           Relayed Echo Reply Mechanism for Label Switched Path (LSP) Ping

Abstract

   In some inter-AS (Autonomous System) and inter-area deployment
   scenarios for RFC 4379 ("Label Switched Path (LSP) Ping and
   Traceroute"), a replying Label Switching Router (LSR) may not have
   the available route to an initiator, and the Echo Reply message sent
   to the initiator would be discarded, resulting in false negatives or
   a complete failure of operation of the LSP Ping and Traceroute. This
   document describes extensions to the LSP Ping mechanism to enable the
   replying LSR to have the capability to relay the Echo Response by a
   set of routable intermediate nodes to the initiator. This document
   updates RFC 4379.

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc7743.

Copyright Notice

Table of Contents

1. Introduction

   This document describes extensions to the Label Switched Path (LSP)
   Ping specified in [RFC4379] by adding a Relayed Echo Reply mechanism
   that could be used to report data-plane failures for inter-AS
   (Autonomous System) and inter-area LSPs. Without these extensions,
   the ping functionality provided by [RFC4379] would fail in many
   deployed inter-AS scenarios, since the replying Label Switching
   Router (LSR) in one AS may not have an available route to the
   initiator in the other AS. The mechanism in this document defines a
   new Message Type referred to as the "Relayed Echo Reply message" and
   a new TLV referred to as the "Relay Node Address Stack TLV".

   This document updates [RFC4379]; it includes updates to the Echo
   Request sending procedure in Section 4.3 of [RFC4379], the Echo
   Request receiving procedure in Section 4.4 of [RFC4379], the Echo
   Reply sending procedure in Section 4.5 of [RFC4379], and the Echo
   Reply receiving procedure in Section 4.6 of [RFC4379].

1.1. 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].

2. Motivation

   LSP Ping [RFC4379] defines a mechanism to detect data-plane failures
   and localize faults. The mechanism specifies that the Echo Reply
   should be sent back to the initiator using a UDP packet with the
   IPv4/IPv6 destination address set to an address of the LSR that
   originated the Echo Request. This works in administrative domains
   where IP-address reachability is allowed among LSRs and every LSR is
   able to route back to the originating LSR. However, in practice,
   this is often not the case due to intra-provider routing policy,
   route hiding, and network address translation at Autonomous System
   Border Routers (ASBRs). In fact, it is almost always the case that
   in inter-AS scenarios, the only node in one AS to which direct
   routing is allowed from the other AS is the ASBR, and routing
   information from within one AS is not distributed into another AS.

   Figure 1 demonstrates a case where an LSP is set up between PE1 and
   PE2. If PE1's IP address is not distributed to AS2, a traceroute
   from PE1 directed towards PE2 can result in a failure because an LSR
   in AS2 may not be able to send the Echo Reply message. For example,
   P2 cannot forward packets back to PE1 given that it is a routable IP
   address in AS1 but not routable in AS2. In this case, PE1 would

   detect a path break, as the Echo Reply messages would not be
   received. Then, localization of the actual fault would not be
   possible.

   Note that throughout the document, "routable address" means that it
   is possible to route an IP packet to this address using the normal
   information exchanged by the IGP and BGP (or EGP) operating in the
   AS.

```
+-------+   +-------+   +------+   +------+   +------+   +------+
|       |   |       |   |      |   |      |   |      |   |      |
|  PE1  +---+  P1   +---+ ASBR1+---+ ASBR2+---+  P2  +---+  PE2 |
|       |   |       |   |      |   |      |   |      |   |      |
+-------+   +-------+   +------+   +------+   +------+   +------+
<--------------AS1------------><--------------AS2------------>
<------------------------- LSP ------------------------------>
```
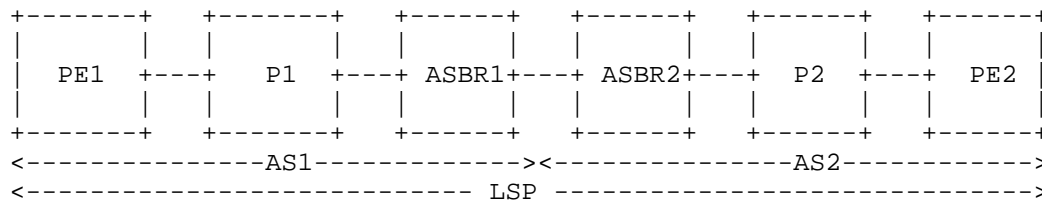
              Figure 1: Simple Inter-AS LSP Configuration

