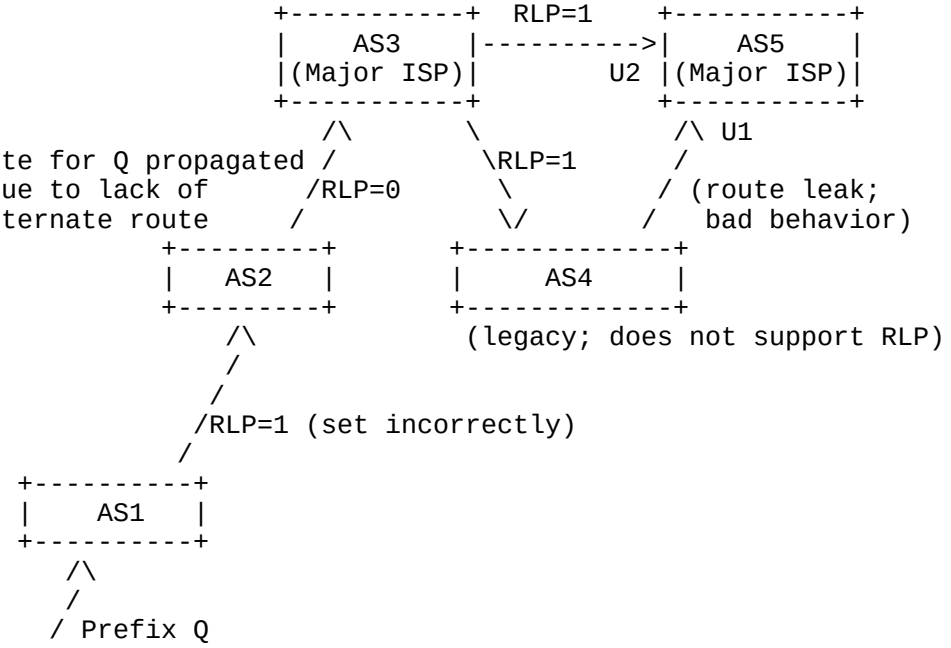

Figure 3: Example for comparison of Method A vs. Method B

If Method A is implemented in the network, the two BGP updates for prefix Q received at AS5 are (note that AS4 is not participating in RLP):

U1A: Q [AS4 AS3 AS2 AS1] {RLP3(AS3)=1, RLP2(AS2)=0, RLP1(AS1)=1} ..... from AS4

U2A: Q [AS3 AS2 AS1] {RLP3(AS3)=1, RLP2(AS2)=0, RLP1(AS1)=1} ..... from AS3

Alternatively, if Method B is implemented in the network, the two BGP updates for prefix Q received at AS5 are:

U1B: Q [AS4 AS3 AS2 AS1] {RLP(AS1)=1} ..... from AS4

U2B: Q [AS3 AS2 AS1] {RLP(AS1)=1} ..... from AS3

All received routes for prefix Q at AS5 are marked as route leak in either case (Method A or B). In the case of Method A, AS5 can use additional information gleaned from the RLP fields in the updates to possibly make a better best path selection. For example, AS5 can determine that U1A update received from its customer AS4 exhibits violation of two RLP fields (those set by AS1 and AS3) and one of them was set just two hops away. But U2A update exhibits that only one RLP field was violated and that was set three hops back. Based on this logic, AS5 may prefer U2A over U1A (even though U1A is a customer route). This would be a good decision. However, Method B does not facilitate this kind of more rational decision process. With Method B, both updates U1B and U2B exhibit that they violated only one RLP field (set by AS1 several hops away). AS5 may then prefer U1B over U2B since U1B is from a customer, and that would be bad decision. This illustrates that, due to more information in per-hop RLP Fields, Method A seems to be operationally more beneficial than Method B.

Further, for detection and notification of neighbor AS’s non-compliance, Method A (per-hop RLP) is better than Method B (single RLP). With Method A, the bad behavior of AS4 would be explicitly evident to AS5 since it violated AS3’s (only two hops away) RLP field as well. AS5 would alert AS4 and also AS2 would alert AS1 about lack of compliance (when Method A is used). With Method B, the alerting process may not be as expeditious.

7. Security Considerations

The proposed Route-Leak Protection (RLP) field requires cryptographic protection in order to prevent malicious route leaks. Since it is proposed that the RLP field be included in the Flags field in the

Secure_Path Segment in BGPsec updates, the cryptographic security mechanisms in BGPsec are expected to also apply to the RLP field. The reader is therefore directed to the security considerations provided in [I-D.ietf-sidr-bgpsec-protocol].

8. IANA Considerations

IANA is requested to register a new optional, non-transitive BGP Path Attribute, named "Preventive Route Leak Protection (pRLP)" in the BGP Path Attributes registry. The attribute type code is TBD. The reference for this new attribute is this document (i.e. the RFC that replaces this draft). The length of this new attribute is 0.

IANA is requested to register a new optional, transitive BGP Path Attribute, named "Route Leak Protection" in the BGP Path Attributes registry. The attribute type code is TBD. The reference for this new attribute is this document (i.e. the RFC that replaces this draft). The length field of this attribute is 2 octets, and the length of the value field of this attribute is variable (see Figure 2) in Section 4.3.1 of this document).

9. Acknowledgements

The authors wish to thank Jared Mauch, Jeff Haas, Job Snijders, Warren Kumari, Amogh Dhamdhere, Jakob Heitz, Geoff Huston, Randy Bush, Alexander Azimov, Ruediger Volk, Sue Hares, Wes George, Job Snijders, Chris Morrow, Sandy Murphy, Danny McPherson, and Eric Osterweil for comments, suggestions, and critique. The authors are also thankful to Padma Krishnaswamy, Oliver Borchert, and Okhee Kim for their review and comments.

10. References

10.1. Normative References

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <http://www.rfc-editor.org/info/rfc4271>.

10.2. Informative References

[Anwar] Anwar, R., Niaz, H., Choffnes, D., Cunha, I., Gill, P., and N. Katz-Bassett, "Investigating Interdomain Routing Policies in the Wild", ACM Internet Measurement Conference (IMC), October 2015, <http://www.cs.usc.edu/assets/007/94928.pdf>.

[Cowie2010]
Cowie, J., "China's 18 Minute Mystery", Dyn Research/Renesys Blog, November 2010, <http://research.dyn.com/2010/11/chinas-18-minute-mystery/>.

[Cowie2013]
Cowie, J., "The New Threat: Targeted Internet Traffic Misdirection", Dyn Research/Renesys Blog, November 2013, <http://research.dyn.com/2013/11/mitm-internet-hijacking/>.

[draft-dickson-sidr-route-leak-solns]
Dickson, B., "Route Leaks -- Proposed Solutions", IETF Internet Draft (expired), March 2012, <https://tools.ietf.org/html/draft-dickson-sidr-route-leak-solns-01>.

[draft-ietf-idr-aspath-orf]
Patel, K. and S. Hares, "AS path Based Outbound Route Filter for BGP-4", IETF Internet Draft (expired), August 2007, <https://tools.ietf.org/html/draft-ietf-idr-aspath-orf-09>.

[draft-kunzinger-idrp-ISO10747-01]
Kunzinger, C., "Inter-Domain Routing Protocol (IDRP)", IETF Internet Draft (expired), November 1994, <https://tools.ietf.org/pdf/draft-kunzinger-idrp-ISO10747-01.pdf>.

[Gao]
Gao, L. and J. Rexford, "Stable Internet routing without global coordination", IEEE/ACM Transactions on Networking, December 2001, <http://www.cs.princeton.edu/~jrex/papers/sigmetrics00.long.pdf>.

