

Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: May 17, 2013

D. Farinacci  
V. Fuller  
D. Meyer  
D. Lewis  
cisco Systems  
November 13, 2012

Locator/ID Separation Protocol (LISP)  
draft-ietf-lisp-24

Abstract

This draft describes a network layer based protocol that enables separation of IP addresses into two new numbering spaces: Endpoint Identifiers (EIDs) and Routing Locators (RLOCs). No changes are required to either host protocol stacks or to the "core" of the Internet infrastructure. LISP can be incrementally deployed, without a "flag day", and offers traffic engineering, multi-homing, and mobility benefits to early adopters, even when there are relatively few LISP-capable sites.

Design and development of LISP was largely motivated by the problem statement produced by the October 2006 IAB Routing and Addressing Workshop.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on May 17, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Requirements Notation . . . . .	5
2. Introduction . . . . .	6
3. Definition of Terms . . . . .	8
4. Basic Overview . . . . .	14
4.1. Packet Flow Sequence . . . . .	16
5. LISP Encapsulation Details . . . . .	18
5.1. LISP IPv4-in-IPv4 Header Format . . . . .	19
5.2. LISP IPv6-in-IPv6 Header Format . . . . .	19
5.3. Tunnel Header Field Descriptions . . . . .	21
5.4. Dealing with Large Encapsulated Packets . . . . .	25
5.4.1. A Stateless Solution to MTU Handling . . . . .	25
5.4.2. A Stateful Solution to MTU Handling . . . . .	26
5.5. Using Virtualization and Segmentation with LISP . . . . .	26
6. EID-to-RLOC Mapping . . . . .	28
6.1. LISP IPv4 and IPv6 Control Plane Packet Formats . . . . .	28
6.1.1. LISP Packet Type Allocations . . . . .	30
6.1.2. Map-Request Message Format . . . . .	30
6.1.3. EID-to-RLOC UDP Map-Request Message . . . . .	33
6.1.4. Map-Reply Message Format . . . . .	34
6.1.5. EID-to-RLOC UDP Map-Reply Message . . . . .	38
6.1.6. Map-Register Message Format . . . . .	40
6.1.7. Map-Notify Message Format . . . . .	42
6.1.8. Encapsulated Control Message Format . . . . .	43
6.2. Routing Locator Selection . . . . .	45
6.3. Routing Locator Reachability . . . . .	47
6.3.1. Echo Nonce Algorithm . . . . .	49
6.3.2. RLOC Probing Algorithm . . . . .	50
6.4. EID Reachability within a LISP Site . . . . .	51
6.5. Routing Locator Hashing . . . . .	52
6.6. Changing the Contents of EID-to-RLOC Mappings . . . . .	53
6.6.1. Clock Sweep . . . . .	54
6.6.2. Solicit-Map-Request (SMR) . . . . .	54
6.6.3. Database Map Versioning . . . . .	56
7. Router Performance Considerations . . . . .	57
8. Deployment Scenarios . . . . .	58

8.1.	First-hop/Last-hop Tunnel Routers . . . . .	59
8.2.	Border/Edge Tunnel Routers . . . . .	59
8.3.	ISP Provider-Edge (PE) Tunnel Routers . . . . .	60
8.4.	LISP Functionality with Conventional NATs . . . . .	60
8.5.	Packets Egressing a LISP Site . . . . .	61
9.	Traceroute Considerations . . . . .	62
9.1.	IPv6 Traceroute . . . . .	63
9.2.	IPv4 Traceroute . . . . .	63
9.3.	Traceroute using Mixed Locators . . . . .	63
10.	Mobility Considerations . . . . .	65
10.1.	Site Mobility . . . . .	65
10.2.	Slow Endpoint Mobility . . . . .	65
10.3.	Fast Endpoint Mobility . . . . .	65
10.4.	Fast Network Mobility . . . . .	67
10.5.	LISP Mobile Node Mobility . . . . .	67
11.	Multicast Considerations . . . . .	69
12.	Security Considerations . . . . .	70
13.	Network Management Considerations . . . . .	72
14.	IANA Considerations . . . . .	73
14.1.	LISP ACT and Flag Fields . . . . .	73
14.2.	LISP Address Type Codes . . . . .	73
14.3.	LISP UDP Port Numbers . . . . .	74
14.4.	LISP Key ID Numbers . . . . .	74
15.	Known Open Issues and Areas of Future Work . . . . .	75
16.	References . . . . .	77
16.1.	Normative References . . . . .	77
16.2.	Informative References . . . . .	78
Appendix A.	Acknowledgments . . . . .	82
Appendix B.	Document Change Log . . . . .	83
B.1.	Changes to draft-ietf-lisp-24.txt . . . . .	83
B.2.	Changes to draft-ietf-lisp-23.txt . . . . .	83
B.3.	Changes to draft-ietf-lisp-22.txt . . . . .	83
B.4.	Changes to draft-ietf-lisp-21.txt . . . . .	83
B.5.	Changes to draft-ietf-lisp-20.txt . . . . .	83
B.6.	Changes to draft-ietf-lisp-19.txt . . . . .	83
B.7.	Changes to draft-ietf-lisp-18.txt . . . . .	83
B.8.	Changes to draft-ietf-lisp-17.txt . . . . .	84
B.9.	Changes to draft-ietf-lisp-16.txt . . . . .	84
B.10.	Changes to draft-ietf-lisp-15.txt . . . . .	84
B.11.	Changes to draft-ietf-lisp-14.txt . . . . .	84
B.12.	Changes to draft-ietf-lisp-13.txt . . . . .	85
B.13.	Changes to draft-ietf-lisp-12.txt . . . . .	85
B.14.	Changes to draft-ietf-lisp-11.txt . . . . .	87
B.15.	Changes to draft-ietf-lisp-10.txt . . . . .	88
B.16.	Changes to draft-ietf-lisp-09.txt . . . . .	88
B.17.	Changes to draft-ietf-lisp-08.txt . . . . .	88
B.18.	Changes to draft-ietf-lisp-07.txt . . . . .	90
B.19.	Changes to draft-ietf-lisp-06.txt . . . . .	92

B.20. Changes to draft-ietf-lisp-05.txt	93
B.21. Changes to draft-ietf-lisp-04.txt	93
B.22. Changes to draft-ietf-lisp-03.txt	95
B.23. Changes to draft-ietf-lisp-02.txt	95
B.24. Changes to draft-ietf-lisp-01.txt	96
B.25. Changes to draft-ietf-lisp-00.txt	96
Authors' Addresses	97

## 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

This document describes the Locator/Identifier Separation Protocol (LISP), which provides a set of functions for routers to exchange information used to map from non globally routeable Endpoint Identifiers (EIDs) to routeable Routing Locators (RLOCs). It also defines a mechanism for these LISP routers to encapsulate IP packets addressed with EIDs for transmission across the Internet that uses RLOCs for routing and forwarding.

