Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 8, 2021

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

July 7, 2020

Route Leak Prevention using Roles in Update and Open messages
draft-ietf-idr-bgp-open-policy-13

Abstract

Route leaks are the propagation of BGP prefixes which violate assumptions of BGP topology relationships; e.g. passing a route learned from one lateral peer to another lateral peer or a transit provider, passing a route learned from one transit provider to another transit provider or a lateral peer. 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 BGP OPEN to establish agreement of the (peer, customer, provider, Route Server, Route Server client) relationship of two neighboring eBGP speakers to enforce appropriate configuration on both sides. Propagated routes are then marked with an Only to Customer (OTC) attribute 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

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 https://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 January 8, 2021.

Table of Contents

1. Introduction

A BGP route leak occurs when a route is learned from a transit provider or lateral peer and then announced to another provider or lateral peer. See [RFC7908]. These are usually the result of misconfigured or absent BGP route filtering or lack of coordination between two eBGP speakers.

The mechanism proposed in [I-D.ietf-grow-route-leak-detection-mitigation] uses large-communities to perform detection and mitigation of route leaks. While signaling using communities is easy to implement and deploy quickly, it normally relies on operator-maintained policy configuration, which is vulnerable to misconfiguration [Streibelt]. The community signal can also be stripped at the ISP boundaries.

This document provides configuration automation using 'BGP roles', which are negotiated using a new BGP Capability Code in OPEN message (see Section 4 in [RFC5492]). Either or both BGP speakers MAY be configured to require that this capability be agreed for the BGP OPEN to succeed.

A new optional, transitive BGP Path Attribute Only to Customer (OTC) is specified that SHOULD be automatically configured using BGP roles. This attribute prevents networks from creating leaks, and detects leaks created by third parties.

In the rest of this document, we use the term "peer" to refer to "lateral peer" for simplicity.

2. Peering Relationships

Despite the use of terms such as "customer", "peer", etc. in this document, these are not necessarily 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 various roles in eBGP peering and the corresponding rules for route propagation:

Provider: MAY send to a customer all available prefixes.

Customer: MAY send to a provider prefixes which the sender originates and prefixes learned from any of their customers. A customer MUST NOT send to a provider prefixes learned from its peers, from other providers, or from Route Servers.

Route Server (RS): MAY send to an Route Server client (RS-client) all available prefixes.

RS-client: MAY send to an RS prefixes which the sender originates and prefixes learned from its customers. An RS-client MUST NOT send to an RS prefixes learned from its peers or providers, or from another RS.

Peer: MAY send to a peer prefixes which the sender originates and prefixes learned from its customers. A peer MUST NOT send to a peer prefixes learned from other peers, from its providers, or from RS(s).

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. Violation of the above rules may result in route leaks and MUST not be allowed. Automatic enforcement of these rules should significantly reduce route leaks that may otherwise occur due to manual configuration mistakes. While enforcing the above rules will address most BGP peering scenarios, their configuration is not part of BGP itself; therefore, configuration of ingress and egress prefix filters is still strongly advised.

3. BGP Role

BGP Role is a new configuration option that is configured on a per-session basis. BGP Roles reflect the agreement between two BGP speakers about their relationship. One of the Roles described below SHOULD be configured on each eBGP session between ISPs that carry IPv4 and(or) IPv6 unicast prefixes.

Allowed Role values for eBGP sessions between ISPs are:

- Provider - sender is a transit provider to neighbor;
- Customer - sender is a transit customer of neighbor;
- RS - sender is a Route Server, usually at an Internet exchange point (IX);
- RS-client - sender is client of an RS;
- Peer - sender and neighbor are peers.

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

4. BGP Role Capability

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

- Type - <TBD1>;
- Length - 1 (byte);
- Value - integer corresponding to speaker's BGP Role (see Table 1).

Value	Role name
0	Sender is Provider
1	Sender is RS
2	Sender is RS-client
3	Sender is Customer
4	Sender is Peer

Table 1: Predefined BGP Role Values

5. Role correctness

Section 3 described how BGP Role encodes the relationship between two eBGP 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 applied with a set of constraints on how speakers' BGP Roles MUST correspond. 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 MUST check the value of the received capability (i.e., the sender's role) with its own BGP Role. The allowed pairings are as follows:

Sender's Role	Receiver's Role
Provider	Customer
Customer	Provider
RS	RS-client
RS-client	RS
Peer	Peer

Table 2: Allowed Pairs of Role Capabilities

If the role of the receiving speaker for the eBGP session in consideration is included in Table 1 and the observed Role pair is not in the above table, then the receiving speaker MUST reject the eBGP connection, send a Role Mismatch Notification (code 2, subcode <TBD2>), and also send a Connection Rejected Notification [RFC4486] (Notification with error code 6, subcode 5).

5.1. 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 a neighbor which does not announce the BGP Role capability in the OPEN message. If a speaker rejects a connection, it MUST send a send a Role Mismatch Notification (code 2, subcode <TBD2>), and also 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.

