Workgroup: Network Working Group
Internet-Draft:
draft-ietf-idr-bgp-open-policy-20
Published: 18 January 2022
Intended Status: Standards Track
Expires: 22 July 2022
Authors: A. Azimov, Qrator Labs & Yandex
E. Bogomazov, Qrator Labs
R. Bush, Internet Initiative Japan & Arrcus, Inc.
K. Patel, Arrcus
K. Sriram, USA NIST

Route Leak Prevention and Detection using Roles in UPDATE and OPEN Messages

Abstract

Route leaks are the propagation of BGP prefixes that violate assumptions of BGP topology relationships, e.g., announcing a route learned from one transit provider to another transit provider or a lateral (i.e., non-transit) peer or announcing a route learned from one lateral peer to another lateral peer or a transit provider. These are usually the result of misconfigured or absent BGP route filtering or lack of coordination between autonomous systems (ASes). Existing approaches to leak prevention rely on marking routes by operator configuration, with no check that the configuration corresponds to that of the eBGP neighbor, or enforcement that the two eBGP speakers agree on the relationship. This document enhances the BGP OPEN message to establish an agreement of the relationship on each eBGP session between autonomous systems in order to enforce appropriate configuration on both sides. Propagated routes are then marked according to the agreed relationship, allowing both prevention and detection of route leaks.

Requirements Language

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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Route leaks are the propagation of BGP prefixes that violate assumptions of BGP topology relationships, e.g., announcing a route learned from one transit provider to another transit provider or a lateral (i.e., non-transit) peer or announcing a route learned from one lateral peer to another lateral peer or a transit provider [RFC7908]. These are usually the result of misconfigured or absent BGP route filtering or lack of coordination between autonomous systems (ASes).

Existing approaches to leak prevention rely on marking routes by operator configuration, with no check that the configuration corresponds to that of the eBGP neighbor, or enforcement that the two eBGP speakers agree on the relationship. This document enhances the BGP OPEN message to establish an agreement of the relationship on each eBGP session between autonomous systems in order to enforce appropriate configuration on both sides. Propagated routes are then marked according to the agreed relationship, allowing both prevention and detection of route leaks.

This document specifies a means of replacing the operator driven configuration-based method of route leak prevention, described above, with a built-in method for route leak prevention and detection.

This method uses a new configuration parameter, BGP Role, which is negotiated using a BGP Role Capability in the OPEN message [RFC5492]. An eBGP speaker may require the use of this capability and confirmation of BGP Role with a neighbor for the BGP OPEN to succeed.

An optional, transitive BGP Path Attribute, called Only to Customer (OTC), is specified in Section 4. It prevents ASes from creating leaks and detects leaks created by the ASes in the middle of an AS path. The main focus/applicability is the Internet (IPv4 and IPv6 unicast route advertisements).

2. Terminology

The terms "local AS" and "remote AS" are used to refer to the two ends of an eBGP session. The "local AS" is the AS where the protocol action being described is to be performed, and "remote AS" is the AS at the other end of the eBGP session in consideration.

The use of the term "route is ineligible" in this document has the same meaning as in [RFC4271], i.e., "route is ineligible to be installed in Loc-RIB and will be excluded from the next phase of route selection."

2.1. Peering Relationships

The terms defined and used in this document (see below) do not necessarily represent business relationships based on payment agreements. These terms are used to represent restrictions on BGP route propagation, sometimes known as the Gao-Rexford model [Gao].

The following is a list of BGP Roles for eBGP peering and the corresponding rules for route propagation:

Provider: MAY propagate any available route to a Customer.

Customer: MAY propagate any route learned from a Customer, or locally originated, to a Provider. All other routes MUST NOT be propagated.

Route Server (RS): MAY propagate any available route to a Route Server Client (RS-Client).

Route Server Client (RS-Client): MAY propagate any route learned from a Customer, or locally originated, to an RS. All other routes MUST NOT be propagated.

Peer: MAY propagate any route learned from a Customer, or locally originated, to a Peer. All other routes MUST NOT be propagated.

If the local AS has one of the above Roles (in the order shown), then the corresponding peering relationship with the remote AS is Provider-to-Customer, Customer-to-Provider, RS-to-RS-Client, RS-Client-to-RS, or Peer-to-Peer (i.e., lateral peers), respectively. These are called normal peering relationships.

If the local AS has more than one peering role with the remote AS such peering relation is called Complex. An example is when the peering relationship is Provider-to-Customer for some prefixes while it is Peer-to-Peer for other prefixes [Gao].