[Gill]
Gill, P., Schapira, M., and S. Goldberg, "A Survey of Interdomain Routing Policies", ACM SIGCOMM Computer Communication Review, January 2014, <https://www.cs.bu.edu/~goldbe/papers/survey.pdf>.

[Giotsas]
Giotsas, V. and S. Zhou, "Valley-free violation in Internet routing - Analysis based on BGP Community data", IEEE ICC 2012, June 2012.

[Hiran] Hiran, R., Carlsson, N., and P. Gill, "Characterizing Large-scale Routing Anomalies: A Case Study of the China Telecom Incident", PAM 2013, March 2013, <http://www3.cs.stonybrook.edu/~phillipa/papers/CTelecom.html>.

[Huston2012] Huston, G., "Leaking Routes", March 2012, <http://labs.apnic.net/blabs/?p=139/>.

[Huston2014] Huston, G., "What's so special about 512?", September 2014, <http://labs.apnic.net/blabs/?p=520/>.

[I-D.ietf-sidr-bgpsec-protocol] Lepinski, M. and K. Sriram, "BGPsec Protocol Specification", draft-ietf-sidr-bgpsec-protocol-22 (work in progress), January 2017.

[I-D.ymbk-idr-bgp-open-policy] Azimov, A., Bogomazov, E., Bush, R., Patel, K., and K. Sriram, "Route Leak Detection and Filtering using Roles in Update and Open messages", draft-ymbk-idr-bgp-open-policy-02 (work in progress), November 2016.

[Kapela-Pilosov] Pilosov, A. and T. Kapela, "Stealing the Internet: An Internet-Scale Man in the Middle Attack", DEFCON-16 Las Vegas, NV, USA, August 2008, <https://www.defcon.org/images/defcon-16/dc16-presentations/defcon-16-pilosov-kapela.pdf>.

[Kephart] Kephart, N., "Route Leak Causes Amazon and AWS Outage", ThousandEyes Blog, June 2015, <https://blog.thousandeyes.com/route-leak-causes-amazon-and-aws-outage>.

[Khare] Khare, V., Ju, Q., and B. Zhang, "Concurrent Prefix Hijacks: Occurrence and Impacts", IMC 2012, Boston, MA, November 2012, <http://www.cs.arizona.edu/~bzhang/paper/12-imc-hijack.pdf>.

[Labovitz] Labovitz, C., "Additional Discussion of the April China BGP Hijack Incident", Arbor Networks IT Security Blog, November 2010, <http://www.arbornetworks.com/asert/2010/11/additional-discussion-of-the-april-china-bgp-hijack-incident/>.

[LRL] Khare, V., Ju, Q., and B. Zhang, "Large Route Leaks", Project web page, 2012, <http://nrl.cs.arizona.edu/projects/lsrl-events-from-2003-to-2009/>.

[Luckie] Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V., and kc. claffy, "AS Relationships, Customer Cones, and Validation", IMC 2013, October 2013, <http://www.caida.org/~amogh/papers/asrank-IMC13.pdf>.

[Madory] Madory, D., "Why Far-Flung Parts of the Internet Broke Today", Dyn Research/Renesys Blog, September 2014, <http://research.dyn.com/2014/09/why-the-internet-broke-today/>.

[Mauch] Mauch, J., "BGP Routing Leak Detection System", Project web page, 2014, <http://puck.nether.net/bgp/leakinfo.cgi/>.

[Mauch-nanog] Mauch, J., "Detecting Routing Leaks by Counting", NANOG-41 Albuquerque, NM, USA, October 2007, <https://www.nanog.org/meetings/nanog41/presentations/mauch-lightning.pdf>.

[Nanog-thread-June2016] "Intra-AS messaging for route leak prevention", NANOG Email List - Discussion Thread , June 2016, <http://mailman.nanog.org/pipermail/nanog/2016-June/thread.html#86348>.

[NIST-800-54] Kuhn, D., Sriram, K., and D. Montgomery, "Border Gateway Protocol Security", NIST Special Publication 800-54, July 2007, <http://csrc.nist.gov/publications/nistpubs/800-54/SP800-54.pdf>.

[Paseka] Paseka, T., "Why Google Went Offline Today and a Bit about How the Internet Works", CloudFare Blog, November 2012, <http://blog.cloudflare.com/why-google-went-offline-today-and-a-bit-about/>.

[proceedings-sixth-ietf] Gross, P., "Proceedings of the April 22-24, 1987 Internet Engineering Task Force", April 1987, <https://www.ietf.org/proceedings/06.pdf>.

[RFC1105-obsolete]
Lougheed, K. and Y. Rekhter, "A Border Gateway Protocol (BGP)", IETF RFC (obsolete), June 1989, <https://tools.ietf.org/html/rfc1105>.

[RFC6811] Mohapatra, P., Scudder, J., Ward, D., Bush, R., and R. Austein, "BGP Prefix Origin Validation", RFC 6811, DOI 10.17487/RFC6811, January 2013, <http://www.rfc-editor.org/info/rfc6811>.

[RFC7454] Durand, J., Pepelnjak, I., and G. Doering, "BGP Operations and Security", BCP 194, RFC 7454, DOI 10.17487/RFC7454, February 2015, <http://www.rfc-editor.org/info/rfc7454>.

[RFC7908] Sriram, K., Montgomery, D., McPherson, D., Osterweil, E., and B. Dickson, "Problem Definition and Classification of BGP Route Leaks", RFC 7908, DOI 10.17487/RFC7908, June 2016, <http://www.rfc-editor.org/info/rfc7908>.

[Snijders]
Snijders, J., "Practical everyday BGP filtering with AS_PATH filters: Peer Locking", NANOG-47 Chicago, IL, USA, June 2016, <https://www.nanog.org/sites/default/files/Snijders_Everyday_Practical_Bgp.pdf>.

[Sriram] Sriram, K., Montgomery, D., Dickson, B., Patel, K., and A. Robachevsky , "Methods for Detection and Mitigation of BGP Route Leaks", IETF-95 IDR WG Meeting), April 2016, <https://www.ietf.org/proceedings/95/slides/slides-95-idr-13.pdf>.

[Toonk] Toonk, A., "What Caused Today’s Internet Hiccup", August 2014, <http://www.bgpmon.net/what-caused-todays-internet-hiccup/>.

[Toonk2015-A]
Toonk, A., "What caused the Google service interruption", March 2015, <http://www.bgpmon.net/what-caused-the-google-service-interruption/>.

[Toonk2015-B]
Toonk, A., "Massive route leak causes Internet slowdown", June 2015, <http://www.bgpmon.net/massive-route-leak-cause-internet-slowdown/>.

[Wijchers]
Wijchers, B. and B. Overeinder, "Quantitative Analysis of BGP Route Leaks", RIPE-69, November 2014, <https://ripe69.ripe.net/presentations/157-RIPE-69-Routing-WG.pdf>.

[Zmijewski]
Zmijewski, E., "Indonesia Hijacks the World", Dyn Research/Renesys Blog, April 2014, <http://research.dyn.com/2014/04/indonesia-hijacks-world/>.

Authors' Addresses

Kotikalapudi Sriram
US NIST

Email: ksriram@nist.gov

Doug Montgomery
US NIST

Email: dougm@nist.gov

Brian Dickson

Email: brian.peter.dickson@gmail.com

Keyur Patel
Arrcus

Email: keyur@arrcus.com

Andrei Robachevsky
Internet Society

Email: robachevsky@isoc.org

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 9, 2017

B. Weis
R. Gagliano
Cisco Systems
K. Patel
Arrcus, Inc.
March 8, 2017

BGPsec Router Certificate Rollover
draft-ietf-sidrops-bgpsec-rollover-00

Abstract

BGPsec will need to address the impact from regular and emergency rollover processes for the BGPsec end-entity (EE) certificates that will be performed by Certificate Authorities (CAs) participating at the Resource Public Key Infrastructure (RPKI). Rollovers of BGPsec EE certificates must be carefully managed in order to synchronize distribution of router public keys and the usage of those public keys by BGPsec routers. This memo provides general recommendations for that process, as well as describing reasons why the rollover of BGPsec EE certificates might be necessary.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 9, 2017.

