Network Working Group
Internet Draft
Expiration Date: September 2005

IJsbrand Wijnands
Bob Thomas
Cisco Systems, Inc.
Yuji Kamite
Hitoshi Fukuda
NTT Communications

March 2005

Multicast Extensions for LDP

draft-wijnands-mpls-ldp-mcast-ext-00.txt

Status of this Memo

Abstract

Forwarding multicast packets efficiently over an MPLS core requires Point-to-Multi-Point (P2MP) or Multi-Point-to-Multi-Point (MP2MP) LSP's between one or more Ingress routers and one or more Egress routers. For efficient forwarding core LSRs need to replicate labeled multicast packets where the branches of the P2MP/MP2MP tree diverge. This draft specifies LDP extensions that enable it to build P2MP/MP2MP LSPs in a receiver initiated manner.

Table of Contents

1. Specification of Requirments

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.

2. Terminology

P2MP - Point to Multipoint Label Switched Tree.
MP2MP - Multipoint to Multipoint Label Switched Tree.
MP-T - Multipoint Tree, either a P2MP or MP2MP Tree.
Ingess - Router that is a sender on the MP-T.
Egress - Router that is a receiver of a MP-T.
Root - Router that acts as the Rendezvous point of the MP-T.

3. Introduction

Multicast trees are built using Protocol Indenpendent Multicast [PIMv2]. PIM supports three different modes of operation: PIM sparse-mode, PIM bidir [BIDIR] and PIM SSM. This draft specifies extensions to LDP that enable it to build point-to-multipoint and multipoint-to-multipoint trees for MPLS packet forwarding. These P2MP and MP2MP MPLS trees are comparable to multicast PIM SSM and PIM Bidir mode trees and may be used to support multicast over MPLS LSP's. This draft does not specify procedures analogous to PIM sparse-mode multicast.

PIM is a receiver driven soft state periodic protocol that builds trees to a source or rendezvous point. Its receiver driven nature allows it to scale well with dynamic multicast group membership and large receiver populations. A router forwards tree membership upstream, but does not forward any information about downstream receivers. LDP extensions in this draft support receiver driven construction of MP-T's. An advantage that the LDP extensions for multicast have over PIM extensions for signaling labels is that LDP has built-in reliability and flow control.

4. Label distribution

The labels which are mapped to MP-T LSPs are distributed by LDP in Downstream Unsolicited Ordered Mode and retained using the Conservative Label Retention mode. An LSR distributes the label for an MP-T LSP to an upstream peer only if the peer is the its next hop for the root of the MP-T.

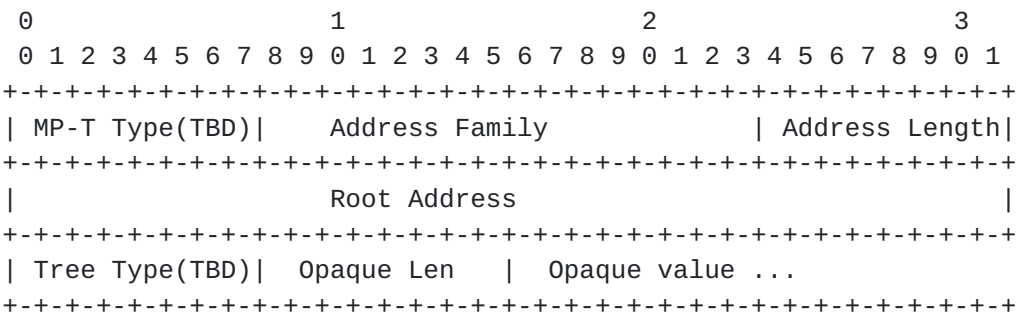
5. Label allocation

Since the extensions for signaling MP-T's use downstream label allocation MP-T LSPs may share the platform wide label space with unicast LSPs. The MPLS forwarding engine is reponsible for deciding to replicate packets using information supplied by LDP. Since MP-T and unicast LSPs share the same label space there is no need for a separate LDP session for MP-T's.

