Network Working Group
Internet Draft
Expiration Date: August 2004

Rahul Aggarwal
Juniper Networks

Cristallo Geoffrey
Jeremy De Clercq
Alcatel

Signaling Tunnel Encapsulation/Deencapsulation Capabilities

draft-raggarwa-ppvpn-tunnel-encap-sig-03.txt

1. Status of this Memo

1. Abstract

This document proposes a mechanism for signaling a PE router's tunnel encapsulation capabilities. One example is its capability to encapsulate MPLS using dynamic GRE and/or IP. This is applicable when a MPLS packet is tunneled using dynamic GRE and/or IP encapsulation [MPLS-IP-GRE] between PE routers. For instance the MPLS packet may be a 2547 based MPLS VPN packet [2547bis], a layer 2 packet transported using MPLS [MARTINI], a MPLS tunneled IPv6 packet or a MPLS IPv6 VPN packet [BGP-VPN-IPv6]. Adding such a mechanism has several benefits. It helps in blackhole avoidance and eases transitioning from MPLS tunneling based Layer 3/Layer 2 VPNs to GRE/IP tunneling based Layer 3/Layer 2 VPNs (and vice versa). Such a mechanism is needed where a network may be using MPLS and GRE (or IP) for

tunneling, simultaneously in different parts of the network. It can help in encapsulation selection when multiple tunneling technologies are supported.

2. Specification of Requirements

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].

3. Summary for Internet Area

3.1. Related documents

See the Reference Section

3.2. Where does it fit in the Picture of the Internet Area Work

This work fits in the L3VPN WG.

3.3. Why is it Targeted at this WG

[2547bis] is a product of the L3VPN WG. This document specifies a mechanism that proposes a lightweight mechanism for signaling a PE router's capability to encapsulate MPLS using dynamic GRE and/or IP. This is applicable when a 2547 based MPLS VPN packet is tunneled using a dynamic GRE and/or IP encapsulation [MPLS-IP-GRE] between PE routers. Since the procedures described in this document are directly related to [2547bis], it would be logical to target this document at the L3PVN WG.

4. Mechanism

Mutliple applications such as 2547 VPNs, Layer 2 VPNs, VPLS, P2P Layer 2 transport over MPLS, IPv6 over IPv4 MPLS or IPv6 VPN over MPLS may use dynamic GRE or IP encapsulation for tunneling traffic across a network backbone. This document uses the term 'soft GRE' to refer to dynamic GRE encapsulation. If a PE router is using soft GRE or IP encapsulation for tunneling traffic for one or more of these applications, across the backbone, it is not possible currently for it to dynamically learn the encapsulation capability of the remote PE router. In the context of 2547 based VPNs this PE router does not know the MPLS in soft GRE or MPLS in IP encapsulation capability of the BGP next-hop to which the traffic is destined. This document proposes a simple signaling mechanism by way of which this PE router can learn the MPLS in soft GRE or MPLS in IP encaspulation capability of the remote PE routers. This is achieved by propagating this information in BGP or in LDP.

4.1 BGP Extension

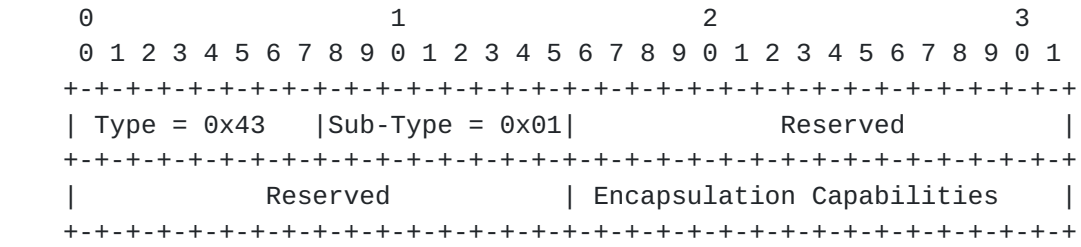
We define a BGP opaque extended community that can be attached to a BGP NLRI advertisement to indicate the MPLS or other encapsulation capabilities of such a NLRI. We define a new subsequent address family identifier (SAFI) to be assigned by IANA, for carrying the tunnel encapsulation capabilities. Typically a PE can advertise a loopback address as an NLRI using a IPv4/IPv6 AFI and the new tunnel encapsulation capability SAFI. The encapsulation capabilities associated with this loopback address can be specified by attaching the new extended community.