Table of Contents

1. Requirements notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Introduction

In BGPsec, a key rollover (or re-key) is the process of changing a router's key pair (or key pairs), issuing the corresponding new end-entity (EE) certificate and (if the old certificate is still valid) revoking the old certificate. This process will need to happen at regular intervals, normally due to the local policies of a network. This document provides general recommendations for that process. Certificate Practice Statements (CPS) documents MAY reference these recommendations. This memo only addresses changing of a router's key pair within the RPKI. Refer to [RFC6489] for a procedure to rollover RPKI Certificate Authority key pairs.

When a router receives or creates a new key pair (using a key provisioning mechanism), this key pair will be used to sign new

BGPsec_Path attributes [I-D.ietf-sidr-bgpsec-protocol] that are originated or that transit through the BGP speaker. Additionally, the BGP speaker MUST refresh its outbound BGPsec Update messages to include a signature using the new key (replacing the replaced key). When the rollover process finishes, the old BGPsec certificate (and its key) will not longer be valid, and thus any BGPsec Update that includes a BGPsec_Path attribute with a signature performed by the old key will be invalid. Consequently, if the router does not refresh its outbound BGPsec Update messages, routing information may be treated as unauthenticated after the rollover process is finished. It is therefore extremely important that the BGPsec router key rollover be performed in such a way that the probability of new router EE certificates have been distributed throughout the RPKI before the router begin signing BGPsec_Path attributes with a new private key.

It is also important for an AS to minimize the BGPsec router key rollover interval (i.e., in between the time an AS distributes an EE certificate with a new public key and the time a BGPsec router begins to use its new private key). This can be due to a need for a BGPsec router to distribute BGPsec_Path attributes signed with a new private key in order to invalidate BGPsec_Path attributes signed with the old private key. In particular, if the AS suspects that a stale BGPsec_Path attribute is being distributed instead of the most recently signed attribute it can cause the stale BGPsec_Path attribute to be invalidated by completing a key rollover procedure. The BGPsec rollover interval can be minimized when an automated certificate provisioning process such as Enrollment over Secure Transport (EST) [RFC7030]) is used.

The Security Requirements for BGP Path Validation [RFC7353] also describes the need for protecting against the replay of BGP UPDATE messages, such as controlling BGPsec’s window of exposure to replay attacks. The BGPsec rollover method in this document can be used to achieve this goal.

In [I-D.ietf-sidr-rtr-keying], the "operator-driven" method is introduced, in which a key pair can be shared among different BGP Speakers. In this scenario, the roll-over of the correspondent BGPsec certificate will impact all the BGP Speakers sharing the same private key.

3. Key rollover in BGPsec

An BGPsec EE certificate SHOULD be replaced when the following events occur, and can be replaced for any other reason at the discretion of the AS responsible for the EE certificate.

BGPsec scheduled rollover: BGPsec certificates have an expiration date (NotValidAfter) that requires a frequent rollover process. The validity period for these certificates is typically expressed at the CA's CPS document.

BGPsec certificate fields changes: Information contained in a BGPsec certificate (such as the ASN or the Subject) may need to be changed.

BGPsec emergency rollover Some special circumstances (such as a compromised key) may require the replacement of a BGPsec certificate.

BGPsec signature anti-replay protection An AS may determine stale BGPsec_Path attributes signed by the AS are being propagated instead of the most recently signed BGPsec_Path attributes. Changing the BGPsec router signing key, distributing a new BGPsec EE certificate for the router, and revoking the old BGPsec EE certificate will invalidate the replayed BGPsec_Path attributes.

In some of these cases it is possible to generate a new certificate without changing the key pair. This practice simplifies the rollover process as the corresponding BGP speakers do not even need to be aware of the changes to its correspondent certificate. However, not replacing the certificate key for a long period of time increases the risk that a compromised router private key may be used by an attacker to deliver unauthorized BGPsec Updates. Distributing the OLD public key in a new certificate is NOT RECOMMENDED when the rollover event is due to the key being compromised, or when stale BGPsec_Path attribute signatures are being distributed.

3.1. A proposed process for BGPsec key rollover

The BGPsec key rollover process will be dependent on the key provisioning mechanisms that are adopted by an AS. The key provisioning mechanisms for BGPsec are not yet fully documented (see [I-D.ietf-sidr-rtr-keying] as a work in progress document). An automatic provisioning mechanism such as EST will allow BGPsec code to include automatic re-keying scripts with minimum development cost.

If we work under the assumption that an automatic mechanism will exist to rollover a BGPsec certificate, a RECOMMENDED process is as follows.

1. New Certificate Publication: The first step in the rollover mechanism is to publish the new public key in a new certificate. In order to accomplish this goal, the new key pair and

certificate will need to be generated and the certificate published at the appropriate RPKI repository publication point. The details of this process will vary as they depend on whether the keys are assigned per-BGP speaker or shared, whether the keys are generated on each BGP speaker or in a central location, and whether the RPKI repository is locally or externally hosted.

2. Staging Period: A staging period will be required from the time a new certificate is published in the RPKI global repository until the time it is fetched by RPKI caches around the globe. The exact minimum staging time will be dictated by the conventional interval chosen between repository fetches. If rollovers will be done more frequently, an administrator can provision two certificates for every router concurrently with different valid start times. In this case when the rollover operation is needed, the relying parties around the globe would already have the new router public keys. A staging period may not be possible to implement during emergency key rollover, in which case routing information may be lost.

3. Twilight: At this moment, the BGP speaker that holds the private key that has been rolled-over will stop using the OLD key for signing and start using the NEW key. Also, the router will generate appropriate BGPsec_Path attributes just as in the typical operation of refreshing out-bound BGP polices. This operation may generate a great number of BGPsec_Path attributes (due to the need to refresh BGP outbound policies). In any given BGP SPEAKER, the Twilight moment may be different for every peer in order to distribute the system load (probably in the order of minutes to avoid reaching any expiration time).

4. Certificate Revocation: This is an optional step, but SHOULD be taken when the goal is to invalidate signatures used with the OLD key. Reasons to invalidate OLD signatures include: when the AS has reason to believe that the router signing key has been compromised, and when the AS needs to invalidate BGPsec_Path attribute signatures used with this key. As part of the rollover process, a CA MAY decide to revoke the OLD certificate by publishing its serial number on the CA's CRL. Alternatively, the CA will just let the OLD certificate to expire and not revoke it. This choice will depend on the reasons that motivated the rollover process.

5. RPKI-Router Protocol Withdrawals: At the expiration of the OLD certificate's validation, the RPKI relying parties around the globe will need to communicate to their router peers that the OLD certificate's public key is not longer valid (e.g., using the RPKI-Router Protocol described in [RFC6810]). A router's

reaction to a message indicating withdrawal of a prefix in the RPKI-Router Protocol SHOULD include the removal of any RIB entry that includes a BGPsec attribute signed with that key and the generation of the correspondent BGP WITHDRAWALs (either implicit or explicit).

The proposed rollover mechanism will depend on the existence of an automatic provisioning process for BGPsec certificates. It will require a staging mechanism based on the RPKI propagation time of around 24 hours, and it will generate BGPsec_Path attributes for all prefixes in the router been re-keyed.