Creation of LISP was initially motivated by discussions during the IAB-sponsored Routing and Addressing Workshop held in Amsterdam in October, 2006 (see [RFC4984]). A key conclusion of the workshop was that the Internet routing and addressing system was not scaling well in the face of the explosive growth of new sites; one reason for this poor scaling is the increasing number of multi-homed and other sites that cannot be addressed as part of topologically- or provider-based aggregated prefixes. Additional work that more completely described the problem statement may be found in [RADIR].

A basic observation, made many years ago in early networking research such as that documented in [CHIAPPA] and [RFC4984], is that using a single address field for both identifying a device and for determining where it is topologically located in the network requires optimization along two conflicting axes: for routing to be efficient, the address must be assigned topologically; for collections of devices to be easily and effectively managed, without the need for renumbering in response to topological change (such as that caused by adding or removing attachment points to the network or by mobility events), the address must explicitly not be tied to the topology.

The approach that LISP takes to solving the routing scalability problem is to replace IP addresses with two new types of numbers: Routing Locators (RLOCs), which are topologically assigned to network attachment points (and are therefore amenable to aggregation) and used for routing and forwarding of packets through the network; and Endpoint Identifiers (EIDs), which are assigned independently from the network topology, are used for numbering devices, and are aggregated along administrative boundaries. LISP then defines functions for mapping between the two numbering spaces and for encapsulating traffic originated by devices using non-routeable EIDs for transport across a network infrastructure that routes and forwards using RLOCs. Both RLOCs and EIDs are syntactically-identical to IP addresses; it is the semantics of how they are used that differs.

This document describes the protocol that implements these functions. The database which stores the mappings between EIDs and RLOCs is

explicitly a separate "module" to facilitate experimentation with a variety of approaches. One database design that is being developed for experimentation as part of the LISP working group work is [ALT]. Others that have been described include [CONS], [EMACS], [NERD]. Finally, [LISP-MS], documents a general-purpose service interface for accessing a mapping database; this interface is intended to make the mapping database modular so that different approaches can be tried without the need to modify installed LISP capable devices in LISP sites.

This experimental specification has areas that require additional experience and measurement. It is NOT RECOMMENDED for deployment beyond experimental situations. Results of experimentation may lead to modifications and enhancements of protocol mechanisms defined in this document. See Section 15 for specific, known issues that are in need of further work during development, implementation, and experimentation.

An examination of the implications of LISP on Internet traffic, applications, routers, and security is for future study. This analysis will explain what role LISP can play in scalable routing and will also look at scalability and levels of state required for encapsulation, decapsulation, liveness, and so on.

### 3. Definition of Terms

**Provider Independent (PI) Addresses:** PI addresses are an address block assigned from a pool where blocks are not associated with any particular location in the network (e.g. from a particular service provider), and is therefore not topologically aggregatable in the routing system.

**Provider Assigned (PA) Addresses:** PA addresses are an address block assigned to a site by each service provider to which a site connects. Typically, each block is sub-block of a service provider Classless Inter-Domain Routing (CIDR) [RFC4632] 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):** A RLOC is an IPv4 [RFC0791] or IPv6 [RFC2460] address of an egress tunnel router (ETR). A RLOC is the output of an 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. Multiple RLOCs can be assigned to the same ETR device or to multiple ETR devices at a site.

**Endpoint ID (EID):** An EID is a 32-bit (for IPv4) or 128-bit (for IPv6) 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 Domain Name System (DNS) [RFC1034] lookup or Session Invitation Protocol (SIP) [RFC3261] exchange. The source EID is obtained via existing mechanisms used to set a host's "local" IP address. An EID used on the public Internet must have the same properties as any other IP address used in that manner; this means, among other things, that it must be globally unique. An EID is allocated to a host from an EID-prefix block associated with the site where the host is located. An EID can be used by a host to refer to other hosts. 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. In theory, the bit string



that represents an EID for one device can represent an RLOC for a different device. As the architecture is realized, if a given bit string is both an RLOC and an EID, it must refer to the same entity in both cases. When used in discussions with other Locator/ID separation proposals, a LISP EID will be called a "LEID". Throughout this document, any references to "EID" refers to an LEID.

**EID-prefix:** An EID-prefix is a power-of-two 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 larger EID-prefix block. A globally routed address block (whether PI or PA) is not inherently an EID-prefix. A globally routed address block MAY be used by its assignee as an EID block. The converse is not supported. That is, a site which receives an explicitly allocated EID-prefix may not use that EID-prefix as a globally routed prefix. This would require coordination and cooperation with the entities managing the mapping infrastructure. Once this has been done, that block could be removed from the globally routed IP system, if other suitable transition and access mechanisms are in place. Discussion of such transition and access mechanisms can be found in [INTERWORK] and [LISP-DEPLOY].

**End-system:** An 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 its routing domain). An end-system can be a host computer, a switch or router device, or any network appliance.

**Ingress Tunnel Router (ITR):** An ITR is a router that resides in a LISP site. Packets sent by sources inside of the LISP site to destinations outside of the site are candidates for encapsulation by the ITR. The ITR treats the 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 closer 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:** A 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):** An ETR is a router that accepts an IP packet where the 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. ETR functionality does not have to be limited to a router device. A server host can be the endpoint of a LISP tunnel as well.

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

**xTR:** A xTR is a reference to an ITR or ETR when direction of data flow is not part of the context description. xTR refers to the router that is the tunnel endpoint. Used synonymously with the term "Tunnel Router". For example, "An xTR can be located at the Customer Edge (CE) router", meaning both ITR and ETR functionality is at the CE router.

**LISP Router:** A LISP router is a router that performs the functions of any or all of ITR, ETR, PITR, or PETR.

**EID-to-RLOC Cache:** The EID-to-RLOC cache is a short-lived, on-demand table 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 full "database" of EID-to-RLOC mappings, it is dynamic, local to the ITR(s), and relatively small while the database is distributed, relatively static, and much more global in scope.

**EID-to-RLOC Database:** The EID-to-RLOC database is a global 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. The same database mapping entries MUST be configured on all ETRs for a given site. In a steady state the EID-prefixes for the site and the locator-set for each EID-prefix MUST be the same on all ETRs. Procedures to enforce and/or verify this are outside the scope of this document. Note that there MAY be transient conditions when the EID-prefix for the site and locator-set for each EID-prefix may not be the same on all ETRs. This has no negative implications since a partial set of locators can be used.

**Recursive Tunneling:** Recursive tunneling occurs 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. Any references to tunnels in this specification refers to dynamic encapsulating tunnels and they are never statically configured.

