Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 3, 2019

Y. Gu
Huawei
J. Chen
Tencent
P. Mi
S. Zhuang
Z. Li
Huawei

July 02, 2018

VPN Label Monitoring Using BMP
draft-gu-grow-bmp-vpn-label-00

Abstract

The BGP Monitoring Protocol (BMP) is designed to monitor BGP running status, such as BGP peer relationship establishment and termination and route updates. This document provides a method of collecting the VPN label using BMP, as well as an implementation example.

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

Table of Contents

1. Introduction

The Border Gateway Protocol (BGP) [RFC4271], as an inter-Autonomous (AS) routing protocol, is used to exchange network reachability information between BGP systems. Later on, [RFC4760] extends BGP to carry not only the routing information for BGP, but also for multiple Network Layer protocols (e.g., IPv6, Multicast, etc.), known as the MP-BGP (Multiprotocol BGP). The MP-BGP is currently widely deployed in case of MPLS L3VPN, to exchange VPN labels learned for the routes from the customer sites over the MPLS network.

The BGP Monitoring Protocol (BMP) [RFC7854] has been proposed around 2006 to monitor the BGP routing information, which includes the monitoring of BGP peer status, BGP route update, and BGP route statistics. BMP is realized through setting up the TCP session between the monitored BGP device and the BMP monitoring station, and then periodic/event-triggerred messages are sent unidirectionaly from the monitored device to the BMP monitoring station. Before BMP was introduced, such information could be only obtained through manual query, such as screen scraping. The introduction of BMP greatly improves the BGP routing monitoring efficiency without interrupting or interfering the ongoing services.

Currently, BMP is mainly utilized to monitor the public BGP routes. There are also cases that the VPN (Virtual Private Network) route/ label information is needed. For example, for the purpose of Traffic Engineering (TE), the network operator may insert explicit routes, subject to certain constraints or optimization ceriteria, into related routers with high local preferences so that these routes will be selected and installed into the routing table. Under the VPN environment, the VPN route labels should be collected from the devices, and be distributed back jointly with the explicit routes to the devices, so that the devices can use the VPN labels to correlate the received routes with the approriate VRFs (VPN Routing and Forwarding tables). The collection and distribution of such labels could be done by an SDN (Software Defined Network) controller, or an route monitoring station equipped with the traffic optimization module.

The VPN routes between CE (Customer Edge) and PE (Provider Edge) can be monitored by BMP using the "RD Instance Peer Type". However, such VPN routes between CE and PE do not include the VPN labels, since labels are allocated at the PE side. On the other hand, the labeled VPN routes are exchanged beween PE and PE, which could have been monitored by BMP but are currently not implemented due to the massive data exchange between the monitored devices and the BMP monitoring station. An existing method to collect the VPN route label, considering the L3VPN scenario, is by setting up BGP VPNv4 peering relationship between the monitored device and the monitoring station/ controller. The label information is then extracted from the collected VPN-IPv4 routes, carried by the BGP NLRI (Network Layer Reachability Information). However, there are several shortcomings of collecting the VPN label using this method.

o The VPN labels, instead of the VPNv4 routes, are the necessary information for fulfilling the traffic optimization purpose. Thus, extracting the label from the VPNv4 route requires extra work compared with directly collecting the label information alone.

o The same VPN label is sometims used for several VPNv4 routes. Depending on the implementation scenarios, there are typically different ways of allocating the VPN route labels: per route per label, per VRF per label, per interface per label, and so on. For example, in the Multi-AS VPN case, the redistribution of labeled VPN-IPv4 routes from one AS to another can be realized through setting up the EBGP peering between ASBRs (Autonomous System Border Routers) of different ASes. In this case, the per route per label allocation method is preferred. However, per route per label allocation can be very consuming as for the label space, thus, in many cases the per VRF per label allocation is adopted.

As a result, repeatedly reporting the same label for several routes wastes network resources.

o The VPN label changes are typically less dynamic compared with the time-varying VPNv4 routes. Thus, acquiring the label information through the real-time monitoring of VPNv4 routes is not quite necessary.

All in all, it's more efficient to collect the VPN label independently than extracting it from the collected VPNv4 routes. In this document, we propose a method to utilize BMP to monitor the VPN label. In Section 2, the VPN label is defined to be encapsulated in the BMP Peer Up Notification message, and in Section 3, a specific implementation example is provided to show case the usage of the collected VPN label.