The first two steps (New Certificate Publication and Staging Period) may happen in advance of the rest of the process. This will allow a network operator to accelerate its subsequent key roll-over.

When a new BGPsec certificate is generated without changing its key, steps 3 (Twilight) and 5 (RPKI-Router Protocol Withdrawals) SHOULD NOT be executed.

4. BGPsec key rollover as a measure against replays attacks in BGPsec

There are two typical generic measures to mitigate replay attacks in any protocol: the addition of a timestamp or the addition of a serial number. However neither BGP nor BGPsec provide either measure. This section discusses the use of BGPsec Rollover as a measure to mitigate replay attacks.

4.1. BGPsec Replay attack window requirement

In [RFC7353] Section 4.3, the need to limit the vulnerability to replay attacks is described. One important comment is that during a window of exposure, a replay attack is effective only in very specific circumstances: there is a downstream topology change that makes the signed AS path no longer current, and the topology change makes the replayed route preferable to the route associated with the new update. In particular, if there have been no topology change at all, then no security threat comes from a replay of a BGPsec_Path attribute because the signed information is still valid.

The BGPsec Ops document [I-D.ietf-sidr-bgpsec-ops] gives some idea of requirements for the size of the BGPsec windows of exposure to replay attacks. It states that the requirement will be in the order of a day or longer.

4.2. BGPsec key rollover as a mechanism to protect against replay attacks

Since the window requirement is in the order of a day (as documented in [I-D.ietf-sidr-bgpsec-ops]) and the BGP speaker re-keying is the edge router of the origin AS, it is feasible for a BGPsec Rollover to mitigate replays. In this case it is important to complete the full process (i.e. the OLD and NEW certificate do not share the same key). By re-keying an AS is letting the BGPsec certificate validation time be a type of "timestamp" against replay attacks. However, the use of frequent key rollovers comes with an additional administrative cost and risks if the process fails. As documented before, re-keying should be supported by automatic tools and for the great majority of the Internet it will be done with good lead time to correct any risk.

For a transit AS that also originates BGPsec_Path attributes for its own prefixes, the key rollover process may generate a large number of UPDATE messages (even the complete Default Free Zone or DFZ). For this reason, it is recommended that routers in a transit AS that also originate BGPsec_Path attributes be provisioned with two certificates: one to sign BGPsec_Path attributes in transit and a second one to sign an BGPsec_Path attribute for prefixes originated in its AS. Only the second certificate (for prefixes originated in its AS) should be rolled-over frequently as a means of limiting replay attack windows. The transit BGPsec certificate is expected to live longer than the origin BGPsec certificate.

Advantage of Re-keying as replay attack protection mechanism:

1. All expiration policies are maintained in the RPKI

2. Much of the additional administrative cost is paid by the provider that wants to protect its infrastructure, as it bears the human cost of creating and initiating distribution of new router key pairs and router EE certificates. (It is true that the cost of relying parties will be affected by the new objects, but their responses should be completely automated or otherwise routine.)

3. The re-keying can be implemented in coordination with planned topology changes by either origin ASes or transit ASes (e.g., if an AS changes providers, it completes a BGP Rollover)

Disadvantage of Re-keying as replay attack protection mechanism:

1. There is more administrative load due to frequent rollover, although how frequent is still not clear. Some initial ideas are found in [I-D.ietf-sidr-bgpsec-ops]

2. The minimum window size is bounded by the propagation time for RPKI caches to obtain the new certificate and CRL (2x propagation time). If provisioning is done ahead of time, the minimum window size is reduced to 1x propagation time for the CRL. However, these bounds will be better understood when RPKI and RPs are well deployed.

3. Re-keying increases the dynamics and size of RPKI repository.

5. IANA Considerations

There are no IANA considerations. This section may be removed upon publication.

6. Security Considerations

Several possible reasons can cause routers participating in BGPsec to rollover their signing keys, which also has the effect of invalidating BGPsec_Path attributes signatures using the current signature verification key. Conventional key management operations my dicate a re-key (e.g.,key exposure, change of certificate attributes, rollover policy). BGPsec routers also may need to change their signing keys and associated certificate as an anti-replay protection.

The BGPsec Rollover method allows for an expedient rollover process when router certificates are distributed through the RPKI, but without causing routing failures due to a receiving router not being able to validate a BGPsec_Path attribute created by a router that is the subject of the rollover.

7. Acknowledgements

We would like to acknowledge Randy Bush, Sriram Kotikalapudi, Stephen Kent and Sandy Murphy.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

8.2. Informative References

[I-D.ietf-sidr-bgpsec-ops]
Bush, R., "BGPsec Operational Considerations", draft-ietf-sidr-bgpsec-ops-16 (work in progress), January 2017.

[I-D.ietf-sidr-bgpsec-protocol]
Lepinski, M. and K. Sriram, "BGPsec Protocol Specification", draft-ietf-sidr-bgpsec-protocol-22 (work in progress), January 2017.

[I-D.ietf-sidr-rtr-keying]
Bush, R., Turner, S., and K. Patel, "Router Keying for BGPsec", draft-ietf-sidr-rtr-keying-12 (work in progress), June 2016.

[RFC6489] Huston, G., Michaelson, G., and S. Kent, "Certification Authority (CA) Key Rollover in the Resource Public Key Infrastructure (RPKI)", BCP 174, RFC 6489, DOI 10.17487/RFC6489, February 2012, <http://www.rfc-editor.org/info/rfc6489>.

[RFC6810] Bush, R. and R. Austein, "The Resource Public Key Infrastructure (RPKI) to Router Protocol", RFC 6810, DOI 10.17487/RFC6810, January 2013, <http://www.rfc-editor.org/info/rfc6810>.

[RFC7030] Pritikin, M., Ed., Yee, P., Ed., and D. Harkins, Ed., "Enrollment over Secure Transport", RFC 7030, DOI 10.17487/RFC7030, October 2013, <http://www.rfc-editor.org/info/rfc7030>.

[RFC7353] Bellovin, S., Bush, R., and D. Ward, "Security Requirements for BGP Path Validation", RFC 7353, DOI 10.17487/RFC7353, August 2014, <http://www.rfc-editor.org/info/rfc7353>.

Authors' Addresses

Brian Weis
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
CA

Email: bew@cisco.com

Roque Gagliano
Cisco Systems
Avenue des Uttins 5
Rolle, VD 1180
Switzerland

Email: rogaglia@cisco.com

Keyur Patel
Arrcus, Inc.

Email: keyur@arrcus.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 14, 2017

A. Azimov
E. Bogomazov
Qrator Labs
R. Bush
Internet Initiative Japan
K. Patel
Arrcus, Inc.

March 13, 2017

Route Leak Detection and Filtering using Roles in Update and Open messages
draft-ymbk-idr-bgp-eotr-policy-00

Abstract

[draft-ymbk-idr-bgp-open-policy] defines a BGP OPEN capability and consequent route marking which enforces a valley-free peering relationship. This document defines an eOTC (external Only To Customer) transitive BGP attribute which propagates the specific marking to automatically detect route leaks. The goal is to allow a distant AS to determine a violation of valley-free peering.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in RFC 2119 [RFC2119] only when they appear in all upper case. They may also appear in lower or mixed case as English words, without normative meaning.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 14, 2017.

Table of Contents

1. Introduction

For the purpose of this document, BGP route leaks are when a BGP route was learned from transit provider or peer is announced to another provider or peer. See [I-D.ietf-grow-route-leak-problem-definition]. These are usually the result of misconfigured or absent BGP route filtering or lack of coordination between two BGP speakers.

