Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 20, 2014

Z. Li
L. Zhang
Huawei Technologies
January 16, 2014

NEXTHOP_PATH ATTIBUTE for BGP
draft-zhang-idr-nexthop-path-attr-01

Abstract

As the BGP is deployed in a single Autonomous System for network convergence such as Seamless MPLS, it is desirable for BGP to carry more information to help select routing more intelligently. It can reduce the cost proposed by complex policy control design on BGP routes and adapt to network change easily. This document proposed a new path attribute for BGP routes that can record the next hop path for the route to help BGP route selection and network management.

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 July 20, 2014.

Copyright Notice

Table of Contents

1. Introduction

[I-D.ietf-mpls-seamless-mpls] describes an architecture which can be used to extend MPLS networks to integrate access and aggregation networks into a single MPLS domain ("Seamless MPLS"). As the mobile backhaul service is deployed widely, the requirement of the integration of mobile backhaul networks and core networks has been proposed. For the reason of scalability , the Seamless MPLS network tends to be divided into multiple IGP areas for access, aggregation, and core networks and IBGPs runs among Area Border Routers (ABRs) which should act as inline RRs to reflect the labeled BGP routes or BGP VPN routes to remote BGP peers with next hop self (NHS).

As the BGP is used in a single Autonomous System for network convergence, it is desirable for BGP to carry more information to help select routing more intelligently. It can reduce the cost proposed by complex policy control design on BGP routes and adapt to network change easily.

This document proposes a new path attribute that can record the next hop path of the route to help BGP route election and network management.

2. Motivation

2.1. Complexity of Route Selection

In the Seamless MPLS network, Area Border Routers (ABRs) which run IBGP should act as inline RRs to reflect the labeled BGP routes or BGP VPN routes to remote BGP peers with next hop self (NHS). Each ABR should process route selection which needs complex route policy to control the BGP route distribution in the Seamless MPLS network , as shown below:

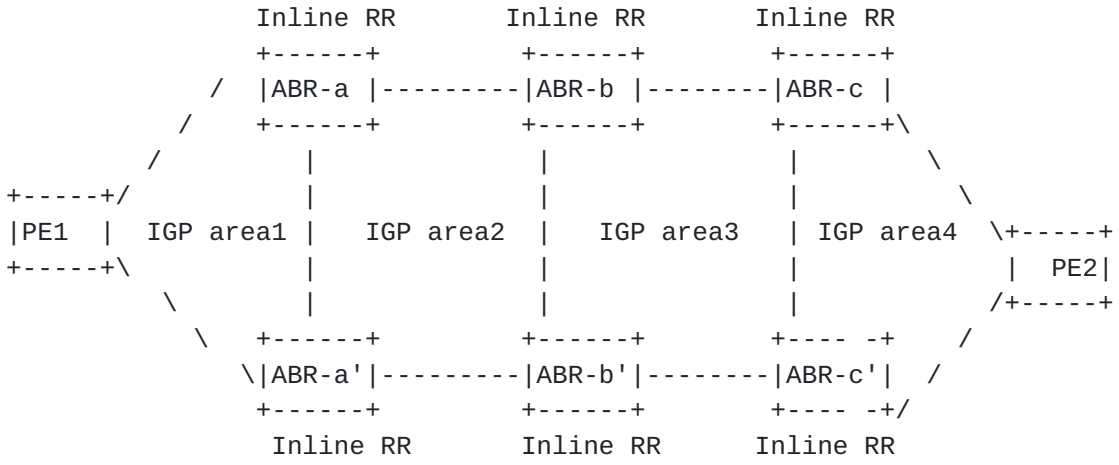

Figure 1 Seamless MPLS Network with Multiple IGP Areas

Just like Figure 1 shown, PE1 and PE2 are BGP VPN service end-point. IBGP peers runs contiguously between ABRs in different IGP areas, and each ABR works as inline RR. When labeled BGP routes or BGP VPN routes originated from PE1 is distributed to the other service end-point PE2, the route can be reflected by the ABRs one by one with next hop self (NHS).

The inline RR will distribute the route to all of the IBGP peers except the IBGP peer from which the route was received. As a result, an ABR may receive routes of the same prefix from different IBGP peers with different next hop. Traditionally the BGP RR should select the best route to reflect to other IBGP peers. But in this network the route selection process will be more complex which needs to introduce complex route policy.

