

Network working group
Internet Draft
Category: Informational

X. Xu
Z. Li
Huawei

Expires: April 2014

October 17, 2013

Multi-domain MPLS Deployment Enhancement

[draft-xu-mpls-multi-domain-deployment-enhancement-00](#)

Abstract

MPLS as a mature technology is increasingly deployed in large-scale networks which consists of multiple domains (e.g., IGP areas/levels and even Autonomous Systems). To scale such multi-domain MPLS deployment, the concept of hierarchical LSPs is usually resorted. This document describes an enhancement to such hierarchical multi-domain MPLS deployment architecture that could further improve the scalability of multi-domain MPLS deployment.

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 April 17, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Conventions used in this document

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](#)].

Table of Contents

1.	Introduction	3
2.	Terminology	3
3.	Deployment Enhancement	3
4.	Conclusions	5
5.	Security Considerations.....	5
6.	IANA Considerations	5
7.	Acknowledgements	5
8.	References	5
8.1.	Normative References	5
8.2.	Informative References	5
	Authors' Addresses	6

1. Introduction

MPLS as a mature technology is increasingly deployed in large-scale networks which consists of multiple IGP areas/levels and even multiple Autonomous Systems (AS's) (e.g., Inter-AS L3VPN option C described in [[RFC4364](#)]). For simplicity, in the rest of this document the term "domain" would be used to refer area/level/AS. To scale such multi-domain MPLS deployment, the concept of hierarchical LSPs is usually resorted. The basic idea behind this concept is the innermost transport LSP which is across domain boundaries is actually transported over multiple outer transport LSPs which are confined within each domain (a.k.a., originated and terminated within the same domain). Such a hierarchical routing and forwarding concept allows exchange of loopback addresses and MPLS label bindings for innermost transport LSPs across these domains while preventing the above information from being flooded into domains or parts of the network that do not need them. In most cases, the innermost transport LSPs are established primarily using labeled BGP [[RFC3107](#)]. In some special cases (e.g., seamless MPLS [[Seamless-MPLS](#)]), the innermost transport LSP could also be a stitched LSP of BGP-signaled LSPs and LDP-signaled LSPs.

Such a hierarchical routing and forwarding concept has greatly improved the scalability of the multi-domain MPLS deployment. However, in the case where the number of PE routers is enormous, a large amount of non-aggregatable labeled BGP routes for those PE routers would have to be advertised across domain boundaries. As stated in the seamless MPLS draft [[Seamless-MPLS](#)], "...this architecture results in carrying all loopbacks of all nodes except pure P nodes (AN, AGN, ABR and core PE) in labeled BGP, e.g., there will be in the order of 100,000 routes in labeled BGP when approaching the stated scalability goal..." Without special implementation and configuration, it would result in tremendous and unnecessary consumption of the BGP RIB and even MPLS forwarding table resources on domain boundary nodes (e.g., ABRs). Therefore, there is still room for improvement in scalability.

2. Terminology

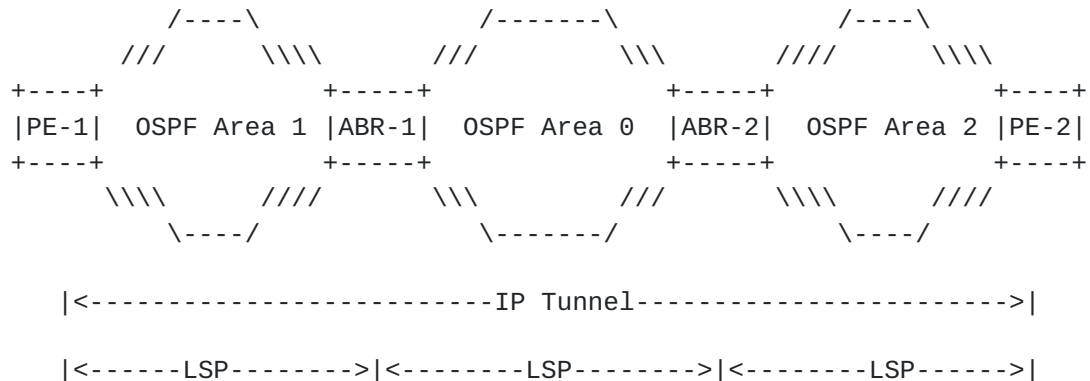
This memo makes use of the terms defined in [[RFC3031](#)] and [[RFC3107](#)].

