Network Working Group
Internet Draft
Category: Standards Track
Expiration Date: October 2007

George Swallow
Cisco Systems, Inc.

Thomas D. Nadeau
Cisco Systems, Inc.

Rahul Aggarwal
Juniper Networks, Inc.

April 2007

Connectivity Verification for Multicast Label Switched Paths

draft-ietf-mpls-mcast-cv-00.txt

Status of this Memo

Abstract

Requirements for MPLS P2MP LSPs extend to hundreds or even thousands of endpoints. This document defines a more scalable approach to verifying connectivity for P2MP LSPs.

Contents

1. Introduction

Requirements for Multi-protocol Label Switching (MPLS) Point-to-multipoint (P2MP) Label Switched Paths (LSPs) call for scaling up to hundreds or even thousands of endpoints. Existing tools such as those defined in [RFC4379] and [MPLS-BFD] generally require explicit acknowledgment to each connectivity probe. Such explicit acknowledgments adversely affect the scalability and/or practicality of performing connectivity verification. That is, the response load at the root would either be overwhelming unless the probing was done infrequently. This document defines a more scalable approach to monitoring P2MP LSP connectivity.

MPLS Echo Request/Reply messages [RFC4379] are used to bootstrap a Bi-directional Forwarding Detection (BFD) session across the P2MP LSP in a manner similar to "BFD For MPLS LSPs" [MPLS-BFD]. The actual monitoring uses extensions to BFD defined in [BFD-MCST].

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].

Based on context the terms leaf, egress and receiver are used somewhat interchangeably. The first two are exactly the same. Egress is used where consistency with [RFC4379] was deemed appropriate. Receiver is used in the context of receiving protocol messages.

2. Overview

In order to scale to large numbers of leaves and to be able to verify connectivity on a frequent basis the protocol defined herein uses BFD packets as unidirectional probes. As specified in [BFD-MCST] BFD packets are sent by the root at a fixed minimum interval. The leaves receive BFD packets and declare a connectivity fault if more than a fixed number of BFD messages are missed.

The session is bootstrapped by an MPLS Echo Request/Reply message exchange. The root periodically sends MPLS Echo Request messages containing a Connectivity Verification Session object which is defined in section 3.1. The Echo Request message contains a FEC stack to identify the LSP. This serves to bind the FEC to a BFD discriminator.

Further discussion on the necessity of bootstrapping the BFD session with with MPLS Echo Request/Reply messages can be found in section 3.2 of [MPLS-BFD].

3. Connectivity Verification Bootstrapping and Maintenance

The root of the multicast tree initiates Connectivity Verification and is responsible for most of the parameters involved in the Connectivity Verification (CV) Session. These parameters are communicated both through MPLS echo request messages and through BFD. The primary role of of the echo request message is to provide the binding between the root's address and chosen BFD discriminator and a particular FEC. It further enables the root to scope the session to a subset of leaves. It also provides a facility for declare some leaves administratively down while maintaining the CV session for the balance of the leaves.

The balance of the session parameters are communicated through BFD.

3.1. Bootstrap and Maintenance Procedures at the Root

The root first selects a discriminator and an IP destination address to be used both in the BFD packets and in the Connectivity Verification Session object. Prior to sending an MPLS Echo Request message, the root SHOULD begin sending BFD packets with the selected Discriminator in the My Discriminator field and destination IP address in-band of the subject LSP. Failure to do this could result in false alarms.

The root then bootstraps the CV Sessions by creating an MPLS Echo Request message containing a Connectivity Verification Session object and a FEC stack which specifies the LSP for which connectivity verification is desired. The Connectivity Verification Session object MUST contain the selected discriminator and destination IP address. For IPv4 the address MUST be in the range 127/8; for IPv6 the address MUST be in the range 0:0:0:0:0:FFFF:127/104.

The Lifetime SHOULD be set to a large value as compared to the BFD Detection Time.

Echo reply messages can be jittered by using the Echo Jitter object defined in [[MCSTPING]. the jitter time is set to value that is a function of the rate at which the root is able to process responses and the expected number of responders to this particular message. Exactly how values are chosen is implementation and platform

dependent. As such, the exact setting of this interval is beyond the scope of this document.