   A second example that illustrates how [RFC4379] would be insufficient
   would be the inter-area situation in a seamless MPLS architecture
   [MPLSARCH] as shown below in Figure 2. In this example, LSRs in the
   core network would not have an IP-reachable route to any of the
   access nodes (ANs). When tracing an LSP from one AN to the remote
   AN, the LSR1/LSR2 node cannot send the Echo Reply either, like the P2
   node in the inter-AS scenario in Figure 1.

```
                +-------+   +-------+   +------+   +------+
                |       |   |       |   |      |   |      |
            +--+ AGN11 +---+ AGN21 +---+ ABR1 +---+ LSR1 +--> to AGN
           /   |       |   | /|    |   |      |   |      |   |
+----+/     +-------+\/ +-------+   +------+  /+------+
| AN |           /\                    \/
+----+\     +-------+  \+------+   +------+/\ +------+
           \   |       |   |       |   |      |   | \|    |
            +--+ AGN12 +---+ AGN22 +---+ ABR2 +---+ LSR2 +--> to AGN
                |       |   |       |   |      |   |      |
                +-------+   +-------+   +------+   +------+
static route    IS-IS L1 LDP              IS-IS L2 LDP
<-Access-><--Aggregation Domain--><---------Core--------->
```
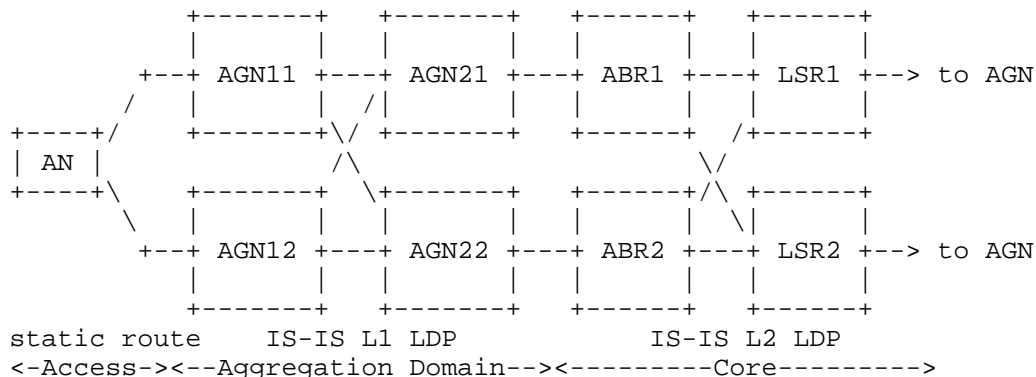
   AGN: aggregation node

                Figure 2: Seamless MPLS Architecture

   This document describes extensions to the LSP Ping mechanism to
   facilitate a response from the replying LSR by defining a mechanism
   that uses a relay node (e.g., ASBR) to relay the message back to the
   initiator. Every designated or learned relay node must be reachable
   to the next relay node or to the initiator. Using a recursive
   approach, a relay node could relay the message to the next relay node
   until the initiator is reached.

   The LSP Ping relay mechanism in this document is defined for unicast.
   How to apply the LSP Ping relay mechanism in multicast is out of
   scope.

3. Extensions

   [RFC4379] defines two Message Types: Echo Request and Echo Reply.
   This document defines a new Message Type: Relayed Echo Reply. The
   Relayed Echo Reply message is used in place of the Echo Reply message
   when an LSR is replying LSR to a relay node.

   A new TLV named the "Relay Node Address Stack TLV" is defined in this
   document to carry the IP addresses of the relay nodes for the
   replying LSR.

   In addition, the Maximum Transmission Unit (MTU) Exceeded Return Code
   is defined to indicate to the initiator that one or more TLVs will
   not be returned due to the MTU size.

