Inter-Domain Routing
Internet-Draft
Intended status: Standards Track
Expires: May 4, 2017

P. Lapukhov
Facebook
E. Aries, Ed.
P. Marques
Juniper Networks
E. Nkposong
Salesforce.com Inc

October 31, 2016

Use of BGP for Opaque Signaling
draft-lapukhov-bgp-opaque-signaling-03

Abstract

Border Gateway Protocol with multi-protocol extensions (MP-BGP) enables the use of the protocol for dissemination of virtually any information. This document proposes a new Address Family/Subsequent Address Family to be used for distribution of opaque data. This functionality is intended to be used by applications other than BGP for exchange of their own data on top of BGP mesh. The structure of such data SHOULD NOT be interpreted by the regular BGP speakers, rather the goal is to use BGP purely as a convenient and scalable communication system.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

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 May 4, 2017.

Table of Contents

1. Introduction

Implementation of Multiprotocol Extensions for BGP-4 [RFC4760] gives the ability to pass arbitrary data in BGP protocol messages. This capability has been leveraged by many for dissemination of non-routing related information over BGP (e.g. "Dissemination of Flow Specification Rules" [RFC5575] as well as "North-Bound Distribution of Link-State and TE Information using BGP" [I-D.ietf-idr-ls-distribution]). However, there has been no channel defined explicitly to disseminate data with arbitrary payload. The intended use case is for applications other than BGP to leverage the

protocol machinery for distribution (broadcasting) of their own state in the network domain. Publishers and consumers will use BGP UPDATE messages over TCP transport to submit and receive opaque data. It is up to the BGP implementation to provide a custom API for message producers or consumers, if needed.

2. BGP Opaque Data AFI

This document introduces a new AFI known as a "BGP Opaque Data AFI" with the actual value to be assigned by IANA. The purpose of this AFI is to exchange opaque information within a BGP network. The propagation scope is to be controlled by the usual means of BGP policy, except that the policy SHOULD not match on NLRI information in any form other than an opaque string.

3. BGP Key-Value SAFI

This document introduces a new SAFI known as "BGP Key-Value SAFI" with the actual value to be assigned by IANA. The purpose of this SAFI is exchange of opaque information structured as Key-Value binding.

4. BGP VPN Key-Value SAFI

This document introduces a new SAFI known as a "BGP VPN Key-Value SAFI" with the actual value to be assigned by IANA. The purpose of this SAFI is exchange of opaque information structured as a Key-Value binding over service provider backbone providing Virtual Private Networks as a service. The [RFC4364] defines a method and procedures for implementing VPNs using BGP as a control plane. All the procedures of [RFC4364] apply to the BGP VPN Key-Value SAFI. Under this SAFI, the NLRI for the opaque information has the mandatory 8 bytes of Route Distinguisher at the beginning of the NLRI field.

5. Capability Advertisement

A BGP speaker that wishes to exchange Opaque Data MUST use the Multiprotocol Extensions Capability Code, as defined in [RFC4760], to advertise the corresponding AFI/SAFI pair.

6. Disseminating Key-Value bindings

This document proposes a distributed, eventually consistent Key-Value store on top of existing BGP protocol transport mechanism. The "Key" and "Value" portions are to be encoded as the NLRI part of MP_REACH_NLRI attribute.

- Publishers advertise keys along with associated values into the routing domain. The BGP network disseminates that state by propagating the encoded data following regular BGP protocol operations.

- Consumers receive the information via BGP protocol UPDATE messages. Only publishers and consumers of the opaque data are supposed to interpret its contents. The rest of the BGP network acts merely as a dissemination system.

Multiple publishers can advertise the same key bound to different values. Only the "Key" part of MP_REACH_NLRI filed MUST be used to differentiate unique advertisements in such case. It is also possible for the advertised binding to have the same Key-Value pairs, but differ in some other BGP attributes. In that case, the BGP implementation MUST follow the regular best-path selection logic to prevent duplicate information in the network. A consumer will receive the value created by the publisher "closest" in terms of BGP best-path selection logic, based on the policies that exist in the routing domain. This document does not propose methods to achieve strong global consensus for all published values of a given key, as it's generally impossible in eventually consistent data distribution systems.

6.1. Publishing a Key-Value binding

The encoding scheme proposed below follows the semantics of a Key-Value binding. The "Key" and "Value" are stored in the NLRI section of the MP_REACH_NLRI attribute, as shown on Figure 1.

