Network Working Group
Internet Draft
Intended Status: Standards Track
Expires: December 2, 2008
Updates: RFCs 3032 and 4023

Toerless Eckert
Eric C. Rosen (editor)
Cisco Systems, Inc.

Rahul Aggarwal
Yakov Rekhter
Juniper Networks, Inc.

June 2, 2008

MPLS Multicast Encapsulations

draft-ietf-mpls-multicast-encaps-10.txt

Status of this Memo

Abstract

RFC 3032 established two data link layer codepoints for MPLS, used to distinguish whether the data link layer frame is carrying an MPLS unicast or an MPLS multicast packet. However, this usage was never deployed. This specification updates RFC 3032 by redefining the meaning of these two codepoints. Both codepoints can now be used to carry multicast packets. The second codepoint (formerly the "multicast codepoint") is now to be used only on multiaccess media,

and it is to mean "the top label of the following label stack is an upstream-assigned label".

RFC 3032 does not specify the destination address to be placed in the "MAC DA" field of an ethernet frame which carries an MPLS multicast packet. This document provides that specification.

This document updates RFC 3032 and RFC 4023.

Contents

1. 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 [RFC2119].

2. Introduction

RFC 3031 [RFC3031] defines the "Next Hop Label Forwarding Entry" (NHLFE). The NHLFE for a particular label maps the label into a next hop (among other things). When an MPLS packet is received, its top label is mapped to an NHLFE, and the packet is sent to the next hop specified by the NHLFE.

We define a particular MPLS label to be a "multicast label" in a particular context if the NHLFE to which it is mapped in that context specifies a set of next hops, with the semantics that the packet is to be replicated, and a copy of the packet sent to each of the specified next hops. Note that this definition accommodates the case where the set of next hops contains a single member. What makes a label a multicast label in a particular context is the semantics attached to the set, i.e., the intention to replicate the packet and transmit to all members of the set if the set has more than one member.

RFC 3032 [RFC3032] established two data link layer codepoints for MPLS: one to indicate that the data link layer frame is carrying an MPLS unicast packet, and the other to indicate that the data link layer frame is carrying an MPLS multicast packet. The term "multicast packet" is not precisely defined in RFC 3032, though one may presume that the "multicast" codepoint is intended to identify the packet's top label as a multicast label. However, the multicast codepoint has never been deployed, and further development of the procedures for MPLS multicast have shown that, while there is a need for two codepoints, the use of the two codepoints is not properly captured by RFC 3032.

In particular, there is no need for the codepoint to indicate whether the top MPLS label is a multicast label. When the receiver of an MPLS packet looks up the top label, the NHLFE will specify whether the label is a multicast label or not.

This document updates RFC 3032 and RFC 4023 by re-specifying the use of the codepoints. The old use of the "multicast codepoint", as specified in those two RFCs, is hereby deprecated.

Note that an implementation that does MPLS multicast according to RFC 3032 and/or 4023 will be unable to interoperate with implementations that do MPLS multicast according to this document. There may be some deployed platforms which support the deprecated use of the codepoints, but those platforms do not support the control plane mechanisms to support MPLS multicast. The absence of the control plane will prevent a system that implements the deprecated use of codepoints from attempting to interoperate with a system that uses

the codepoints as specified herein. (If an MPLS multicast control plane were to be implemented on a platform that only supports the deprecated codepoint, interoperability problems such as black holes and/or misrouting would arise. This does not seem like a potential problem in practice.)

While RFC 3032 allows an MPLS packet to be carried in an ethernet multicast frame, it fails to specify how the Medium Access Layer Destination Address (MAC DA) field is to be set in that case. This document provides that specification.

3. Upstream-Assigned vs. Downstream-Assigned

Suppose a labeled packet P is sent from LSR R1 to LSR R2, where R1 puts label L on the packet's label stack, and R2 has to look up label L in order to determine the corresponding Forwarding Equivalence Class (FEC), call it F.

If the binding between L and F was made by R2 and advertised to R1, then the label binding is known as "downstream-assigned". RFC 3031 only discusses downstream-assigned label bindings.

If the binding between L and F was made by R1 and advertised to R2, then the label binding is known as "upstream-assigned".

If the binding between L and F was made by a third party, say R3, and then advertised to both R1 and R2, we also refer to the label binding as "upstream-assigned".

Upstream-assigned labels are not required to come from the same "label space" as downstream-assigned labels. See [RFC3031], section 3.14, and especially [UPSTREAM] for a discussion of the notion of "label space". The procedures for properly interpreting an upstream-assigned label are given in [UPSTREAM].

