BGP Flow-Spec Redirect to Tunnel Action
draft-hao-idr-flowspec-redirect-tunnel-01.txt

Abstract

This draft defines a new flow-spec action, Redirect-to-Tunnel, and a new sub-TLV for Redirect-to-Tunnel extended community. A BGP UPDATE for a flow-spec NLRI can contain the extended community. When activated, the corresponding flow packets will be encapsulated and carried via a tunnel. The redirect tunnel information is encoded in BGP Path Attribute or extended community [TUNNELENCAPS][MPP] that is carried in the BGP flow-spec UPDATE. The draft expands the tunnel encapsulation attribute [TUNNELENCAPS] to apply to flow-spec SAFI, i.e., 133 and 134.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Copyright Notice

Hao, et al Expires September 18, 2016 [Page 1]
1. Introduction

BGP Flow-spec is an extension to BGP that allows for the dissemination of traffic flow specification rules. It leverages the BGP Control Plane to simplify the distribution of ACLs, new filter rules can be injected to all BGP peers simultaneously without changing router configuration. The typical application of BGP Flow-spec is to automate the distribution of traffic filter lists to routers for DDOS mitigation.

Every flow-spec route consists of a matching part (encoded in the NLRI field) and an action part (encoded in one or more BGP extended communities). The flow-spec standard [RFC 5575] defines widely-used filter actions such as discard and rate limit; it also defines a redirect-to-VRF action for policy-based forwarding. [Redirect to IP] defines a new redirect-to-IP flow-spec action that provides a simpler method of policy-based forwarding. In some cases like
service chaining, traffic steering and etc, the traffic needs to be redirected to a tunnel directly. Redirect-to-VRF action or redirect-to-IP action can’t service this purpose.

This draft proposes a new redirect-to-tunnel flow-spec action that provides a straightforward policy-based forwarding. The details of the redirect tunnel information are encoded in BGP Path Attributes or extended communities.

2. Redirect-to-Tunnel Extended Community

To support Redirect-to-Tunnel action, besides the extended communities in below per RFC5575, a Redirect-to-Tunnel extended community is defined by this draft. This extended community conveys redirecting tunnel action; the tunnel information is specified in BGP Tunnel Encapsulation Attribute [TUNNELENCAPS] and/or BGP Extended Unicast Tunnel Attribute [MPP].

<table>
<thead>
<tr>
<th>type</th>
<th>extended community</th>
<th>RFC or Draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x8006</td>
<td>traffic-rate</td>
<td>RFC5575</td>
</tr>
<tr>
<td>0x8007</td>
<td>traffic-action</td>
<td>RFC5575</td>
</tr>
<tr>
<td>0x8008</td>
<td>redirect</td>
<td>RFC5575</td>
</tr>
<tr>
<td>0x8009</td>
<td>traffic-marking</td>
<td>RFC5575</td>
</tr>
<tr>
<td>TBD</td>
<td>redirect-to-tunnel</td>
<td>This draft</td>
</tr>
</tbody>
</table>

The Redirect-to-Tunnel extended community has a type indicating it is transitive and Redirect-to-Tunnel [to be assigned by IANA]. The sub-TLV has following format.

```
+---+---+---+---+---+---+---+---+
|        reserved           | C |
+---+---+---+---+---+---+---+---+
```

In this value field (6 bytes) the least-significant bit is defined as the ‘C’ (or copy) bit. When the ‘C’ bit is set the redirection applies to copies of the matching packets and not to the original traffic stream. All bits other than the ‘C’ bit MUST be set to 0 by the originating BGP speaker and ignored by the receiving BGP speakers.
This draft extends BGP Tunnel Encapsulation Attribute to apply to BGP flow-spec SAFI, i.e., SAFI=133,134. When a tunnel is specified by BGP Tunnel Encapsulation Attribute [TUNNELENCAPs], the tunnel type and encapsulation information such as VXLAN, NVGRE, VXLAN-GPE are encoded in the Tunnel Encapsulation Attribute Sub-TLVs. When applying it to flow-spec safi, the target IP address, IPv4 or IPv6 MUST be encoded in the Remote Endpoint Sub-TLV with the corresponding AFI. The AS number in the sub-TLV MUST be the number of the AS to which the target IP address in the sub-TLV belongs. If the redirect to tunnel end point is the BGP next hop, the AFI in the sub-TLV should be filled with zero, and the address in the sub-TLV should be omitted, and AS field should be filled with zero.