   It should be noted that this document focuses only on detecting the
   LSP that is set up using a uniform IP address family type. That is,
   all hops between the source and destination node use the same address
   family type for their LSP Ping control planes. This does not
   preclude nodes that support both IPv6 and IPv4 addresses
   simultaneously, but the entire path must be addressable using only
   one address family type. Support for mixed IPv4-only and IPv6-only
   is beyond the scope of this document.

3.1. Relayed Echo Reply Message

   The Relayed Echo Reply message is a UDP packet, and the UDP payload
   has the same format with Echo Request/Reply message. A new Message
   Type is requested from IANA.

   New Message Type:
       Value   Meaning
       -----   -------
       5   MPLS Relayed Echo Reply

The use of TCP and UDP port number 3503 is described in [RFC4379] and
has been allocated by IANA for LSP Ping messages. The Relayed Echo
Reply message will use the same port number.

3.2. Relay Node Address Stack

The Relay Node Address Stack TLV is an optional TLV. It MUST be
carried in the Echo Request, Echo Reply, and Relayed Echo Reply
messages if the Echo Reply relayed mechanism described in this
document is required. Figure 3 illustrates the TLV format.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Initiator Source Port     | Reply Add Type|   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Source Address of Replying Router (0, 4, or 16 octets)     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Destination Address Offset   |  Number of Relayed Addresses  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                 Stack of Relayed Addresses                    ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
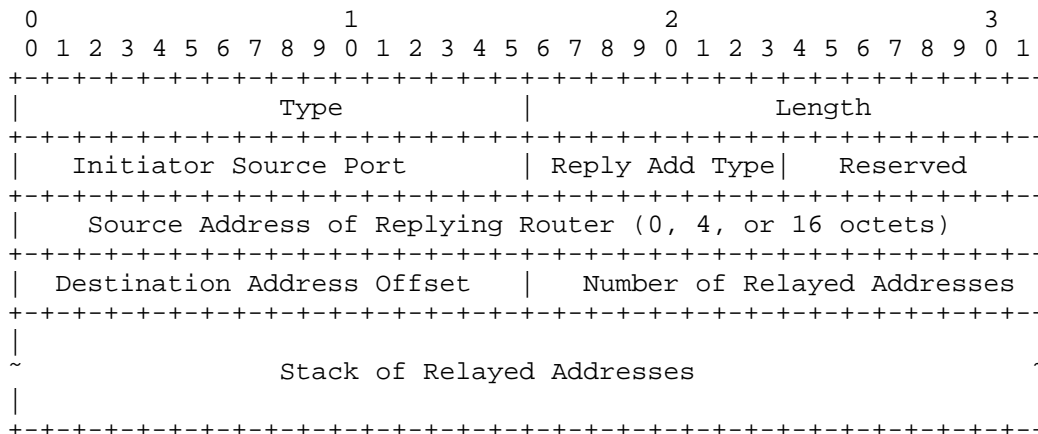
                  Figure 3: Relay Node Address Stack TLV

   - Type: Value is 32768. The value has been assigned by IANA from
     the 32768-49161 range as suggested by [RFC4379], Section 3.

   - Length: The length of the value field in octets.

   - Initiator Source Port: The source UDP port that the initiator uses
     in the Echo Request message, and the port that is expected to
     receive the Echo Reply message.

   - Reply Address Type: Address type of replying router. This value
     also implies the length of the address field as shown below.

   Type#   Address Type   Address Length
   ----   ------------   ------------
   0   Null   0
   1   IPv4   4
   2   IPv6   16

   - Reserved: This field is reserved and MUST be set to zero.

   - Source Address of Replying Router: Source IP address of the
      originator of Echo Reply or Relay Echo Reply message.

   - Destination Address Offset: The offset in octets from the top-of-
      stack to the destination address entry. Each entry size has been
      listed in this section. Please also refer to Section 4.2 for more
      detail of the operation.

   - Number of Relayed Addresses: An integer indicating the number of
      relayed addresses in the stack.