**Reencapsulating Tunnels:** Reencapsulating tunneling occurs when an ETR removes a LISP header, then acts as an ITR to prepend another LISP header. Doing this allows a packet to be re-routed by the re-encapsulating router without adding the overhead of additional tunnel headers. Any references to tunnels in this specification refers to dynamic encapsulating tunnels and they are never statically configured. When using multiple mapping database systems, care must be taken to not create reencapsulation loops through misconfiguration.

**LISP Header:** a term used in this document to refer to the outer IPv4 or IPv6 header, a UDP header, and a LISP-specific 8-octet header that follows the UDP header, an ITR prepends or an ETR strips.

**Address Family Identifier (AFI):** a term used to describe an address encoding in a packet. An address family currently pertains to an IPv4 or IPv6 address. See [AFI]/[AFI-REGISTRY] and [RFC3232] for details. An AFI value of 0 used in this specification indicates an unspecified encoded address where the length of the address is 0 octets following the 16-bit AFI value of 0.

**Negative Mapping Entry:** A negative mapping entry, also known as a negative cache entry, is an EID-to-RLOC entry where an EID-prefix is advertised or stored with no RLOCs. That is, the locator-set for the EID-to-RLOC entry is empty or has an encoded locator count of 0. This type of entry could be used to describe a prefix from a non-LISP site, which is explicitly not in the mapping database. There are a set of well defined actions that are encoded in a

Negative Map-Reply (Section 6.1.5).

**Data Probe:** A data-probe is a LISP-encapsulated data packet where the inner header destination address equals the outer header destination address used to trigger a Map-Reply by a decapsulating ETR. In addition, the original packet is decapsulated and delivered to the destination host if the destination EID is in the EID-prefix range configured on the ETR. Otherwise, the packet is discarded. A Data Probe is used in some of the mapping database designs to "probe" or request a Map-Reply from an ETR; in other cases, Map-Requests are used. See each mapping database design for details. When using Data Probes, by sending Map-Requests on the underlying routing system, EID-prefixes must be advertised. However, this is discouraged if the core is to scale by having less EID-prefixes stored in the core router's routing tables.

**Proxy ITR (PITR):** A PITR is defined and described in [INTERWORK], a PITR acts like an ITR but does so on behalf of non-LISP sites which send packets to destinations at LISP sites.

**Proxy ETR (PETR):** A PETR is defined and described in [INTERWORK], a PETR acts like an ETR but does so on behalf of LISP sites which send packets to destinations at non-LISP sites.

**Route-returnability:** is an assumption that the underlying routing system will deliver packets to the destination. When combined with a nonce that is provided by a sender and returned by a receiver, this limits off-path data insertion. A route-returnability check is verified when a message is sent with a nonce, another message is returned with the same nonce, and the destination of the original message appears as the source of the returned message.

**LISP site:** is a set of routers in an edge network that are under a single technical administration. LISP routers which reside in the edge network are the demarcation points to separate the edge network from the core network.

**Client-side:** a term used in this document to indicate a connection initiation attempt by an EID. The ITR(s) at the LISP site are the first to get involved in obtaining database map cache entries by sending Map-Request messages.

**Server-side:** a term used in this document to indicate a connection initiation attempt is being accepted for a destination EID. The ETR(s) at the destination LISP site are the first to send Map-Replies to the source site initiating the connection. The ETR(s) at this destination site can obtain mappings by gleaning

information from Map-Requests, Data-Probes, or encapsulated packets.

**Locator Status Bits (LSBs):** Locator status bits are present in the LISP header. They are used by ITRs to inform ETRs about the up/down status of all ETRs at the local site. These bits are used as a hint to convey up/down router status and not path reachability status. The LSBs can be verified by use of one of the Locator Reachability Algorithms described in Section 6.3.

**Anycast Address:** a term used in this document to refer to the same IPv4 or IPv6 address configured and used on multiple systems at the same time. An EID or RLOC can be an anycast address in each of their own address spaces.

#### 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. When a packet is LISP encapsulated, 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 the destination addresses. For routers between the source host and the ITR as well as routers from the ETR to the destination host, the destination address is an EID. For the routers between the ITR and the ETR, the destination address is an RLOC.

Another key LISP concept is the "Tunnel Router". A tunnel router prepends LISP headers on host-originated packets and strips 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 ETR, which has the RLOC as one of its IP addresses.

Some basic rules governing LISP are:

- o End-systems (hosts) only send to addresses which are EIDs. They don't know addresses are EIDs versus RLOCs but assume packets get to their intended destinations. In a system where LISP is deployed, LISP routers intercept EID addressed packets and assist in delivering them across the network core where EIDs cannot be routed. The procedure a host uses to send IP packets does not change.
- o EIDs are always IP addresses assigned to hosts.
- o LISP 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 (Classless Inter-Domain Routing) blocks.
- o When a router originates packets it may use as a source address either an EID or RLOC. When acting as a host (e.g. when terminating a transport session such as SSH, TELNET, or SNMP), it

may use an EID that is explicitly assigned for that purpose. An EID that identifies the router as a host MUST NOT be used as an RLOC; an EID is only routable within the scope of a site. A typical BGP configuration might demonstrate this "hybrid" EID/RLOC usage where a router could use its "host-like" EID to terminate iBGP sessions to other routers in a site while at the same time using RLOCs to terminate eBGP sessions to routers outside the site.

- o Packets with EIDs in them are not expected to be delivered end-to-end in the absence of an EID-to-RLOC mapping operation. They are expected to be used locally for intra-site communication or to be encapsulated for inter-site communication.
- 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 allocation hierarchy which is independent of the network topology.
- o EIDs may also be structured (subnetted) in a manner suitable for local routing within an autonomous system.

An additional LISP header MAY be prepended to packets by a TE-ITR when re-routing of the path for a packet is desired. A potential use-case for this would be an ISP router that needs to perform traffic engineering for packets flowing 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 a TE-ETR within the ISP (along intra-ISP traffic engineered path) or a TE-ETR within another ISP (an inter-ISP traffic engineered path, where an agreement to build such a path exists).

In order to avoid excessive packet overhead as well as possible encapsulation loops, this document mandates that a maximum of two LISP headers can be prepended to a packet. For initial LISP deployments, it is assumed two headers is sufficient, where the first prepended header is used at a site for Location/Identity separation and second prepended header is used inside a service provider for Traffic Engineering purposes.

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 packet flow with the following conditions:

- o Source host "host1.abc.example.com" is sending a packet to "host2.xyz.example.com", exactly what host1 would do if the site was not using LISP.
- o Each site is multi-homed, so each tunnel router has an address (RLOC) assigned from the service provider address block for each provider to which that particular tunnel router is attached.
- o The ITR(s) and ETR(s) are directly connected to the source and destination, respectively, but the source and destination can be located anywhere in LISP site.
- o Map-Requests can be sent on the underlying routing system topology, to a mapping database system, or directly over an alternative topology [ALT]. A Map-Request is sent for an external destination when the destination is not found in the forwarding table or matches a default route.
- o Map-Replies are sent on the underlying routing system topology.