The new BGP extended community is referred to as the Tunnel Encapsulation Capabilities extended community. It is non-transitive across the Autonomous System boundary. It should not be propagated by EBGP when the next hop associated with the NLRI is changed. However in certain cases it may be desirable to propagate this extended community in EBGP if the next hop is unchanged.

As a note since the extended community attribute itself is optional and transitive, a BGP speaker that does not understand an extended community attribute will set the partial bit in the attribute. Hence a BGP peer that understands the Tunnel Encapsulation Capabilities extended community may not use this extended community if it is received as part of an extended community attribute that has the partial bit set. This is because the next hop may have been changed by a router that did not understand the extended community attribute.

The Tunnel Encapsulation Capabilities community is of an extended type. The value of the high-order octet of the Type Field is 0x43. The value of the low-order octet of the Type field of this extended community is 0x01, subject to IANA approval [BGP-EXT-COM].

The Tunnel encapsulation extended community has the following format:



The encapsulation capabilities bit-mask indicates all the encapsulations supported by the BGP speaker for the advertised NLRI and is encoded in the two least significant octets. The following encapsulation capabilities are defined as of now:

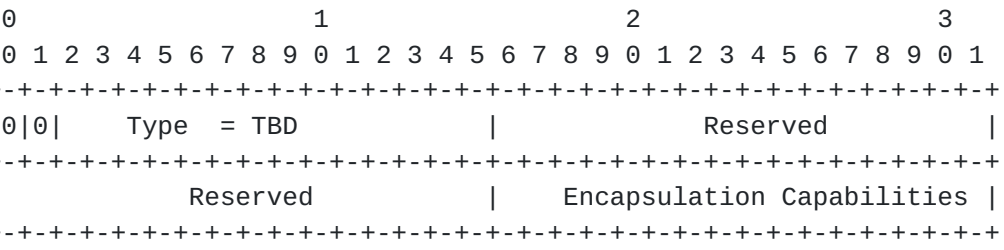
```
0x0001 - IPv4 in IPv4
0x0002 - IPv4 in GRE
0x0004 - IPv4 in IPSec
0x0008 - GRE in IPSec
0x0010 - MPLS in soft GRE
0x0020 - MPLS in IPv4
0x0040 - MPLS in IPv6
0x0080 - IPv6 in IPv4
0x0100 - IPv6 in MPLS
```

For IPv4 in GRE and for MPLS in soft GRE, the reserved 32 bits can be used to signal the GRE key.

4.2 LDP Extension

MPLS in soft GRE or MPLS in IP encapsulation capability may need to be advertised when LDP signaling is used for establishing pseudo wires [MARTINI] or for Layer 2 VPNs [LDP-SIG]. When BGP is used as a discovery mechanism for Layer 2 VPNs BGP extensions proposed in 4.1 should be sufficient for determining the right encapsulation to use. If this is not the case, the encapsulation capability is advertised in LDP. This is done at the time of LDP session establishment. We define a LDP Tunnel Encapsulation Capabilities Session TLV for this purpose.

This TLV is advertised as an optional parameter in the LDP Initialization message. The type of this optional paramenter has to be assigned by IANA and has the following format:



Encapsulation bit-mask indicates all the encapsulations supported by the PE originating the Initialization message. They are encoded in the two least significant octets:

```
0x01 - MPLS in soft GRE
0x02 - MPLS in IPv4
0x04 - MPLS in IPv6
```

For MPLS in soft GRE, the reserved 32 bits can be used to signal the GRE key.