The source and destination IP address of the MPLS echo request packet MUST be the same as those used in the BFD packets. The message is then sent in-band of the LSP.

The root (assuming the root does not want the session to time-out) MUST refresh the session within Lifetime milliseconds. It is RECOMMENDED that the root refresh the CV Session at approximately one third of the Lifetime.

If the root wishes to increase the Lifetime, it should behave as if it were first bootstrapping the session. That is it should seek Echo reply messages from all receivers.

If the entire CV Session is administratively taken down, this SHOULD be handled through BFD. If, however, a subset of the egress nodes is to be administratively taken down, this is accomplished by including the Administratively Down Nodes sub-object listing the subject nodes. This list may be modified on any refresh message to indicate additional nodes being taken down or to indicate certain nodes as no-longer administratively down. Note that refresh messages MAY be sent at any time to accomplish this.

3.1.1. Special Considerations for RSVP-TE P2MP Tunnels

For RSVP P2MP tunnels the root knows all of the leaves. When bootstrapping a session, the root can know when all the leaves have responded. Suppose that an initial bootstrap message has been sent and sufficient time for responses have been allowed. If the root has not received MPLS Echo Reply messages from all of the leaves, the root MAY send a subsequent bootstrap message immediately using the scoping techniques of [MCSTPING] to limit the responses.

If a new leaf is added to the tree, the root MAY send a refresh message immediately. Further it MAY use the scoping techniques of [MCSTPING] to limit the response to just the new leaf.

3.1.2. Special Considerations for mLDP P2MP Tunnels

For Multicast LDP P2MP tunnels the root generally does not know all of the leaves. It is therefore RECOMMENDED that the initial bootstrapping messages be retransmitted several times at relatively short intervals. The number of times SHOULD be equal to or greater than the value of bfd.DetectMult of the associated BFD MultipointHead

session.

Note that the root can learn who the leaves are from the MPLS Echo Reply messages. It is RECOMMENDED that the root keep a list of active leaves. When the any of the parameters in section 3.2 above are changed, the root can then use the technique in section 3.2.1 to ensure that state is updated, noting however, that some leaves may have ceased connectivity to the tree, while others may have joined.

3.2. Procedures at an Egress

[Note: this section needs to be brought into in sync with [BFD-MCST]]

BFD packets which have the M bit set and are addressed to IPv4 addresses in the range 127/8 or IPv6 addresses in the range 0:0:0:0:0:FFFF:127/104 SHOULD be ignored if no MPLS Echo Request has been received containing the associated IP source address and discriminator combination.

When a node receives a MPLS Echo Request containing a Connectivity Verification object, it begins by processing the message as it would any other MPLS Echo Request message. If the result of that processing is error free and this node is an egress for the FEC at the bottom of the FEC stack, it checks to see if it has CV session state matching the source IP address, discriminator and FEC stack. If not it creates state as specified in section 3.2.1 below. If it does it updates that state as specified in section 3.2.2. Normal response processing for the received MPLS Echo is then done.

3.2.1. Creating Egress Connectivity Verification State

CV session state is created keyed on the source IP address and Discriminator value. This state is set to expire in Lifetime milliseconds. The session is considered to have expired if not refreshed prior to the expiration of this timer. Included in this state is the FEC and CV Session state, initially set to Init. The egress SHOULD now process BFD packets with this source IP address and Discriminator value.

When a BFD packet is received that matches the source IP address and Discriminator it is processed and a BFD session is created. The BFD session is linked to this CV state. In particular the CV session is informed of the BFD state transitions. The CV Session state is changed to UP.

3.2.2. Updating Egress Connectivity Verification State

[Note, this section will be updated when the Egress CV Session State Machine is added].

If a Connectivity Verification Session object is received which matches the Source_IP_Addr, Discriminator and FEC Stack of existing CV state the Lifetime is reset and the message is examined to determine if there have been any changes in parameters. If the IP address of the egress has been added to the Administratively down nodes, the egress MUST change the CV session state to Administratively Down.