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

1. host1.abc.example.com wants to open a TCP connection to host2.xyz.example.com. It does a DNS lookup on host2.xyz.example.com. An A/AAAA record is returned. This address is the destination EID. The locally-assigned address of host1.abc.example.com is used as the source EID. An IPv4 or IPv6 packet is built and forwarded through the LISP site as a normal IP packet until it reaches a LISP ITR.
2. The LISP ITR must be able to map the destination EID to an RLOC of one of the ETRs at the destination site. The specific method used to do this is not described in this example. See [ALT] or [CONS] for possible solutions.
3. The ITR will send a LISP Map-Request. Map-Requests SHOULD be rate-limited.



4. When an alternate mapping system is not in use, the Map-Request packet is routed through the underlying routing system. Otherwise, the Map-Request packet is routed on an alternate logical topology, for example the [ALT] database mapping system. In either case, when the Map-Request arrives at one of the ETRs at the destination site, it will process the packet as a control message.
5. The ETR looks at the destination EID of the Map-Request and matches it against the prefixes in the ETR's configured EID-to-RLOC mapping database. This is the list of EID-prefixes the ETR is supporting for the site it resides in. If there is no match, the Map-Request is dropped. Otherwise, a LISP Map-Reply is returned to the ITR.
6. The ITR receives the Map-Reply message, parses the message (to check for format validity) and stores the mapping information from the packet. This information is stored in the ITR's EID-to-RLOC mapping cache. Note that the map cache is an on-demand cache. An ITR will manage its map cache in such a way that optimizes for its resource constraints.
7. Subsequent packets from host1.abc.example.com to host2.xyz.example.com will have a LISP header prepended by the ITR using the appropriate RLOC as the LISP header destination address learned from the ETR. Note the packet MAY be sent to a different ETR than the one which returned the Map-Reply due to the source site's hashing policy or the destination site's locator-set policy.
8. The ETR receives these packets directly (since the destination address is one of its assigned IP addresses), checks the validity of the addresses, strips the LISP header, and forwards packets to the attached destination host.

In order to defer the need for a mapping lookup in the reverse direction, an ETR MAY create a cache entry that maps the source EID (inner header source IP address) to the source RLOC (outer header source IP address) in a received LISP packet. Such a cache entry is termed a "gleaned" mapping and only contains a single RLOC for the EID in question. More complete information about additional RLOCs SHOULD be verified by sending a LISP Map-Request for that EID. Both ITR and the ETR may also influence the decision the other makes in selecting an RLOC. See Section 6 for more details.

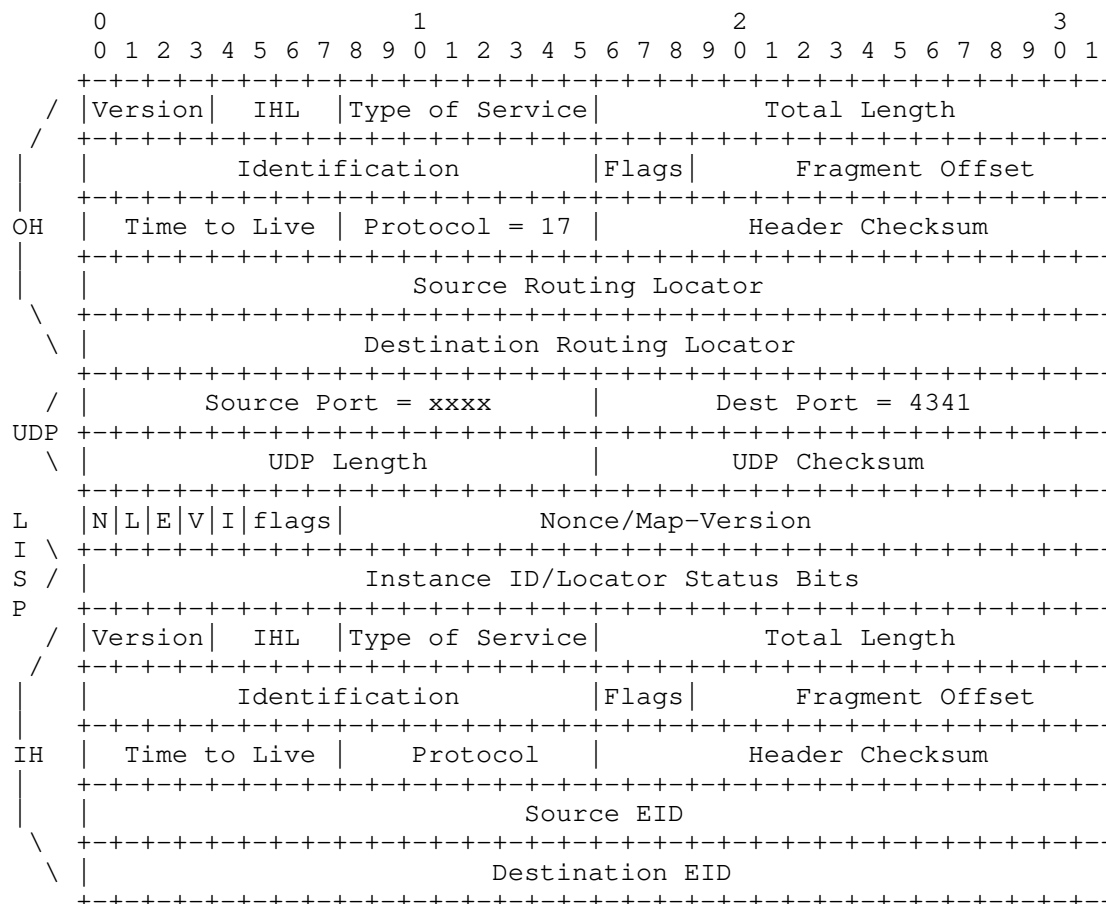
## 5. LISP Encapsulation Details

Since additional tunnel headers are prepended, the packet becomes larger and can exceed the MTU of any link traversed from the ITR to the ETR. It is RECOMMENDED in IPv4 that packets do not get fragmented as they are encapsulated by the ITR. Instead, the packet is dropped and an ICMP Too Big message is returned to the source.

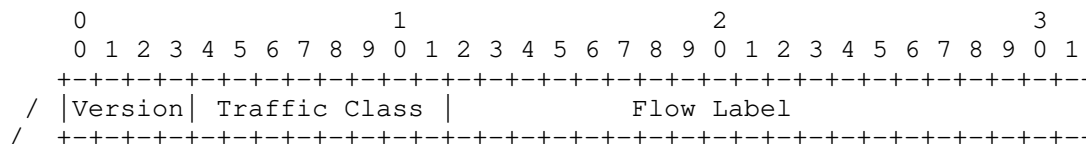
This specification RECOMMENDS that implementations provide support for one of the proposed fragmentation and reassembly schemes. Two existing schemes are detailed in Section 5.4.