2. Extension of BMP Peer Up Message

The Peer Up message of BMP, defined in [RFC7854], is used to indicate the come-up of a peering session. The VPN route label can be carried in the Peer Up message and reported to the BMP monitoring station in the TLV format. The Information TLV defined in [RFC7854] can be used to encode the label, and new Information Types are defined. Each Information TLV contains at most one label, and one or more Information TLVs can be included in the Peer Up Notification when necessary.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-------------------------------+-------------------------------+
|       Information  Type       |      Information  Length      |
+-------------------------------+-------------------------------+
+                 Information      (variable)                   +
~                                                               ~
+---------------------------------------------------------------+
```

o Information Type (2 bytes): indicates the type of the Inoformation TLV. Depending on the label allocation method, the following new types are defined:

* Type = TBD1: VPN Label, allocated per VRF per label.

* Type = TBD2: VPN Label, allocated per interface per label.

* Type = TBD3: VPN Label, allocated per route per label.

* Type = TBD4: VPN Label, allocated per next hop per label.

- o Information Length (2 bytes): indicates the length of the following Inforamtion field, in bytes.

- o Information (variable): specifies the Label information according to the Information Type.

  * If the Information Type is VPN Label, allocated per VRF per label, the Information field should be the VPN label (20 bits), padded with zeros to 24 bits (3 bytes). The corresponding Length field should be set to 3.

  * If the Information Type is VPN Label, allocated per interface per label, the Information field should include the VPN label (20 bits label and 4 bits zero padding) and the corresponding interface address, with the total length specified in the Information Length field. One label and one interface address is allowed for each Information TLV.

  * If the Information Type is VPN Label, allocated per route per label, the Information includes the VPN label (20 bits label and 4 bits zero padding) and the corresponding route prefix, with the total length specified in the Information Length field. The prefix should be in VPNv4 address format. One label and one prefix is allowed for each Information TLV.

  * If the Information Type is VPN Label, allocated per next hop per label, the Information should include the VPN label (20 bits label and 4 bits zero padding) and the corresponding next hop address, with the total length specified in the Information Length field. One label and one next hop address is allowed for each Information TLV.

Considering the per VRF per label allocation, instead of extracting this same label information from all the monitored VPNv4 routes, it an be reported only once to save both device and network resources. Similarly, for the per interface per label and per next hop per label, label reporting frequencies can be reduced compared with the VPNv4 routes minotoring. Even for the per route per label case, reporting only the label information can be immune from the update of route changes, and reduce the reported information size.

3. Operation

In this section, we use an example of traffic optimization applicaiton to more specifically explain how the BMP VPN label collection functions. An Internet content provider (ICP) may own a Backbone network as the DCI (Data Center Interconnection) and Internet access solutions. Such backbone network may implement

different VPNs as the bearer networks for different services, and the granularity depends on specific service requirement. Each VPN, piggybacking on the backbone network, may connect to the Internet through other ISPs' (Internet Service Providers') networks. Different Internet Exchange (IX) devices are deployed for the Internet traffic exchange between the ICP and different ISPs.

Suppose two ISPs are considered in this example, ISP A and ISP B, as shown in the following figure.The ICP backbone network, implements VPN 1 for a specific service. This VPN exchanges Internet traffic with ISP A and ISP B through IX device A and IX device B, respetively. Prefixes are advertised from ISP A (considered as CE A of VPN 1) and ISP B (CE B) to the IX A (PE A) and IX B (PE B), repectively. Consider the case that ISP B advertises a more specific prefix (20.1.128.0/17) than ISP A (20.1.0.0/16). Both routes would be learnt by the PE devices of VPN 1, and be installed on both PE A's and PE B's routing tables. Now suppose there's a packet with destination 20.1.128.1, then according to the Longest prefix match (LPM) rule, PE B will be used as the ICP's exit for this packet. Similarly, more traffic with such prefixes may choose to exit the ICP to other ISPs through PE B, while PE A is lightly burdened, which leads to unbalanced traffic load and even traffic congestion at PE B.

```
+---------+                  +---------------------+
|  ISP A  |   20.1.0.0/16   +----+                 |
|         | +-------------> |IX A|                 |
+---------+                 +-+--+                 |
                              |                    |
                              |       ICP VPN 1    |
                              |                    |
                              |                    |
+---------+  20.1.128.0/17  +-+--+                 |
|  ISP B  | +-------------> |IX B|                 |
|         |                 +----+                 |
+---------+                  +---------------------+
```

The above mentioned issue can be solved as follows. Through traffic monitoring, the SDN controller can reoptimize the traffic load through explicit routes installation into PE A and PE B. The Next Hop field is indicated explicitly by the controller for the routes that need to be adjusted. For example, for the destination prefix 20.1.128.1, its next hop in the explicit route sent to PE A is set to the router's address in ISP A, while the next hop in the explicit route sent to PE B is set to PE A. Simutainiously, BMP is used to collect VPN 1's route labels from PE A (Label: 1000) and PE B (Label: 2000). Assume in this example, the labels are allocated per VRF per label, then Label 1000 is the label allocated to PE A for VRF 1, and Label 2000 is the label allocated to PE B for VRF 1. The explicit

routes distributed to PE A and PE B are specified in the following figures, respectively. After receiving the explicit route, PE A/B may use the label information to correlate the route to the correct VRF and then install it into VRF 1. Thus, part of the traffic may exit VPN 1 through PE A to balance the traffic load.

Dest. Addr./Mask	NH	Label	Local Pref.
20.1.128.0/17	ISP A	1000	200

Dest. Addr./Mask	NH	Label	Local Pref.
20.1.128.0/17	PE A	1000	200

4. Acknowledgements

TBD.

5. IANA Considerations

TBD.

6. Security Considerations

TBD.

7. 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>.

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

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

Authors' Addresses

Yunan Gu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: guyunan@huawei.com

Jie Chen
Tencent

Email: jasonjchen@tencent.com

Penghui Mi
Huawei
Shenzhen, Guangdong
China

Email: mipenghui@huawei.com

Shunwan Zhuang
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: zhuangshunwan@huawei.com

Zhenbin Li
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com