Here is an example for complex route policy. ABR-b may receive routes of the same prefix originated from PE1 from different three IBGP peers, ABR-a, ABR-a', ABR-c, and ABR-c'. The route policy should guarantee that the route from ABR-a or ABR-a' is selected as the best one. At the same time, routes of the same prefix originated from PE2 may be received from ABR-a, ABR-a', ABR-c, and ABR-c'. The route policy should guarantee that the route from ABR-c or ABR-c' is selected as the best one. To satisfy the different best route selection requirements, each IBGP speaker has to configure complex route policy.

2.2. New Role of BGP in Seamless MPLS Network

When Seamless MPLS makes integration of mobile backhaul networks and core networks, BGP in Seamless MPLS network act more like an "Interior Gateway Protocol (IGP)". As the whole Seamless MPLS network is in a single AS for uniform administration, the security requirement proposed for traditional BGP can be reduced. At the same time some path attributes for BGP route such as AS_PATH is no use in this scenario. As the BGP is deployed from implementing network convergence, it is desirable for BGP to carry more information to help select route more intelligently. It can simplify policy control design on BGP routes and adapt to network change easily. Moreover the additional path information may facilitate the network operation and maintenance. [I-D.ietf-idr-aigp] is the example which can help BGP route selection by advertising IGP metric information with BGP route. In this document, we propose a new method to record next hop list for the BGP route, which can be used for automatic BGP route selection and facilitating network operation and maintenance. The new attribute, NEXTHOP_PATH ATTRIBUTE, is defined for the BGP route to record the next hop path. It can work as AS_PATH ATTRIBUTE.

3. Definition of NEXTHOP_PATH ATTRIBUTE

The NEXTHOP_PATH ATTRIBUTE is an optional transitive BGP Path Attribute. The NEXTHOP_PATH ATTRIBUTE type is defined as below (refer to [RFC4271]):

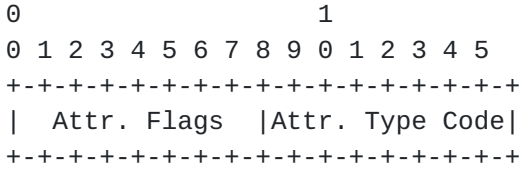

Figure 2 NEXTHOP_PATH ATTRIBUTE Type definition

Attr. Flags

SHOULD be optional transitive

Attr. Type Code

SHOULD be allocated by IANA

NEXTHOP_PATH is composed of a sequence of next hop path segments. Each next hop path segment is represented by a triple <path segment type, path segment length, path segment value>. The format of the next hop path segmen is shown in the figure 3.

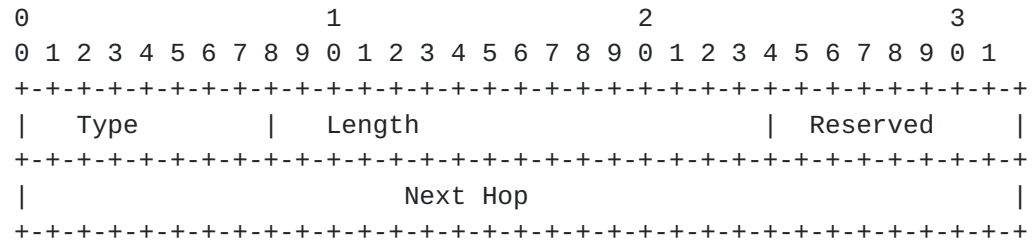

Figure 3 NH_SEQUENCE_V4 TLV

- Type: A single octet encoding the TLV Type. The Type of "NH_SEQUENCE_V4" is defined in this document, which needs to be allocated by IANA. The procedure for next hop path segement usage for IPv6 or other extensions will be described in the future version.

- Length: Two octets encoding the length in octets of the TLV, including the type and length fields. The length is encoded as an unsigned binary integer.

- Reserved: A single octet that must be zero now.

- NextHop: four octets encoding for the route next hop address.

4. Process of NEXTHOP_PATH ATTRIBUTE

The NEXTHOP_PATH ATTRIBUTE defined here is an optional transitive BGP Path Attribute, the process of this attribute MUSTaccord with the procedures in [RFC4271].

4.1. Creating and Modifying the NEXTHOP_PATH Attribute

When a BGP speaker distributes a route to its BGP peer within UPDATE message, the NEXTHOP_PATH ATTRIBUTE should be processed based on different route states:

1. If the route is originated in this BGP peaker

* If the NEXTHOP_PATH ATTRIBUTE is supported, the NEXTHOP_PATH ATTRIBUTE SHOULD be originated including the BGP speaker's own next hop address in a next hop path segment. In this case, the next hop address of the originating BGP speaker will be the only entry of the next hop path segment, and this path segment will be the only segment in NEXTHOP_PATH ATTRIBUTE.

