Network Working Group
Request for Comments: 4950
Category: Standards Track

R. Bonica
Juniper Networks
D. Gan
D. Tappan
Consultant
C. Pignataro
Cisco Systems, Inc.

August 2007

ICMP Extensions for Multiprotocol Label Switching

Status of This Memo

Copyright Notice

Abstract

This memo defines an extension object that can be appended to selected multi-part ICMP messages. This extension permits Label Switching Routers to append MPLS information to ICMP messages, and has already been widely deployed.

Table of Contents

1. Introduction

IP routers use the Internet Control Message Protocol, ICMPv4 [RFC0792] and ICMPv6 [RFC4443], to convey control information to source hosts. Network operators use this information to diagnose routing problems.

When a router receives an undeliverable IP datagram, it can send an ICMP message to the host that originated the datagram. The ICMP message indicates why the datagram could not be delivered. It also contains the IP header and leading payload octets of the "original datagram" to which the ICMP message is a response.

MPLS Label Switching Routers (LSR) also use ICMP to convey control information to source hosts. Section 2.3 of [RFC3032] describes the interaction between MPLS and ICMP, and Sections 2.4 and 3 of [RFC3032] provide applications of that interaction.

When an LSR receives an undeliverable MPLS-encapsulated datagram, it removes the entire MPLS label stack, exposing the previously encapsulated IP datagram. The LSR then submits the IP datagram to an error processing module. Error processing can include ICMP message generation.

The ICMP message indicates why the original datagram could not be delivered. It also contains the IP header and leading octets of the original datagram.

The ICMP message, however, contains no information regarding the MPLS label stack that encapsulated the original datagram when it arrived at the LSR. This omission is significant because the LSR would have forwarded the original datagram based upon information contained by the MPLS label stack.

This memo defines an ICMP extension object that permits an LSR to append MPLS information to ICMP messages. Selected ICMP messages SHOULD include the MPLS label stack, as it arrived at the router that is sending the ICMP message. The ICMP message MUST also include the IP header and leading payload octets of the original datagram.

The ICMP extensions defined in this document must be preceded by an ICMP Extension Structure Header and an ICMP Object Header. Both are defined in [RFC4884].

The ICMP extension defined in this document is equally applicable to ICMPv4 [RFC0792] and ICMPv6 [RFC4443]. Throughout this document, unless otherwise specified, the acronym ICMP refers to multi-part ICMP messages, encompassing both ICMPv4 and ICMPv6.

2. Conventions Used in This Document

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 [RFC2119].

3. Application to TRACEROUTE

The ICMP extension defined in this memo supports enhancements to TRACEROUTE. Enhanced TRACEROUTE applications, like older implementations, indicate which nodes the original datagram visited en route to its destination. They differ from older implementations in that they also reflect the original datagram's MPLS encapsulation status as it arrived at each node.

Figure 1 contains sample output from an enhanced TRACEROUTE implementation.

```
> traceroute 192.0.2.1

  traceroute to 192.0.2.1 (192.0.2.1), 30 hops max, 40 byte packets

   1  192.0.2.13 (192.0.2.13)  0.661 ms  0.618 ms  0.579 ms

   2  192.0.2.9 (192.0.2.9)  0.861 ms  0.718 ms  0.679 ms

     MPLS Label=100048 Exp=0 TTL=1 S=1

   3  192.0.2.5 (192.0.2.5)  0.822 ms  0.731 ms  0.708 ms

     MPLS Label=100016 Exp=0 TTL=1 S=1

   4  192.0.2.1 (192.0.2.1)  0.961 ms  8.676 ms  0.875 ms
```

Figure 1: Enhanced TRACEROUTE Sample Output

4. Disclaimer

This memo does not define the general relationship between ICMP and MPLS. Section 2.3 of [RFC3032] defines this relationship.

The current memo does not define encapsulation-specific TTL (Time to Live) manipulation procedures. It defers to Section 5.4 of RFC 3034 [RFC3034] and Section 10 of [RFC3035] in this matter.

When encapsulation-specific TTL manipulation procedures defeat the basic TRACEROUTE mechanism, they will also defeat enhanced TRACEROUTE implementations.