If the IP address of the egress has been removed from the Administratively Down nodes, then if the BFD session state is Down or the BFD session has been deleted, the CV state is set to INIT; if the BFD session state is UP the CV session state is set to UP.

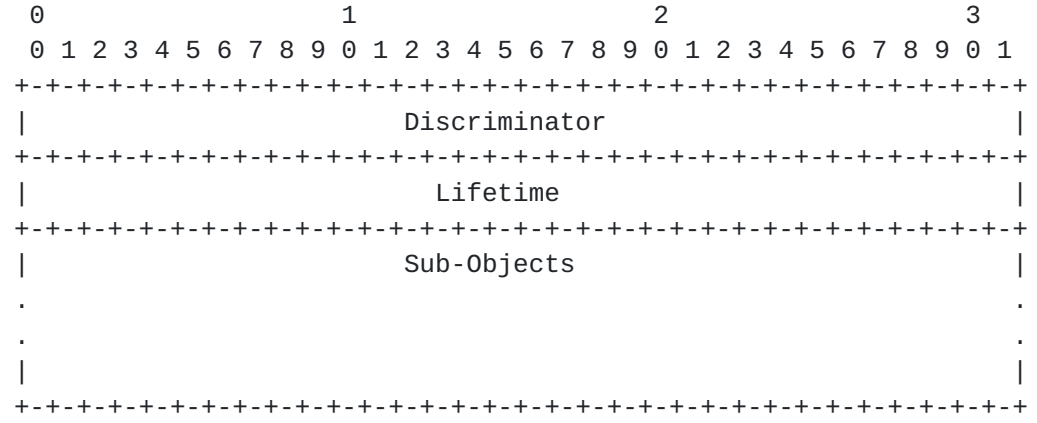
3.2.3. CV Session State Machine

[To be written]

4. Connectivity Verification Session Object

The Connectivity Verification Session object is used to notify leaves that connectivity verification will be performed on the LSP and to set the connectivity verification parameters.

The Connectivity Verification Session object has the following format:



Discriminator

The unique, nonzero discriminator value generated by the transmitting system, which will be used to identify this BFD session.

Lifetime

This is the minimum period before a refresh message is sent in milliseconds.

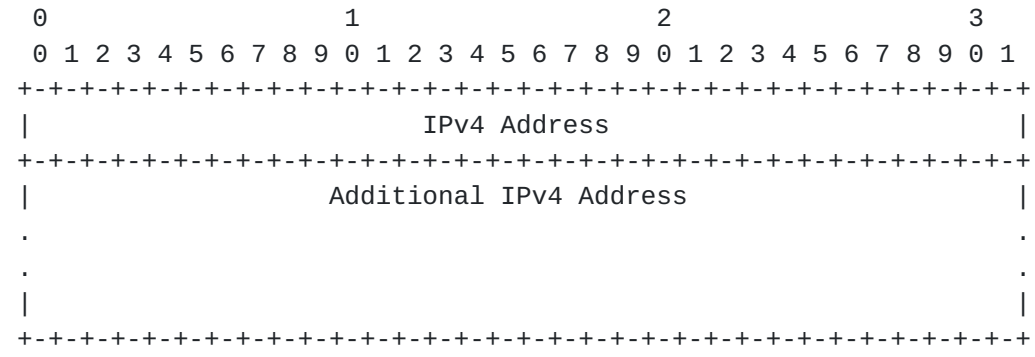
Sub-Objects

Two sub-objects are defined

Sub-Type	Length	Value Field
1	4+	Administratively Down IPv4 Nodes
2	16+	Administratively Down IPv6 Nodes