Since IPv4 or IPv6 addresses can be either EIDs or RLOCs, the LISP architecture supports IPv4 EIDs with IPv6 RLOCs (where the inner header is in IPv4 packet format and the other header is in IPv6 packet format) or IPv6 EIDs with IPv4 RLOCs (where the inner header is in IPv6 packet format and the other header is in IPv4 packet format). The next sub-sections illustrate packet formats for the homogeneous case (IPv4-in-IPv4 and IPv6-in-IPv6) but all 4 combinations MUST be supported.

## 5.1. LISP IPv4-in-IPv4 Header Format



## 5.2. LISP IPv6-in-IPv6 Header Format





### 5.3. Tunnel Header Field Descriptions

Inner Header (IH): The inner header is the header on the datagram received from the originating host. The source and destination IP addresses are EIDs, [RFC0791], [RFC2460].

Outer Header: (OH) The outer header is a new 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]. The setting of the DF bit Flags field is according to rules in Section 5.4.1 and Section 5.4.2.

UDP Header: The UDP header contains an ITR selected source port when encapsulating a packet. See Section 6.5 for details on the hash algorithm used to select a source port based on the 5-tuple of the inner header. The destination port MUST be set to the well-known IANA assigned port value 4341.

UDP Checksum: The UDP checksum field SHOULD be transmitted as zero by an ITR for either IPv4 [RFC0768] or IPv6 encapsulation [UDP-TUNNELS] [UDP-ZERO]. When a packet with a zero UDP checksum is received by an ETR, the ETR MUST accept the packet for decapsulation. When an ITR transmits a non-zero value for the UDP checksum, it MUST send a correctly computed value in this field. When an ETR receives a packet with a non-zero UDP checksum, it MAY choose to verify the checksum value. If it chooses to perform such verification, and the verification fails, the packet MUST be silently dropped. If the ETR chooses not to perform the verification, or performs the verification successfully, the packet MUST be accepted for decapsulation. The handling of UDP checksums for all tunneling protocols, including LISP, is under active discussion within the IETF. When that discussion concludes, any necessary changes will be made to align LISP with the outcome of the broader discussion.

UDP Length: The UDP length field is set for an IPv4 encapsulated packet to be the sum of the inner header IPv4 Total Length plus the UDP and LISP header lengths. For an IPv6 encapsulated packet, the UDP length field is the sum of the inner header IPv6 Payload Length, the size of the IPv6 header (40 octets), and the size of the UDP and LISP headers.

N: The N bit is the nonce-present bit. When this bit is set to 1, the low-order 24-bits of the first 32-bits of the LISP header contains a Nonce. See Section 6.3.1 for details. Both N and V bits MUST NOT be set in the same packet. If they are, a decapsulating ETR MUST treat the "Nonce/Map-Version" field as having a Nonce value present.

L: The L bit is the Locator Status Bits field enabled bit. When this bit is set to 1, the Locator Status Bits in the second 32-bits of the LISP header are in use.

```

  x 1 x x 0 x x x
+-----+
|N|L|E|V|I|flags|      Nonce/Map-Version      |
+-----+
|                                     Locator Status Bits                                     |
+-----+

```

E: The E bit is the echo-nonce-request bit. This bit MUST be ignored and has no meaning when the N bit is set to 0. When the N bit is set to 1 and this bit is set to 1, means an ITR is requesting for the nonce value in the Nonce field to be echoed back in LISP encapsulated packets when the ITR is also an ETR. See Section 6.3.1 for details.

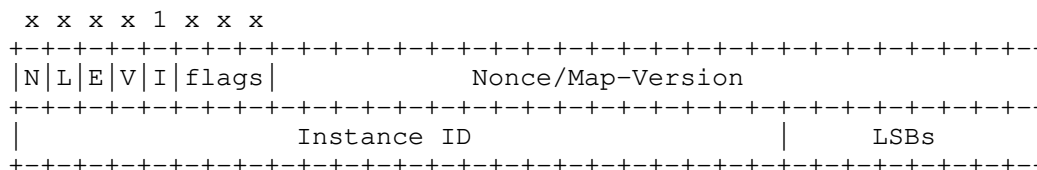
V: The V bit is the Map-Version present bit. When this bit is set to 1, the N bit MUST be 0. Refer to Section 6.6.3 for more details. This bit indicates that the LISP header is encoded in this case as:

```

  0 x 0 1 x x x x
+-----+
|N|L|E|V|I|flags|  Source Map-Version  |  Dest Map-Version  |
+-----+
|                                     Instance ID/Locator Status Bits                                     |
+-----+

```

I: The I bit is the Instance ID bit. See Section 5.5 for more details. When this bit is set to 1, the Locator Status Bits field is reduced to 8-bits and the high-order 24-bits are used as an Instance ID. If the L-bit is set to 0, then the low-order 8 bits are transmitted as zero and ignored on receipt. The format of the LISP header would look like in this case:



flags: The flags field is a 3-bit field is reserved for future flag use. It MUST be set to 0 on transmit and MUST be ignored on receipt.

LISP Nonce: The LISP nonce field is a 24-bit value that is randomly generated by an ITR when the N-bit is set to 1. Nonce generation algorithms are an implementation matter but are required to generate different nonces when sending to different destinations. However, the same nonce can be used for a period of time to the same destination. The nonce is also used when the E-bit is set to request the nonce value to be echoed by the other side when packets are returned. When the E-bit is clear but the N-bit is set, a remote ITR is either echoing a previously requested echo-nonce or providing a random nonce. See Section 6.3.1 for more details.

LISP Locator Status Bits (LSBs): When the L-bit is also set, the locator status bits field in the LISP header is set by an ITR to indicate to an ETR the up/down status 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 Status Bits are numbered from 0 to n-1 from the least significant bit of field. The field is 32-bits when the I-bit is set to 0 and is 8 bits when the I-bit is set to 1. When a Locator Status Bit is set to 1, the ITR is indicating to the ETR the RLOC associated with the bit ordinal has up status. See Section 6.3 for details on how an ITR can determine the status of the ETRs at the same site. When a site has multiple EID-prefixes which result in multiple mappings (where each could have a different locator-set), the Locator Status Bits setting in an encapsulated packet MUST reflect the mapping for the EID-prefix that the inner-header source EID address matches. If the LSB for an anycast locator is set to 1, then there is at least one RLOC with that address the ETR is considered 'up'.

When doing ITR/PITR encapsulation:

- o The outer header Time to Live field (or Hop Limit field, in case of IPv6) SHOULD be copied from the inner header Time to Live field.