6. BGP Only to Customer (OTC) Attribute

Newly defined here, the Only to Customer (OTC) is an optional, 4 bytes long, transitive BGP Path attribute with the Type Code <TBD3>. The purpose of this attribute is to guarantee that once a route is sent to customer, peer, or RS-client, it will subsequently go only to customers. The value of OTC is an AS number determined by policy as described below. The semantics and usage of the OTC attribute are made clear by the ingress and egress policies described below.

The following ingress policy applies to the OTC attribute:

1. If a route with OTC attribute is received from a Customer or RS-client, then it is a route leak and MUST be rejected.

2. If a route with OTC attribute is received from a Peer and its value is not equal to the sending neighbor's Autonomous System (AS) number, then it is a route leak and MUST be rejected.

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 value equal to the sending neighbor's AS number.

The egress policy MUST be:

1. A route with the OTC attribute set MUST NOT be sent to Providers, Peers, or RS(s).

2. If route is sent to a Customer or Peer, or an RS-client (when the sender is an RS) and the OTC attribute is not present, then it MUST be added with value equal to AS number of the sender.

Once the OTC attribute has been set, it MUST be preserved unchanged.

7. Enforcement

Having the relationship unequivocally agreed between the two peers in BGP OPEN is critical; BGP implementations MUST enforce the relationship/role establishment rules (see Section 5) in order to ameliorate operator policy configuration errors (if any).

Similarly, the application of that relationship on prefix propagation using OTC MUST be enforced by the BGP implementations, and not exposed to user misconfiguration.

As opposed to communities, BGP attributes may not be generally modified or stripped by the operator; BGP router implementations enforce such treatment. This is the desired property for the OTC marking. Hence, this document specifies OTC as an attribute.

8. Additional Considerations

There are peering relationships that are 'complex', i.e., both parties are intentionally sending prefixes received from each other to their non-transit peers and/or transit providers. If multiple BGP peerings can segregate the 'complex' parts of the relationship, the complex peering roles can be segregated into different normal BGP sessions, and BGP Roles MUST be used on each of the resulting normal (non-complex) BGP sessions.

No Roles SHOULD be configured on a 'complex' BGP session (assuming it is not segregated) and in that case, OTC MUST be set by configuration on a per-prefix basis. However, there are no built-in measures to check correctness of OTC use if BGP Role is not configured.

The incorrect setting of BGP Roles and/or OTC attributes may affect prefix propagation. Further, this document doesn't specify any special handling of incorrect/private ASNs in OTC attribute; such errors should not happen with proper configuration.

As the BGP Role reflects the peering relationship between neighbors, it might have other uses beyond the route leak solution discussed so far. For example, BGP Role might affect route priority, or be used to distinguish borders of a network if a network consists of multiple ASs. Though such uses may be worthwhile, they are not the goal of

this document. Note that such uses would require local policy control.

The use of BGP Roles are specified for unicast IPv4 and IPv6 address families. While BGP roles can be configured on other address families its applicability for these cases is out of scope of this document.

As BGP role configuration results in automatic creation of inbound/ outbound filters, existence of roles should be treated as existence of Import and Export policy [RFC8212].

9. IANA Considerations

This document defines a new Capability Codes option [to be removed upon publication: https://www.iana.org/assignments/capability-codes/capability-codes.xhtml] [RFC5492], named "BGP Role" with an 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]. The allocation policy for new entries up to and including value 127 is "Expert Review" [RFC5226]. The allocation policy for values 128 through 251 is "First Come First Served". The values from 252 through 255 are for "Experimental Use".

This document defines a new subcode, "Role Mismatch" with an 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, transitive BGP Path Attributes option, named "Only to Customer (OTC)" with an 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 four bytes.

10. Security Considerations

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

A misconfiguration in OTC setup may affect prefix propagation. But the automation that is provided by BGP roles should make such misconfiguration unlikely.

11. References

11.1. 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>.

[RFC4486] Chen, E. and V. Gillet, "Subcodes for BGP Cease Notification Message", RFC 4486, DOI 10.17487/RFC4486, April 2006, <https://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, <https://www.rfc-editor.org/info/rfc5492>.

[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>.

11.2. 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>.

[I-D.ietf-grow-route-leak-detection-mitigation] Sriram, K. and A. Azimov, "Methods for Detection and Mitigation of BGP Route Leaks", draft-ietf-grow-route-leak-detection-mitigation-02 (work in progress), January 2020.

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

[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>.

[RFC8212] Mauch, J., Snijders, J., and G. Hankins, "Default External BGP (EBGP) Route Propagation Behavior without Policies", RFC 8212, DOI 10.17487/RFC8212, July 2017, <https://www.rfc-editor.org/info/rfc8212>.

[Streibelt] Streibelt, F., Lichtblau, F., Beverly, R., Feldmann, A., Cristel, C., Smaragdakis, G., and R. Bush, "BGP Communities: Even more Worms in the Routing Can", <https://people.mpi-inf.mpg.de/~fstreibelt/preprint/communities-imc2018.pdf>.

Acknowledgements

The authors wish to thank 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
US

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