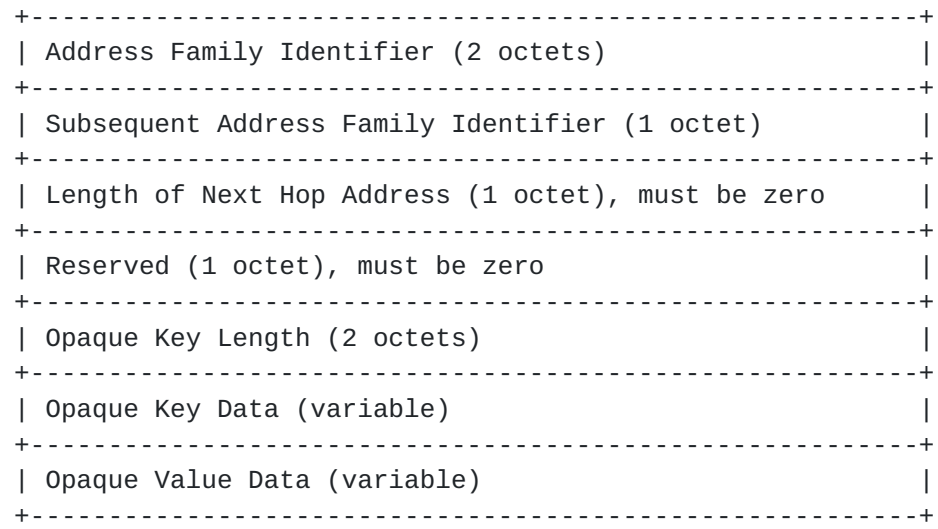

Figure 1: MP_REACH_NLRI Layout

- The AFI/SAFI values are to be allocated by IANA.

- Length of Next Hop Address: must be zero, indicating empty next-hop.

- Opaque Key Length: identifies the size of the Key field in octets, an unsigned integer value. The field MUST have a value of at least one octet under the Key-Value SAFI and at least 9 octets under the VPN Key-Value SAFI. Violating this requirement MUST cause the receiver to ignore the advertised Key-Value binding.

- Opaque Key Data: the byte string representing the opaque key contents.

- Opaque Value Data: The length of this field is determined by subtracting the length of all previous fields from the total length of MP_REACH_NLRI attribute. This field MAY be empty.

The maximum size of the Opaque "Key" and "Value" fields together is limited by the BGP UPDATE message size. With the default BGP protocol implementation is may not exceed 4096 octets (see [RFC4271] Section 4). However, if [I-D.ietf-idr-bgp-extended-messages] is implemented, the UPDATE message size could be as large as 65536 octets.

6.2. Removing a Key-Value binding

The removal procedure follows the regular MP-BGP route withdrawal, using the MP_UNREACH_NLRI attribute. This section defines the attribute structure for the new AFI/SAFI.

The specific MP_UNREACH_NLRI format is shown on Figure 2. This message instructs the receiving BGP speaker to delete the N bindings corresponding to Key 1, Key 2 ... Key N if the keys have been previously learned from the withdrawing speaker. If any of the keys could not be found in the LocRIB or has not been previously received from the withdrawing BGP peer, such key removal request MUST be ignored and the event MAY be logged. For the Key-Value SAFI, each key length field must have the value of at least "1". For the VPN Key-Value SAFI, each key length must be at least 9 octets long. Violation of of these constraints MUST cause the receiver of the UPDATE message to ignore the corresponding key withdrawal.

