Networking Working Group
Internet Draft

A. Dolganow (ed.)
M. Bocci (ed.)
Alcatel-Lucent

L. Martini (ed.)
Cisco

Frederic Jounay (ed.)
France Telecom

Yuji Kamite (ed.)
NTT Communications

Intended status: Standards Track
April 15, 2009
Expires: October 15, 2009

OSPF Extensions for Dynamic Placement of Multi-Segment Pseudowires
draft-dolganow-pwe3-ospf-ms-pw-ext-03.txt

Status of this Memo

Abstract

Multi-segment pseudowires have been defined to enable emulated layer 1 and layer 2 services to be delivered from an IP based packet

switched network over a sparse mesh of PSN tunnels and PW control protocol sessions. MS-PWs can be used to scale PW based networks

over both a single AS, or between multiple ASs, and there is a particular need to be able to dynamically route MS-PWs through a given AS to reach PEs at or beyond the edge of the AS, where the route of the PW through each AS needs to be automatically determined.

This draft proposes extensions to OSPF to enable the automatic advertisement of summarized PW FECs, thus enabling the dynamic routing of MS-PWs across an OSPF domain.

Copyright Notice

Table of Contents

1. Introduction

Multi-segment pseudowires have been defined to enable emulated layer 2 services to be delivered from an IP based packet switched network over a sparse mesh of PSN tunnels and PW control protocol adjacencies. MS-PWs can be used to scale PW based networks over both a single AS, and multiple ASs. Requirements for MS-PWs are detailed in [8].

A basic approach to MS-PWs, where the switching points are statically placed, is described in [10]. This is extended in [11]to allow the automatic placement of the MS-PWs. This draft uses FEC 129 with AII type II to summarize the PW end points that are reachable through a given PE, and to provide a layer 2 address for the S-PEs. MP-BGP is used to distribute FECs.

The use of MP-BGP is primarily focused on scenarios where each PWE3 domain is a separate AS, and S-PEs are used to switch PWs between adjacent ASs. MP-BGP; therefore, provides the BGP-enabled T-PE or S-PE at the ingress of the AS with reachability information for AIIs at or beyond the border of the AS. This provides sufficient information to dynamically route the PW across the AS when there is a direct PSN tunnel between the ingress and egress S-PE or T-PE. When there is no direct PSN tunnel, a mechanisms must be provided to determine an indirect route for the PW via some intermediate S-PE within an AS domain.

A second important case is where MS-PWs are deployed in service provider access and metro networks. Pseudowires in these networks typically span only a single IGP domain or AS. Furthermore, the nodes contain a minimal routing implementation to cut on the operational complexity. In such networks MP-BGP is not typically deployed on MTUs and full MP-BGP functionality may not be required. This prevents methods defined in [11] to be employed; however, the need to be able to dynamically route MS-PWs through these topologies still exists.

To enable automatic placement of MS-PWs in the above described cases, it is possible to leverage the mechanisms of the PSN IGP to distribute MS-PW endpoint reachability information. This draft proposes extensions to OSPF to enable the automatic advertisement of summarized PW layer 2 addresses within a single AS, thus enabling the automatic placement of MS-PWs across an OSPF domain. The advertised information is used by T-PEs and S-PEs to derive the MS-PW next hop route which is used then to signal the next-hop S-PE or T-PE, as described in [11]. The draft also describes mechanisms to isolate OSPF advertisements used for MS-PWs from those used for routing in the underlying PSN to avoid any overloading of the IGP routing by the mechanisms proposed.

1.1. Terminology

The terminology defined in [9] applies.

1.2. Architecture

```
+-----+                                                  +-----+
|T-PE1+--------------------------------------------------+T-PE2|
+-----+                                                  +-----+

|<---------------------------PW----------------------------->|

+-----+             +-----+           +-----+            +-----+
|T-PE1+-------------+S-PE1+-----------+S-PE2+------------+T-PE2|
+-----+             +-----+           +-----+            +-----+

 <---------------------- Single AS -------------------------->
```

Figure 1 MS-PW Routing Model for a Single AS

Figure 1 illustrates the MS-PW routing model for a single AS. ACs attached to T-PE2 are associated with the OSPF Router_ID or any locally assigned routable address. Each S-PE / T-PE is also assigned its own layer 2 address in the form of an AII as described in [11].

