Network Working Group
Internet-Draft
Intended status: Experimental
Expires: January 10, 2008

D. Farinacci
V. Fuller
D. Oran
D. Meyer
cisco
Systems

July 9,
2007

Locator/ID Separation Protocol (LISP)
draft-farinacci-lisp-01.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months
and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on January 10, 2008.

Copyright Notice

Abstract

This draft describes a simple, incremental, network-based protocol to implement separation of Internet addresses into Endpoint Identifiers (EIDs) and Routing Locators (RLOCs). This mechanism requires no changes to host stacks and no major changes to existing database infrastructures. The proposed protocol can be implemented in a relatively small number of routers.

This proposal was stimulated by the problem statement effort at the Amsterdam IAB Routing and Addressing Workshop (RAWS), which took place in October 2006.

Table of Contents

1. Requirements Notation

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

Many years of discussion about the current IP routing and addressing architecture have noted that its use of a single numbering space (the "IP address") for both host transport session identification and network routing creates scaling issues (see [CHIAPPA] and [RFC1498]). A number of scaling benefits would be realized by separating the current IP address into separate spaces for Endpoint Identifiers (EIDs) and Routing Locators (RLOCs); among them are:

1. Reduction of routing table size in the "default-free zone" (DFZ). Use of a separate numbering space for RLOCs will allow them to be assigned topologically (in today's Internet, RLOCs would be assigned by providers at client network attachment points), greatly improving aggregation and reducing the number of globally-visible, routable prefixes.

2. Easing of renumbering burden when clients change providers. Because host EIDs are numbered from a separate, non-provider-assigned and non-topologically-bound space, they do not need to be renumbered when a client site changes its attachment points to the network.

3. Mobility with session survivability. Because session state is associated with a persistent host EID, it should be possible for a host (or a collection of hosts) to move to a different point in the network topology (whether by changing providers or by physically moving) without disruption of connectivity.

4. Traffic engineering capabilities that can be performed by network elements and do not depend on injecting additional state into the routing system. This will fall out of the mechanism that is used to implement the EID/RLOC split (see Section 4).

This draft describes protocol mechanisms to achieve the desired functional separation. For flexibility, the document decouples the mechanism used for forwarding packets from that used to determine EID to RLOC mappings. This work is in response to and intended to address the problem statement that came out of the RAWS effort [RAWS].

This draft focuses on a router-based solution. Building the solution into the network should facilitate incremental deployment of the technology on the Internet. Note that while the detailed protocol specification and examples in this document assume IP version 4 (IPv4), there is nothing in the design that precludes use of the same techniques and mechanisms for IPv6. It should be possible for IPv4 packets to use IPv6 RLOCs and for IPv6 EIDs to be mapped to IPv4

RLOCs.

Related work on host-based solutions is described in Shim6 [SHIM6] and HIP [RFC4423]. Related work on other router-based solutons is described in GSE [GSE]. This draft attempts to not compete or overlap with such solutions and the proposed protocol changes are expected to complement a host-based mechanism when Traffic Engineering functionality is desired.

Some of the design goals of this proposal include:

1. Minimize required changes to Internet infrastructure.

2. Require no hardware or software changes to end-systems (hosts).

3. Be incrementally deployable.

4. Require no router hardware changes.

5. Minimize router software changes.

6. Avoid or minimize packet loss when EID-to-RLOC mappings need to be performed.

There are 4 variants of LISP, which differ along a spectrum of strong to weak dependence on the topological nature and possible need for routability of EIDs. The variants are:

LISP 1: where EIDs are routable through the RLOC topology for bootstrapping EID-to-RLOC mappings. [LISP1]

LISP 1.5: where EIDs are routable for bootstrapping EID-to-RLOC mappings; such routing is via a separate topology.

LISP 2: where EIDS are not routable and EID-to-RLOC mappings are implemented within the DNS. [LISP2]

LISP 3: where non-routable EIDs are used as lookup keys for a new EID-to-RLOC mapping database. Use of Distributed Hash Tables [DHTs] to implement such a database would be an area to explore. Other examples of new mapping database services are [CONS], [NERD], and [APT].

This document will focus on LISP 1 and LISP 1.5, both of which rely on a router-based distributed cache and database for EID-to-RLOC mappings. The LISP 2 and LISP 3 mechanisms, which require separate EID-to-RLOC infrastructure, will be documented in additional drafts.

3. Definition of Terms

Provider Independent (PI) Addresses: an address block assigned from a pool that is not associated with any service provider and is therefore not topologically-aggregatable in the routing system.

Provider Assigned (PA) Addresses: a block of IP addresses that are assigned to a site by each service provider to which a site connects. Typically, each block is sub-block of a service provider CIDR block and is aggregated into the larger block before being advertised into the global Internet. Traditionally, IP multihoming has been implemented by each multi-homed site acquiring its own, globally-visible prefix. LISP uses only topologically-assigned and aggregatable address blocks for RLOCs, eliminating this demonstrably non-scalable practice.

Routing Locator (RLOC): the IPv4 or IPv6 address of an egress tunnel router (ETR). It is the output of a EID-to-RLOC mapping lookup. An EID maps to one or more RLOCs. Typically, RLOCs are numbered from topologically-aggregatable blocks that are assigned to a site at each point to which it attaches to the global Internet; where the topology is defined by the connectivity of provider networks, RLOCs can be thought of as PA addresses.

Endpoint ID (EID): a 32- or 128-bit value used in the source and destination address fields of the first (most inner) LISP header of a packet. The host obtains a destination EID the same way it obtains an destination address today, for example through a DNS lookup or SIP exchange. The source EID is obtained via existing mechanisms used to set a hosts "local" IP address. LISP uses PI blocks for EIDs; such EIDs MUST NOT be used as LISP RLOCs. Note that EID blocks may be assigned in a hierarchical manner, independent of the network topology, to facilitate scaling of the mapping database. In addition, an EID block assigned to a site may have site-local structure (subnetting) for routing within the site; this structure is not visible to the global routing system.