   - Stack of Relayed Addresses: A list of relay node address entries.
      This stack grows downward, with relay node addresses that are
      further along the LSP appearing lower down in the stack. Please
      refer to Section 4.2 for the relay node discovery mechanism.

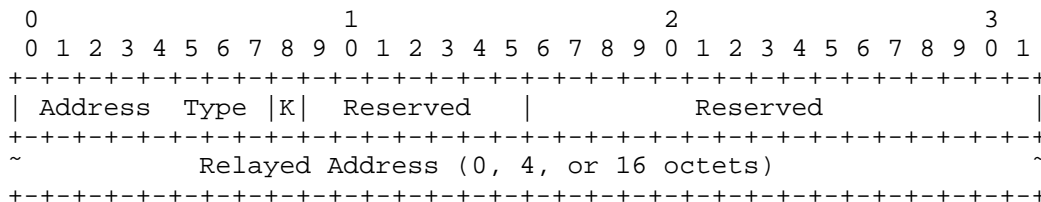   The format of each relay node address entry is as below:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Address   Type |K|  Reserved   |           Reserved          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   ~             Relayed Address (0, 4, or 16 octets)             ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

                 Figure 4: Relay Node Address Entry

Type#	Address Type	Address Length	Size of the Entry
0	Null	0	4
1	IPv4	4	8
2	IPv6	16	20

   Reserved: The two fields are reserved and MUST be set to zero.

   K bit: If the K bit is set to 1, then this address stack entry MUST
   NOT be stripped from the Relay Node Address Stack during the
   processing described in Section 4.2. If the K bit is clear, the
   entry might be stripped according to the processing described in
   Section 4.2.

   Having the K bit set in the relay node address entry causes that
   entry to be preserved in the Relay Node Address Stack TLV for the
   entire traceroute operation. A responder node MAY set the K bit to
   ensure its relay node address entry remains as one of the relay nodes

   in the Relay Node Address Stack TLV. The address with the K bit set
   will always be a relay node address for the Relayed Echo Reply, see
   Section 4.3.

   Relayed Address: This field specifies the node address: either IPv4
   or IPv6.

3.3. MTU Exceeded Return Code

   This Return Code is defined to indicate that one or more TLVs were
   omitted from the Echo Reply or Relayed Echo Reply message to avoid
   exceeding the message's effective MTU size. These TLVs MAY be
   included in an Errored TLV's Object with their lengths set to 0 and
   no value. The return sub-code MUST be set to the value that
   otherwise would have been sent. For more detail, please refer to
   Section 4.2.

   MTU Exceeded Return Code:
      Value   Meaning
      -----   -------
      20   One or more TLVs not returned due to MTU size

   This document updates step 7 in Section 4.4 of [RFC4379] to integrate
   the processing of the MTU Exceeded Return Code by adding the
   following text:

      Before sending Echo Reply, the new packet size should be checked.
      If Best-return-code is 3 ("Replying router is an egress for the
      FEC at stack depth"), or 8 ("Label switched at stack-depth"), and
      if the packet size exceeds MTU size, then Best-return-code is 20
      ("One or more TLVs not returned due to MTU size").

4. Procedures

   To perform a ping operation, the initiator first discovers the relay
   nodes. Once those nodes have been discovered, the initiator includes
   the Relay Node Address Stack TLV into any Echo Request message. The
   node can then ping as normal. Note that, in some cases, the repeated
   lack of replies to Echo Request messages may be due to a route change
   that has impacted the necessary stack of relay nodes. In this case,
   the initiator may need to rediscover the relay nodes. The following
   sections describe the procedures for sending and receiving Echo
   Request messages with the Relay Node Address Stack TLV. These
   procedures can be used in traceroute mode to discover the relay
   nodes.

4.1. Sending an Echo Request

   In addition to the procedures described in Section 4.3 of [RFC4379],
   a Relay Node Address Stack TLV MUST be carried in the Echo Request
   message if the relay functionality is required. The relay function
   initiation is out of the scope of this document. One possible way to
   do this is that the operator can explicitly add an option to the ping
   command.