If Ru and Rd are LSP adjacencies, then they transmit a MPLS packet to each other through one of the following mechanisms:

1. by putting the MPLS packet in a data link layer frame and transmitting the frame,

2. by transmitting the MPLS packet through an MPLS tunnel, i.e., by pushing an additional label (or labels) onto the label stack, and then invoking mechanism 1,

3. by transmitting the MPLS packet through an IP-based tunnel (e.g., via RFC 4023 [RFC4023]), and then invoking mechanisms 1 and/or 2.

In short, an MPLS packet is transmitted either through a data link or through an MPLS tunnel or through an IP tunnel. In any of those cases, when the packet emerges through the tunnel, the downstream LSR must know whether the label that now appears at the top of the label stack has an upstream-assigned label binding or a downstream-assigned label binding. For convenience, we will speak of a label with an upstream-assigned label binding as an "upstream-assigned label".

Under certain conditions, specified below, multicast labels MAY be upstream-assigned. The ability to use upstream-assigned labels is an OPTIONAL feature. Upstream-assigned labels MUST NOT be used unless it is known that the downstream LSR supports them. How this is known is outside the scope of this document.

This document makes no changes to the procedures regarding unicast labels.

We discuss three different types of data link or tunnel:

- Point-to-Point. A point-to-point data link or tunnel associates two systems, such that transmissions on that link or tunnel made by the one are received by the other, and only by the other.

  For a given direction of a given point-to-point data link or tunnel, the following MUST be the case: either every MPLS packet will carry an upstream-assigned label, or else every MPLS packet will carry a downstream-assigned label. The procedures for determining whether upstream-assigned or downstream-assigned labels are being used are outside the scope of this specification. However, in the absence of any other information, the use of downstream-assigned labels MUST be presumed by default.

- Point-to-Multipoint. A point-to-multipoint link or tunnel associates n systems, such that only one of them can transmit onto the link or tunnel, and the transmissions may be received by the other n-1 systems.

  The top labels (before applying the data link or tunnel encapsulation) of all MPLS packets which are transmitted on a particular point-to-multipoint data link or tunnel MUST be of the same type; either all upstream-assigned or all downstream-assigned. This means that all the receivers on the MPLS or IP tunnel must know a priori whether upstream-assigned or

downstream-assigned labels are being used in the tunnel. How this is known is outside the scope of this document.

- Multipoint-to-Multipoint. A multipoint-to-multipoint link or tunnel associates n systems, such that any of them can transmit on the link or tunnel, and the transmissions may be received by the other n-1 systems.

  If MPLS packets are transmitted on a particular multipoint-to-multipoint link or tunnel, one of the following scenarios applies:

  1. It is known (by methods outside the scope of this document) that the top label of every MPLS packet on the link or tunnel is downstream-assigned.

  2. It is known (by methods outside the scope of this document) that the top label of every MPLS packet on the link or tunnel is upstream-assigned.

  3. Some MPLS packets on the link may have upstream-assigned top labels while some may have downstream-assigned top labels.

  If (and only if) the third scenario applies, the data link or tunnel encapsulation MUST provide a codepoint which specifies whether the top label of the encapsulated MPLS packet is upstream-assigned or downstream-assigned. If a particular type of data link or tunnel does not provide such a codepoint, then the third scenario MUST NOT be used.

The remainder of this document specifies procedures for setting the data link layer codepoints and address fields.

4. Ethernet Codepoints

Ethernet is an example of a multipoint-to-multipoint data link.

Ethertype 0x8847 is used whenever a unicast ethernet frame carries an MPLS packet.

Ethertype 0x8847 is also used whenever a multicast ethernet frame carries an MPLS packet, EXCEPT for the case where the top label of the MPLS packet has been upstream-assigned.

Ethertype 0x8848, formerly known as the "MPLS multicast codepoint", is to be used only when an MPLS packet whose top label is upstream-

assigned is carried in a multicast ethernet frame.

5. PPP Protocol Field

PPP is an example of a point-to-point data link. When a PPP frame is carrying an MPLS packet, the PPP Protocol field is always set to 0x0281.

6. GRE Protocol Type

RFC 4023 is modified as described below.

If the IP destination address of the GRE encapsulation is a unicast IP address, then the ethertype value 0x8847 MUST be used in all cases for the MPLS-in-GRE encapsulation.

If the IP destination address of the GRE encapsulation is a multicast IP address, then:

- the ethertype value 0x8847 MUST be used when the top label of the encapsulated MPLS packet is downstream-assigned,

- the ethertype value 0x8848 MUST be used when the top label of the encapsulated MPLS packet is upstream-assigned.

Through procedures which are outside the scope of this specification, it may be known that if the destination address of a GRE packet is a multicast IP address, then the top label of the GRE payload is upstream-assigned. In such a case, the occurrence of the 8847 codepoint in a GRE packet with a multicast destination IP address MUST be considered an error, and the packet MUST be discarded.

