BGP Tunnel Address Prefix Attribute and Tunnel Address Prefix Ext. Community

draft-xu-idr-tunnel-address-prefix-01

Xiaohu Xu (Huawei)

Kai Lee (China Telecom)

IETF84, Vancouver

Problem Statement

- **There are some MPLS-based L2VPN or L3VPN application scenarios where the underlying networks are IP enabled, rather than MPLS enabled (e.g., multi-tenant cloud data center networks).**
  - Moreover, load-balancing is much desirable in these scenarios (e.g., to maximize the bisection bandwidth between servers within or across data centers).
  - However, since distinct customer traffic flows between a given PE pair would be encapsulated with the same IP/GRE tunnel as per normal operations, P routers (i.e., core routers) could not achieve an ideal load-balancing for these tunneled traffic flows due to the lack of adequate entropy information.

Problem Statement (cont.)

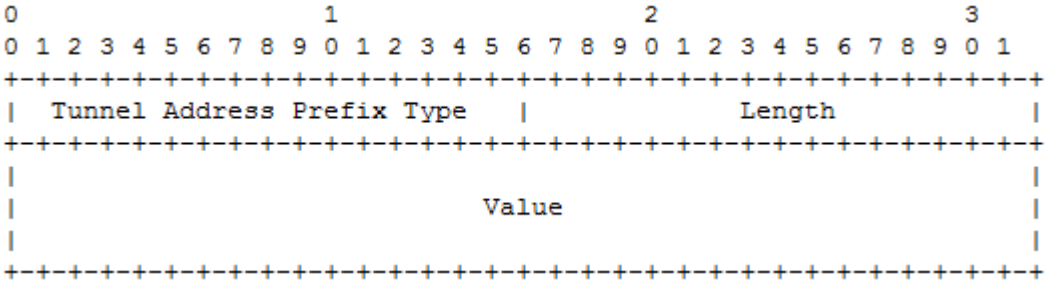
- **The existing method [RFC5640] requires a change to the data plane of core routers.**
  - Core routers is required to be capable of performing hash calculation on the specific "load-balancing" field contained in the L2TPv3 or GRE tunnel header.
- **Such requirement can not be met in some cases.**
  - For example, some deployed core routers could only perform hash calculation on the five tuple of TCP/UDP packets or some fields in the IP header of non-TCP/UDP packets.

Solution Overview

- **For a given egress PE router, it could tell ingress PE routers more than one tunnel destination address (in the form of a prefix) to be used when tunneling traffic flows to it.**
  - Meanwhile, it would create the corresponding loopback interface for each IP address and advertise a route for that prefix via IGP.
- **As such, distinct customer traffic flows between a given PE pair would be encapsulated with different destination IP addresses if possible.**
  - P routers could accordingly achieve a good load-balancing for those IP/GRE tunneled traffic flows between a given PE pair.

Tunnel Address Prefix Attribute

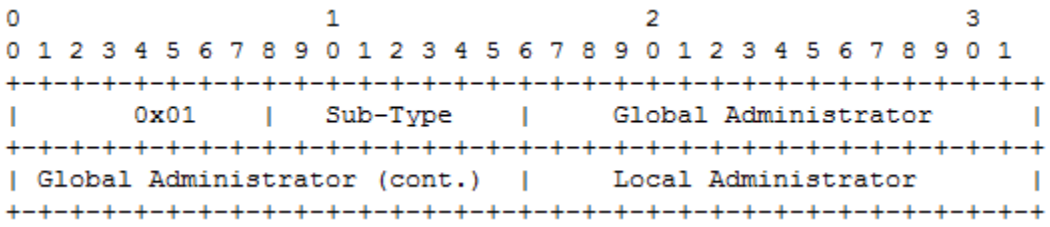
- **Tunnel Address Prefix Attribute (contained in the Encapsulation SAFI [RFC5512]) is used by a given egress PE router to indicate available tunnel destination addresses to be used by ingress PE routers when tunneling traffic flows towards it.**



  - If the AFI of the Encapsulation SAFI is IPv4, the length value is set to 64; otherwise if the AFI is IPv6, the length value is set to 256.
  - The Value (variable) field contains the tunnel address prefix information in the form of IP address and subnet mask pair.

Tunnel Address Prefix Ext. Community

- **This extended community is an alternative which is useful in the cases where the Encapsulation SAFI capability is not supported or one really wants to specify different tunnel destination address prefixes for distinct sets of traffic flows**.



  - The Global Administrator sub-field contains an IPv4 unicast address.
  - The Local Administrator sub-field contains the corresponding Prefix Length.

Applicability

- **This approach is applicable to many technologies such as**
  - L3VPN [RFC4364]
  - 6PE [RFC4798]
  - Softwire mesh [RFC5565]
  - BGP free core
  - L2VPN including VPLS [RFC4761, RFC4762] and E-VPN [E-VPN].

Next-steps

- Solicit more comments and suggestions.
- WG adoption?