A BGP speaker may apply policy to reduce what is announced, and a recipient may apply policy to reduce the set of routes they accept. Violation of the above rules may result in route leaks. Automatic enforcement of these rules should significantly reduce route leaks that may otherwise occur due to manual configuration mistakes.

As specified in Section 4, the Only to Customer (OTC) Attribute is used to identify all the routes in the AS that have been received from a Peer, Provider, or RS.

3. BGP Role

The BGP Role characterizes the relationship between the eBGP speakers forming a session. One of the Roles described below SHOULD be configured at the local AS for each eBGP session (see definitions in Section 2) based on the local AS's knowledge of its Role. The only exception is when the eBGP connection is Complex (see Section 5). BGP Roles are mutually confirmed using the BGP Role Capability (described in Section 3.1) on each eBGP session.

Allowed Roles for eBGP sessions are:

* Provider - the local AS is a transit Provider of the remote AS;
* Customer - the local AS is a transit Customer of the remote AS;
* RS - the local AS is a Route Server (usually at an Internet exchange point) and the remote AS is its RS-Client;
* RS-Client - the local AS is a client of an RS and the RS is the remote AS;
* Peer - the local and remote ASes are Peers (i.e., have a lateral peering relationship).

3.1. BGP Role Capability

The BGP Role Capability is defined as follows:

* Code - 9
* Length - 1 (octet)
* Value - integer corresponding to speaker's BGP Role (see Table 1).

Value	Role name (for the local AS)
0	Provider
1	RS
2	RS-Client
3	Customer
4	Peer (Lateral Peer)
5-255	Unassigned

Table 1: Predefined BGP Role Values

If BGP Role is locally configured, the eBGP speaker MUST advertise BGP Role Capability in the BGP OPEN message. An eBGP speaker MUST NOT advertise multiple versions of the BGP Role Capability.

3.2. Role Correctness

Section 3.1 described how BGP Role encodes the relationship on each eBGP session between autonomous systems (ASes).

The mere receipt of BGP Role Capability does not automatically guarantee the Role agreement between two eBGP neighbors. If the BGP Role Capability is advertised, and one is also received from the peer, the Roles MUST correspond to the relationships in Table 2. If the Roles do not correspond, the BGP speaker MUST reject the connection using the Role Mismatch Notification (code 2, subcode TBD).

Local AS Role	Remote AS Role
Provider	Customer
Customer	Provider
RS	RS-Client
RS-Client	RS
Peer	Peer

Table 2: Allowed Pairs of Role Capabilities

For backward compatibility, if the BGP Role Capability is sent but one is not received, the BGP Speaker SHOULD ignore the absence of the BGP Role Capability and proceed with session establishment. The locally configured BGP Role is used for the procedures described in Section 4.

An operator may choose to apply a "strict mode" in which the receipt of a BGP Role Capability from the remote AS is required. When operating in the "strict mode", if the BGP Role Capability is sent, but one is not received, then the connection is rejected using the Role Mismatch Notification (code 2, subcode TBD). See comments in Section 7.

If an eBGP speaker receives multiple but identical BGP Role Capabilities with the same value in each, then the speaker must consider it to be a single BGP Role Capability and proceed [RFC5492]. If multiple BGP Role Capabilities are received and not all of them have the same value, then the BGP speaker MUST reject the connection using the Role Mismatch Notification (code 2, subcode TBD).

The BGP Role value for the local AS (in conjunction with the OTC Attribute in the received UPDATE message) is used in the route leak prevention and detection procedures described in Section 4.

4. BGP Only to Customer (OTC) Attribute

The Only to Customer (OTC) Attribute is an optional transitive path attribute of the UPDATE message with Attribute Type Code 35 and a length of 4 octets. The purpose of this attribute is to guarantee that once a route is sent to a Customer, Peer, or RS-Client, it will subsequently go only to Customers. The attribute value is an AS number (ASN) determined by the procedures described below.

The following ingress procedure applies to the processing of the OTC Attribute on route receipt:

1. If a route with the OTC Attribute is received from a Customer or RS-Client, then it is a route leak and MUST be considered ineligible (see Section 2).

2. If a route with the OTC Attribute is received from a Peer and the Attribute has a value that is not equal to the remote (i.e., Peer's) AS number, then it is a route leak and MUST be considered ineligible.

3. If a route is received from a Provider, Peer, or RS, and the OTC Attribute is not present, then it MUST be added with a value equal to the AS number of the remote AS.