4.3 Applicability of BGP and LDP Extensions

The decision to use BGP or LDP for advertising the tunnel encapsulation capability depends on the application. For 2547 based VPNs this information is advertised in BGP advertisements using the tunnel encapsulation extended community. This is also true for BGP based Layer 2 VPNs [BGP-L2VPN]. BGP may be used as an auto-discovery mechanism for Layer 2 VPNs established using LDP signaling [BGP-AUTO]. In this case tunnel encapsulation extended community can be added to the BGP auto-discovery advertisements to convey the encapsulation capability. For IPv6 tunneling and IPv6 VPN applications BGP tunnel extended community can be used [BGP-VPN-IPv6].

There may be cases when LDP is used for establishing pseudo wires and Layer 2 VPNs [MARTINI, LDP-SIG], and BGP is not used as an auto-discovery protocol. In this case the encapsulation capability can be advertised using the Encapsulation TLV in the LDP Initialization message.

5. Usage

We describe the usage of this signaling enhancement in the context of 2547, though its equally applicable to Layer 2 VPNs and other tunneling applications. With this mechanism a PE can 'signal' its tunnel enapsulation capabilities including MPLS in IP or MPLS in soft GRE encapsulation capability to other PEs. A PE (say PE1) now has two pieces of information while determining if VPN routes learned from a remote PE (say PE2) are eligible for MPLS in IP or MPLS in soft GRE encapsulation:

- Is PE1 configured to support MPLS in IP or MPLS in soft GRE encapsulation
- Does PE2 support MPLS in IP or MPLS in soft GRE encapsulation. This is learned via the mechanism described above.

If both the above are true, the VPN route can be installed in the VRF and tunneled using IP or soft GRE. Else the VPN route cannot be tunneled using IP or soft GRE. However it can still be tunneled using MPLS or some other tunneling mechanism. If PE2 supports multiple encapsulations, this mechanism can be used to pick one of the encapsulations based on local policy at PE1. In certain implementations BGP may propagate the capability of PE2 to the local RIB. Hence the local RIB can determine if a particular next-hop is eligible for MPLS in IP or MPLS in soft GRE enapsulation.

6. Benefits

This mechanism adds several benefits:

6.1. Blackhole Avoidance

Without this mechanism its possible in certain cases for a local PE to tunnel packets to a remote PE using an encapsulation that is not supported by the remote PE. For instance without knowing the MPLS in IP or MPLS in soft GRE capability of the remote PE, the local PE, if configured for MPLS in IP or MPLS in soft GRE, can start sending IP or GRE encapsulated MPLS traffic to the remote PE even if the remote PE doesn't support MPLS in IP or MPLS in soft GRE encapsulation. This can happen if the remote PE is running a software version that is not capable of performing the corresponding de-encapsulation or if its simply not configured to support the expected de-encapsulation. This can result in blackholing the MPLS traffic.

This mechanism avoids that as the local PE will never send IP or GRE encapsulated VPN traffic to the remote PE unless the remote PE advertises that its MPLS in IP or MPLS in soft GRE capable.

6.2. Co-existing MPLS and IP or Soft GRE tunneling

It is conceivable that in a network providing 2547 based VPN service some of the PEs are attached to a part of the backbone which runs MPLS while other PEs are attached to a part of the backbone where MPLS is not running. Thus some of the PEs may support MPLS in soft GRE or MPLS in IP while others may support only MPLS tunneling. Further still its conceivable that one may wish to use soft GRE tunneling for certain VPN routes and MPLS tunneling for other VPN routes destined to the same PE. An example would be a co-existing IPSec over GRE and MPLS tunneling service for VPN-routes.

Hence if LDP is used for MPLS tunneling, a given PE (say PE1) may be configured to run LDP and support soft GRE at the same time. The reason being that some of the remote PEs can only use MPLS tunneling. However currently there is no way for a remote PE (say PE2) that supports soft GRE to know the tunneling tecnology to use while sending MPLS VPN traffic to PE1. If it prefers using soft GRE it cannot be sure that PE1 supports soft GRE and it cannot rely on the LDP FECs received from PE1 to make this decision. The mechanism proposed in this document solves this problem as PE2 can learn the soft GRE capability of PE1.