6. MP-T FEC element

The MP-T FEC element identifes an MP-T by means of the tree's root address, the tree type and information that is opaque to core LSRs.

MP-T type FEC Element encoding:



MP-T Type
MP-T type FEC element, value to be assigned by IANA.

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

Address Length
Length of the Root address value in octets.

Root Address
The root address of the MP-T. Used by receiving LSR to determine the next-hop toward the MP-T root.

Tree Type
1 octet that identifies the tree type ie.
- P2MP LSP.

- MP2MP downstream LSP.
- MP2MP upstream LSP.

Opaque Len
Length of the opaque value in octets.

Opaque value
Variable length opaque value that uniquely identifies the MP-T.

The triple <Root Address, Tree Type, Opaque Value> uniquely identifies the MP-T. LDP uses the Root Address to determine the upstream LSR toward the MP-T; the Tree Type determines the nature of LDP protocol interactions required to establish the MP-T LSP; and, the Opaque Value carries information that may be meaningful to edge LSRs.

7. In-band signaling using LDP

LDP is used to build Label Switched paths through a network. The packets that traverse the LSP are not of interest to LDP. Edge LSRs may use the Opaque field of the MP-T FEC element to encode multicast stream information. Egress LSRs may encode the source and group for a multicast stream in the Opaque field. Of course, different Egress LSRs which receive the same multicast stream must use the same source/group encoding. Such an opaque value could be used to signal the Root LSR which multicast stream is to be forwarded on the MP-T. Specification of such encodings is outside the scope of this draft.

The multicast component that wishes to receive multicast packets over a LDP created MP-T creates the opaque FEC value. Depending on the different applications there will be different Opaque FEC encodings. Different FEC encodings are to be documented elsewhere.

8. Out-of-band signaling with LDP

When an egress router wishes to receive a multicast stream over a MP-T it needs to know the identifier of that MP-T. Using in-band signaling the egress router can create the MP-T identifier (Opaque FEC) using a pre-defined algorithm, so there no need for other signaling. If the egress router is not able to use in-band signaling, for example when different multicast streams are aggregated over the same MP-T then an out-of-band form of signaling is required to learn the MP-T identifier. Such out-of-band signaling is beyond the scope of this document.

9. Building a P2MP LSP tree

In order for a set of LSRs to become egress LSRs of the same P2MP or MP2MP LSP, they must encode the same "root node" and "opaque identifier" values into the FEC element of the LDP Mapping messages that they send. How they determine what the proper root node and opaque identifier values are is outside the scope of this specification.

9.1. Label mapping

An LSR which is setting up a particular MP-T only sends a Label Mapping Message for a P2MP LSP to the LSR which is to be its "upstream neighbor" on that LSP. The upstream neighbor is the one which is the next hop on the best path to the root. If there are multiple paths to the root which are equally good, one is chosen. However, once a label for a given P2MP LSP has been advertised to an upstream LSR, no further label mapping messages for that LSP are sent upstream, until such time as the label has been withdrawn, released, or the LDP connection has failed.

When an LSR receives a Label Mapping it determines if it already has state for this MP-T. If it does, it updates its MPLS forwarding engine to reflect the new MP-T branch.

If the receiving LSR does not have state and is not the the Root LSR for the MP-T it allocates a label, sends a Label Mapping for the MP-T toward the Root LSR, and installs the binding on its chosen upstream interface. This process is repeated until a Label Mapping message for the MP-T reaches the Root LSR.

If the receiving LSR is the Root, it simply informs any subscribing applicaton that the P2MP tree exists. How this is done is beyond the scope of this document.

9.2. Label withdraw

When an MP-T egress LSR is no longer interested in an MP-T it withdraws its label for the MP-T by means of a Label Withdraw message using the MP-T FEC Element to identify the MP-T.