EID-prefix: A power-of-2 block of EIDs which are allocated to a site by an address allocation authority. EID-prefixes are associated with a set of RLOC addresses which make up a "database mapping". EID-prefix allocations can be broken up into smaller blocks when an RLOC set is to be associated with the smaller EID-prefix.

End-system: is an IPv4 or IPv6 device that originates packets with a single IPv4 or IPv6 header. The end-system supplies an EID value for the destination address field of the IP header when communicating globally (i.e. outside of it's routing domain). An end-system can be a host computer, a switch or router device, or any network appliance. An iPhone.

Ingress Tunnel Router (ITR):   a router which accepts an IP packet with a single IP header (more precisely, an IP packet that does not contain a LISP header). The router treats this "inner" IP destination address as an EID and performs an EID-to-RLOC mapping lookup. The router then prepends an "outer" IP header with one of its globally-routable RLOCs in the source address field and the result of the mapping lookup in the destination address field. Note that this destination RLOC may be an intermediate, proxy device that has better knowledge of the EID-to-RLOC mapping closest to the destination EID. In general, an ITR receives IP packets from site end-systems on one side and sends LISP-encapsulated IP packets toward the Internet on the other side.

Specifically, when a service provider prepends a LISP header for Traffic Engineering purposes, the router that does this is also regarded as an ITR. The outer RLOC the ISP ITR uses can be based on the outer destination address (the originating ITR's supplied RLOC) or the inner destination address (the originating hosts supplied EID).

TE-ITR: is an ITR that is deployed in a service provider network that prepends an additional LISP header for Traffic Engineering purposes.

Egress Tunnel Router (ETR):   a router that accepts an IP packet where destination address in the "outer" IP header is one of its own RLOCs. The router strips the "outer" header and forwards the packet based on the next IP header found. In general, an ETR receives LISP-encapsulated IP packets from the Internet on one side and sends decapsulated IP packets to site end-systems on the other side.

TE-ETR: is an ETR that is deployed in a service provider network that strips an outer LISP header for Traffic Engineering purposes.

EID-to-RLOC Cache: a short-lived, on-demand database in an ITR that stores, tracks, and is responsible for timing-out and otherwise validating EID-to-RLOC mappings. This cache is distinct from the "database", the cache is dynamic, local, and relatively small while and the database is distributed, relatively static, and much global in scope.

EID-to-RLOC Database: a globally, distributed database that contains all known EID-prefix to RLOC mappings. Each potential ETR typically contains a small piece of the database: the EID-to-RLOC mappings for the EID prefixes "behind" the router. These map to one of the router's own, globally-visible, IP addresses.

Recursive Tunneling: when a packet has more than one LISP IP header. Additional layers of tunneling may be employed to implement traffic engineering or other re-routing as needed. When this is done, an additional "outer" LISP header is added and the original RLOCs are preserved in the "inner" header.

Reencapsulating Tunnels: when a packet has no more than one LISP IP header (two IP headers total) and when it needs to be diverted to new RLOC, an ETR can decapsulate the packet (remove the LISP header) and prepend a new tunnel header, with new RLOC, on to the packet. Doing this allows a packet to be re-routed by the re-encapsulating router without adding the overhead of additional tunnel headers.

LISP Header: a term used in this document to refer to the outer IPv4 or IPv6 header, a UDP header, and a LISP header, an ITR prepends or an ETR strips.

4. Basic Overview

One key concept of LISP is that end-systems (hosts) operate the same way they do today. The IP addresses that hosts use for tracking sockets, connections, and for sending and receiving packets do not change. In LISP terminology, these IP addresses are called Endpoint Identifiers (EIDs).

Routers continue to forward packets based on IP destination addresses. These addresses are referred to as Routing Locators (RLOCs). Most routers along a path between two hosts will not change; they continue to perform routing/forwarding lookups on addresses (RLOCs) in the IP header.

This design introduces "Tunnel Routers", which prepend LISP headers on host-originated packets and strip them prior to final delivery to their destination. The IP addresses in this "outer header" are RLOCs. During end-to-end packet exchange between two Internet hosts, an ITR prepends a new LISP header to each packet and an egress tunnel router strips the new header. The ITR performs EID-to-RLOC lookups to determine the routing path to the the ETR, which has the RLOC as one of its IP addresses.

Some basic rules governing LISP are:

- o End-systems (hosts) only know about EIDs.

- o EIDs are always IP addresses assigned to hosts.

- o Routers mostly deal with Routing Locator addresses. See details later in Section 4.1 to clarify what is meant by "mostly".

- o RLOCs are always IP addresses assigned to routers; preferably, topologically-oriented addresses from provider CIDR blocks.

- o Routers can use their RLOCs as EIDs but can also be assigned EIDs when performing host functions. Those EIDs MUST NOT be used as RLOCs. When EIDs are used the routeability of them is scoped to within the site. A hybrid use of this, for example is when a router runs the BGP protocol where iBGP peerings may use EIDs and eBGP peerings may use RLOCs.

- o EIDs are not expected to be usable for end-to-end communication in the absence of an EID-to-RLOC mapping operation.

- o EID prefixes are likely to be hierarchically assigned in a manner

which is optimized for administrative convenience and to
facilitate scaling of the EID-to-RLOC mapping database. The

hierarchy is based on a address alocation hierarchy which is not dependent on the network toplogy.

- EIDs may also be structured (subnetted) in a manner suitable for local routing within an autonomous system.

An additional LISP header may be pre-pended to packets by a transit router (i.e. TE-ITR) when re-routing of the end-to-end path for a packet is desired. An obvious instance of this would be an ISP router that needs to perform traffic engineering for packets in flow through its network. In such a situation, termed Recursive Tunneling, an ISP transit acts as an additional ingress tunnel router and the RLOC it uses for the new prepended header would be either an TE-ETR within the ISP (along intra-ISP traffic engineered path) or in an TE-ETR within another ISP (an inter-ISP traffic engineered path, where an agreement to build such a path exists).