VPN routes advertised by a PE may be advertised with different next-hops if this PE wants the remote PEs to use different tunneling technologies for different next-hops. Hence this PE may wish to receive GRE encapsulated VPN traffic for some VPN routes and MPLS encapsulated VPN traffic for other VPN routes. It is possible to advertise the soft GRE capability only for certain VPN routes, associated with a particular next-hop.

6.3. Transitioning

An operator may wish to transition some or all of the routers in a 2547 based network from using MPLS based tunneling to soft GRE or IP based tunneling and vice-versa. This approach greatly simplifies this transition. Once the remote soft GRE or IP encapsulation capability is known a PE can determine if it wishes to use MPLS or GRE or IP to encapsulate the traffic. Without this mechanism an operator transitioning certain routers from MPLS based tunneling to GRE based tunneling needs to enable soft GRE on all such routers before MPLS can be turned off on any of the routers. Similarly, without this mechanism, an operator transitioning certain routers from soft GRE based tunneling to MPLS tunneling needs to enable MPLS on all such routers before soft GRE can be turned off on any of the routers.

7. Deployment Considerations

It is recommended that an implementation provide a configuration option to trigger the announcement of a PE's encapsulation capabilities in BGP or LDP. This will help in selective deployment of this mechanism.

8. IANA Considerations

This document requires the use of a new BGP SAFI, a new BGP opaque extended community sub-type and a LDP Tunnel encapsulation TLV. These values have to be assigned by IANA.

9. Security Considerations

This document does not introduce any new security issues. The security issues identified in [BGP-EXT-COM], [RFC3036], [MPLS-IP-GRE] and [2547bis] are still relevant.

10. Acknowledgements

We would like to thank Enke Chen, Jenny Yuan, Naiming Shen, Acee Lindem, and Ravi Chandra for their valuable contributions to this document and for helping in evolving this mechanism.

Thanks to Yakov Rekhter for his comments and valuable suggestions. We would also like to thank Eric Rosen and Pedro Roque Marques for their comments.

11. References

[BGP-EXT-COM] S.R. Sangli et. al., "BGP Extended Communities Attribute", draft-ietf-idr-bgp-ext-communities-05.txt.

[RFC3036] L. Andersson et. al., "LDP Specification", Request For Comments 3036.

[MPLS-IP-GRE] T. Worster et. al., "Encapsulating MPLS in IP or GRE", draft-rosen-mpls-in-ip-or-gre-00.txt.

[2547bis] Rosen, E. et. al., "BGP/MPLS VPNs," Internet-draft draft-ietf-ppvpn-rfc2547bis-04.txt, January 2002.

[MARTINI] L. Martini. et. al., "Transport of Layer 2 Frames Over MPLS", draft-ietf-pwe3-control-protocol-00.txt.

[BGP-L2VPN] K. Kompella et. al., "Layer 2 VPNs over Tunnels", draft-kompella-ppvpn-l2vpn-02.txt.

[BGP-AUTO] Ould-Brahim et. al., "Using BGP as an Auto-Discovery Mechanism for Network based VPNs", draft-ietf-ppvpn-bgpvpn-auto-05.txt.

[LDP-SIG] E. Rosen, "LDP-based Signaling for L2VPNs", draft-rosen-ppvpn-l2-signaling-02.txt.

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

[BGP-VPN-IPv6] J. De Clercq et. al., "BGP-MPLS VPN extension for IPv6 VPN", draft-ietf-ppvpn-bgp-ipv6-vpn-03.txt.

12. Author Information

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
Email: rahul@juniper.net

Cristallo Geoffrey
Alcatel
Fr. Wellesplein 1, 2018 Antwerp, Belgium
Email: geoffrey.cristallo@alcatel.be

Jeremy De Clercq
Alcatel
Fr. Wellesplein 1, 2018 Antwerpen, Belgium.
Email: Jeremy.De_Clercq@alcatel.be