5. MPLS Label Stack Object

The MPLS Label Stack Object can be appended to the ICMP Time Exceeded and Destination Unreachable messages. A single instance of the MPLS Label Stack Object represents the entire MPLS label stack, formatted exactly as it was when it arrived at the LSR that sends the ICMP message.

Figure 2 depicts the MPLS Label Stack Object. It must be preceded by an ICMP Extension Structure Header and an ICMP Object Header. Both are defined in [RFC4884].

In the object payload, octets 0-3 depict the first member of the MPLS label stack. Each remaining member of the MPLS label stack is represented by another 4 octets that share the same format.

```
                Class-Num = 1, MPLS Label Stack Class
                C-Type = 1, Incoming MPLS Label Stack
                Length = 4 + 4 * (number of MPLS LSEs)

        0             1             2            3
   +-------------+-------------+-------------+-------------+
   |              Label                 |EXP |S|    TTL     |
   +-------------+-------------+-------------+-------------+
   |                                                        |
   |       // Remaining MPLS Label Stack Entries //         |
   |                                                        |
   +-------------+-------------+-------------+-------------+
```

Figure 2: MPLS Label Stack Object

Label: 20 bits

Exp: Experimental Use, 3 bits

S: Bottom of Stack, 1 bit

TTL: Time to Live, 8 bits

6. Security Considerations

This memo does not specify the conditions that trigger the generation of ICMP Messages for Labeled IP Packets. It does not define the interaction between MPLS and ICMP. However, this document defines an extension that allows an MPLS router to append MPLS information to multi-part ICMP messages, and therefore can provide the user of the TRACEROUTE application with additional information. Consequently, a network operator may wish to provide this information selectively based on some policy; for example, only include the MPLS extensions in ICMP messages destined to addresses within the network management blocks with administrative control over the router. An implementation could determine whether to include the MPLS Label Stack extensions based upon the destination address of the ICMP message, or based on a global configuration option in the router. Alternatively, an implementation may determine whether to include these MPLS extensions when TTL expires based on the number of label stack entries (depth of the label stack) of the incoming packet. Finally, an operator can make use of the TTL treatment on MPLS Pipe Model LSPs defined in [RFC3443] for a TTL-transparent mode of operation that would prevent ICMP Time Exceeded altogether when tunneled over the MPLS LSP.

7. IANA Considerations

IANA has assigned the following object Class-num in the ICMP Extension Object registry:

Class-Num	Description
1	MPLS Label Stack Class

IANA has established a registry for the corresponding class sub-type (C-Type) space, as follows:

MPLS Label Stack Class Sub-types:

C-Type	Description
0	Reserved
1	Incoming MPLS Label Stack
0x02-0xF6	Available for assignment
0xF7-0xFF	Reserved for private use

C-Type values are assignable on a first-come-first-serve (FCFS) basis [RFC2434].

8. References

8.1. Normative References

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 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.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

[RFC4884] Bonica, R., Gan, D., Tappan, D., and C. Pignataro, "Extended ICMP to Support Multi-Part Messages", RFC 4884, April 2007.

8.2. Informative References

[RFC3034] Conta, A., Doolan, P., and A. Malis, "Use of Label Switching on Frame Relay Networks Specification", RFC 3034, January 2001.

[RFC3035] Davie, B., Lawrence, J., McCloghrie, K., Rosen, E., Swallow, G., Rekhter, Y., and P. Doolan, "MPLS using LDP and ATM VC Switching", RFC 3035, January 2001.

[RFC3443] Agarwal, P. and B. Akyol, "Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks", RFC 3443, January 2003.

Authors' Addresses

Ronald P. Bonica
Juniper Networks
2251 Corporate Park Drive
Herndon, VA 20171
US

EMail: rbonica@juniper.net

Der-Hwa Gan
Consultant

EMail: derhwagan@yahoo.com

Daniel C. Tappan
Consultant

EMail: Dan.Tappan@gmail.com

Carlos Pignataro
Cisco Systems, Inc.
7025 Kit Creek Road
Research Triangle Park, NC 27709
US

EMail: cpignata@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.