[I-D.ietf-idr-route-leak-detection-mitigation] describes a method of marking and detecting leaks which relies on operator maintained markings. Unfortunately, in most cases, a leaking router will likely also be misconfigured to mark incorrectly.

It has been suggested to use white list filtering, relying on knowing the prefixes in the peer's customer cone as import filtering, in order to detect route leaks. Unfortunately, a large number of medium transit operators use a single prefix list as only the ACL for export filtering, without community tagging and without paying attention to the source of a learned route. So, if they learn a customer's route from their provider or peer - they will announce it in all

directions, including other providers or peers. This misconfiguration affects a limited number of prefixes; but such route leaks will obviously bypass customer cone import filtering made by upper level upstream providers.

This document specifies a way to to create automatic filters for detection of route leaks via new BGP Path Attribute which is set according to BGP Roles ([I-D.ymbk-idr-bgp-open-policy]) . While iOTC provides strong vendor-code-based enforcement of route leak prevention, route leaks could still exist as result of misconfigured old BGP implementations. Route leaks could also be result of malicious activity such as MITM attacks or DoS. The goal of this proposal is to allow a distant AS to determine a violation of valley-free peering that is made by mistake or by purpose.

2. BGP External Only To Customer attribute

The External Only To Customer (eOTC) attribute is a new optional, transitive BGP Path attribute with the Type Code <TBD1>. This attribute is four bytes and contains an AS number of the AS that added the attribute to the route.

There are four rules for setting the eOTC attribute:

1. If eOTC is not set and the sender’s Role is Provider or Peer, the eOTC attribute MUST be added with value equal to the sender’s AS number.

2. If eOTC is not set and the sender’s Role is Complex and the prefix role is Provider or Peer, the eOTC attribute MUST be added with value equal to to the sender’s AS number.

3. If eOTC is set, the receiver’s Role is Provider or Peer, and its value is not the neighbor’s AS number then the incoming route is route leak and MUST be given a lower local preference, or MAY be dropped.

4. If eOTC is set, the receiver’s Role is Complex, the prefix role Role is Provider or Peer, and the eOTC value is not equal to the neighbor’s AS number, then the incoming route is a route leak and MUST be given a lower local preference, or they MAY be dropped.

These four rules provide mechanism for route leak detection that is created by a distant party in the AS_Path.

3. Compatibility with BGPsec

For BGPsec [I-D.ietf-sidr-bgpsec-protocol] enabled routers, the Flags field will have a bit added to indicate that an eOTC attribute exists. The eOTC value will be automatically carried in AS field of the added Secure_Path Segment.

When a route is translated from a BGPsec enabled router to a non-BGPsec router, in addition to AS_PATH reconstruction, reconstruction MUST be performed for the eOTC attribute. If Flag bit was set in one of Secure_Path Segments, the eOTC attribute SHOULD be added with the AS number of the segment in which it appears for the first time.

4. IANA Considerations

This document defines a new optional, transitive BGP Path Attributes option, named "External Only To Customer", assigned value <TBD1> [To be removed upon publication: http://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-2] [RFC4271]. The length of this attribute is 4.

5. Security Considerations

This document proposes a mechanism for detection of route leaks that are the result of BGP policy misconfiguration. If BGPSec is enabled it will also provide mechanism to detect leaks that are result of malicious activity.

Deliberate mis-marking of the eOTC flag could be used to affect the BGP decision process, but could not sabotage a route’s propagation.

eOTC is a transitive BGP AS_PATH attribute which reveals a information about a BGP speaker’s peering relationship. It will give a strong hint that some link isn’t customer to provider, but will not help to distinguish if it is provider to customer or peer to peer. In addition it could reveal sequence of p2c to downstream ISPs. If eOTC is BGPsec signed, it can not be removed for peering confidentiality.

Still, any Tier-1 number in AS_PATH could be used in the same way to reveal possible p2c sequence.

6. References

6.1. Normative References

[I-D.ymbk-idr-bgp-open-policy]
Azimov, A., Bogomazov, E., Bush, R., Patel, K., and K. Sriram, "Route Leak Detection and Filtering using Roles in Update and Open messages", draft-ymbk-idr-bgp-open-policy-02 (work in progress), November 2016.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <http://www.rfc-editor.org/info/rfc4271>.

6.2. Informative References

[I-D.ietf-grow-route-leak-problem-definition]
Sriram, K., Montgomery, D., McPherson, D., Osterweil, E., and B. Dickson, "Problem Definition and Classification of BGP Route Leaks", draft-ietf-grow-route-leak-problem-definition-06 (work in progress), May 2016.

[I-D.ietf-idr-route-leak-detection-mitigation]
Sriram, K., Montgomery, D., Dickson, B., Patel, K., and A. Robachevsky, "Methods for Detection and Mitigation of BGP Route Leaks", draft-ietf-idr-route-leak-detection-mitigation-03 (work in progress), May 2016.

[I-D.ietf-sidr-bgpsec-protocol]
Lepinski, M. and K. Sriram, "BGPsec Protocol Specification", draft-ietf-sidr-bgpsec-protocol-15 (work in progress), March 2016.

Authors' Addresses

Alexander Azimov
Qrator Labs

Email: aa@qrator.net

Eugene Bogomazov
Qrator Labs

Email: eb@qrator.net

Randy Bush
Internet Initiative Japan

Email: randy@psg.com

Keyur Patel
Arrcus, Inc.

Email: keyur@arrcus.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 14, 2017

A. Azimov
E. Bogomazov
Qrator Labs
R. Bush
Internet Initiative Japan
K. Patel
Arrcus, Inc.
K. Sriram
US NIST

March 13, 2017

Route Leak Prevention using Roles in Update and Open messages
draft-ymbk-idr-bgp-open-policy-03

Abstract

Route Leaks are the propagation of BGP prefixes which violate assumptions of BGP topology relationships; e.g. passing a route learned from one peer to another peer or to a transit provider, passing a route learned from one transit provider to another transit provider or to a peer. Today, approaches to leak prevention rely on marking routes according to operator configuration options without any check that the configuration corresponds to that of the BGP neighbor, or enforcement that the two BGP speakers agree on the relationship. This document enhances BGP Open to establish agreement of the (peer, customer, provider, internal) relationship of two neighboring BGP speakers to enforce appropriate configuration on both sides. Propagated routes are then marked with an iOTC attribute according to agreed relationship allowing prevention of route leaks.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in RFC 2119 [RFC2119] only when they appear in all upper case. They may also appear in lower or mixed case as English words, without normative meaning.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Copyright Notice

Table of Contents

1. Preamble

1.1. Peering Relationships

Despite uses of words such as "Customer," "Peer." etc. the intent is not business relationships, who pays whom, etc. These are common terms to represent restrictions on BGP propagation, some times known as Gao/Rexford. E.g. if A is a "peer" of B and C, A does not propagate B's prefixes to C. If D is a "customer" of E and F, D does not propagate prefixes learned from E to F.

As the whole point of route leak detection and prevention is to prevent vioation of these relationships, they are inescapable.

2. Introduction

This document specifies a new BGP Capability Code, [RFC5492] Sec 4, which two BGP speakers MAY use to ensure that they MUST agree on their relationship; i.e. customer and provider or peers. Either or both may optionally be configured to require that this option be exchanged for the BGP Open to succeed.

Also this document specifies a way to mark routes according to BGP Roles established in OPEN and a way to create double-boundary filters for prevention of route leaks via new BGP Path Attribute.

For the purpose of this document, BGP route leaks are when a BGP route was learned from transit provider or peer and is announced to another provider or peer. See [I-D.ietf-grow-route-leak-problem-definition]. These are usually the result of misconfigured or absent BGP route filtering or lack of coordination between two BGP speakers.