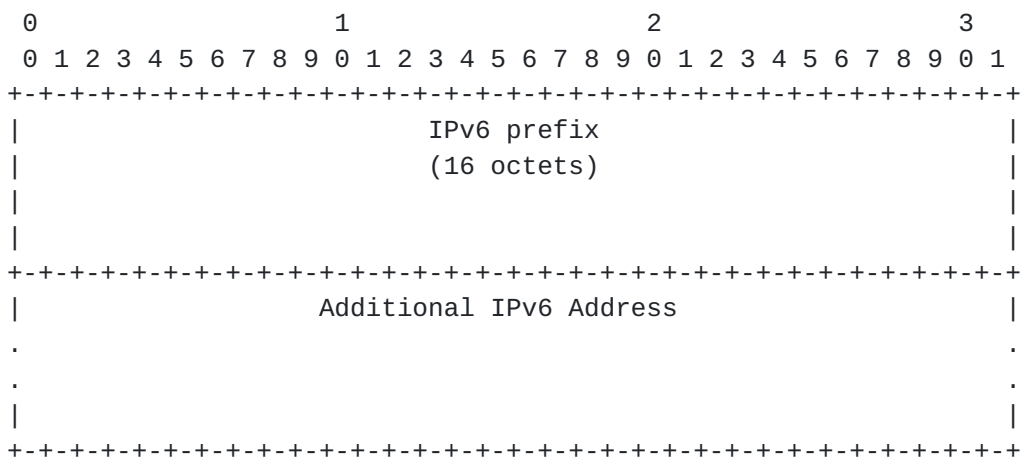
4.1. Administratively Down IPv4 Nodes

The Administratively Down IPv4 Nodes sub-object is used to suppress alarms from specific nodes.



4.2. Administratively Down IPv6 Nodes

The Administratively Down IPv6 Nodes sub-object is used to suppress alarms from specific nodes.



5. Security Considerations

Security considerations discussed in [BFD], [BFD-MHOP] and [RFC4379] apply to this document.

6. IANA Considerations

This document makes the following codepoint assignments from the LSP Ping Object Type registry (pending IANA action):

Object Sub-objects	Codepoint
Connectivity Verification Session	tba
Administratively Down IPv4 Nodes	1
Administratively Down IPv6 Nodes	2

7. Acknowledgments

The authors would like to thank Dave Katz, Dave Ward, and Vanson Lim for their comments and suggestions.

8. References

8.1. Normative References

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.

[BFD-MCST] Katz, D. and D. Ward, "BFD for Multipoint Networks", draft-katz-ward-bfd-multipoint-00.txt, February 2007.

[BFD] Katz, D., and Ward, D., "Bidirectional Forwarding Detection", draft-ietf-bfd-base-05.txt, June 2006.

[BFD-MHOP] D. Katz, D. Ward, "BFD for Multihop Paths", draft-ietf-bfd-multihop-04.txt, June 2006.

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

[MCSTPING] Farrel, A. et al, "Detecting Data Plane Failures in Point-to-Multipoint MPLS Traffic Engineering - Extensions to LSP Ping", draft-ietf-mpls-p2mp-lsp-ping-02.txt, September 2006.

8.2. Informative References

[MPLS-BFD] Aggarwal, R., et al., "BFD For MPLS LSPs", draft-ietf-bfd-mpls-03.txt, June 2006.

9. Authors' Addresses

George Swallow
Cisco Systems, Inc.

Email: swallow@cisco.com

Tom Nadeau
Cisco Systems, Inc.

Email: tnadeau@cisco.com

Rahul Aggarwal
Juniper Networks, Inc.

Email: rahul@juniper.net

Full Copyright Notice

Notes:

Destination address in CV object

The destination address is included to allow out of band pings to solicit responses from individual destinations. Is this desirable?

Alarm mode

Would it be useful to have any configuration of alarm mode? I.e. syslog vs NMS. Notification back to the root is covered in BFD-MCST

Think about later:

Finer control on how individual nodes alarm

Individual tails could be configured via LSP Ping so that they never send BFD control packets to the head, even when the head wishes notification of path failure from the tail. Such tails will never be known to the head, but will still be able to detect multipoint path failures from the head. Is such a thing useful?

Automatic authentication configuration (is that an oxymoron?)

If authentication is in use, all tails must be configured to have a common authentication key in order to receive the multipoint BFD Control packets. The bootstrap *could* be used to configure the BFD auth info, but I'm not at all sure that can be done securely.

Updating section needs to have change of FEC covered.

Configuration at the egress needs to be discussed.