The LSR receiving a Label Withdraw message for an MP-T from a peer updates its MPLS forwarding engine by removing the output information for the MP-T that corresponds to the peer and sends a Label Release message in reply. If no output information for the MP-T remains in its MPLS forwarding engine the LSR sends a Label Withdraw for the

MP-T upstream.

10. Building a MP2MP tree

A MP2MP LSP is much like a P2MP LSP, is has a Root node and multiple LSP-egress routers. The big difference compared with a P2MP tree is that there will be multiple senders that can forward MPLS packets upstream in the direction of the tree root. Each LSP-egress can be a LSP-ingress router for the same tree. The root node for a P2MP tree is always the same as the LSP-ingress router. This is not the case for MP2MP trees, it can be an LSP-ingress router but can also be a P-router in the core.

MPLS forwards packets based on the incoming label. The incomming label identifies the next-hop(s) to which the MPLS packet should be sent. For any MP-T tree there can be multiple next-hops. For an P2MP tree there is only one interface that receives (i.e has forwarding state for) packets belonging to the tree. For MP2MP there is some set of interfaces which have state for the tree. Packets received on any of these interfaces are replicated to each of the other members of this set.

From the perspective of a given LSR, a MP2MP "tree" consists of n P2MP LSPs, where n is the number of interfaces (upstream + downstream) on the tree. Thus the forwarding state if O(number of interfaces). If you instead used one P2MP LSP for each member of the tree, your state would be O(number of members), which scales much more poorly.

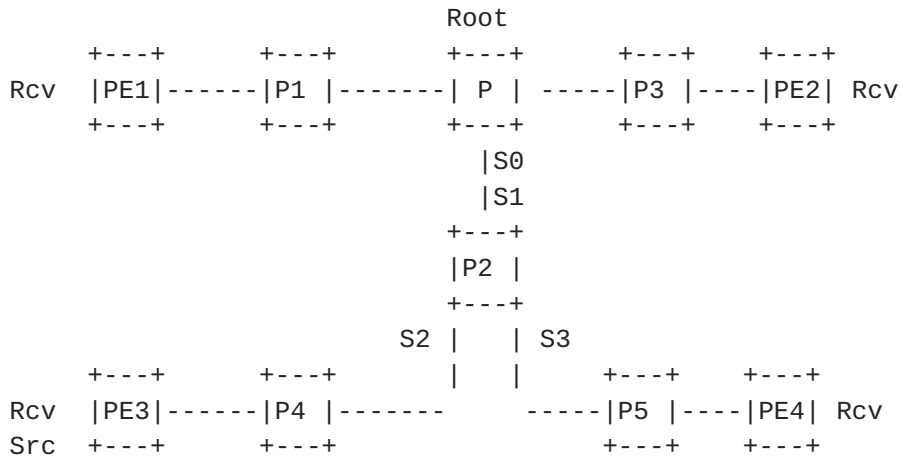


Figure 1.

Router P4 sends a label mapping to P2 and tells P2 to use label L2 for downstream traffic. Router P5 does the same and makes P2 assigned L3 for downstream traffic. If we look at P2 the downstream state is as follows.

Incoming	Outgoing
L1, S1	L2, S2
	L3, S3

Table 1.

Based on the label mapping send from P4 router P2 will send a label mapping reply to P4 with a label that is required for upstream traffic to P2. Same will happen for the upstream state from P5. Router P2 tells router P5 to use label L5 for upstream traffic.

From P2 to the root of the tree (P) the same protocol operations occur. We send a label mapping including a downstream label L1 (see table 1) and we receive an upstream label L6 mapping back for the upstream state. The upstream label we received in the label mapping reply is shared between the 2 upstream states since they both need to go to the same root via the same path, so we merge them together. Router P2 has the following upstream states:

Incoming	Outgoing
L4, S2	L6, S1

Table 2.

Incoming	Outgoing
L5, S3	L6, S1

Table 3.