* If the NEXTHOP_PATH ATTRIBUTE is not supported, the route will be distributed without NEXTHOP_PATH ATTRIBUTE.

2. if the route is received from one BGP speaker's UPDATE message

* If the NEXTHOP_PATH ATTRIBUTE is NULL and the local BGP speaker support NEXTHOP_PATH ATTRIBUTE, when the route is propagated to another IBGP speaker with next hop self (NHS), the NEXTHOP_PATH ATTRIBUTE SHOULD be originated including the BGP speaker's own next hop address in a next hop path segment. In this case, the next hop address of this BGP speaker will be the only entry to the next hop path segment, and this path segment will be the only segment in NEXTHOP_PATH ATTRIBUTE

* If the NEXTHOP_PATH ATTRIBUTE is non-NULL and the local BGP speaker support NEXTHOP_PATH ATTRIBUTE, when the route is propagated to another IBGP speaker with next hop self (NHS), the BGP speaker MUST appends its own next hop address as the last one of the next hop path segments.

* If the NEXTHOP_PATH ATTRIBUTE is NULL and the local BGP speaker support NEXTHOP_PATH ATTRIBUTE, when the route is propagated to another BGP speaker without changing the next hop by the BGP speaker, the BGP speaker MUST NOT originate the NEXTHOP_PATH ATTRIBUTE.

* If the NEXTHOP_PATH ATTRIBUTE is non-NULL and the local BGP speaker support NEXTHOP_PATH ATTRIBUTE, when the route is propagated to another BGP speaker without changing the next hop by the BGP speaker, the BGP speaker MUST NOT change the next hop path sequence.

* If the BGP speaker does not support NEXTHOP_PATH ATTRIBUTE, it SHOULD keep the NEXTHOP_PATH ATTRIBUTE unchanged whether the route is distribute with next hop self or not.

4.2. Decision Process

Support for the NEXTHOP_PATH ATTRIBUTE involves several modifications to the tie breaking procedures of the "phase 2" decision of BGP route selection, described in section 9.1.2.2 of [RFC4271].

If the NEXTHOP_PATH ATTRIBUTE of a BGP route contains a next hop path loop, the BGP route MUST be excluded from the Phase 2 decision function. The next hop path loop detection is done by scanning the full next hop path (as specified in the NEXTHOP_PATH ATTRIBUTE), and checking if the local BGP speaker appears in the next hop path.

The NEXTHOP_PATH ATTRIBUTE can be used for BGP route selection. The priority of the NEXTHOP_PATH ATTRIBUTE for route selection is the same as the AS_PATH attribute.

When a route is received from different IBGP speakers, if the best route cannot acquired through the higher priority rules, the NEXTHOP_PATH ATTRIBUTE SHOULD be used for route selection, and the route with least nexthops will be selected. If the lengths of the next hop lists are the same, the rest rules SHOULD be used for route selection.

5. IANA Considerations

IANA need to assign the codepoint in the "BGP Path Attributes" registry to the NEXTHOP_PATH ATTRIBUTE.

IANA shall create a registry for "next hop path segment". The type field consists of a single octet, with possible values from 0 to 255. The allocation policy for this field is to be "Standards Action with Early Allocation". A new Type should be defined as "NH_SEQUENCE_V4".

6. Security Considerations

Note that, the NEXTHOP_PATH ATTRIBUTE is defined as a optional transitive BGP Path attribute. Both the IBGP and EBGP speaker can use this attribute. When an ASBR propagates the route receive from a IBGP peer to an EBGP peer, the NEXTHOP_PATH ATTRIBUTE will be distribute to the EBGP Speaker which may be controlled by other Service Provider. If the EBGP speaker can support the NEXTHOP_PATH ATTRIBUTE, it can parse the NEXTHOP_PATH ATTRIBUTE to get the inner network architecture of the other network.

In order to prevent this possible security problem, the NEXTHOP_PATH ATTRIBUTE capability should be disabled for specific BGP speaker, such as EBGP. This can reduce the security risk.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

7.2. Informative References

[I-D.ietf-idr-aigp] Mohapatra, P., Fernando, R., Rosen, E., and J. Uttaro, "The Accumulated IGP Metric Attribute for BGP", draft-ietf-idr-aigp-14 (work in progress), December 2013.

[I-D.ietf-mpls-seamless-mpls] Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz, M., and D. Steinberg, "Seamless MPLS Architecture", draft-ietf-mpls-seamless-mpls-04 (work in progress), July 2013.

Authors' Addresses

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

Email: lizhenbin@huawei.com

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

Email: monica.zhangli@huawei.com