The following egress procedure applies to the processing of the OTC Attribute on route advertisement:

1. If a route is to be advertised to a Customer, Peer, or RS-Client (when the sender is an RS), and the OTC Attribute is not present, then an OTC Attribute MUST be added with a value equal to the AS number of the local AS.

2. If a route already contains the OTC Attribute, it MUST NOT be propagated to Providers, Peers, or RS(s).

The above-described procedures provide both leak prevention for the local AS and leak detection and mitigation multiple hops away. In the case of prevention at the local AS, the presence of an OTC Attribute indicates to the egress router that the route was learned from a Peer, Provider, or RS, and it can be advertised only to the customers. The same OTC Attribute which is set locally also provides a way to detect route leaks by an AS multiple hops away if a route is received from a Customer, Peer, or RS-Client. For example, if an AS sets the OTC Attribute on a route sent to a Peer and the route is subsequently received by a compliant AS from a Customer, then the receiving AS detects (based on the presence of the OTC Attribute) that the route is a leak.

The OTC Attribute might be set at the egress of the remote AS or at the ingress of the local AS, i.e., if the remote AS is noncompliant with this specification, then the local AS will have to set the OTC Attribute if it is absent. In both scenarios, the OTC value will be the same. This makes the scheme more robust and benefits early adopters.

If an eBGP speaker receives an UPDATE with an OTC Attribute with a length different from 4 octets, then the UPDATE SHALL be considered malformed. If malformed, the UPDATE message SHALL be handled using the approach of "treat-as-withdraw" [RFC7606].

The procedures specified in this document are NOT RECOMMENDED to be used between autonomous systems in an AS Confederation [RFC5065]. If an OTC Attribute is added on egress from the AS Confederation, its value MUST equal the AS Confederation Identifier. Also, on egress from the AS Confederation, an UPDATE MUST NOT contain an OTC Attribute with a value corresponding to any Member-AS Number other than the AS Confederation Identifier.

The procedures specified in this document in scenarios that use private AS numbers behind an Internet-facing ASN (e.g., a data center network [RFC7938] or stub customer) may be used, but any details are outside the scope of this document. On egress from the Internet-facing AS, the OTC Attribute MUST NOT contain a value other than the Internet-facing ASN.

Once the OTC Attribute has been set, it MUST be preserved unchanged (this also applies to an AS Confederation).

The described ingress and egress procedures are applicable only for unicast IPv4 and IPv6 address families and MUST NOT be applied to other address families by default. The operator MUST NOT have the ability to modify the procedures defined in this section.

5. Additional Considerations

Roles MUST NOT be configured on an eBGP session with a Complex peering relationship. If multiple eBGP sessions can segregate the Complex peering relationship into eBGP sessions with normal peering relationships, BGP Roles SHOULD be used on each of the resulting eBGP sessions.

An operator may want to achieve an equivalent outcome by configuring policies on a per-prefix basis to follow the definitions of peering relations as described in Section 2.1. However, in this case, there are no built-in measures to check the correctness of the per-prefix peering configuration.

The incorrect setting of BGP Roles and/or OTC Attributes may affect prefix propagation. Further, this document does not specify any special handling of incorrect AS numbers in the OTC Attribute.

In AS migration scenarios [RFC7705], a given router may represent itself as any one of several different ASes. This should not be a problem since the egress procedures in Section 4 specify that the OTC Attribute is to be attached as part of route transmission. Therefore, a router is expected to set the OTC value equal to the ASN it is currently representing itself as.

6. IANA Considerations

IANA has registered a new BGP Capability (Section 3.1) in the "Capability Codes" registry's "IETF Review" range [RFC5492]. The description for the new capability is "BGP Role". IANA has assigned the value 9 [to be removed upon publication: https://www.iana.org/assignments/capability-codes/capability-codes.xhtml]. This document is the reference for the new capability.

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" in the Capability Codes registry. Assignments consist of a Value and a corresponding Role name. Initially, this registry is to be populated with the data contained in Table 1 found in Section 3.1. Future assignments may be made by the "IETF Review" policy as defined in [RFC8126]. The registry is as shown in Table 3.

Value	Role name (for the local AS)	Reference
0	Provider	This document
1	RS	This document
2	RS-Client	This document
3	Customer	This document
4	Peer (i.e., Lateral Peer)	This document
5-255	To be assigned by IETF Review	

Table 3: IANA Registry for BGP Role

IANA has registered a new OPEN Message Error subcode named the "Role Mismatch" (see Section 3.2) in the OPEN Message Error subcodes registry. IANA has assigned the value TBD [to be removed upon publication: https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-6]. This document is the reference for the new subcode.