Tunnel Routers can be placed fairly flexibly in a multi-AS topology. For example, the ITR for a particular end-to-end packet exchange might be the first-hop or default router within a site for the source host. Similarly, the egress tunnel router might be the last-hop router directly-connected to the destination host. Another example, perhaps for a VPN service out-sourced to an ISP by a site, the ITR could be the site's border router at the service provider attachment point. Mixing and matching of site-operated, ISP-operated, and other tunnel routers is allowed for maximum flexibility. See Section 8 for more details.

4.1. Packet Flow Sequence

This section provides an example of the unicast unicast packet flow with the following parameters:

- Source host "host1.abc.com" is sending a packet to "host2.xyz.com".

- Each site is multi-homed, so each tunnel router has an address (RLOC) assigned from each of the site's attached service provider address blocks.

- The ITR and ETR are directly connected to the source and destination, respectively.

Client host1.abc.com wants to communicate with server host2.xyz.com:

1. host1.abc.com wants to open a TCP connection to host2.xyz.com. It does a DNS lookup on host2.xyz.com. An A record is returned. This address is used as the destination EID and the locally-

assigned address of host1.abc.com is used as the source EID. An IP packet is built using the EIDs in the IP header and sent to the default router.

2. The default router is configured as an ITR. It prepends a LISP header to the packet, with one of its RLOCs as the source IP address and uses the destination EID from the original packet header as the destination IP address. Subsequent packets continue to behave the same way until a mapping is learned.

3. In LISP 1, the packet is routed through the Internet as it is today. In LISP 1.5, the packet is routed on a different topology which may have EID prefixes distributed and advertised in an aggregatable fashion. In either case, the packet arrives at the ETR. The router is configured to "punt" the packet to the router's control-plane processor. See Section 7 for more details.

4. The LISP header is stripped so that the packet can be forwarded by the router control-plane. The router looks up the destination EID in the router's EID-to-RLOC database (not the cache, but the configured data structure of RLOCs). An EID-to-RLOC Map-Reply message is originated by the egress router and is addressed to the source RLOC from the LISP header of the original packet (this is the ITR). The source RLOC in the IP header of the UDP message is one of the ETR's RLOCs (one of the RLOCs that is embedded in the UDP payload).

5. The ITR receives the UDP message, parses the message (to check for format validity) and stores the EID-to-RLOC information from the packet. This information is put in the ITR's EID-to-RLOC mapping cache (this is the on-demand cache, the cache where entries time out due to inactivity).

6. Subsequent packets from host1.abc.com to host2.xyz.com will have a LISP header prepended with the RLOCs learned from the ETR.

7. The egress tunnel receives these packets directly (since the destination address is one of its assigned IP addresses), strips the LISP header and delivers the packets to the attached destination host.

In order to eliminate the need for a mapping lookup in the reverse direction, the ETR gleans RLOC information from the LISP header.

Both ITR and the ETR may also influence the decision the other makes in selecting an RLOC. See section Section 6 for more details.

5. Tunneling Details

This section describes the LISP Data Message which defines the tunneling header used to encapsulate IPv4 and IPv6 packets which contain EID addresses. Even though the following formats illustrate IPv4-in-IPv4 and IPv6-in-IPv6 encapsulations, the other 2 combinations are supported as well.

5.1. LISP IPv4-in-IPv4 Header 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
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |Version|  IHL  |Type of Service|          Total
Length         |
     /  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
    /   |         Identification        |Flags|      Fragment
Offset    |
   /    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
OH     |  Time to Live | Protocol = 17 |         Header Checksum       |
   \    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
    \   |                    Source Routing
Locator                     |
     \  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      \ |                 Destination Routing
Locator                   |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |       Source Port = xxxx      |       Dest Port =
4342        |
   UDP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      \ |           UDP length          |        UDP
Checksum           |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / | Type  |  Locator Reach Bits   |
Nonce ...                |
LISP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      \ |                            ...
Nonce                            |
```

```
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |Version|  IHL  |Type of Service|          Total
Length         |
     /  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
    /   |         Identification        |Flags|      Fragment
Offset    |
   /    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
IH     |  Time to Live |    Protocol   |         Header Checksum       |
   \    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
    \   |                           Source
EID                          |
     \  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      \ |                         Destination
EID                       |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
```

5.2. LISP IPv6-in-IPv6 Header Format

```
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |Version| Traffic Class |           Flow
Label                 |
     /  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
    /   |         Payload Length        | Next Header=17|   Hop
Limit   |
   /    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+

|                                                               |
O     +                                                               +
u     |                                                               |
t     +                     Source Routing Locator                    +
e     |                                                               |
r     +                                                               +

|                                                               |
H     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
d     |                                                               |
r     +                                                               +

|                                                               |
   \    +                    Destination Routing
Locator                 +
    \
|                                                               |
     \
+                                                               +
      \
|                                                               |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |       Source Port = xxxx      |       Dest Port =
4342        |
   UDP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      \ |           UDP length          |        UDP
Checksum           |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |Type=1 |  Locator Reach Bits   |
Nonce ...             |
LISP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      \ |                          ...
Nonce                         |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

```
+-+
      / |Version| Traffic Class |           Flow
Label                   |
     /  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
    /   |         Payload Length        |  Next Header  |   Hop
Limit   |
   /    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+

|                                                               |
I     +                                                               +
n     |                                                               |
n     +                          Source EID                           +
e     |                                                               |
r     +                                                               +

|                                                               |
H     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
d     |                                                               |
r     +                                                               +

|                                                               |
   \    +                        Destination
EID                        +
    \
|                                                               |
     \
+                                                               +
      \
|                                                               |
```

```
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
```

5.3. Tunnel Header Field Descriptions

IH Header: is the inner header, preserved from the datagram received from the originating host. The source and destination IP addresses are EIDs.