   When the initiator sends the first Echo Request with a Relay Node
   Address Stack TLV, the TLV MUST contain the initiator address as the
   first entry of the stack of relayed addresses, the Destination
   Address Offset set to this entry, and the source address of the
   replying router set to null. The Initiator Source Port field MUST be
   set to the source UDP port. Note that the first relay node address
   in the stack will always be the initiator's address.

   When sending a subsequent Echo Request message, refer to Section 4.6.

4.2. Receiving an Echo Request

   The Type of the Relay Node Address Stack TLV (32768) falls within the
   range defined in [RFC4379] as "optional TLVs" that can be silently
   dropped if not recognized. An LSR that does not recognize the TLV
   SHOULD ignore it.

   In addition to the processes in Section 4.4 of [RFC4379], the
   procedures of the Relay Node Address Stack TLV are defined here.

   Upon receiving a Relay Node Address Stack TLV in an Echo Request
   message, the receiver updates the "Source Address of Replying
   Router". The address MUST be the same as the source IP address of
   Relay Echo Reply (Section 4.3) or Echo Reply message (Section 4.5)
   being sent.

   Those address entries with the K bit set to 1 MUST be kept in the
   stack. The receiver MUST check the addresses of the stack in
   sequence from bottom to top to find the last address in the stack
   with the K bit set (or the top of the stack if no K bit was found).
   The receiver then checks the stack beginning with this entry,
   proceeding towards the bottom to find the first routable address IP
   address. The Destination Address Offset MUST be set to this entry,
   which is also the resolved destination address. Address entries
   below the first routable IP address MUST be deleted. At least one
   address entry of the replying LSR MUST be added at the bottom of the
   stack. A second address entry (or more) MAY also be added if
   necessary, depending on implementation. The final address added MUST
   be an address that is reachable through the interface that the Echo

Request message would have been forwarded to if its TTL had not
expired at this node. The updated Relay Node Address Stack TLV MUST
be carried in the response message.

If the replying LSR is configured to hide its routable address
information, the address entry added in the stack MUST be a NIL entry
with Address Type set to NULL.

If a node spans two addressing domains (with respect to this message)
where nodes on either side may not be able to reach nodes in the
other domain, then the final address added SHOULD set the K bit. One
example of spanning two address domains is the ASBR node.

K bit applies in the case of a NULL address, to serve as a warning to
the initiator that further Echo Request messages may not result in
receiving Echo Reply messages.

If the full reply message would exceed the MTU size, the Relay Node
Address Stack TLV SHOULD be included in the Echo Reply message (i.e.,
other optional TLVs are excluded).

4.3. Originating a Relayed Echo Reply

The destination address determined as described in Section 4.2 is
used as the next relay node address. If the resolved next relay node
address is not routable, then the sending of the Relayed Echo Reply
or Echo Reply will fail.

If the first IP address in the Relay Node Address Stack TLV is not
the next relay node address, the replying LSR SHOULD send a Relayed
Echo Reply message to the next relay node. The processing of Relayed
Echo Reply is the same with the procedure of the Echo Reply described
in Section 4.5 of [RFC4379], except the destination IP address and
the destination UDP port. The destination IP address of the Relayed
Echo Reply is set to the next relay node address from the Relay Node
Address Stack TLV, and both the source and destination UDP port are
set to 3503. The updated Relay Node Address Stack TLV described in
Section 4.2 MUST be carried in the Relayed Echo Reply message. The
Source Address of the Replying Router field is kept unmodified, and
the Source IP address field of the IP header is set to an address of
the sending node.

When the next relay node address is the first one in the address
list, please refer to Section 4.5.

4.4. Relaying a Relayed Echo Reply

   Upon receiving a Relayed Echo Reply message with its own address as
   the destination address in the IP header, the relay node MUST
   determine the next relay node address as described in Section 4.2,
   with the modification that the location of the received destination
   address is used instead of the bottom of stack in the algorithm. The
   Destination Address Offset in Relay Node Address Stack TLV will be
   set to the next relay node address. Note that unlike in Section 4.2,
   no changes are made to the Stack of Relayed Addresses.