- o The outer header Type of Service field (or the Traffic Class field, in the case of IPv6) SHOULD be copied from the inner header Type of Service field (with one exception, see below).

When doing ETR/PETR decapsulation:

- o The inner header Time to Live field (or Hop Limit field, in case of IPv6) SHOULD be copied from the outer header Time to Live field, when the Time to Live field of the outer header is less than the Time to Live of the inner header. Failing to perform this check can cause the Time to Live of the inner header to increment across encapsulation/decapsulation cycle. This check is also performed when doing initial encapsulation when a packet comes to an ITR or PITR destined for a LISP site.
- o The inner header Type of Service field (or the Traffic Class field, in the case of IPv6) SHOULD be copied from the outer header Type of Service field (with one exception, see below).

Note if an ETR/PETR is also an ITR/PITR and choose to reencapsulate after decapsulating, the net effect of this is that the new outer header will carry the same Time to Live as the old outer header minus 1.

Copying the TTL serves two purposes: first, it preserves the distance the host intended the packet to travel; second, and more importantly, it provides for suppression of looping packets in the event there is a loop of concatenated tunnels due to misconfiguration. See Section 9.3 for TTL exception handling for traceroute packets.

The ECN field occupies bits 6 and 7 of both the IPv4 Type of Service field and the IPv6 Traffic Class field [RFC3168]. The ECN field requires special treatment in order to avoid discarding indications of congestion [RFC3168]. ITR encapsulation MUST copy the 2-bit ECN field from the inner header to the outer header. Re-encapsulation MUST copy the 2-bit ECN field from the stripped outer header to the new outer header. If the ECN field contains a congestion indication codepoint (the value is '11', the Congestion Experienced (CE) codepoint), then ETR decapsulation MUST copy the 2-bit ECN field from the stripped outer header to the surviving inner header that is used to forward the packet beyond the ETR. These requirements preserve Congestion Experienced (CE) indications when a packet that uses ECN traverses a LISP tunnel and becomes marked with a CE indication due to congestion between the tunnel endpoints.



#### 5.4. Dealing with Large Encapsulated Packets

This section proposes two mechanisms to deal with packets that exceed the path MTU between the ITR and ETR.

It is left to the implementor to decide if the stateless or stateful mechanism should be implemented. Both or neither can be used since it is a local decision in the ITR regarding how to deal with MTU issues, and sites can interoperate with differing mechanisms.

Both stateless and stateful mechanisms also apply to Reencapsulating and Recursive Tunneling. So any actions below referring to an ITR also apply to an TE-ITR.

##### 5.4.1. A Stateless Solution to MTU Handling

An ITR stateless solution to handle MTU issues is described as follows:

1. Define H to be the size, in octets, of the outer header an ITR prepends to a packet. This includes the UDP and LISP header lengths.
2. Define L to be the size, in octets, of the maximum sized packet an ITR can send to an ETR without the need for the ITR or any intermediate routers to fragment the packet.
3. Define an architectural constant S for the maximum size of a packet, in octets, an ITR must receive so the effective MTU can be met. That is,  $S = L - H$ .

When an ITR receives a packet from a site-facing interface and adds H octets worth of encapsulation to yield a packet size greater than L octets, it resolves the MTU issue by first splitting the original packet into 2 equal-sized fragments. A LISP header is then prepended to each fragment. The size of the encapsulated fragments is then  $(S/2 + H)$ , which is less than the ITR's estimate of the path MTU between the ITR and its correspondent ETR.

When an ETR receives encapsulated fragments, it treats them as two individually encapsulated packets. It strips the LISP headers then forwards each fragment to the destination host of the destination site. The two fragments are reassembled at the destination host into the single IP datagram that was originated by the source host. Note that reassembly can happen at the ETR if the encapsulated packet was fragmented at or after the ITR.

This behavior is performed by the ITR when the source host originates

a packet with the DF field of the IP header is set to 0. When the DF field of the IP header is set to 1, or the packet is an IPv6 packet originated by the source host, the ITR will drop the packet when the size is greater than L, and sends an ICMP Too Big message to the source with a value of S, where S is  $(L - H)$ .

When the outer header encapsulation uses an IPv4 header, an implementation SHOULD set the DF bit to 1 so ETR fragment reassembly can be avoided. An implementation MAY set the DF bit in such headers to 0 if it has good reason to believe there are unresolvable path MTU issues between the sending ITR and the receiving ETR.

This specification RECOMMENDS that L be defined as 1500.

#### 5.4.2. A Stateful Solution to MTU Handling

An ITR stateful solution to handle MTU issues is described as follows and was first introduced in [OPENLISP]:

1. The ITR will keep state of the effective MTU for each locator per mapping cache entry. The effective MTU is what the core network can deliver along the path between ITR and ETR.
2. When an IPv6 encapsulated packet or an IPv4 encapsulated packet with DF bit set to 1, exceeds what the core network can deliver, one of the intermediate routers on the path will send an ICMP Too Big message to the ITR. The ITR will parse the ICMP message to determine which locator is affected by the effective MTU change and then record the new effective MTU value in the mapping cache entry.
3. When a packet is received by the ITR from a source inside of the site and the size of the packet is greater than the effective MTU stored with the mapping cache entry associated with the destination EID the packet is for, the ITR will send an ICMP Too Big message back to the source. The packet size advertised by the ITR in the ICMP Too Big message is the effective MTU minus the LISP encapsulation length.

Even though this mechanism is stateful, it has advantages over the stateless IP fragmentation mechanism, by not involving the destination host with reassembly of ITR fragmented packets.

#### 5.5. Using Virtualization and Segmentation with LISP

When multiple organizations inside of a LISP site are using private addresses [RFC1918] as EID-prefixes, their address spaces MUST remain segregated due to possible address duplication. An Instance ID in

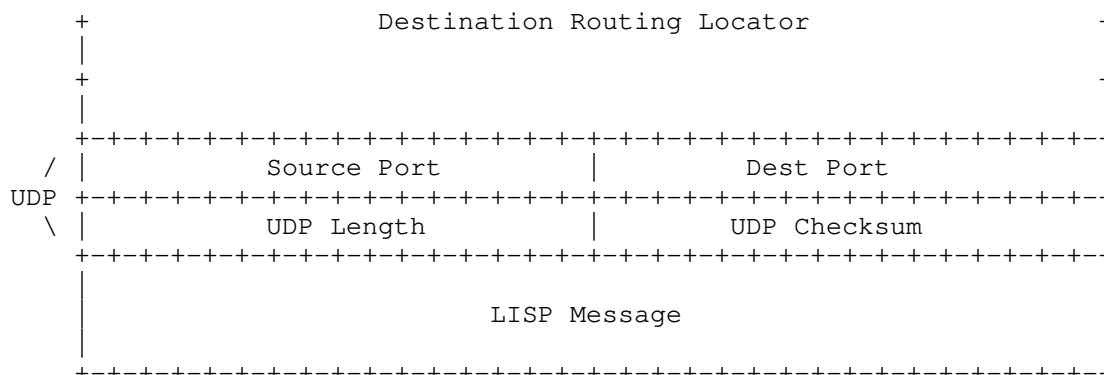
the address encoding can aid in making the entire AFI based address unique. See IANA Considerations Section 14.2 for details for possible address encodings.