OH Header: is the outer header prepended by an ITR. The address fields contain RLOCs obtained from the ingress router's EID-to-RLOC cache. The IP protocol number is "UDP (17)" from [RFC0768].

UDP Header: contains a random source port allocated by the ITR when encapsulating a packet. The destination port MUST be set to the well-known IANA assigned port value 4342. The UDP checksum field MUST be transmitted as 0 and not ignore by the ETR.

UDP Length: field contains the original packet's length. For an IPv4 encapsulated packet, the inner header Total Length is copied. For an IPv6 encapsualted packet, the inner header Payload Length plus the size of the IPv6 header (40 bytes) is copied.

LISP Type: set to 1 to encode a LISP Data Message.

LISP Nonce: is an ITR randomly generated 6-byte value which tests return routability of an ETR echoing back the none in a Map-Reply message.

LISP Locator Reach Bits: in the LISP header are set by an ITR to indicate to an ETR the reachability of the Locators in the source site. Each RLOC in a Map-Reply is assigned an ordinal value from 0 to n-1 (when there are n RLOCs in a mapping entry). The Locator Reach Bits are number from 0 to n-1 from the right significant bit of the 12-bit field. When a bit is set to 1, the ITR is indicating to the ETR the RLOC associated with the bit ordinal is reachable. See Section 6.3 for details on how an ITR can determine other site ITRs are reachable.

When doing Recursive Tunneling:

- o The OH header Time to Live field MUST be copied from the IH header Time to Live field.

o The OH header Type of Service field SHOULD be copied from the IH header Type of Service field.

When doing Re-encapsulated Tunneling:

- o The new OH header Time to Live field SHOULD be copied from the stripped OH header Time to Live field.

- o The new OH header Type of Service field SHOULD be copied from the stripped OH header Type of Service field.

Copying the TTL serves two purposes. First it preserves the distance the host intended the packet to travel. And more importantly, it provides for suppression of looping packets in the event there is a loop of concatenated tunnels due to misconfiguration.

6. EID-to-RLOC Mapping

6.1. Control-Plane Packet Format

When LISP 1 or LISP 1.5 are used, new UDP packet types encode the EID-to-RLOC mappings:

```
         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
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |Version|  IHL  |Type of Service|          Total
Length         |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |         Identification        |Flags|      Fragment
Offset    |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |  Time to Live | Protocol = 17 |         Header
Checksum       |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |                    Source Routing
Locator                  |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |                 Destination Routing
Locator                |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |           Source Port         |         Dest
Port             |
   UDP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      \ |           UDP length          |        UDP
Checksum           |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+

|                                                               |
        |                         LISP
Message                      |

|                                                               |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
```

```
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |Version| Traffic Class |           Flow
Label                |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |         Payload Length        | Next Header=17|   Hop
Limit   |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+

|                                                               |

+                                                               +

|                                                               |
        +                     Source Routing
Locator                   +

|                                                               |

+                                                               +

|                                                               |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+

|                                                               |

+                                                               +

|                                                               |
        +                  Destination Routing
Locator                 +
```

```
|                                                                 |
+                                                                 +
|                                                                 |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      / |           Source Port         |         Dest
Port            |
   UDP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
      \ |           UDP length          |        UDP
Checksum          |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
|                                                                 |
        |                         LISP
Message                         |
|                                                                 |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
```

The LISP UDP-based messages are the Map-Request and Map-Reply messages. These message formats are also used by LISP-CONS [CONS] but are sent over TCP connections instead. However, this specification is the authoritative source for message format definitions for the Map-Request and Map-Reply messages.

6.1.1. Map-Request Message Format

```
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        | Type  |       Reserved        |
Checksum            |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        | Record count  |
Nonce ...                   |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |        ... Nonce              |A|
Reserved            |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

```
+-+
        |           ITR-AFI              |            CAR-
AFI             |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |                   Originating ITR RLOC
Address                 |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |                   Originating CAR EID-
Prefix                   |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
Rec -> | EID mask-len  |    EID-AFI    |         EID-prefix ...        |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |                      Path Vector
List                       |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
```

Packet field descriptions:

Type: 2 (Map-Request)

Reserved: Set to 0 on transmission and ignored on receipt.

Checksum: A complement of the 1-complements sum of the LISP packet. The checksum MUST be computed and the UDP checksum MUST be set to 0.

Record count: The number of records in this request message. A record comprises of what is labeled 'Rec" above and occurs the a number of times equal to Record count.

RLOC Count: The number of RLOCs associated with this EID prefix.

EID Mask Len: The mask length of the EID prefix. By encoding an EID prefix, a set of RLOCs can be associated with a block of EIDs. Values are between 0 and 32 inclusive.

Nonce: A 6-byte random value created by the sender of the Map-Request.

A: This is an authoritative bit, which is set to 0 for UDP-based Map-Requests sent by an ITR. See [CONS] for TCP-based Map-Requests.

ITR-AFI: Address family of the "Originating ITR RLOC Address" field.

CAR-AFI: Address family of the "Originating ITR RLOC Address" field.

Originating ITR RLOC Address: Set to 0 for UDP-based messages. See [CONS] for TCP-based Map-Requests.

Originating CAR EID-Prefix: Set to 0 for UDP-based messages by an ITR. See [CONS] for TCP-based Map-Requests.

EID Mask-length: Mask length for EID prefix.

EID-AFI: Address family of EID-prefix according to [RFC2434]

EID-prefix: 4 bytes if an IPv4 address-family, 16 bytes if an IPv6 address-family.

6.1.2. EID-to-RLOC UDP Map-Request Message

A Map-Request contains one or more EIDs encoded in prefix format with a Locator count of 0. The EID-prefix MUST NOT be more specific than a cache entry stored from a previously-received Map-Reply.

A Map-Request is sent from an ITR when it wants to test an RLOC for reachability. This is performed by using the RLOC as the destination address for Map-Request message with a randomly allocated source UDP port number and the well-known destination port number 4342. A successful Map-Reply updates the cached set of RLOCs associated with the EID prefix range.

Map-Requests MUST be rate-limited. It is recommended that a Map-Request for the same EID-prefix be sent no more than once per second.

6.1.3. Map-Reply Message Format