The proposed model requires the existence of PSN tunnels between T-PE/S-PE, S-PE/S-PE and/or S-PE/T-PE. PWs are established on these tunnels using Targeted LDP signaling [11].

When a PSN tunnel exists and can be used for the establishment of MS-PW or it segment, T-PE advertises the set of exterior AIIs reachable via this tunnel. This is done using summarized Type 2 AIIs so that a separate advertisement is not required for every AC reachable via

that tunnel. AIIs may be summarized using the aggregation rules for AII Type 2 described in [6]. When an advertisement is received by an S-PE and the S-PE has a PSN tunnel connectivity to the advertising PE, the advertisement is installed in the S-PE's routing information and the S-PE summarizes AIIs at the far end of the tunnel in its own advertisement for propagation in the AS backbone and further propagation into other areas. When an advertisement is received by a T-PE and the T-PE has PSN tunnel connectivity to the advertising PE, the advertisement is installed in the PE's routing information base. As an alternative to this summarization-based population of an advertisement at S-PEs, a static (through configuration) method may be employed at the S-PE in order to populate the AIIs reachable through it.

Based on the advertised information, each PE builds a routing information base containing all exterior AIIs reachable through a given next hop S-PE. When creating a MS-PW, the PE looks up the AII and determines the next hop S-PE or T-PE for LDP signaling, as described in [11].

Figure 1 depicts the simple model of a one-to-one relationship of T-PE to S-PE and S-PE to S-PE, and of a single S-PE to S-PE segment. In the general case, multiple S-PE segments will exist, and the relationship between two S-PEs and/or the T-PE and S-PEs will be one-to-many or many-to-one. Processing in these cases follows that of the general case illustrated in Figure 1. Selection of an S-PE from a set of multiple available next-hop S-PEs may be achieved by comparing the IGP metrics for the route to the terminating T-PE (TT-PE) via each of these next-hop S-PEs.

2. Conventions used in this document

In examples, "C:" and "S:" indicate lines sent by the client and server respectively.

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 [1].

3. Applicability

The proposed OSPFv2 and OSPFv3 protocol extensions are intended for domains where MP-BGP is not used and/or only a partial mesh of PSN tunnels exists. In many cases, this will apply to routing MS-PWs across a single AS, where the source T-PE (ST-PE), the Terminating T-PE (TT-PE) and all of the intermediate S-PEs reside in the same AS.

However, the above application may also be used where OSPF is used to route one portion of a MS-PW across a given AS where the ST-PE and the TT-PE reside in different ASs. Here, OSPF is used to advertise the AIIs reachable through S-PEs corresponding to ASBRs. This enables the ingress S-PEs and intermediate S-PEs in an AS to route MS-PWs to the correct egress S-PE in the AS to reach a TT-PE in another AS. This draft does not define how the egress S-PE learns what AIIs are externally reachable through it, but this could be by configuration, or by learning the exterior reachable addresses from an exterior gateway protocol such as BGP.