   If the next relay node address is not the first one in the address
   list, i.e., another intermediate relay node, the relay node MUST send
   a Relayed Echo Reply message to the determined upstream node with the
   payload unchanged other than the Relay Node Address Stack TLV. The
   TTL SHOULD be copied from the received Relay Echo Reply and
   decremented by 1. The Source Address of the Replying Router field is
   kept unmodified and the Source IP address field of the IP header is
   set to an address of the sending node.

   When the next relay node address is the first one in the address
   list, please refer to Section 4.5.

4.5. Sending an Echo Reply

   The Echo Reply is sent in two cases:

   1. When the replying LSR receives an Echo Request, and the first IP
       address in the Relay Node Address Stack TLV is the next relay
       node address (Section 4.3), the replying LSR would send an Echo
       Reply to the initiator. In addition to the procedure of the Echo
       Reply described in Section 4.5 of [RFC4379], the updated Relay
       Node Address Stack TLV described in Section 4.2 MUST be carried
       in the Echo Reply.

       If the receiver does not recognize the Relay Node Address Stack
       TLV, as per Sections 3 and 4.5 of [RFC4379], it will send an Echo
       Reply without including the TLV.

   2. When the intermediate relay node receives a Relayed Echo Reply,
       and the first IP address in the Relay Node Address Stack TLV is
       the next relay node address (Section 4.4), the intermediate relay
       node will send the Echo Reply to the initiator, and update the
       Message Type field from type of "Relayed Echo Reply" to "Echo
       Reply". The updated Relay Node Address Stack TLV described in
       Section 4.4 MUST be carried in the Echo Reply. The destination
       IP address of the Echo Reply is set to the first IP address in

the stack, and the destination UDP port will be copied from the
Initiator Source Port field of the Relay Node Address Stack TLV.
The source UDP port should be 3503. The TTL of the Echo Reply
SHOULD be copied from the received Relay Echo Reply and
decremented by 1. The Source Address of the Replying Router
field is kept unmodified, and the Source IP address field of the
IP header is set to an address of the sending node.

4.6. Sending Subsequent Echo Requests

   During a traceroute operation, multiple Echo Request messages are
   sent. Each time the TTL is increased, the initiator SHOULD copy the
   Relay Node Address Stack TLV received in the previous Echo Reply to
   the next Echo Request. The Relay Node Address Stack TLV MUST NOT be
   modified except as follows. A NIL entry that does not have the K bit
   set MAY be removed. A NIL entry with the K bit serves as a warning
   that further Echo Request messages are likely not to result in a
   reply. If, however, the initiator decides to continue a traceroute
   operation, the NIL entry with the K bit set MUST be removed. The
   Source Address of the Replying Router and Destination Address Offset
   fields may be preserved or reset since these fields are ignored in
   the received MPLS Echo Request.

4.7. Impact on Traceroute

   The Source IP address in Echo Reply and Relay Echo Reply should be
   the address of the node sending those packets, not the original
   responding node. Then the traceroute address output module will
   print the source IP address as below:

      if (Relay Node Address Stack TLV exists) {
   Print the Source Address of Replying Router in
   Relay Node Address Stack TLV;
      } else {
   Print the source IP address of Echo Reply message;
      }

5. LSP Ping Relayed Echo Reply Example

   Considering the inter-AS scenario in Figure 5 below, AS1 and AS2 are
   two independent address domains. In the example, an LSP has been
   created between PE1 to PE2, but PE1 in AS1 is not reachable by P2 in
   AS2.

```
   +-------+   +-------+   +------+   +------+   +------+   +------+
   |       |   |       |   |      |   |      |   |      |   |      |
   |  PE1  +---+  P1   +---+ ASBR1+---+ ASBR2+---+  P2  +---+  PE2 |
   |       |   |       |   |      |   |      |   |      |   |      |
   +-------+   +-------+   +------+   +------+   +------+   +------+
   <--------------AS1------------><--------------AS2----------->
   <----------------------- LSP ----------------------------->
```
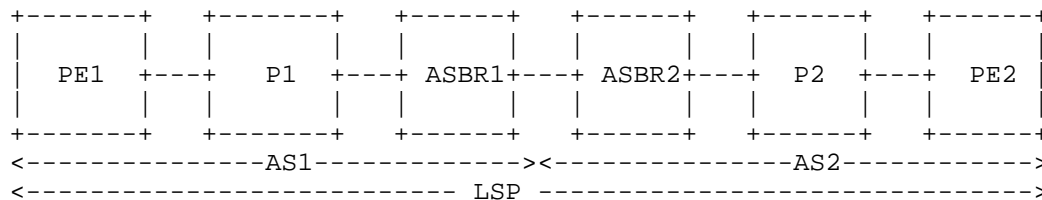
                  Figure 5: Example Inter-AS LSP