```
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        | Type  |       Reserved        |
Checksum              |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        | Record count  |
Nonce ...                  |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
        |        ... Nonce              |
Reserved              |
+----> +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      |                         Record  TTL                            |
|      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      | Locator count | EID mask-len  |A|        Reserved              |
|      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
R      |           ITR-AFI             |            EID-AFI             |
e      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
c      |                  Originating ITR RLOC Address                  |
o      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
r      |                         EID-prefix                             |
d      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     /|    Priority   |    Weight     |    Unused     |    Loc-AFI     |
|  Loc +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     \|                            Locator                             |
+--->  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |                        Path Vector
List                        |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
+-+
```

Packet field descriptions:

Type: 3 (Map-Reply)

Reserved: Set to 0 on transmission and ignored on receipt.

Checksum: A complement of the 1-complements sum of the LISP packet. The checksum MUST be computed and the UDP checksum MUST be set to 0.

Record count: The number of records in this request message. A record comprises of what is labeled 'Record' above and occurs the number of times equal to Record count.

Record TTL: The time in minutes the recipient of the Map-Reply will store the mapping. If the TTL is 0, the entry should be removed from the cache immediately. If the value is 0xffffffff, the recipient can decide locally how long to store the mapping.

Locator count: The number of Locator entries. A locator entry comprises what is labeled above as 'Loc'.

EID mask-len: Mask length for EID prefix.

A: The Authoritative bit, when sent by a UDP-based message is always set by the ETR. See [CONS] for TCP-based Map-Replies.

ITR-AFI: Address family of the "Originating ITR RLOC Address" field.

EID-AFI: Address family of EID-prefix according to [RFC2434].

Originating ITR RLOC Address: Set to 0 for UDP-based messages. See [CONS] for TCP-based Map-Replies.

EID-prefix: 4 bytes if an IPv4 address-family, 16 bytes if an IPv6 address-family.

Priority: each RLOC is assigned a priority. Lower values are more preferable. When multiple RLOCs have the same priority, they are used in a load-split fashion. A value of 255 means the RLOC MUST NOT be used.

Weight: when priorities are the same for multiple RLOCs, the weight indicates how to balance traffic between them. Weight is encoded as a percentage of total packets that match the mapping entry. If a non-zero weight value is used for any RLOC, then all RLOCs must use a non-zero weight value and then the sum of all weight values MUST equal 100. What did the 3rd grader say after Steve Jobs gave an iPhone demo to the class? If a zero value is used for any RLOC weight, then all weights MUST be zero and the receiver of the Map-Reply will decide how to load-split traffic.

Locator: an IPv4 or IPv6 address (as encoded by the 'Loc-AFI' field) assigned to an ETR or router acting as a proxy replier for the EID-prefix. Note that the destination RLOC address MAY be an anycast address if the tunnel egress point may be via more than one physical device. A souce RLOC can be an anycast address as well. The source or destination RLOC MUST NOT be the broadcast address (255.255.255.255 or any subnet broadcast address known to the router), and MUST NOT be a link-local multicast address. The source RLOC MUST NOT be a multicast address. The destination RLOC SHOULD be a multicast address if it is being mapped from a multicast destination EID.

6.1.4. EID-to-RLOC UDP Map-Reply Message

When a data packet triggers a Map-Reply to be sent, the RLOCs associated with the EID-prefix matched by the EID in the original packet destination IP address field will be returned. The RLOCs in the Map-Reply are the globally-routable IP addresses of the ETR but are not necessarily reachable; separate testing of reachability is required.

Note that a Map-Reply may contain different EID-prefix granularity (prefix + length) than the Map-Request which triggers it. This might occur if a Map-Request were for a prefix that had been returned by an earlier Map-Reply. In such a case, the requester updates its cache with the new prefix information and granularity. For example, a requester with two cached EID-prefixes that are covered by a Map-Reply containing one, less-specific prefix, replaces the entry with the less-specific EID-prefix. Note that the reverse, replacement of one less-specific prefix with multiple more-specific prefixes, can also occur but not by removing the less-specific prefix rather by adding the more-specific prefixes which during a lookup will override the less-specific prefix.

Replies SHOULD be sent for an EID-prefix no more often than once per second to the same requesting router. For scalability, it is expected that aggregation of EID addresses into EID-prefixes will allow one Map-Reply to satisfy a mapping for the EID addresses in the prefix range thereby reducing the number of Map-Request messages.

The addresses for a Data message or Map-Request message are swapped and used for sending the Map-Reply. The UDP source and destination ports are swapped as well. That is, the source port in the UDP header for the Map-Reply is set to the well-known UDP port number 4342.

6.2. Routing Locator Selection

Both client-side and server-side may need control over the selection of RLOCs for conversations between them. This control is achieved by manipulating the Priority and Weight fields in EID-to-RLOC Map-Reply messages. Alternatively, RLOC information may be gleaned from received tunneled packets or EID-to-RLOC Map-Request messages.

The following enumerates different scenarios for choosing RLOCs and the controls that are available:

- o Server-side returns one RLOC. Client-side can only use one RLOC. Server-side has complete control of the selection.

- o Server-side returns a list of RLOC where a subset of the list has the same best priority. Client can only use the subset list according to the weighting assigned by the server-side. In this case, the server-side controls both the subset list and load-splitting across its members. The client-side can use RLOCs outside of the subset list if it determines that the subset list is unreachable (unless RLOCs are set to a Priority of 255). Some sharing of control exists: the server-side determines the destination RLOC list and load distribution while the client-side has the option of using alternatives to this list if RLOCs in the list are unreachable.

