

Network Working Group
Internet Draft
Category: Standards Track
Expiration Date: May 2008

George Swallow
Cisco Systems, Inc.

Jim Guichard
Cisco Systems, Inc.

November 2007

Network Scaling with Aggregate LSPs

[draft-swallow-mpls-aggregate-fec-00.txt](#)

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of 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>

Abstract

This document defines a means for an Multiprotocol Label Switched network to summarize routes while maintaining end-to-end LSP connectivity thereby reducing the number of host routes that need to be carried within the interior gateway protocol.

Contents

1	Introduction	3
1.1	Conventions	3
1.2	Terminology	3
2	Overview	4
2.1	Inter-area LSP Construction	4
2.2	Label forwarding operation	6
3	Aggregate FEC	7
3.1	Aggregate FEC Element	7
3.2	Label distribution procedures	7
4	Algorithmically derived de-aggregation label	8
5	Security Considerations	8
6	IANA Considerations	8
7	References	8
7.1	Normative References	8
7.2	Informative References	9
8	Authors' Addresses	9

1. Introduction

The growth of next-generation Multiprotocol Label Switched (MPLS)-based services such as l2vpn, l3vpn and so on, have introduced an explosion in the number of edge devices that are needed to deploy such services. Typically these services require an end-to-end label-switched path LSP, from ingress Label-Switching Router (LSR) to egress LSR. These LSPs are maintained between the host IP addresses of the LSRs. This requirement forces the operator to carry each host address for every edge device within their interior gateway protocol (IGP) which has a negative impact on the scale and convergence goals of their IGP.

This document defines extensions to the Label Distribution Protocol (LDP) to provide a means for the end-to-end LSP path to be maintained between edge devices without the necessity of carry each and every host address within the IGP or to distribute labels throughout a domain for host addresses. This is achieved by defining an Aggregate Forwarding Equivalence Class (FEC) and an algorithmically derived label per aggregated host route which stacked together form end-to-end LSPs.

1.1. Conventions

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](#) [[KEYWORDS](#)].

1.2. Terminology

ABR	Area border router
FEC	Forwarding equivalence class
IGP	Interior gateway protocol
LDP	Label Distribution Protocol
LIB	Label information base
LSP	Label-switched path
LSR	label-switching router
NHLFE	Next-hop label forwarding entry
PE	Provider edge [LSR]
RIB	Routing information base
VPN	Virtual private network

2. Overview

In a typical MPLS deployment today, Service Providers are either running a flat IGP or leaking the loopback addresses of their PEs between the area of their IGPs. LDP then creates LSPs based on all of these host routes. Our objective here is to eliminate the need for a full mesh of LSPs between all PE pairs. This will enable Service Providers to scale and improve the convergence of their IGPs.

The basic idea is to use an LSP bound to a summarized IP prefix to deliver MPLS packets from the ingress PE, across the core, to the ABRs serving the egress PEs. Nested within these LSPs are LSPs for each of the specific hosts covered by that prefix. This is accomplished by defining two new FECs, the Aggregate FEC and the De-aggregation FEC.

The Aggregate FEC is exactly like the Prefix FEC as far as routing is concerned. An Aggregate FEC differs in that at the end of the LSP, it forms the context for interpreting the next label which is bound to a De-aggregation FEC. In that regard, an aggregate LSP never uses penultimate-hop popping. Further, to ensure that an LSP exists all the way to an LSR capable of de-aggregating the FEC, labels bound to these FECs are always distributed in ordered mode.

A De-aggregation FEC represents a specific host route. It is exactly the same as a Host Address FEC except that labels bound to these FECs are not distributed in LDP. Instead, the label value bound to a particular host is algorithmically derived from the host address and the prefix length of the aggregate FEC using a simple algorithm described in [section 4](#) below.

The use of these FECs is illustrated in the example below.