An Instance ID can be carried in a LISP encapsulated packet. An ITR that prepends a LISP header, will copy a 24-bit value, used by the LISP router to uniquely identify the address space. The value is copied to the Instance ID field of the LISP header and the I-bit is set to 1.

When an ETR decapsulates a packet, the Instance ID from the LISP header is used as a table identifier to locate the forwarding table to use for the inner destination EID lookup.

For example, a 802.1Q VLAN tag or VPN identifier could be used as a 24-bit Instance ID.





The LISP UDP-based messages are the Map-Request and Map-Reply messages. When a UDP Map-Request is sent, the UDP source port is chosen by the sender and the destination UDP port number is set to 4342. When a UDP Map-Reply is sent, the source UDP port number is set to 4342 and the destination UDP port number is copied from the source port of either the Map-Request or the invoking data packet. Implementations MUST be prepared to accept packets when either the source port or destination UDP port is set to 4342 due to NATs changing port number values.

The UDP Length field will reflect the length of the UDP header and the LISP Message payload.

The UDP Checksum is computed and set to non-zero for Map-Request, Map-Reply, Map-Register and ECM control messages. It MUST be checked on receipt and if the checksum fails, the packet MUST be dropped.

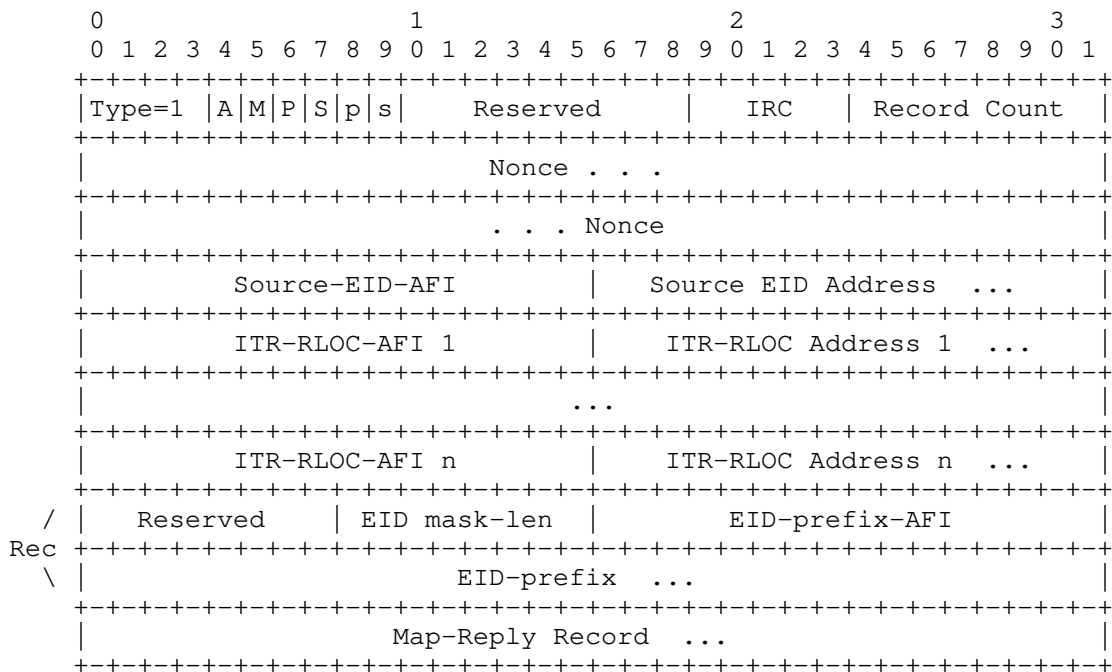
The format of control messages includes the UDP header so the checksum and length fields can be used to protect and delimit message boundaries.

### 6.1.1. LISP Packet Type Allocations

This section will be the authoritative source for allocating LISP Type values and for defining LISP control message formats. Current allocations are:

Reserved:	0	b'0000'
LISP Map-Request:	1	b'0001'
LISP Map-Reply:	2	b'0010'
LISP Map-Register:	3	b'0011'
LISP Map-Notify:	4	b'0100'
LISP Encapsulated Control Message:	8	b'1000'

### 6.1.2. Map-Request Message Format



Packet field descriptions:

Type: 1 (Map-Request)

A: This is an authoritative bit, which is set to 0 for UDP-based Map-Requests sent by an ITR. Set to 1 when an ITR wants the destination site to return the Map-Reply rather than the mapping database system.

M: This is the map-data-present bit, when set, it indicates a Map-Reply Record segment is included in the Map-Request.

P: This is the probe-bit which indicates that a Map-Request SHOULD be treated as a locator reachability probe. The receiver SHOULD respond with a Map-Reply with the probe-bit set, indicating the Map-Reply is a locator reachability probe reply, with the nonce copied from the Map-Request. See Section 6.3.2 for more details.

S: This is the Solicit-Map-Request (SMR) bit. See Section 6.6.2 for details.

p: This is the PITR bit. This bit is set to 1 when a PITR sends a Map-Request.

s: This is the SMR-invoked bit. This bit is set to 1 when an xTR is sending a Map-Request in response to a received SMR-based Map-Request.

Reserved: It MUST be set to 0 on transmit and MUST be ignored on receipt.

IRC: This 5-bit field is the ITR-RLOC Count which encodes the additional number of (ITR-RLOC-AFI, ITR-RLOC Address) fields present in this message. At least one (ITR-RLOC-AFI, ITR-RLOC-Address) pair MUST be encoded. Multiple ITR-RLOC Address fields are used so a Map-Replier can select which destination address to use for a Map-Reply. The IRC value ranges from 0 to 31. For a value of 0, there is 1 ITR-RLOC address encoded, and for a value of 1, there are 2 ITR-RLOC addresses encoded and so on up to 31 which encodes a total of 32 ITR-RLOC addresses.

Record Count: The number of records in this Map-Request message. A record is comprised of the portion of the packet that is labeled 'Rec' above and occurs the number of times equal to Record Count. For this version of the protocol, a receiver MUST accept and process Map-Requests that contain one or more records, but a sender MUST only send Map-Requests containing one record. Support for requesting multiple EIDs in a single Map-Request message will be specified in a future version of the protocol.