- o Server-side sets weight of 0 for the RLOC subset list. In this case, the client-side can choose how the traffic load is spread across the subset list. Control is shared by the server-side determining the list and the client determining load distribution. Again, the client can use alternative RLOCs if the server-provided list of RLOCs are unreachable.

- o Either side (more likely on the server-side ETR) decides not to send an Map-Request. For example, if the server-side ETR does not send Map-Requests, it gleans RLOCs from the client-side ITR, giving the client-side ITR responsibility for bidirectional RLOC reachability and preferability. Server-side ETR gleaning of the client-side ITR RLOC is done by caching the inner header source EID and the outer header source RLOC of received packets. The client-side ITR controls how traffic is returned and can alternate using an outer header source RLOC, which then can be added to the list the server-side ETR uses to return traffic. Since no Priority or Weights are provided using this method, the server-side ETR must assume each client-side ITR RLOC uses the same best Priority with a Weight of zero. In addition, since EID-prefix encoding cannot be conveyed in data packets, the EID-to-RLOC cache on tunnel routers can grow to be very large.

RLOCs that appear in EID-to-RLOC Map-Reply messages are considered reachable. The Map-Reply and the database mapping service does not provide any reachability status for Locators. This is done outside of the mapping service. See next section for details.

6.3. Routing Locator Reachability

There are 4 methods for determining when a Locator is either reachable or has become unreachable:

1. Locator reachability is determined by an ETR by examining the Loc-Reach-Bits from a LISP header of a Data Message which is provided by an ITR when an ITR encapsulates data.

2. Locator unreachability is determined by an ITR by receiving ICMP Network or Host Unreachable messages.

3. ETR unreachability is determined when a host sends an ICMP Port Unreachable message.

4. Locator reachability is determined by receiving a Map-Reply message from a ETR's Locator address in response to a previously sent Map-Request.

When determining Locator reachability by examining the Loc-Reach-Bits from the LISP Data Message, an ETR will receive up to date status from the ITR closest to the Locators at the source site. The ITRs at the source site can determine reachability when running their IGP at the site. When the ITRs are deployed on CE routers, typically a default route is injected into the site's IGP from each of the ITRs. If an ITR goes down, the CE-PE link goes down, or the PE router goes down, the CE router withdraws the default route. This allows the other ITRs at the site to determine one of the Locators has gone unreachable.

The Locators listed in a Map-Reply are numbered with ordinals 0 to n-1. The Loc-Reach-Bits in a LISP Data Message are numbered from 0 to n-1 starting with the least signfiicant bit numbered as 0. So, for example, if the ITR with locator listed as the 3rd Locator position in the Map-Reply goes down, all other ITRs at the site will have the 3rd bit from the right cleared (the bit that corresponds to ordinal 2).

When an ETR decapsulates a packet, it will look for a change in the Loc-Reach-Bits value. When a bit goes from 1 to 0, the ETR will refrain from encapsulating packets to the Locator that has just gone unreachable. It can start using the Locator again when the bit that corresponds to the Locator goes from 0 to 1.

When ITRs at the site are not deployed in CE routers, the IGP can still be used to determine the reachability of Locators provided they

are injected a stub links into the IGP. This is typically done when
a /32 address is configured on a loopback interface.

When ITRs receive ICMP Network or Host Unreachable messages as a method to determine unreachability, they will refrain from using Locators which are described in Locator lists of Map-Replies. However, using this approach is unreliable because many network operators turn off generation of ICMP Unreachable messages.

Optionally, an ITR can send a Map-Request to a Locator and if a Map-Reply is returned, reachability of the Locator has been achieved. Obviously, sending such probes increases the number of control messages originated by tunnel routers for active flows, so Locators are assumed to be reachable when they are advertised.

This assumption does create a dependency: Locator unreachability is detected by the receipt of ICMP Host Unreachable messages. When an Locator has been determined unreachable, it is not used for active traffic; this is the same as if it were listed in a Map-Reply with priority 255.

The ITR can later test the reachability of the unreachable Locator by sending periodic Requests. Both Requests and Replies MUST be rate-limited. Locator reachability testing is never done with data packets since that increases the risk of packet loss for end-to-end sessions.

7. Router Performance Considerations

LISP is designed to be very hardware-based forwarding friendly. By doing tunnel header prepending [RFC1955] and stripping instead of re-writing addresses, existing hardware could support the forwarding model with little or no modification. Where modifications are required, they should be limited to re-programming existing hardware rather than requiring expensive design changes to hard-coded algorithms in silicon.

A few implementation techniques can be used to incrementally implement LISP:

- When a tunnel encapsulated packet is received by an ETR, the outer destination address may not be the address of the router. This makes it challenging for the control-plane to get packets from the hardware. This may be mitigated by creating special FIB entries for the EID-prefixes of EIDs served by the ETR (those for which the router provides an RLOC translation). These FIB entries are marked with a flag indicating that control-plane processing should be performed. The forwarding logic of testing for particular IP protocol number value is not necessary. No changes to existing, deployed hardware should be needed to support this.

- On an ITR, prepending a new IP header is as simple as adding more bytes to a MAC rewrite string and prepending the string as part of the outgoing encapsulation procedure. Many routers that support GRE tunneling [RFC3056] or 6to4 tunneling [RFC2784] can already support this action.

- When a received packet's outer destination address contains an EID which is not intended to be forwarded on the routable topology (i.e. LISP 1.5), the source address of a data packet or the router interface with which the source is associated (the interface from which it was received) can be associated with a VRF (Virtual Routing/Forwarding), in which a different (i.e. non-congruent) topology can be used to find EID-to-RLOC mappings.