2.1. Inter-area LSP Construction

To illustrate the construction of inter-area LSPs consider the following simple topology with one backbone area and two stub areas. Using the mechanisms described within this document, we describe an LSP is built from ingress PE4 to egress PE1 with traffic destined to 192.169.0.22/32 in VPN_1:

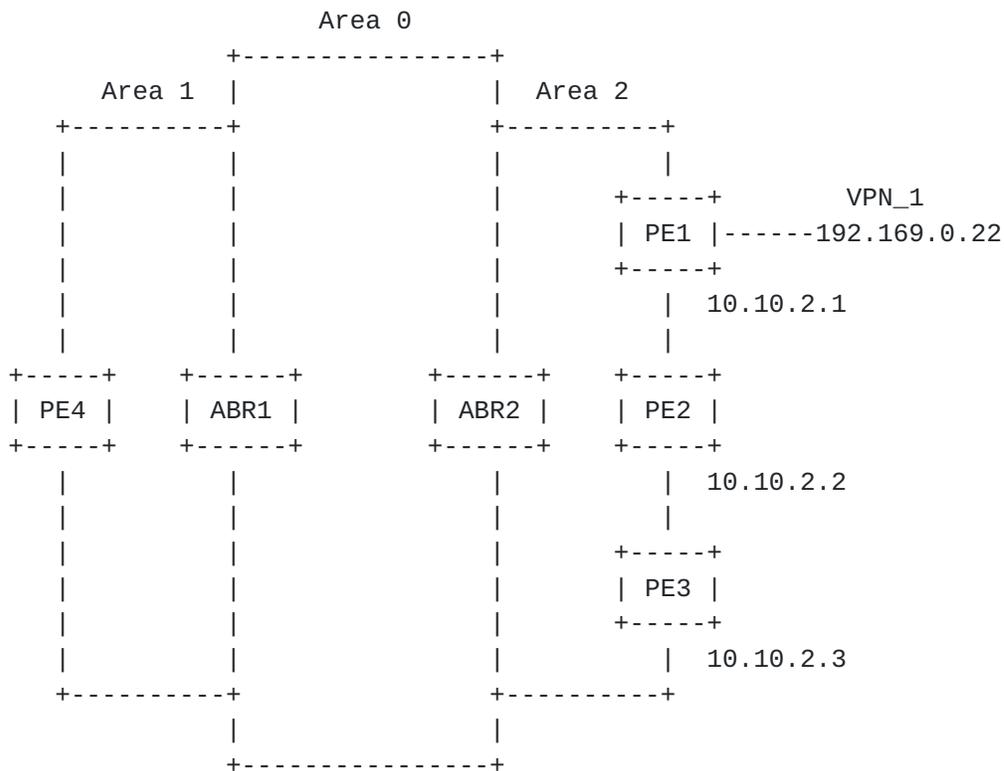


Figure 1: Example network

All routers are MPLS enabled and MPLS connectivity is required between all PE routers to facilitate edge services.

In the egress area (area 2), the records available are:

RIB	LIB
10.10.2.1/32	Labels bound to: 10.10.2.1/32
10.10.2.2/32	10.10.2.2/32
10.10.2.3/32	10.10.2.3/32

The area border router ABR2 advertises in the backbone area:

- the IP prefix 10.10.2/24 (which covers all available PE routers in area 2) in the IGP
- a label (say 51) bound to the Aggregate FEC 10.10.2/24 in LDP to its neighbors in area 0.

ABR2 algorithmically creates a label entry for each host route covered by the summary 10.10.2/24 (see [section 4](#)). It then creates NHLFEs for each of these binding them to the next-hop labels it received for each of the specific routes.

In the backbone (area 0), the records available are:

RIB	LIB
10.10.2/24	Labels bound to: 10.10.2/24

The area border router ABR1 receives the route 10.10.2/24 via the IGP and labels from its neighbors in area 0 bound to the Aggregate FEC 10.10.2/24 {22} via LDP from the next-hop router towards the route 10.10.2/24. It redistributes both the route and a label bound to the FEC into area 1.

The routers in area 1, including PE4, receive the route 10.10.2/24 via the IGP and labels from their neighbors bound to the Aggregate FEC 10.10.2/24 via LDP.