[I-D.ietf-idr-route-leak-detection-mitigation] The mechanism proposed in that draft provides the opportunity to detect route leaks made by third parties but provides no support to strongly prevent route leak creation.

Also, route tagging which relies on operator maintained policy configuration is too easily and too often misconfigured.

3. Role Definitions

As many of these terms are used differently in various contexts, it is worth being explicit.

A Provider: sends their own routes and (possibly) a subset of routes learned from their other customers, peers, and transit providers to their customer.

A Customer: accepts 'transit routes' from its provider(s) and announces their own routes and the routes they have learned from the transitive closure of their customers (AKA their 'customer cone') to their provider(s).

A Peer: announces their routes and the routes from their customer cone to other Peers.

An Internal: announces all routes, accepts all routes.

A Complex: BGP relationship is an attempt to allow those whose policy may vary by prefix. It is aptly named and the authors question its real utility.

Of course, any BGP speaker may apply policy to reduce what is announced, and a recipient may apply policy to reduce the set of routes they accept.

4. BGP Role

BGP Role is new mandatory configuration option. It reflects the real-world agreement between two BGP speakers about their peering relationship.

Allowed Role values are:

- Provider - sender is a transit provider to neighbor;
- Customer - sender is customer of neighbor;
- Peer - sender and neighbor are peers;
- Internal - sender and neighbor is part of same organization. This includes but is not limited to situation when sender and neighbor are in same AS.
- Complex - sender has a non-standard relationship and wants to use manual per-prefix based role policies.

Since BGP Role reflects the relationship between two BGP speakers, it could also be used for more than route leak mitigation.

5. Role capability

The TLV (type, length, value) of the BGP Role capability are:

- Type - <TBD1>;
- Length - 1 (octet);
- Value - integer corresponding to speaker' BGP Role.

Value	Role name
0	Undefined
1	Sender is Peer
2	Sender is Provider
3	Sender is Customer
4	Sender is Internal
5	Sender is Complex

Table 1: Predefined BGP Role Values

6. Role correctness

Section 4 described how BGP Role is a reflection of the relationship between two BGP speakers. But the mere presence of BGP Role doesn't automatically guarantee role agreement between two BGP peers.

To enforce correctness, the BGP Role check is used with a set of constrains on how speakers' BGP Roles MUST corresponded. Of course, each speaker MUST announce and accept the BGP Role capability in the BGP OPEN message exchange.

If a speaker receives a BGP Role capability, it SHOULD check value of the received capability with its own BGP Role. The allowed pairings are (first a sender's Role, second the receiver's Role):

Sender Role	Receiver Role
Peer	Peer
Provider	Customer
Customer	Provider
Internal	Internal
Complex	Complex

Table 2: Allowed Role Capabilities

In all other cases speaker MUST send a Role Mismatch Notification (code 2, sub-code <TBD2>).

6.1. Strict mode

A new BGP configuration option "strict mode" is defined with values of true or false. If set to true, then the speaker MUST refuse to establish a BGP session with peers which do not announce the BGP Role capability in their OPEN message. If a speaker rejects a connection, it MUST send a Connection Rejected Notification [RFC4486] (Notification with error code 6, subcode 5). By default strict mode SHOULD be set to false for backward compatibility with BGP speakers, that do not yet support this mechanism.

7. Restrictions on the Complex role

The Complex role should be set only if the relationship between BGP neighbors can not be described using simple Customer/Provider/Peer roles. For a example, if neighbor is literal peer, but for some prefixes it provides full transit; the complex role SHOULD be set on both sides. In this case roles Customer/Provider/Peer should be set on per-prefix basis, keeping the abstraction from filtering mechanisms (Section 8).

If role is not Complex all per-prefix role settings MUST be ignored.

8. BGP Internal Only To Customer attribute

The Internal Only To Customer (iOTC) attribute is a new optional, non-transitive BGP Path attribute with the Type Code <TBD3>. This attribute has zero length as it is used only as a flag.

There are four rules for setting the iOTC attribute:

1. The iOTC attribute MUST be added to all incoming routes if the receiver’s Role is Customer or Peer;

2. The iOTC attribute MUST be added to all incoming routes if the receiver’s Role is Complex and the prefix Role is Customer or Peer;

3. Routes with the iOTC attribute set MUST NOT be announced by a sender whose Role is Customer or Peer;

4. Routes with the iOTC attribute set MUST NOT be announced if by a sender whose Role is Complex and the prefix Role is Customer or Peer;

These four rules provide mechanism that strongly prevents route leak creation by an AS.

9. Compatibility with BGPsec

As the iOTC field is non-transitive, it is not seen by or signed by BGPsec [I-D.ietf-sidr-bgpsec-protocol].

10. Additional Considerations

As the BGP Role reflects the relationship between neighbors, it can also have other uses. As an example, BGP Role might affect route priority, or be used to distinguish borders of a network if a network consists of multiple AS.

Though such uses may be worthwhile, they are not the goal of this document. Note that such uses would require local policy control.

This document doesn’t provide any security measures to check correctness of per-prefix roles, so the Complex role should be used with great caution. It is as dangerous as current BGP peering.

11. IANA Considerations

This document defines a new Capability Codes option [to be removed upon publication: http://www.iana.org/assignments/capability-codes/capability-codes.xhtml] [RFC5492], named "BGP Role", assigned value <TBD1> . The length of this capability is 1.

The BGP Role capability includes a Value field, for which IANA is requested to create and maintain a new sub-registry called "BGP Role Value". Assignments consist of Value and corresponding Role name. Initially this registry is to be populated with the data in Table 1. Future assignments may be made by a standard action procedure [RFC5226].

This document defines new subcode, "Role Mismatch", assigned value <TBD2> in the OPEN Message Error subcodes registry [to be removed upon publication: http://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-6] [RFC4271].

This document defines a new optional, non-transitive BGP Path Attributes option, named "Internal Only To Customer", assigned value <TBD3> [To be removed upon publication: http://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-2] [RFC4271]. The length of this attribute is 0.

12. Security Considerations

This document proposes a mechanism for prevention of route leaks that are the result of BGP policy misconfiguration.

Deliberate sending of a known conflicting BGP Role could be used to sabotage a BGP connection. This is easily detectable.

BGP Role is disclosed only to an immediate BGP neighbor, so it will not itself reveal any sensitive information to third parties.

13. Acknowledgments

The authors wish to thank Douglas Montgomery, Brian Dickson, and Andrei Robachevsky for their contributions to a variant of this work.

14. References

14.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <http://www.rfc-editor.org/info/rfc4271>.

[RFC4486] Chen, E. and V. Gillet, "Subcodes for BGP Cease Notification Message", RFC 4486, DOI 10.17487/RFC4486, April 2006, <http://www.rfc-editor.org/info/rfc4486>.

[RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <http://www.rfc-editor.org/info/rfc5492>.

14.2. Informative References

[I-D.ietf-grow-route-leak-problem-definition]
Sriram, K., Montgomery, D., McPherson, D., Osterweil, E., and B. Dickson, "Problem Definition and Classification of BGP Route Leaks", draft-ietf-grow-route-leak-problem-definition-06 (work in progress), May 2016.

[I-D.ietf-idr-route-leak-detection-mitigation]
Sriram, K., Montgomery, D., Dickson, B., Patel, K., and A. Robachevsky, "Methods for Detection and Mitigation of BGP Route Leaks", draft-ietf-idr-route-leak-detection-mitigation-03 (work in progress), May 2016.

