MPLS Working Group R. Torvi
Internet-Draft R. Bonica
Intended status: Informational Juniper Networks
Expires: April 26, 2015 M. Conn
D. Pacella
L. Tomotaki
M. Wygant
Verizon
October 23, 2014

LSP Self-Ping
draft-bonica-mpls-self-ping-02

Abstract

This memo describes LSP Self-ping. Ingress LSR's can use LSP Self-ping to verify that an LSP is ready to carry traffic.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

An ingress Label Switching Router (LSR) can use RSVP-TE [RFC3209] to establish an MPLS Label Switched Path [RFC3032]. The following paragraphs provide an overview of RSVP-TE procedures.

The ingress LSR calculates an explicit path between itself and an egress LSR. It then formats an RSVP PATH message, including an Explicit Route Object (ERO). The ERO represents the explicit path between the ingress and egress LSRs.

The ingress LSR forwards the PATH message in the direction of the egress LSR, following the path defined by the ERO. Each transit LSR that receives the PATH message executes admission control procedures. If the transit LSR admits the LSP, it reserves bandwidth (if necessary) and sends the PATH message downstream, to the next node in the ERO.

When the egress LSR receives the PATH message, it binds a label to the LSP. The label can be implicit null, explicit null, or non-null. The egress LSR then installs forwarding state (if necessary), and constructs an RSVP RESV message. The RESV message includes a Label Object containing the label that has been bound to the LSP.

The egress LSR sends the RESV message upstream towards the ingress LSR. The RESV message visits the same transit LSRs that the PATH message visited, but in reverse order. Each transit LSR binds a label to the LSP, updates its forwarding state and updates the RESV message. As a result, the RESV message contains a Label Object and the Label Object contains the label that has been bound to the LSP. Next, the transit LSR sends the RESV message upstream, along the explicit path.

The ingress LSR receives the RESV message and installs forwarding state. Once the ingress LSR installs forwarding state it can forward traffic through the LSP.

An implementation can optimize the procedure described above by allowing LSRs to send a RESV messages upstream before installing forwarding state. This optimization is desirable, because it allows LSRs to install forwarding state in parallel, thus accelerating the process of LSP signaling and setup. However, this optimization creates a race condition. When the ingress LSR receives a RESV message, some downstream LSRs may have not yet completed the process of forwarding state installation. If the ingress sends traffic over the LSP, the traffic will be black-holed until forwarding state has been installed on all downstream LSRs.

The ingress LSP can prevent back-holing by verifying the LSPs readiness to carry traffic before forwarding traffic through it. Ingress LSRs can use LSP Self-Ping to verify that an LSP is ready to carry traffic.

LSP Self-ping is an extremely lightweight mechanism, designed to perform well when control plane resources are scarce. Therefore, LSP Self-ping consumes no control plane resources on transit or egress LSRs.

This memo describes LSP Self-ping.

2. LSP Self Ping Procedures

In order to verify that an LSP is ready to carry traffic, the ingress LSR creates a short-lived LSP Self-ping session. All session state is maintained locally on the ingress LSR. Session state includes the following:

- Session-id: A 32-bit number that identifies the session
- verification-status: A boolean variable indicating whether LSP readiness has been verified. The initial value of this variable is FALSE.
- retries: The number of times that the ingress LSR probes the LSP before giving up. The initial value of this variable is determined by configuration.
- retry-timer: The number of milliseconds that the LSR waits after probing the LSP. The initial value of this variable is determined by configuration.

The ingress LSR executes the following procedure until verification-status equals TRUE or retries is less than 1:

- o Format a MPLS Echo [RFC4379] message
- o Send the MPLS Echo message through the LSP under test
- o Set a timer to expire in retry-timer milliseconds
- o Wait until either a) a MPLS Echo message associated with the session returns or b) the timer expires. If an MPLS Echo message associated with the session returns, set verification-status to TRUE. Otherwise, decrement retries. Optionally, increase the value of retry-timer according to an appropriate back off algorithm.

As per [RFC4379], the MPLS Echo message is encapsulate in a User Datagram Protocol (UDP) [RFC0768] header. If the protocol messages used to establish the LSP were delivered over IPv4 [RFC0791], the UDP datagram is encapsulated in an IPv4 header. If the protocol messages used to establish the LSP were delivered over IPv6 [RFC2460], the UDP datagram is encapsulated in an IPv6 header.