IANA has also registered a new path attribute named "Only to Customer (OTC)" (see Section 4) in the "BGP Path Attributes" registry. IANA has assigned code value 35 [To be removed upon publication: http://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-2]. This document is the reference for the new attribute.

7. Security Considerations

The security considerations of BGP (as specified in [RFC4271] and [RFC4272]) apply.

This document proposes a mechanism using BGP Role for the prevention and detection of route leaks that are the result of BGP policy misconfiguration. A misconfiguration of the BGP Role may affect prefix propagation. For example, if a downstream (i.e., towards a Customer) peering link were misconfigured with a Provider or Peer Role, this will limit the number of prefixes that can be advertised in this direction. On the other hand, if an upstream provider were misconfigured (by a local AS) with the Customer Role, this may result in propagating routes that are received from other Providers or Peers. But the BGP Role negotiation and the resulting confirmation of Roles make such misconfigurations unlikely.

Setting the strict mode of operation for BGP Role negotiation as the default may result in a situation where the eBGP session will not come up after a software update. Implementations with such default behavior are strongly discouraged.

Removing the OTC Attribute or changing its value can limit the opportunity of route leak detection. Such activity can be done on purpose as part of an on-path attack. For example, an AS can remove the OTC Attribute on a received route and then leak the route to its transit provider. This kind of threat is not new in BGP and it may affect any Attribute (Note: BGPsec [RFC8205] offers protection only for the AS_PATH Attribute).

Adding an OTC Attribute when the route is advertised from Customer to Provider will limit the propagation of the route. Such a route may be considered as ineligible by the immediate Provider or its Peers or upper layer Providers. This kind of OTC Attribute addition is unlikely to happen on the Provider side because it will limit the traffic volume towards its Customer. On the Customer side, adding an OTC Attribute for traffic engineering purposes is also discouraged because it will limit route propagation in an unpredictable way.

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, <https://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, <https://www.rfc-editor.org/info/rfc4271>.

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, DOI 10.17487/RFC5065, August 2007, <https://www.rfc-editor.org/info/rfc5065>.

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

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[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, <https://www.rfc-editor.org/info/rfc7908>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

8.2. Informative References

[Gao] Gao, L. and J. Rexford, "Stable Internet routing without global coordination", IEEE/ACM Transactions on Networking, Volume 9, Issue 6, pp 689-692, DOI 10.1109/90.974523, December 2001, <https://ieeexplore.ieee.org/document/974523>.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <https://www.rfc-editor.org/info/rfc4272>.

[RFC7705] George, W. and S. Amante, "Autonomous System Migration Mechanisms and Their Effects on the BGP AS_PATH Attribute", RFC 7705, DOI 10.17487/RFC7705, November 2015, <https://www.rfc-editor.org/info/rfc7705>.

[RFC7938] Lapukhov, P., Premji, A., and J. Mitchell, Ed., "Use of BGP for Routing in Large-Scale Data Centers", RFC 7938, DOI 10.17487/RFC7938, August 2016, <https://www.rfc-editor.org/info/rfc7938>.

[RFC8205] Lepinski, M., Ed. and K. Sriram, Ed., "BGPsec Protocol Specification", RFC 8205, DOI 10.17487/RFC8205, September 2017, <https://www.rfc-editor.org/info/rfc8205>.

Acknowledgments

The authors wish to thank Alvaro Retana, Andrei Robachevsky, Daniel Ginsburg, Jeff Haas, Ruediger Volk, Pavel Lunin, Gyan Mishra, Ignas Bagdonas, Sue Hares, and John Scudder for comments, suggestions, and critique.

Contributors

Brian Dickson
Independent
Email: brian.peter.dickson@gmail.com

Doug Montgomery
USA National Institute of Standards and Technology
Email: dougm@nist.gov

Authors' Addresses

Alexander Azimov
Qrator Labs & Yandex
Ulitsa Lva Tolstogo 16
Moscow
119021
Russian Federation

Email: a.e.azimov@gmail.com

Eugene Bogomazov
Qrator Labs
1-y Magistralnyy tupik 5A
Moscow
123290
Russian Federation

Email: eb@qrator.net

Randy Bush
Internet Initiative Japan & Arrcus, Inc.
5147 Crystal Springs
Bainbridge Island, Washington 98110
United States of America

Email: randy@psg.com

Keyur Patel
Arrcus
2077 Gateway Place, Suite #400

San Jose, CA 95119
United States of America

Email: keyur@arrcus.com

Kotikalapudi Sriram
USA National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899
United States of America

Email: ksriram@nist.gov