When a tunnel is specified by BGP Extended Unicast Tunnel Attribute [MPP], the tunnel type such as RSVP-TE, LDP, Segment Routing Path and encapsulation information are encoded in BGP Extended Unicast Tunnel Attributes (See section 5.1 of [MPP]). Note that BGP Extended Unicast Tunnel Attribute is used in Centralized Controller Environment [MPP].

The flow-spec UPDATE carries the Redirect-to-Tunnel extended community MUST have at least one BGP Path Attribute that specifies a set of tunnel(s) that the flow packets can be redirected to.

When a BGP speaker receives a flow-spec route with a Redirect-to-Tunnel extended community and BGP tunnel encapsulation attribute [TUNNELENCAPS], if this route represents the one and only best path, it installs a traffic filtering rule that matches the packets described by the NLRI field; the packets matching the rules will be redirected (C=0) or copied (C=1) via the IP tunnel with remote endpoint address encoded in Remote Endpoint sub-TLV of Tunnel Encapsulation Attribute. If the ‘target address’ is invalid or unreachable then the extended community and the tunnel attribute SHOULD be ignored.

When a BGP speaker receives a flow-spec route with a Redirect-to-Tunnel extended community and extended unicast tunnel attribute, it installs traffic filtering rules that matches the packets described by the NLRI field and the tunnel info. If BGP speaker can’t resolve the tunnel locally according to the unicast tunnel attribute, then the extended community and the tunnel attribute SHOULD be ignored.

If a BGP speaker receives a flow-spec route with one Redirect-to-Tunnel extended community and one BGP Tunnel Encapsulation Attribute that represents a set of tunnels to the same target address, and all of them are considered best and usable paths according to the BGP speaker’s multipath configuration, the BGP speaker SHOULD load-share
the redirected packets across all the tunnels. If the BGP speaker is not capable of redirecting and copying the same packet it SHOULD ignore the extended community with C=0. If the BGP speaker is not capable of redirecting/copying a packet towards multiple tunnels it SHOULD deterministically select one tunnel to the ‘target address’ and ignore the others.

If a BGP speaker receives multiple flow-spec routes for the same flow-spec NLRI and all of them are considered best and usable paths according to the BGP speaker’s multipath configuration and each one carries one Redirect-to-Tunnel extended community and one Tunnel Encapsulation Attribute, the BGP speaker SHOULD load-share the tunneled redirected/copied packets across all the tunnels, with the same fallback rules as discussed in the previous paragraph. Note that this situation does not require the BGP speaker to have multiple peers - i.e. Add-Paths could be used for the flow-spec address family.

If a BGP speaker receives a flow-spec route with one Redirect-to-Tunnel and one ‘redirect to VRF’ extended community, and this route represents the one and only best path, the Redirect-to-Tunnel actions described above should be applied in the context of the ‘target VRF’ matching the ‘redirect to VRF’ extended community, i.e. the ‘target addresses’ should be looked up in the FIB of the ‘target VRF’. If the BGP speaker is not capable of ‘redirect to VRF’ followed by Redirect-to-Tunnel then it SHOULD give preference to performing the ‘redirect to VRF’ action and doing only longest-prefix-match forwarding in the ‘target VRF’.

If a BGP speaker receives multiple flow-spec routes for the same flow-spec NLRI and all of them are considered best and usable paths according to the BGP speaker’s multipath configuration and they carry a combination of Redirect-to-Tunnel and ‘redirect to VRF’ extended communities, the BGP speaker SHOULD apply the Redirect-to-Tunnel actions in the context of the ‘target VRF’ as described above. Note that this situation does not require the BGP speaker to have multiple peers - i.e. Add-Paths could be used for the flow-spec address family.