7. IP Protocol Number

RFC 4023 is modified as follows: the IPv4 Protocol Number field or the IPv6 Next Header field is always set to 137, whether or not the encapsulated MPLS packet is an MPLS multicast packet.

If the IP destination address of the IP encapsulation is an IP multicast address, the IP tunnel may be considered to be a point-to-multipoint tunnel or a multipoint-to-multipoint tunnel. In either case, either all encapsulated MPLS packets in the particular tunnel have a downstream-assigned label at the top of the stack, or all encapsulated MPLS packets in that tunnel have an upstream-assigned label at the top of the stack. The means by which this is determined

for a particular tunnel is outside the scope of this specification.

8. Ethernet MAC DA for Multicast MPLS

When an LSR transmits a multicast MPLS packet in a multicast ethernet frame, it MUST set the Destination MAC Address to the value 01-00-5e-8v-wx-yz, where vwxyz is a 20-bit (five-nibble) value set as follows:

1. vwxyz MAY be set to 0

2. vwxyz MAY be set to the value of one of the MPLS labels on the packet's label stack.

Which of these procedures is the default procedure in any particular LSR is implementation-dependent. However, LSRs using the two different procedures MUST interoperate. That is, an LSR MUST NOT filter packets for which vwxyz has been set to zero, and it MUST NOT indiscriminately filter all packets for which vwxyz has not been set to zero.

If an LSR follows the procedure of setting vwxyz to the value of one of the MPLS labels on the packet's label stack, and if that label stack contains two or more labels, then by default, vwxyz MUST be set to the value of the second MPLS label on the packet's label stack. By "the second label", we mean the label that is in the label stack entry that immediately follows the topmost label stack entry. The LSR MAY, if configured to do so, allow a a label other than the second to be used for this purpose. If the MPLS packet has only one label, the value of that label will be used instead of the value of the (non-existent) second label.

It is expected that the LSR will follow the procedures of [UPSTREAM], pushing on two labels, with the topmost label being a "context label" that is the same for all MPLS packets being transmitted by the LSR onto the ethernet, but with the second label being different for different LSPs. Thus if the MAC DA value is a function of the second label, more of the LSP-specific information about the packet appears in the MAC DA field. This can be used to filter multicast packets with "unexpected" non-zero values of vwxyz. Further discussion of such filtering or its uses is outside the scope of this document.

The use of ethernet and/or IP broadcast addresses (as distinguished from multicast addresses) does not fall within the scope of this specification.

9. IANA Considerations

IANA already owns the set of ethernet multicast addresses in the range 01-00-5e-00-00-00 to 01-00-5e-ff-ff-ff. Addresses in the range 01-00-5e-00-00-00 to 01-00-5e-7f-ff-ff are already reserved for use when an ethernet multicast frame carries an IP multicast packet.

When this document is accepted, IANA shall reserve ethernet addresses in the range 01-00-5e-80-00-00 to 01-00-5e-8f-ff-ff for use when an ethernet multicast frame carries an MPLS multicast packet. Addresses in this range are to be valid when used with ethertype 8847 or 8848.

As this document modifies the usage of ethertypes 8847 and 8848, IANA shall change the description of these ethertypes as follows. Ethertype 8847 shall be defined as "MPLS", as defined in RFC 3032 and in this document. Ethertype 8848 shall be defined as "MPLS with upstream-assigned label", as defined in this document.

10. Security Considerations

The security considerations of RFC 3032 and RFC 4023 apply.

Malicious changing of the codepoint may result in loss or misrouting of packets. However, altering the codepoint without also altering the label does not result in a predictable effect.

Malicious alteration of the MAC DA on an ethernet can result in packets being received by a third party, rather than by the intended recipient.

11. Normative References

[RFC2119] "Key words for use in RFCs to Indicate Requirement Levels.", Bradner, March 1997

[RFC3031] "Multiprotocol Label Switching Architecture", Rosen, Viswanathan, Callon, January 2001

[RFC3032] "MPLS Label Stack Encoding", Rosen, et. al., January 2001

[RFC4023] "Encapsulating MPLS in IP or GRE", Worster, Rekhter, Rosen, March 2005

[UPSTREAM] "MPLS Upstream Label Assignment and Context Specific Label Space", Aggarwal, Rekhter, Rosen, draft-ietf-mpls-upstream-

label-05.txt, March 2008.

12. Authors' Addresses

Toerless Eckert
Cisco Systems, Inc.
170 Tasman Drive
San Jose, CA, 95134
Email: eckert@cisco.com

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719
Email: erosen@cisco.com

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

Yakov Rekhter
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
Email: yakov@juniper.net

13. Full Copyright Statement

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.

14. 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.