In either case, message contents are as follows:

- o IP Source Address is configurable. By default, it is the address of the egress LSR
- o IP Destination Address is the address of the ingress LSR
- o IP Time to Live (TTL) / Hop Count is 255
- o IP DSCP is configurable. By default, it is equal to CS6 (Ox48) [RFC4594]
- o UDP Source Port is 3503
- o UDP Destination Port is 3503
- o MPLS Echo Global Flags are clear (i.e., set to 0)
- o MPLS Echo Type is equal to "MPLS Echo Reply" (2)
- o MPLS Echo Reply Mode is "Reply via an IPv4/IPv6 UDP packet" (2)
- o MPLS Echo Senders Handle is equal to the Session-ID
- o MPLS Echo Sequnce Number is equal to retries

The reader should note that the ingress LSR probes the LSP by sending an MPLS Echo message, addressed to itself, through the LSP. The egress LSR forwards the MPLS Echo message back to the ingress LSR, exactly as it would forward any other packet.

If the LSP under test is ready to carry traffic, the egress LSR receives the MPLS Echo message. The MPLS Echo message can arrive at the egress LSR with or without an MPLS header, depending on whether the LSP under test executes penultimate hop-popping procedures. If the MPLS Echo message arrives at the egress LSR with an MPLS header, the egress LSR removes that header.

The egress LSR forwards the MPLS Echo message to its destination, the ingress LSR. The egress LSR forwards the MPLS Echo message exactly as it would forward any other packet. If the egress LSR's most preferred route to the ingress LSR is through an LSP, the egress LSR forwards the MPLS Echo message through that LSP. However, if the egress LSR's most preferred route to the ingress LSR is not through an LSP, the egress LSR forwards the MPLS Echo message without MPLS encapsulation.

If the ingress LSR receives an MPLS Echo message with Senders Handle equal to the Session-ID, it sets the verification-status to TRUE. The Sequence Number does not have to match the last Sequence Number sent.

When an LSP Self-ping session terminates, it returns the value of verification-status to the invoking protocol. For example, assume that RSVP-TE invokes LSP Self-ping as part of the LSP set-up procedure. If LSP Self-ping returns TRUE, RSVP-TE makes the LSP under test available for forwarding. However, if LSP Self-ping returns FALSE, RSVP-TE takes appropriate remedial actions.

LSP Self-ping fails if all of the following conditions are true:

o The Source Address of the MPLS Echo message is equal to its default value (that is, the address of the egress LSR)

o The penultimate hop pops the MPLS label

o The egress LSR executes Unicast Reverse Path Forwarding (uRPF) procedures

In this scenario and in similar scenarios, the egress LSR discards the MPLS Echo message rather than forwarding it. In such scenarios, the calling application can set the source address to a more appropriate value.

3. Rejected Approaches

In a rejected approach, the ingress LSR uses LSP-Ping, exactly as described in [RFC4379] to verify LSP readiness to carry traffic. This approach was rejected for the following reasons.

While an ingress LSR can control its control plane overhead due to LSP Ping, an egress LSR has no such control. This is because each ingress LSR can, on its own, control the rate of the LSP Ping originated by the LSR, while an egress LSR must respond to all the LSP Pings originated by various ingresses. Furthermore, when an MPLS Echo Request reaches an egress LSR it is sent to the control plane of the egress LSR, which makes egress LSR processing overhead of LSP Ping well above the overhead of its data plane (MPLS/IP forwarding). These factors make LSP Ping problematic as a tool for detecting LSP readiness to carry traffic when dealing with a large number of LSPs.

By contrast, LSP Self-ping does not consume any control plane resources at the egress LSR, and relies solely on the data plane of the egress LSR, making it more suitable as a tool for checking LSP readiness when dealing with a large number of LSPs.

In another rejected approach, the ingress LSR does not verify LSP readiness. Alternatively, it sets a timer when it receives an RSVP RESV message and does not forward traffic through the LSP until the timer expires. This approach was rejected because it is impossible to determine the optimal setting for this timer. If the timer value is set too low, it does not prevent black-holing. If the timer value is set too high, it slows down the process of LSP signalling and setup.

Moreover, the above-mentioned timer is configured on a per-router basis. However, its optimum value is determined by a network-wide behavior. Therefore, changes in the network could require changes to the value of the timer, making the optimal setting of this timer a moving target.

4. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

5. Security Considerations

MPLS Echo messages are easily forged. Therefore, an attacker can send the ingress LSR a forged MPLS Echo message, causing the ingress LSR to terminate the LSP Self-ping session prematurely.

6. Acknowledgements

Thanks to Yakov Rekhter, Ravi Singh, Eric Rosen, Eric Osborne and Nobo Akiya for their contributions to this document.

7. Normative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

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

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, August 2006.

Authors' Addresses

Ravi Torvi
Juniper Networks

Email: rtorvi@juniper.net

Ron Bonica
Juniper Networks

Email: rbonica@juniper.net

Michael Conn
Verizon

Email: michael.e.conn@verizon.com

Dante Pacella
Verizon

Email: dante.j.pacella@verizon.com

Luis Tomotaki
Verizon

Email: luis.tomotaki@verizon.com

Mark Wygant
Verizon

Email: mark.wygant@verizon.com