The redirected/copied flow packets will be encapsulated first. The outer src address on the encapsulated packets MUST be filled with the IP address of the forwarding router; the outer dst address on the packets MUST be filled with the target IP address. If the flow has multiple tunnels that have the ‘target address’ as remote tunnel endpoint, the redirected/copied packets MAY be encapsulated according to tunnel type and be load-shared across these tunnels according to the router’s ECMP configuration.
If the ‘target route’ has one or more tunnel next-hops then, in turn, the tunneled redirect/copy packets SHOULD be encapsulated appropriately again.

3. Usage Rules for Redirect-to-Tunnel Action

3.1. Matching Filters for Redirect Tunnel Action

Redirect-to-Tunnel action can apply to different types of flow spec rules described in the NLRI field. Here are the types of flow spec rules that can have the Redirect-to-Tunnel action. Applicability for Other types of flow spec rules are for further study.

- IPv4 or IPv6
- L3VPN
- L2VPN
- NVO3

3.2. Other Actions Considerations

Flow spec rules in a NLRI can associate with one or more actions that are specified in extended communities, which means that flow spec packets will subject to a sequence of actions. [COMBO] specified default ordering precedence of actions. However some actions do not make a sense to be used with Redirect-to-Tunnel action, i.e. they have to be used in mutually exclusive.

In general Redirect-to-Tunnel action can work with traffic rate in Bite, traffic rate in packet, traffic action, redirect to VRF, interface set, time actions. The use cases for Redirect-to-Tunnel action to work with other actions are for further study. Note that: the two actions that can be used with Redirect-to-Tunnel action may be in mutually exclusive usage.

Memo: need a standard way to document these rules for a flow spec action.

3.3. Validation Procedures

The validation check described in [RFC 5575] and revised in [VALIDATE] SHOULD be applied by default to received flow-spec routes with a Redirect-to-Tunnel extended community, as it is to all types of flow-spec routes and the validation check described in
[TUNNELENCAPS] SHOULD be applied to the tunnel encapsulation attribute. This means that a flow-spec route with a destination prefix subcomponent SHOULD NOT be accepted from an EBGP peer unless that peer also advertised the best path for the matching unicast route.

BGP speakers that support the extended community defined in this draft MUST also, by default, enforce the following check when receiving a flow-spec route from an EBGP peer: if the received flow-spec route has a Redirect-to-Tunnel extended community with a ‘target address’ X (in the remote endpoint sub-TLV) and the best matching route to X is not a BGP route with origin AS matching the peer AS then the extended community should be discarded and not propagated along with the flow-spec route to other peers. It MUST be possible to disable this additional validation check on a per-EBGP session basis.

4. Security Considerations

A system that originates a flow-spec route with a ‘redirect to tunnel’ extended community can cause many receivers of the flow-spec route to send traffic to a single next-hop, overwhelming that next-hop and resulting in inadvertent or deliberate denial-of-service. This is particularly a concern when the ‘redirect to tunnel’ extended community is allowed to cross AS boundaries. The validation check described in section 2.1 significantly reduces this risk.

5. IANA Considerations

IANA is requested to update the reference for the following assignment in the "BGP Extended Communities Type/sub-Type for Redirect-to-Tunnel that is specified in this draft.

5.1. Normative References


5.2. Informative References


6. Acknowledgments

The authors wish to acknowledge the important contributions of Shunwan Zhuang, Qiandeng Liang.
Authors’ Addresses

Weiguo Hao  
Huawei Technologies  
101 Software Avenue,  
Nanjing 210012  
China  
Email: haoweiguo@huawei.com

Zhenbin Li  
Huawei Technologies  
Huawei Bld., No.156 Beiqing Rd.  
Beijing 100095  
China  
Email: lizhenbin@huawei.com

Lucy Yong  
Huawei Technologies  
Phone: +1-918-808-1918  
Email: lucy.yong@huawei.com