   When performing LSP traceroute on the LSP, the first Echo Request
   sent by PE1 with outermost label TTL=1 contains the Relay Node
   Address Stack TLV with PE1's address as the first relayed address.

   After being processed by P1, P1's interface address facing ASBR1
   without the K bit set will be added in the Relay Node Address Stack
   TLV address list following PE1's address in the Echo Reply.

   PE1 copies the Relay Node Address Stack TLV into the next Echo
   Request when receiving the Echo Reply.

   Upon receiving the Echo Request, ASBR1 checks the address list in the
   Relay Node Address Stack TLV and determines that PE1's address is the
   next relay address. Then it deletes P1's address, and adds its
   interface address facing ASBR2 with the K bit set. As a result,
   there will be PE1's address followed by ASBR1's interface address
   facing ASBR2 in the Relay Node Address Stack TLV of the Echo Reply
   sent by ASBR1.

   PE1 then sends an Echo Request with outermost label TTL=3, containing
   the Relay Node Address Stack TLV copied from the received Echo Reply
   message. Upon receiving the Echo Request message, ASBR2 checks the
   address list in the Relay Node Address Stack TLV and determines that
   ASBR1's interface address is the next relay address in the stack TLV.
   ASBR2 adds its interface address facing P2 with the K bit set. Then
   ASBR2 sets the next relay address as the destination address of the
   Relayed Echo Reply and sends the Relayed Echo Reply to ASBR1.

Upon receiving the Relayed Echo Reply from ASBR2, ASBR1 checks the
address list in the Relay Node Address Stack TLV and determines that
PE1's address is the next relay node. Then ASBR1 sends an Echo Reply
to PE1.

For the Echo Request with outermost label TTL=4, P2 checks the
address list in the Relay Node Address Stack TLV and determines that
ASBR2's interface address is the next relay address. Then P2 sends a
Relayed Echo Reply to ASBR2 with the Relay Node Address Stack TLV
containing four addresses: PE1's, ASBR1's interface address, ASBR2's
interface address, and P2's interface address facing PE2 in sequence.

Then, according to the process described in Section 4.4, ASBR2 sends
the Relayed Echo Reply to ASBR1. Upon receiving the Relayed Echo
Reply, ASBR1 sends an Echo Reply to PE1. And, as relayed by ASBR2
and ASBR1, the Echo Reply would finally be sent to the initiator PE1.

For the Echo Request with outermost label TTL=5, the Echo Reply would
relayed to PE1 by ASBR2 and ASBR1, similar to the case of TTL=4.

The Echo Reply from the replying node that has no IP reachable route
to the initiator is thus transmitted to the initiator by multiple
relay nodes.

6. Security Considerations

The Relayed Echo Reply mechanism for LSP Ping creates an increased
risk of DoS by putting the IP address of a target router in the Relay
Node Address Stack. These messages could then be used to attack the
control plane of an LSR by overwhelming it with these packets. A
rate limiter SHOULD be applied to the well-known UDP port on the
relay node as suggested in [RFC4379]. The node that acts as a relay
node SHOULD validate the relay reply against a set of valid source
addresses and discard packets from untrusted border router addresses.
An implementation SHOULD provide such filtering capabilities.

If an operator wants to obscure their nodes, it is RECOMMENDED that
they replace the replying node address that originated the Echo Reply
with the NIL address entry in the Relay Node Address Stack TLV.

A receiver of an MPLS Echo Request could verify that the first
address in the Relay Node Address Stack TLV is the same address as
the source IP address field of the received IP header.