4. OSPF Extensions

4.1. Attachment Circuit Addressing

As in [11], attachment circuit addressing is derived from AII type 2 [2], as shown in the following figure:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  AII Type=02  |    Length     |         Global ID             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |      Global ID (contd.)       |           Prefix              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |      Prefix (contd.)          |            AC ID              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |      AC ID (contd.)           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                  Figure 2 Attachment Circuit Addressing
```

Implementations of this procedure MUST interpret the AII as described in [11].

4.2. OSPFv2 LSAs

This extension makes use of the opaque LSA.

One new LSA is defined: the PW Switching LSA. This LSA describes the S-PEs/T-PEs, PSN tunnel(s) between peer S-PEs or T-PEs, and AII addresses reachable.

4.2.1. Pseudowire Switching LSA

OSPFv2 routers behaving as S-PEs or T-PEs advertise the layer 2 addresses reachable through them. This advertisement MUST be in an AS-scoped or Area-scoped opaque LSA described in [3].

The 'O' bit in the LSA Option field MUST be set to 1. The Opaque LSA type is TBD.

[Note: IANA will assign the Opaque LSA value]

The LSA Information field is formatted as described in Section 4.4. below.

4.3. OSPFv3 LSAs

4.3.1. Pseudowire Switching LSA

The OSPFv3 PW switching LSA has a function code of TBD. The S1/S2 bit are set to indicate an AS flooding scope for the LSA. The U bit is set indicating the OSPFv3 PW switching LSA should be flooded even if it is not understood. The LSA Information field is formatted as described in Section 4.4. below.

4.4. LSA Information Field

The LSA information consists of two or more nested Type/Length/Value (TLV) triplets.

The LSA MUST contain a TLV for the IP address of the advertising router. For OSPF v2 routers, this is the Router Address TLV defined in Section 2.4.1 of RFC 3630 [4]. For OSPF v3 routers, this is the Router IPv6 Address TLV specified in Section 3 of draft-ietf-ospf-ospfv3-traffic-07.txt [5]. In each of these, the router address MUST be set to the IP address of the advertising T-PE/S-PE.

4.4.1. Exterior AII TLV

The Exterior AII TLV is used to describe addresses attached of attached T-PEs or those routable through S-PEs to T-PEs in another AS. If the LSA information is of type AII, then the value field contains one or more AII Type 2 TLVs, as described above.

Exterior AIIs correspond to the AII (or AC) configured on a T-PE, which must contain at least the prefix and global ID to be used in FEC129 for signaling the PW endpoint. The prefix may or may not be directly related to the loopback address of the T-PE.

5. LSA Processing Procedures

Nodes capable of pseudowire switching on either side of a PSN tunnel exchange PW switching information using the Pseudowire Switching opaque LSA. These Pseudowire Switching LSAs are processed and flooded as described in Section 5.2.

5.1. P Routers

OSPF routers that receive LSAs described in this draft and that are not S-PEs or T-PEs MUST flood them according to the rules of OSPFv2 or OSPFv3, as applicable.

5.2. PE Routers

S-PEs and T-PEs that are OSPF routers and that receive LSAs described in this draft MUST flood them according to the rules of OSPFv2 or OSPFv3, as applicable. These LSAs are also installed in a PW routing database. This routing database MAY be used by S-PEs and T-PEs to calculate a PW routing table. The PW routing table has the structure described in Section 7 of [11], and is used to determine the next signaling hop when a S-PE receives a PW setup message as described in that draft. PW static routes may also be provisioned, as described in [11].

S-PEs that are also OSPF ABR MAY opt to summarize the PW routing information receive in type 10 area scoped opaque LSA. The summarization can be done in a similar way as for Ipv4 or ipv6 routes.

S-PEs that are ASBRs that receive type 10 LSAs described in this draft MAY summarize Exterior AII TLVs received in non-backbone advertisements in that S-PEs' own backbone advertisement, and MAY summarize Exterior AII TLVs received in backbone advertisements in that S-PEs non-backbone area advertisements.

6. Deployment Considerations

6.1. Addition and Removal of ACs, S-PEs and T-PEs

It is important that the impact of PW switching information advertised on the underlying OSPF routing is minimized. To achieve that it is RECOMMENDED that:

1. The Exterior AII TLV only contains the prefix and global ID of the T-PE. Such summarization allows the content of the Exterior AII TLV to remain unchanged when an AC is added or removed, thus removing a need to re-advertise the Exterior AII TLV. Likewise, no new LSA is advertised when AC connectivity flaps or when a pseudowire is established/provisioned.