In the ingress area (area 1), the records available are:

RIB	LIB
10.10.2/24	Labels bound to: 10.10.2/24

2.2. Label forwarding operation

Using the information presented in the previous section, the label forwarding from ingress PE4 to egress PE1 is as follows:

PE4 has a VPN destination 192.169.0.22/32 reachable via remote PE1 whose next-hop is 10.10.2.1/32 which is bound to say 47. PE4 only has a summary route 10.10.2/24 covering the more specific next-hop 10.10.2.1/32, but has label bindings for that aggregate FEC. Using the same algorithm as the egress ABR2, PE4 determines that the de-aggregation label for PE1 is 17. PE1 creates a VRF entry for 192.169.0.22/32 which includes a three label stack consisting of its next-hop label for 10.10.2/24, say 22, stacked upon the de-aggregation label (17) stacked upon the VPN label (47).

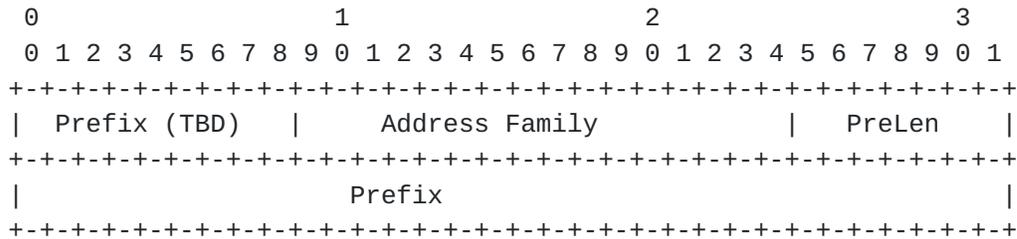
Thus to forward a packet 192.169.0.22/32 in VPN_1, PE4 pushes on a label stack {22, 17, 47}. This is forwarded via normal label-switching to ABR2 where it arrives with the label stack {51, 17, 47}. ABR2 recognizes label 51 as context label and pops it. It then looks up label 17 in the context associated 51. ABR2 then pushes onto the label stack the LDP derived label for the local PE. Forwarding continues at this point as per normal [RFC4364](#) procedures.

3. Aggregate FEC

A FEC TLV is defined in [LDP] [section 3.4.1](#). This TLV serves as a container for FEC Elements. To enable distribution of the Aggregate FEC, we define a new FEC Element.

3.1. Aggregate FEC Element

Aggregate FEC Element value encoding:



Address Family

Two octet quantity containing a value from ADDRESS FAMILY NUMBERS in [RFC1700] that encodes the address family for the address prefix in the Prefix field.

PreLen

One octet unsigned integer containing the length in bits of the address prefix that follows. A length of zero indicates a prefix that matches all addresses (the default destination); in this case the Prefix itself is zero octets).

Prefix

An address prefix encoded according to the Address Family field, whose length, in bits, was specified in the PreLen field, padded to a byte boundary.

3.2. Label distribution procedures

LSRs MUST NOT assign a NULL label value to an Aggregate FEC.

4. Algorithmically derived de-aggregation label

In order to avoid having to distribute de-aggregation labels, they are algorithmically derived from the host address with the following procedure:

The network mask is inverted and applied to mask off the high-order bits of the address. The remaining bits are treated as an integer to which 16 is added. This value represented as a 20-bit integer becomes the label value.

The value 16 is added in the algorithm in order to bypass the reserved label range.

Issue: This algorithm may not be appropriate for IPv6.

5. Security Considerations

[to be written]

6. IANA Considerations

[to be written]

7. References

7.1. Normative References

[LDP] Andersson, L. et al., "LDP Specification", [RFC 3036](#), January 2001.

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

7.2. Informative References

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

8. Authors' Addresses

George Swallow
Cisco Systems, Inc.
1414 Massachusetts Ave.
Boxborough, MA 01719

Email: swallow@cisco.com

Jim Guichard
Cisco Systems, Inc.
1414 Massachusetts Ave
Boxborough, MA 01719

Email: jguichar@cisco.com

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement

this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Full Copyright Notice

Copyright (C) The IETF Trust (2007). This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