3. Deployment Enhancement

In the hierarchical LSP case as mentioned in [Section 1](#), the innermost transport LSP only represents a logical connectivity to the final tunnel endpoint (e.g., egress PE routers). As such, it's no problem to replace such innermost transport LSP with an IP tunnel while

keeping the remaining outer MPLS LSPs unchanged. In this way, there is no need for advertising no-aggregatable labeled BGP host routes across domain boundaries anymore. Instead, it only requires advertising aggregated non-labeled BGP routes across domain boundaries.

To clearly understand the concept of the multi-domain MPLS deployment enhancement as suggested above, a multi-area MPLS deployment example with enhancement is illustrated as follows:



In the above example, iBGP sessions are established between PEs (i.e., PE-1 and PE-2) and ABRs (e.g., ABR-1 and ABR-2). Assume loopback addresses of all PEs within area 1 are within 10.1.0.0/16 while loopback addresses of all PEs within area 2 are within 10.2.0.0/16. ABR1 would advertise a route for 10.1.0.0/16 to ABR-2 which in turn advertises that route upon receiving to PE-2. Similarly, ABR-2 would advertise a route for 10.2.0.0/16 to ABR-1 which in turn advertises that route upon receiving to PE-1. In addition, intra-domain LSPs have been established between PEs and ABRs.

Assume PE-1 needs to send a packet P1 to PE-2, PE-1 would encapsulate such packet into an IP tunnel with tunnel source of PE-1's loopback address and tunnel destination of PE-2's loopback address. For example, if the packet is a MPLS IP VPN packet, the packet would be encapsulated using any IP-based encapsulation method for MPLS (e.g., MPLS-in-IP). PE-1 then performs IP forwarding lookup for the encapsulated packet P2. Since the BGP next-hop of the best route (i.e., 10.2.0.0/16) for the packet P2's destination (i.e., PE-2's loopback address) is ABR-1 and PE-1 has a LSP towards ABR-1, PE-1 therefore would transport that encapsulated packet P2 over that LSP. Upon receipt of that encapsulated packet P2 via that LSP, ABR-1 would in turn perform IP forwarding lookup for the encapsulated packet P2. Since the BGP next-hop of the best route for that packet is ABR-2 and ABR1 has a LSP towards ABR-2, ABR-1 would transport that encapsulated packet P2 via the LSP towards ABR-2. When that encapsulated packet P2

arrives at ABR-2, ABR-2 would also perform IP forwarding lookup and then forward that packet P2 via a LSP towards PE-2. PE-2 decapsulates the received packet P2 and then process the resulting decapsualted packet P1 accordingly.

4. Conclusions

By simply replacing the innermost transport LSP with an IP tunnel, the need for advertising non-aggregatable BGP labeled host routes across domains is eliminated. Instead, it only requires advertising aggregated non-labeled BGP routes across domains. As a result, the requirement for BGP RIB and MPLS forwarding table resources are largely reduced. Furthermore, in the multi-area/level MPLS deployment case where MPLS-TE shortcut or Forwarding Adjacency (FA) feature is enabled between ABRs, the need for running BGP between ABRs can be eliminated further. Instead, IGP route summary across area boundaries is good enough.

5. Security Considerations

TBD.

6. IANA Considerations

No action is required for IANA.

7. Acknowledgements

Thanks to.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3031] Rosen, E., Viswanathan, A. and R. Callon, "Multiprotocol Label Switching Architecture", [RFC 3031](#), January 2001.
- [RFC4023] Worster, T., Rekhter, Y., and E. Rosen, "Encapsulating MPLS in IP or GRE", [RFC4023](#), March 2005.

8.2. Informative References

- [RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", [RFC 3107](#), May 2001.

[RFC4364] Rosen. E and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", [RFC 4364](#), February 2006.

[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

Xiaohu Xu
Huawei Technologies,
Beijing, China
Phone: +86-10-60610041
Email: xuxiaohu@huawei.com

Zhenbin Li
Huawei Technologies,
Beijing, China
Phone: +86-10-60613676
Email: lizhenbin@huawei.com