[I-D.ietf-sidr-bgpsec-protocol]
Lepinski, M. and K. Sriram, "BGPsec Protocol Specification", draft-ietf-sidr-bgpsec-protocol-15 (work in progress), March 2016.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, DOI 10.17487/RFC5226, May 2008, <http://www.rfc-editor.org/info/rfc5226>.

Authors' Addresses

Alexander Azimov
Qrator Labs

Email: aa@qrator.net

Eugene Bogomazov
Qrator Labs

Email: eb@qrator.net

Randy Bush
Internet Initiative Japan

Email: randy@psg.com

Keyur Patel
Arrcus, Inc.

Email: keyur@arrcus.com

Kotikalapudi Sriram
US NIST

Email: ksriram@nist.gov

Network Working Group
Internet-Draft
Intended status: Best Current Practice
Expires: September 14, 2017

Y. Gilad
S. Goldberg
Boston University
K. Sriram
NIST

March 13, 2017

The Use of Maxlength in the RPKI
draft-yossigi-rpkimaxlen-00

Abstract

This document recommends that operators avoid using the maxLength attribute when issuing Route Origin Authorizations (ROAs) in the Resource Public Key Infrastructure (RPKI). These recommendations complement those in [RFC7115].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 14, 2017.

Table of Contents

1. Introduction

The RPKI [RFC6480] uses Route Origin Authorizations (ROAs) to create a trusted mapping from an IP prefix to a set of autonomous systems (ASes) that are authorized to originate this prefix. Each ROA contains a set of IP prefixes, and an AS number of an AS authorized originate all the IP prefixes in the set [RFC6482]. Each ROA is cryptographically signed by the party that is authorized to allocate the set of IP prefixes.

The RPKI also supports a maxLength attribute. According to [RFC6482], "When present, the maxLength specifies the maximum length of the IP address prefix that the AS is authorized to advertise." Thus, rather than requiring the ROA to explictly list each prefix the AS is authorized to originate, the maxLength attribute provides a shorthand that authorizes an AS to announce a set of IP prefixes.

However, measurements of current RPKI deployments have found that use of the maxLength in ROAs tends to lead to security problems. Specifically, as of September 2016, 89% of the prefixes specified in ROAs that use the maxLength attribute, are vulnerable to a forged-origin subprefix hijack. The forged-origin subprefix hijack affects any IP prefix that is authorized in ROA but is not announced in BGP. The impact of such an attack is the same as standard subprefix hijack on an IP prefix that is unprotected by a ROA in the RPKI.

For this reason, this document recommends that operators avoid using the maxLength attribute in their ROAs as a best current practice. Instead, ROAs should be consist of explicit lists of the IP prefixes that an AS is authorized to announce, without using the maxLength

attribute. Whenever possible, this ROA should also be "miminal", in that it includes only the list of IP prefixes that are actually originated in BGP. The recommendations in this document clarify and extend the following recommendation from [RFC7115]:

> One advantage of minimal ROA length is that the forged origin attack does not work for sub-prefixes that are not covered by overly long max length. For example, if, instead of 10.0.0.0/16-24, one issues 10.0.0.0/16 and 10.0.42.0/24, a forged origin attack cannot succeed against 10.0.666.0/24. They must attack the whole /16, which is more likely to be noticed because of its size.

These recommendations requires no changes to the RPKI specification and will not increase the number of signed ROAs in the RPKI, because ROAs already support lists of IP prefixes [RFC6482].

1.1. Requirements

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 [RFC2119].

2. Suggested Reading

It is assumed that the reader understands BGP [RFC4271], the RPKI [RFC6480] Route Origin Authorizations (ROAs) [RFC6482], RPKI-based Prefix Validation [RFC6811], and BGPSEC [I-D.ietf-sidr-bgpsec-protocol].

3. Forged Origin Subprefix Hijack

The forged-origin subprefix hijack is relevant to a scenario in which (1) the RPKI [RFC6480] is deployed, and (2) routers use RPKI origin validation to drop invalid routes [RFC6811], but (3) BGPSEC [I-D.ietf-sidr-bgpsec-protocol] is not deployed.

We describe the forged-origin subprefix hijack [RFC7115] [GCHSS] using a running example.

Consider the IP prefix 168.122.0.0/16 which is allocated to an organization that also operates AS 111. In BGP, AS 111 announces the IP prefix 168.122.0.0/16 as well as its subprefix 168.122.225.0/24. Therefore, the RPKI should contain a ROA authorizing AS 111 to originate these two IP prefixes. That is, the ROA should be

> ROA:(168.122.0.0/16,168.122.225.0/24, AS 111)

This ROA is "minimal" because it includes only those two prefixes that are actually originated by AS 111 in BGP. [RFC6907]

Now suppose an attacking AS 666 originates a BGP announcement for a subprefix 168.122.0.0/24. This is a standard "subprefix hijack".

In the absence of the minimal ROA above, AS 666 could intercept traffic for the addresses in 168.122.0.0/24. This is because routers perform a longest-prefix match when deciding where to forward IP packets, and 168.122.0.0/24 originated by AS 666 is a longer prefix than 168.122.0.0/16 originated by AS 111.

However, the ROA above renders AS 666's BGP announcement invalid, because (1) this ROA "covers" the attacker's announcement (since 168.122.0.0/24 is a subprefix of 168.122.0.0/16), and (2) there is no ROA "matching" the attacker's announcement (there is no ROA for AS 666 and IP prefix 168.122.0.0/24) [RFC6811]. If routers ignore invalid BGP announcements, the minimal ROA above ensures that the subprefix hijack will fail.

Now suppose that instead the ROA above was replaced with a "loose ROA" that used maxLength as a shorthand for set of IP prefixes that AS 111 is authorized to announce. The ROA would be:

> ROA:(168.122.0.0/16-24, AS 111)

This ROA authorizes AS 111 to originate any subprefix of 168.122.0.0/16, up to length /24. That is, AS 111 could originate 168.122.225.0/24 as well as all of 168.122.0.0/17, 168.122.128.0/17, ..., 168.122.255.0/24 but not 168.122.0.0/25.

However, AS 111 only originates two prefixes in BGP: 168.122.0.0/16 and 168.122.255.0/24. This means that all other prefixes authorized by the loose ROA (for instance, 168.122.0.0/24), are vulnerable to the following forged-origin subprefix hijack [RFC7115,[GCHSS]]:

> The hijacker AS 666 sends a BGP announcement "168.122.0.0/24: AS 666, AS 111", falsely claiming that AS 666 is a neighbor of AS 111 and falsely claiming that AS 111 originates the IP prefix 168.122.0.0/24. In fact, the IP prefix 168.122.0.0/24 is not announced by AS 111.

The hijacker's BGP announcement is valid according the RPKI, since the ROA (168.122.0.0/16-24, AS 111) authorizes AS 111 to originate BGP routes for 168.122.0.0/24. Becaue AS 111 does not actually originate a route for 168.122.0.0/24, the hijacker's route is the *only* route to the 168.122.0.0/24. Longest-prefix-match routing ensures that the hijacker's route to the subprefix 168.122.0.0/24 is

always preferred over the legitimate route to 168.122.0.0/16 announced by AS 111. Thus, if the hijacker's route propagates through the Internet, the hijacker will intercept traffic destined for IP addresses in 168.122.0.0/24.

The forged origin *subprefix* hijack would have failed if "minimal ROA" described above was used instead of the "loose ROA". If the "minimal ROA" had been used instead, the attacker would be forced to launch a forged origin *prefix* hijack in order to attract traffic, as follows:

> The hijacker AS 666 sends a BGP announcement "168.122.0.0/16: AS 666, AS 111", falsely claiming that AS 666 is a neighbor of AS 111.