8. Deployment Scenarios

This section will explore how and where ingress and ETRs can be deployed and will discuss the pros and cons of each deployment scenario. There are two basic deployment tradeoffs to consider: centralized versus distributed caches and flat, recursive, or re-encapsulating tunneling.

When deciding on centralized versus distributed caching, the following issues should be considered:

- o Are the tunnel routers spread out so that the caches are spread across all the memories of each router?

- o Should management "touch points" be minimized by choosing few tunnel routers, just enough for redundancy?

- o In general, using more ITRs doesn't increase management load, since caches are built and stored dynamically. On the other hand, more ETRs does require more management since EID-prefix-to-RLOC mappings need to be explicitly configured.

When deciding on flat, recursive, or re-encapsulation tunneling, the following issues should be considered:

- o Flat tunneling implements a single tunnel between source site and destination site. This generally offers better paths between sources and destinations with a single tunnel path.

- o Recursive tunneling is when tunneled traffic is again further encapsulated in another tunnel, either to implement VPNs or to perform Traffic Engineering. When doing VPN-based tunneling, the site has some control since the site is prepending a new tunnel header. In the case of TE-based tunneling, the site may have control if it is prepending a new tunnel header, but if the site's ISP is doing the TE, then the site has no control. Recursive tunneling generally will result in suboptimal paths but at the benefit of steering traffic to resource available parts of the network.

- o The technique of re-encapsulation ensures that packets only require one tunnel header. So if a packet needs to be rerouted, it is first decapsulated by the ETR and then re-encapsulated with a new tunnel header using a new RLOC.

The next sub-sections will describe where tunnel routers can reside in the network.

8.1. First-hop/Last-hop Tunnel Routers

By locating tunnel routers close to hosts, the EID-prefix set is at the granularity of an IP subnet. So at the expense of more EID-prefix-to-RLOC sets for the site, the caches in each tunnel router can remain relatively small. But caches always depend on the number of non-aggregated EID destination flows active through these tunnel routers.

With more tunnel routers doing encapsulation, the increase in control traffic grows as well: since the EID-granularity is greater, more Map-Requests and replies are traveling between more routers.

The advantage of placing the caches and databases at these stub routers is that the products deployed in this part of the network have better price-memory ratios then their core router counterparts. Memory is typically less expensive in these devices and fewer routes are stored (only IGP routes). These devices tend to have excess capacity, both for forwarding and routing state.

LISP functionality can also be deployed in edge switches. These devices generally have layer-2 ports facing hosts and layer-3 ports facing the Internet. Spare capacity is also often available in these devices as well.

8.2. Border/Edge Tunnel Routers

Using customer-edge (CE) routers for tunnel endpoints allows the EID space associated with a site to be reachable via a small set of RLOCs assigned to the CE routers for that site.

This offers the opposite benefit of the first-hop/last-hop tunnel router scenario: the number of mapping entries and network management touch points are reduced, allowing better scaling.

One disadvantage is that less of the network's resources are used to reach host endpoints thereby centralizing the point-of-failure domain and creating network choke points at the CE router.

Note that more than one CE router at a site can be configured with the same IP address. In this case an RLOC is an anycast address. This allows resiliency between the CE routers. That is, if a CE router fails, traffic is automatically routed to the other routers using the same anycast address. However, this comes with the disadvantage where the site cannot control the entrance point when the anycast route is advertised out from all border routers.

8.3. ISP Provider-Edge (PE) Tunnel Routers

Use of ISP PE routers as tunnel endpoint routers gives an ISP control over the location of the egress tunnel endpoints. That is, the ISP can decide if the tunnel endpoints are in the destination site (in either CE routers or last-hop routers within a site) or at other PE edges. The advantage of this case is that two or more tunnel headers can be avoided. By having the PE be the first router on the path to encapsulate, it can choose a TE path first, and the ETR can decapsulate and re-encapsulate for a tunnel to the destination end site.

An obvious disadvantage is that the end site has no control over where its packets flow or the RLOCs used.

As mentioned in earlier sections a combination of these scenarios is possible at the expense of extra packet header overhead, if both site and provider want control, then recursive or re-encapsulating tunnels are used.

9. Multicast Considerations

A multicast group address, as defined in the original Internet architecture is an identifier of a grouping of topologically independent receiver host locations. The address encoding itself does not determine the location of the receiver(s). The multicast routing protocol, and the network-based state the protocol creates, determines where the receivers are located.

In the context of LISP, a multicast group address is both an EID and a Routing Locator. Therefore, no specific semantic or action needs to be taken for a destination address, as it would appear in an IP header. Therefore, a group address that appears in an inner IP header built by a source host will be used as the destination EID. And the outer IP header (the destination Routing Locator address), prepended by a LISP router, will use the same group address as the destination Routing Locator.

Having said that, only the source EID and source Routing Locator needs to be dealt with. Therefore, an ITR merely needs to put its own IP address in the source Routing Locator field when prepending the outer IP header. This source Routing Locator address, like any other Routing Locator address MUST be globally routable.

Therefore, an EID-to-RLOC mapping does not need to be performed by an ITR when a received data packet is a multicast data packet or when processing a source-specific Join (either by IGMPv3 or PIM). But the source Routing Locator is decided by the multicast routing protocol in a receiver site. That is, an EID to Routing Locator translation is done at control-time.