2. S-PEs use AII summarization that minimizes the impact on the S-PEs' advertisements into backbone on changes to Exterior AII TLV received in a non-backbone advertisements.

6.2. Impact on Existing P-Routers

P-routers supporting Opaque LSA processing procedures must exist along the flooding path in the AS to ensure propagation of the information required for dynamic pseudowire routing. Ideally, multiple "Opaque LSA" flooding paths exist, so a failure of a router along a path does not isolate subset of a network.

Routers supporting Opaque LSA processing as described in [3] will flood the LSAs as specified in [3]. The impact of this additional Opaque LSA flooding load MAY be constrained through appropriate levels of aggregation of AIIs as described in Section 6.1. Section 6.3. describes complementary methods for limiting the impact of any additional flooding.

6.3. Congestion in the Underlying PSN Routing

This draft describes the use of the underlying interior gateway protocol in an IP network to advertise routing information for the automatic placement of MS-PWs. Congestion may occur in the routing plane of the PSN if a large amount of pseudowire LSAs are flooded. It is therefore important to ensure that this does not degrade the performance of the IGP for the underlying PSN.

Implementations MAY use a number of methods to avoid routing congestion, including:

- o Prioritization of PSN LSAs over PW Switching LSAs.

- o Rate limiting PW Switching LSAs so that they do not consume excessive bandwidth or route processor capacity.

- o AII summarization methods as described in section 6.1. above

- o OSPF Refresh and Flooding reduction mechanisms as defined in [7].

It is RECOMMENDED that implementations either:

- o Use all of the above mentioned techniques to minimize the impact of pseudowire switching advertisements on the underlying IGP routing when a single routing instance is used, or

- o Use a separate transport instance for pseudowire switching advertisements.

7. Security Considerations

The security discussion in Section 6 of [3]is also applicable to PW routing information.

8. IANA Considerations

This draft requests that the following allocations be made from existing registries:

1. The OSPFv2 opaque LSA type TBD for the PW switching opaque LSA.

2. The OSPFv3 LSA type function code TBD for the PW switching LSA

9. Acknowledgments

The authors gratefully acknowledge the contributions of Vach Kompella, Devendra Raut and Yuichi Ikejiri.

This document was prepared using 2-Word-v2.0.template.dot.

10. References

10.1. Normative References

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

[2] Metz, C., et al, "AII Types for Aggregation", Internet Draft, draft-metz-aii-aggregate-01.txt, October 2006.

[3] L. Berger, et al, " The OSPF Opaque LSA Option", RFC 5250, July 2008

[4] Katz D. et al., "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003

[5] Ishiguro K. et al., "Traffic Engineering Extensions to OSPF version 3", RFC 5329, September 2008

[6] Metz C. et al., "Attachment Individual Identifier (AII) Types for Aggregation", RFC 5003, September 2007

10.2. Informative References

[7] Pillay-Esnault P., "OSPF Refresh and Flooding Reduction in Stable Topologies", RFC 4136, July 2005

[8] Bitar, N. at al. "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", RFC 5254, October 2008

[9] Bocci, M., and Bryant, S.,T., "An Architecture for Multi-Segment Pseudo Wire Emulation Edge-to-Edge", Internet Draft, draft-ietf-pwe3-ms-pw-arch-06.txt, February 2009

[10] Martini et al, "Segmented Pseudo Wire", Internet Draft, draft-ietf-pwe3-segmented-pw-11.txt, February 2009

[11] Martini, L., Bocci, M., Balus, F., et al, "Dynamic Placement of Multi Segment Pseudo Wires", Internet Draft, draft-ietf-pwe3-dynamic-ms-pw-09.txt, March 2009

Author's Addresses

Matthew Bocci
Alcatel-Lucent
Voyager Place,
Shoppenhangers Road
Maidenhead
Berks, UK
Email: matthew.bocci@alcatel-lucent.co.uk

Dimitri Papadimitriou
Alcatel-Lucent
Copernicuslaan 50
2018 ANTWERP
BELGIUM
Email: dimitri.papadimitriou@alcatel-lucent.be

Alex Zinin
ALCATEL-Lucent.
701 East Middlefield Road
M/S MOUNT-HRPB6
MOUNTAIN VIEW, CA 94043
USA
Email: alex.zinin@alcatel-lucent.com

Mustapha Aissaoui
Alcatel-Lucent
600 March Road
OTTAWA, ON K2K 2E6
CANADA
Email: mustapha.aissaoui@alcatel-lucent.com

Andrew Dolganow
Alcatel-Lucent
600 March Road
OTTAWA, ON K2K 2E6
CANADA
Email: andrew.dolganow@alcatel-lucent.com

Yuji Kamite
NTT Communcations
Email: y.kamite@ntt.com

Luca Martini
Cisco
Email: lmartini@cisco.com

Frederic Jounay
France Telecom
Email : Frederic.jounay@orange-ftgroup.com