Notice, however, that this hijack is significantly less effective for the hijacker, since AS 111 is actually originating 168.122.0.0/16 in BGP. In contrast to the forged-origin subprefix hijack, with this hijack AS 666 is not presenting the *only* route to 168.122.0.0/16. Moreover, the path originated by AS 666 is one hop longer than the path originated by the legitimate origin AS 111. As discussed in [LSG16], this means that the hijacker will attract less traffic than he would have in the forged origin *subprefix* hijack.

In sum, a forged-origin subprefix hijack has exactly the same impact as a regular subprefix hijack. A forged-origin subprefix hijack is also more damaging than than forged-origin prefix hijack.

Any ROA with maxLength m longer than the prefix length p is vulnerable to a forged-origin subprefix hijack, unless every subprefix of prefix p of length m is legitimately announced in BGP.

4. Measurements of Today's RPKI

Network measurements from September 13, 2016 show that 16% of the IP prefixes authorized in ROAs have a maxLength longer than their prefix length. The vast majority of these (89%) of these are vulnerable to forged-origin subprefix hijacks. Even large providers are vulnerable to these attacks. See [GSG16] for details.

These measurements suggest that operators commonly misconfigure the maxLength attribute, and unwittingly open themselves up to forged-origin subprefix hijacks.

5. Use Minimal ROAs without Maxlength

This document recommends that operators avoid using the maxLength attribute in their ROAs.

Operators should use "minimal ROAs" whenever possible. A minimal ROA enumerates the exact list of IP prefixes that are actually originated by an AS in BGP, as described in the running example of Section 3.

Sometimes, it is not possible to use a "minimal ROA", because an operator wants to issue a ROA that includes an IP prefix that is sometimes (but not always) announced in BGP. In this case the ROA should still consist of an explicit list of IP prefixes, including those prefixes that are sometimes, but not always announced in BGP. The list of prefixes should still avoid the use of the maxLength attribute.

This practice requires no changes to the RPKI specification and will not increase the number of signed ROAs in the RPKI, because ROAs already support lists of IP prefixes [RFC6482]. See also [GSG16] for further discussion of why this practice will have minimal impact on the performance of the RPKI ecosystem.

5.1. When a Minimal ROA Cannot Be Used?

We now extend our running example to illustrate one situation where where it is not possible to issue a minimal ROA.

Suppose AS 111 has a contract with a DDoS mitigation service provider that holds AS 222. When a DDoS attack is detected, AS 222 immediately originates 168.122.0.0/17 and 168.122.128.0/17, thus attracting all the DDoS traffic to itself. The traffic is scrubbed at AS 222 and then and sent back to AS 111 over a backhaul data link. Notice that, during a DDoS attack, the DDoS mitigation service provider AS 222 originates two /17 prefixes that are longer than than AS 111’s /16 prefix, and so all the traffic that normally goes to AS 111 goes to AS 222 instead.

First, suppose the RPKI only had the minimal ROA for AS 111, as described in Section 3. But, if there is no ROA authorizing AS 222 to announce the two /17 prefixes, then the traffic-scrubbing scheme would not work. That is, if AS 222 originates the two /17 prefixes in BGP during a DDoS attack, the announcement would be invalid [RFC6811].

Instead, the RPKI should have two ROAs: one for AS 111 and one for AS 222.

ROA:(168.122.0.0/16,168.122.225.0/24, AS 111)

ROA:(168.122.0.0/17,168.122.128.0/17, AS 222)

Neither ROA uses the maxLength attribute. But, the second ROA is not "minimal" because it contains two /17 prefixes that are not announced by anyone in BGP during normal operations. These two /17 prefixes are only announced by AS 222 as part of its DDoS mitigation service during a DDoS attack.

Notice, however, that this scheme does not come without risks. Namely, all of the IP addresses in 168.122.0.0/16 (except those in 68.122.225.0/24) are vulnerable to a forged-origin subprefix hijack during normal operations, when the two /17 prefixes are not announced. (The hijacker AS 666 would send the BGP announcement ‘168.122.0.0/17: AS 666, AS 222’’, falsely claiming that AS 666 is a neighbor of AS 222 and falsely claiming that AS 222 originates 168.122.0.0/17.)

Thus, a better approach would be to limit the address space in the ROA for AS 222, so it includes only those IP addresses that must actively be protected by the DDoS mitigation service provider. For instance, if DDoS protection is contracted only for those servers in AS 111 that have addresses in 168.122.0.0/23, then the following ROAs suffice:

ROA:(168.122.0.0/16,168.122.225.0/24, AS 111)

ROA:(168.122.0.0/23, AS 222)

Now, fewer IP addresses (namely, only those addresses in 168.122.0.0/23) are vulnerable to forged origin subprefix hijacks, and DDoS mitigation service could still protect these addresses during DDoS attacks.

6. Contributors

This document would not be possible without the work of Omar Sagga (Boston University).

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <http://www.rfc-editor.org/info/rfc4271>.

[RFC6480] Lepinski, M. and S. Kent, "An Infrastructure to Support Secure Internet Routing", RFC 6480, DOI 10.17487/RFC6480, February 2012, <http://www.rfc-editor.org/info/rfc6480>.

[RFC6482] Lepinski, M., Kent, S., and D. Kong, "A Profile for Route Origin Authorizations (ROAs)", RFC 6482, DOI 10.17487/RFC6482, February 2012, <http://www.rfc-editor.org/info/rfc6482>.

[RFC6811] Mohapatra, P., Scudder, J., Ward, D., Bush, R., and R. Austein, "BGP Prefix Origin Validation", RFC 6811, DOI 10.17487/RFC6811, January 2013, <http://www.rfc-editor.org/info/rfc6811>.

7.2. Informative References

[GSG16] Gilad, Y., Sagga, O., and S. Goldberg, "Maxlength Considered Harmful to the RPKI", in ePrint Cryptology Archive 2016/1015, February 2017, <https://eprint.iacr.org/2016/1015.pdf>.

[LSG16] Lychev, R., Shapira, M., and S. Goldberg, "Rethinking Security for Internet Routing", in Communications of the ACM, October 2016, <http://cacm.acm.org/magazines/2016/10/207763-rethinking-security-for-internet-routing/>.

[GCHSS] Gilad, Y., Cohen, A., Herzberg, A., Schapira, M., and H. Shulman, "Are We There Yet? On RPKI's Deployment and Security", in NDSS 2017, February 2017, <https://eprint.iacr.org/2016/1010.pdf>.

[RFC6907] Manderson, T., Sriram, K., and R. White, "Use Cases and Interpretations of Resource Public Key Infrastructure (RPKI) Objects for Issuers and Relying Parties", RFC 6907, DOI 10.17487/RFC6907, March 2013, <http://www.rfc-editor.org/info/rfc6907>.

[RFC7115] Bush, R., "Origin Validation Operation Based on the Resource Public Key Infrastructure (RPKI)", BCP 185, RFC 7115, DOI 10.17487/RFC7115, January 2014, <http://www.rfc-editor.org/info/rfc7115>.

[I-D.ietf-sidr-bgpsec-protocol]
Lepinski, M. and K. Sriram, "BGPsec Protocol Specification", draft-ietf-sidr-bgpsec-protocol-22 (work in progress), January 2017.

Authors' Addresses

Yossi Gilad
Boston University
111 Cummington St, MCS135
Boston, MA 02215
USA

EMail: yossigi@bu.edu

Sharon Goldberg
Boston University
111 Cummington St, MCS135
Boston, MA 02215
USA

EMail: goldbe@cs.bu.edu

Kotikalapudi Sriram
NIST
100 Bureau Drive
Gaithersburg, MD 20899
USA

EMail: kotikalapudi.sriram@nist.gov