Another approach is to have the ITR not encapsulate a multicast packet and allow the the host built packet to flow into the core even if the source address is allocated out of the EID namespace. If the RPF-Vector TLV [RPFV] is used by PIM in the core, then core routers can RPF to the ITR (the Locator address which is injected into core routing) rather than the host source address (the EID address which is not injected into core routing).

10. Security Considerations

We believe that most of the security mechanisms will be part of the mapping database service when using control-plane procedures for obtaining EID-to-RLOC mappings. For data-plane triggered mappings, as described in this specification, protection is provided against ETR spoofing by using Return- Routeability mechanisms evidenced by the use of a 6-byte Nonce field in the LISP encapsulation header. The nonce, coupled with the ITR accepting only solicited Map-Replies goes a long way towards providing decent authentication.

LISP does not rely on a PKI infrastructure or a more heavy weight authentication system. These systems challenge the scalability of LISP which was a primary design goal.

DoS attack prevention will depend on implementations rate- limiting of Map-Requests and Map-Replies to the control-plane as well as rate-limiting the number of data triggered Map-Replies.

11. Prototype Plans and Status

The operator community has requested that the IETF take a practical approach to solving the scaling problems associated with global routing state growth. This document offers a simple solution which is intended for use in a pilot program to gain experience in working on this problem.

The authors hope that publishing this specification will allow the rapid implementation of multiple vendor prototypes and deployment on a small scale. Doing this will help the community:

- Decide whether a new EID-to-RLOC mapping database infrastructure is needed or if a simple, UDP-based, data-triggered approach is flexible and robust enough.

- Experiment with provider-independent assignment of EIDs while at the same time decreasing the size of DFZ routing tables through the use of topologically-aligned, provider-based RLOCs.

- Determine whether multiple levels of tunneling can be used by ISPs to achieve their Traffic Engineering goals while simultaneously removing the more specific routes currently injected into the global routing system for this purpose.

- Experiment with mobility to determine if both acceptable convergence and session survivability properties can be scalably implemented to support both individual device roaming and site service provider changes.

Here are a rough set of milestones:

1. Stabilize this draft by Summer 2007 Chicago IETF.

2. Start implementations to report on by Summer 2007 Chicago IETF.

3. Start pilot deployment between summer and fall IETFs. Report on deployment at Fall 2007 Vancouver IETF.

4. Achieve multi-vendor interoperability by Fall 2007 Vancouver IETF.

5. Consider prototyping other database lookup schemes, be it DNS, DHTs, CONS, NERD, or other mechanisms by Fall 2007 IETF.

As of this writing the following accomplishments have been achieved:

1. A unit tested software switching implementation has been completed for both IPv4 and IPv6 encapsulations for LISP 1 and LISP 1.5 functionality.

2. Dave Meyer and Vince Fuller are testing the implementation this summer.

3. An implementation of LISP-CONS is under way.

Please contact authors if interested in doing an implementation and want to interoperability test with our implementation.

12. References

12.1. Normative References

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

[RFC1498] Saltzer, J., "On the Naming and Binding of Network Destinations", RFC 1498, August 1993.

[RFC1955] Hinden, R., "New Scheme for Internet Routing and Addressing (ENCAPS) for IPNG", RFC 1955, June 1996.

[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.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, March 2000.

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[RFC4423] Moskowitz, R. and P. Nikander, "Host Identity Protocol (HIP) Architecture", RFC 4423, May 2006.

12.2. Informative References

[APT] Jen, D., Meisel, M., Massey, D., Wang, L., Zhang, B., and L. Zhang, "APT: A Practical Transit Mapping Service", draft-jen-apt-00.txt (work in progress), July 2007.

[CHIAPPA] Chiappa, J., "Endpoints and Endpoint names: A Proposed Enhancement to the Internet Architecture", Internet-Draft http://www.chiappa.net/~jnc/tech/endpoints.txt, 1999.

[CONS] Farinacci, D., Fuller, V., and D. Meyer, "LISP-CONS: A Content distribution Overlay Network Service for LISP", draft-meyer-lisp-cons-00.txt (work in progress), June 2007.

[DHTs] Ratnasamy, S., Shenker, S., and I. Stoica, "Routing Algorithms for DHTs: Some Open Questions", PDF

file http://www.cs.rice.edu/Conferences/IPTPS02/174.pdf.

[GSE] "GSE - An Alternate Addressing Architecture for IPv6", draft-ietf-ipngwg-gseaddr-00.txt (work in progress), 1997.

[LISP1] Farinacci, D., Oran, D., Fuller, V., and J. Schiller, "Locator/ID Separation Protocol (LISP1) [Routable ID Version]", Slide-set http://www.dinof.net/~dino/ietf/lisp1.ppt, October 2006.

[LISP2] Farinacci, D., Oran, D., Fuller, V., and J. Schiller, "Locator/ID Separation Protocol (LISP2) [DNS-based Version]", Slide-set http://www.dinof.net/~dino/ietf/lisp2.ppt, November 2006.

[NERD] Lear, E., "NERD: A Not-so-novel EID to RLOC Database", draft-lear-lisp-nerd-01.txt (work in progress), June 2007.

[RAWS] Meyer, D., Zhang, L., and K. Fall, "Report from the IAB Workshop on Routing and Addressing", draft-iab-raws-report-00.txt (work in progress), November 2006.

[RPFV] Wijnands, IJ., Boers, A., and E. Rosen, "The RPF Vector TLV", draft-ietf-pim-rpf-vector-03.txt (work in progress), October 2006.

[SHIM6] Nordmark, E. and M. Bagnulo, "Level 3 multihoming shim protocol", draft-ietf-shim6-proto-06.txt (work in progress), October 2006.

Appendix A. Acknowledgments

The authors would like to gratefully acknowledge many people who have contributed discussion and ideas to the making of this proposal. They include Jason Schiller, Lixia Zhang, Dorian Kim, Peter Schoenmaker, Darrel Lewis, Vijay Gill, Geoff Huston, David Conrad, Ron Bonica, Ted Seely, Mark Townsley, Chris Morrow, Brian Weis, Dave McGrew, Peter Lothberg, Dave Thaler, Scott Brim, Eliot Lear, and Shane Amante.

In particular, we would like to thank Dave Meyer for his clever suggestion for the name "LISP". ;-)

Authors' Addresses

Dino Farinacci
cisco Systems
Tasman Drive
San Jose, CA 95134
USA

Email: dino@cisco.com

Vince Fuller
cisco Systems
Tasman Drive
San Jose, CA 95134
USA

Email: vaf@cisco.com

Dave Oran
cisco Systems
7 Ladyslipper Lane
Acton, MA
USA

Email: oran@cisco.com

Dave Meyer
cisco Systems
170 Tasman Drive
San Jose, CA
USA

Email: dmm@cisco.com

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).