The P (root) router will have the following upstream state.

Incoming	Outgoing
L6, S0	

Table 4.

The upstream state has been setup and packets can travel to the root of the tree. However, while forwarding on a MP2MP tree we want to send packets down the tree on intermediate nodes while packets travel upstream. That means packets received from PE3 and PE4 need to be sent to P5 via P2. We don't need to depend on the root to send the packets downstream. We do this by merging the downstream state and the upstream states. Each upstream state will copy the interfaces from the downstream state replication list, except if it's the same as its incoming interface. So the upstream state's on router P2 look like this:

Incoming	Outgoing
L4, S2	L6, S1 L3, S3

Table 5.

Incoming	Outgoing
L5, S3	L6, S1 L2, S2

Table 6.

Using the technique of creating specific upstream states in combination with merging the downstream replication list we are able to build a full feature MP2MP LSP tree. The LSP tree does contain more then one LSP path which costs extra labels, but the advantage is that the forwarding logic does not need to deal with any specific forwarding exceptions like we have in PIM bidir trees (forwarding packet that are received on an Outgoing Interface List).

11. Assigning Labels for MP2MP upstream Traffic

To support MP2MP (bidirectional tree) LSP's we need to setup both a downstream and a upstream LSP. The downstream path has the same protocol operation as is used for P2MP LSP's, the upstream path is different. The upstream path is setup in response to the downstream path that is build. For a label mapping we sent to an upstream router, the upstream router sends a label mapping in the opposite direction to assign us a label for upstream traffic for this MP2MP FEC.

The label mapping for the upstream path has the same encoding as the downstream path. To be able to distinguish between the two we use the

Tree type encoded in the MP-T FEC. If an LSR receives a label mapping of the type "upstream LSP" then this router will respond with a label mapping of type "downstream LSP". When the downstream router receives this label mapping it knows what this labbel mapping is for the upstream path.

12. Acknowledgments

The authors would like to thank Arjen Boers, Eric Rosen, Nidhi Bhaskar, Toerless Eckert and George Swallow for their contribution.

13. References

13.1. Normative References

[MPLS] "Multiprotocol Label Switching Architecture", Rosen, E., Viswanathan, A. and R. Callon, RFC 3031, January 2001.

[BIDIR] "Bi-directional Protocol Independent Multicast", Handley, Kouvelas, Speakman, Vicisano, June 2002, <draft-ietf-pim-bidir-04.txt>

[PIMv2] "Protocol Independent Multicast - Sparse Mode (PIM-SM)", Fenner, Handley, Holbrook, Kouvelas, December 2002, draft-ietf-pim-sm-v2-new-06.txt.

[LDP] "LDP Specification", Andersson, Doolan, Feldman, Fredette, Thomas, January 2001, rfc3036.

13.2. Informational References

[MPLS-PIM] "Using PIM to Distribute MPLS Labels for Multicast Routes", Farinacci, Rekhter, Rosen, Qian, November 2000, <draft-farinacci-mpls-multicast-03.txt>

[RFC2547bis] "BGP/MPLS VPNs", Rosen, et. al., November 2002, draft-ietf-ppvpn-rfc2547bis-03.txt

14. Authors' Addresses

IJsbrand Wijnands
Cisco Systems, Inc.
De kleetlaan 6a
1831 Diegem
Belgium
E-mail: ice@cisco.com

Bob Thomas
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA, 01719
E-mail: rhthomas@cisco.com

Yuji Kamite
NTT Communications Corporation
Tokyo Opera City Tower
3-20-2 Nishi Shinjuku, Shinjuku-ku,
Tokyo 163-1421,
Japan
Email: y.kamite@ntt.com

Hitoshi Fukuda
NTT Communications Corporation
1-1-6, Uchisaiwai-cho, Chiyoda-ku
Tokyo 100-8019,
Japan
Email: hitoshi.fukuda@ntt.com

15. Full Copyright Statement

16. 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.