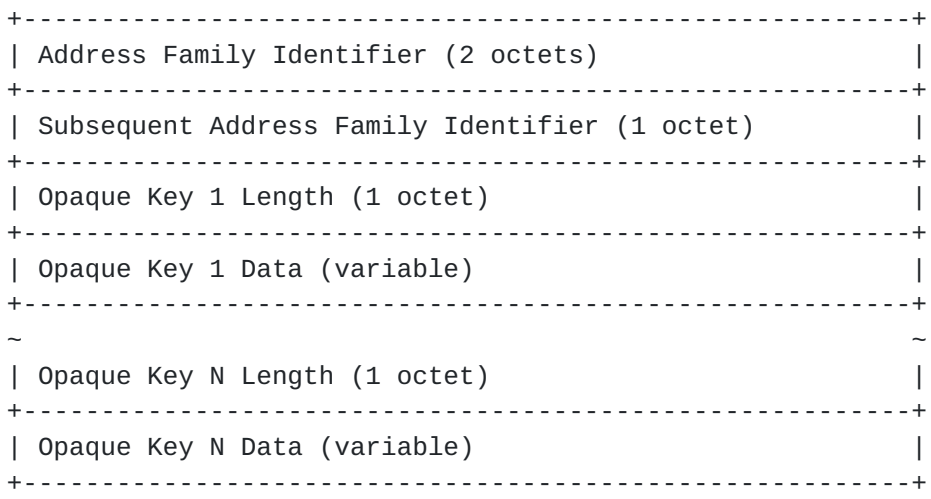

Figure 2: MP_UNREACH_NLRI attribute layout

7. Manageability Considerations

7.1. Propagating multiple values for the same key

It is possible to propagate multiple values associated with the same key using the Add-Path extension defined in [I-D.ietf-idr-add-paths]. However, this document recommends that instead unique key values SHOULD be used for this purpose. It is up to the consumers and publishers of the opaque data to settle on single unique value using some kind of consensus protocol.

As a recommendation, the originators of key-value pairs may use the origin ASN and the IPv4 or IPv6 address assigned to the originating BGP speaker to create a unique key prefix. Alternatively, UUIDs could be used to generate the unique key names, see [RFC4122]

7.2. Automated filtering

One can leverage mechanics described in [RFC4684] and use the route-target extended community attribute to identify "channels" where key-value bindings are published. The consumers would signal their interest in particular "channel" by advertising the corresponding router-target membership. The publications then need to carry the router-target extended community attribute to restrict information propagation.

7.3. Filtering via policy

Ad-doc message filtering could be implemented using BGP standard (see [RFC4271]) or extended community attributes (see [RFC4360]). The semantic of these attributes is to determined by the policy and

publishers/consumers. Filtering could be done locally on receiving BGP speaker, or on remote BGP speaker, by using outbound route filtering feature defined in [RFC5291].

8. IANA Considerations

For the purpose of this work, IANA would be asked to allocate values for the new AFI and SAFIs.

9. Security Considerations

This document does not introduce any changes in terms of BGP security. The usual set of issues that arise from running multiple AFI/SAFI's over single BGP session would apply in this case. Additional concerns may be raised due to increase of the volume and rate of change of the information distributed by means of opaque signaling.

10. Acknowledgements

Keyur Patel provided useful feedback and suggested a practical implementation of unique key semantic and support for VPN Key-Value SAFI.

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, <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>.

11.2. Informative References

[RFC4122] Leach, P., Mealling, M., and R. Salz, "A Universally Unique IDentifier (UUID) URN Namespace", RFC 4122, DOI 10.17487/RFC4122, July 2005, <http://www.rfc-editor.org/info/rfc4122>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <http://www.rfc-editor.org/info/rfc4360>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <http://www.rfc-editor.org/info/rfc4364>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <http://www.rfc-editor.org/info/rfc4684>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <http://www.rfc-editor.org/info/rfc4760>.

[RFC5291] Chen, E. and Y. Rekhter, "Outbound Route Filtering Capability for BGP-4", RFC 5291, DOI 10.17487/RFC5291, August 2008, <http://www.rfc-editor.org/info/rfc5291>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <http://www.rfc-editor.org/info/rfc5575>.

[I-D.ietf-idr-add-paths]
Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", draft-ietf-idr-add-paths-15 (work in progress), May 2016.

[I-D.ietf-idr-ls-distribution]
Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE Information using BGP", draft-ietf-idr-ls-distribution-13 (work in progress), October 2015.

[I-D.ietf-idr-bgp-extended-messages]
Bush, R., Patel, K., and D. Ward, "Extended Message support for BGP", draft-ietf-idr-bgp-extended-messages-13 (work in progress), June 2016.

Authors' Addresses

Petr Lapukhov
Facebook
1 Hacker Way
Menlo Park, CA 94025
US

Email: petr@fb.com

Ebben Aries (editor)
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
US

Email: exa@juniper.net

Pedro Marques
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
US

Email: roque@juniper.net

Edet Nkposong
Salesforce.com Inc
The Landmark @ One Market, ST 300
San Francisco, CA 94105
US

Email: enkposong@salesforce.com