The Relay Node Address Stack TLV has the path information of the LSP,
and such information may be maliciously used by any uncontrolled LSR/
LER. We have two ways to reduce the path information in the TLV:

o It is recommended to clear the K bit in the relay address entry
   unless it must be set (e.g., as listed in Section 4.2).

o It is recommended to use the NIL address entry to hide node
   information, if possible.

Other security considerations discussed in [RFC4379] are also
applicable to this document.

7. Backward Compatibility

When one of the nodes along the LSP does not support the mechanism
specified in this document, the node will ignore the Relay Node
Address Stack TLV as described in Section 4.2. Then the initiator
may not receive the Relay Node Address Stack TLV in Echo Reply
message from that node. In this case, an indication should be
reported to the operator, and the Relay Node Address Stack TLV in the
next Echo Request message should be copied from the previous Echo
Request, and continue the ping process. If the node described above
is located between the initiator and the first relay node, the ping
process could continue without interruption.

8. IANA Considerations

IANA has assigned one new Message Type, one new TLV, and one Return
Code.

8.1. MPLS Relayed Echo Reply

One new Message Type from the "Multi-Protocol Label Switching (MPLS)
Label Switched Paths (LSPs) Ping Parameters" registry in the "Message
Type" subregistry has been allocated:

        Value   Meaning
        -----   -------
        5   MPLS Relayed Echo Reply

The value has been assigned from the "Standards Action" [RFC5226]
range (0-191) using the lowest free value within this range.

8.2. Relay Node Address Stack TLV

   One new TLV from the "Multi-Protocol Label Switching (MPLS) Label
   Switched Paths (LSPs) Ping Parameters" registry in the "TLVs"
   subregistry has been allocated:

        Type   TLV Name
        ----   --------
        32768   Relay Node Address Stack TLV

   The value has been assigned from the "Standards Action" range
   (32768-49161) as suggested by [RFC4379] Sections 3 and 7.2 using the
   first free value within this range.

8.3. MTU Exceeded Return Code

   The MTU Exceeded return code from the "Multi-Protocol Label Switching
   (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry in the
   "Return Codes"subregistry has been allocated:

        Value   Meaning
        -----   -------
        20   One or more TLVs not returned due to MTU size

   The value has been assigned from the "Standards Action" range (0-191)
   using the lowest free value within this range.

9. References

9.1. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119,
              DOI 10.17487/RFC2119, March 1997,
              <http://www.rfc-editor.org/info/rfc2119>.

   [RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol
              Label Switched (MPLS) Data Plane Failures", RFC 4379,
              DOI 10.17487/RFC4379, February 2006,
              <http://www.rfc-editor.org/info/rfc4379>.

9.2. Informative References

   [MPLSARCH] Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz,
              M., and D. Steinberg, "Seamless MPLS Architecture", Work
              in Progress, draft-ietf-mpls-seamless-mpls-07, June 2014.

   [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
              IANA Considerations Section in RFCs", BCP 26, RFC 5226,
              DOI 10.17487/RFC5226, May 2008,
              <http://www.rfc-editor.org/info/rfc5226>.

Acknowledgements

   The authors would like to thank Carlos Pignataro, Xinwen Jiao, Manuel
   Paul, Loa Andersson, Wim Henderickx, Mach Chen, Thomas Morin, Gregory
   Mirsky, Nobo Akiya, and Joel M. Halpern for their valuable comments
   and suggestions.

Contributors

   Ryan Zheng
   JSPTPD
   371, Zhongshan South Road
   Nanjing 210006
   China

   Email: ryan.zhi.zheng@gmail.com

Authors' Addresses

   Jian Luo (editor)
   ZTE
   50, Ruanjian Avenue
   Nanjing 210012
   China

   Email: luo.jian@zte.com.cn

   Lizhong Jin (editor)
   Shanghai
   China

   Email: lizho.jin@gmail.com

   Thomas Nadeau (editor)
   Brocade

   Email: tnadeau@lucidvision.com

   George Swallow (editor)
   Cisco Systems

   Email: swallow@cisco.com