**Nonce:** An 8-octet random value created by the sender of the Map-Request. This nonce will be returned in the Map-Reply. The security of the LISP mapping protocol depends critically on the strength of the nonce in the Map-Request message. The nonce SHOULD be generated by a properly seeded pseudo-random (or strong random) source. See [RFC4086] for advice on generating security-sensitive random data.

**Source-EID-AFI:** Address family of the "Source EID Address" field.

**Source EID Address:** This is the EID of the source host which originated the packet which is caused the Map-Request. When Map-Requests are used for refreshing a map-cache entry or for RLOC-probing, an AFI value 0 is used and this field is of zero length.

**ITR-RLOC-AFI:** Address family of the "ITR-RLOC Address" field that follows this field.

**ITR-RLOC Address:** Used to give the ETR the option of selecting the destination address from any address family for the Map-Reply message. This address MUST be a routable RLOC address of the sender of the Map-Request message.

**EID mask-len:** Mask length for EID prefix.

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

**EID-prefix:** 4 octets if an IPv4 address-family, 16 octets if an IPv6 address-family. When a Map-Request is sent by an ITR because a data packet is received for a destination where there is no mapping entry, the EID-prefix is set to the destination IP address of the data packet. And the 'EID mask-len' is set to 32 or 128 for IPv4 or IPv6, respectively. When an xTR wants to query a site about the status of a mapping it already has cached, the EID-prefix used in the Map-Request has the same mask-length as the EID-prefix returned from the site when it sent a Map-Reply message.

**Map-Reply Record:** When the M bit is set, this field is the size of a single "Record" in the Map-Reply format. This Map-Reply record contains the EID-to-RLOC mapping entry associated with the Source EID. This allows the ETR which will receive this Map-Request to cache the data if it chooses to do so.



### 6.1.3. EID-to-RLOC UDP Map-Request Message

A Map-Request is sent from an ITR when it needs a mapping for an EID, wants to test an RLOC for reachability, or wants to refresh a mapping before TTL expiration. For the initial case, the destination IP address used for the Map-Request is the data packet's destination address (i.e. the destination-EID) which had a mapping cache lookup failure. For the latter two cases, the destination IP address used for the Map-Request is one of the RLOC addresses from the locator-set of the map cache entry. The source address is either an IPv4 or IPv6 RLOC address depending if the Map-Request is using an IPv4 versus IPv6 header, respectively. In all cases, the UDP source port number for the Map-Request message is an ITR/PITR selected 16-bit value and the UDP destination port number is set to the well-known destination port number 4342. A successful Map-Reply, which is one that has a nonce that matches an outstanding Map-Request nonce, will update the cached set of RLOCs associated with the EID prefix range.

One or more Map-Request (ITR-RLOC-AFI, ITR-RLOC-Address) fields MUST be filled in by the ITR. The number of fields (minus 1) encoded MUST be placed in the IRC field. The ITR MAY include all locally configured locators in this list or just provide one locator address from each address family it supports. If the ITR erroneously provides no ITR-RLOC addresses, the Map-Replier MUST drop the Map-Request.

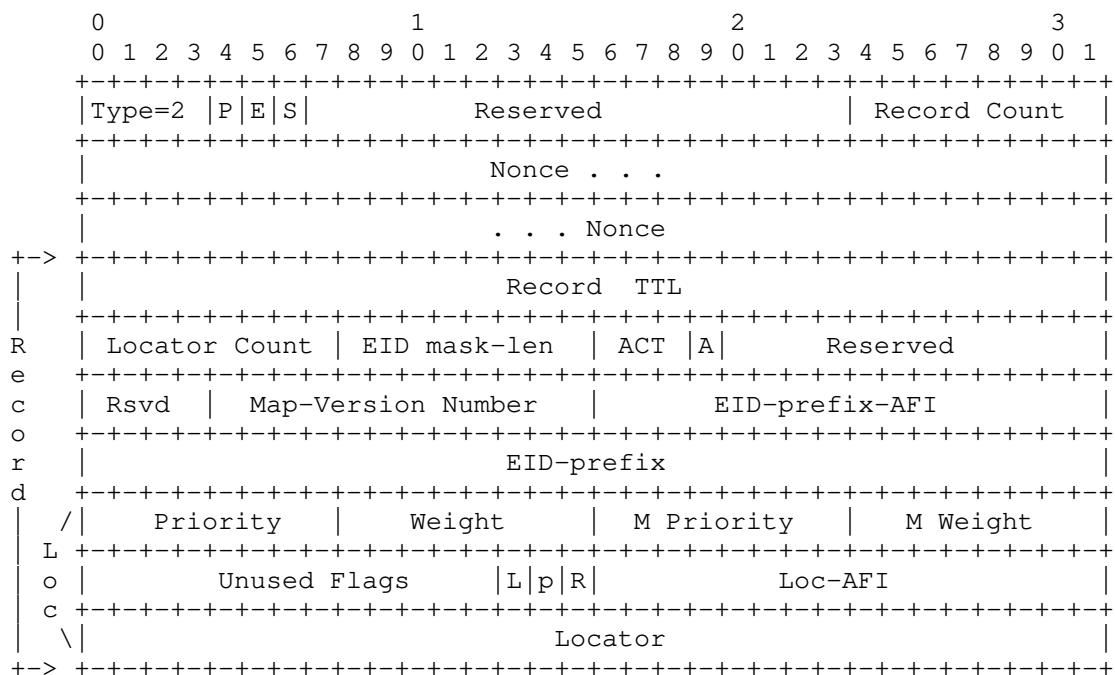
Map-Requests can also be LISP encapsulated using UDP destination port 4342 with a LISP type value set to "Encapsulated Control Message", when sent from an ITR to a Map-Resolver. Likewise, Map-Requests are LISP encapsulated the same way from a Map-Server to an ETR. Details on encapsulated Map-Requests and Map-Resolvers can be found in [LISP-MS].

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.

An ITR that is configured with mapping database information (i.e. it is also an ETR) MAY optionally include those mappings in a Map-Request. When an ETR configured to accept and verify such "piggybacked" mapping data receives such a Map-Request and it does not have this mapping in the map-cache, it MAY originate a "verifying Map-Request", addressed to the map-requesting ITR and the ETR MAY add a map-cache entry. If the ETR has a map-cache entry that matches the "piggybacked" EID and the RLOC is in the locator-set for the entry, then it may send the "verifying Map-Request" directly to the originating Map-Request source. If the RLOC is not in the locator-set, then the ETR MUST send the "verifying Map-Request" to the "piggybacked" EID. Doing this forces the "verifying Map-Request" to

go through the mapping database system to reach the authoritative source of information about that EID, guarding against RLOC-spoofing in in the "piggybacked" mapping data.

#### 6.1.4. Map-Reply Message Format



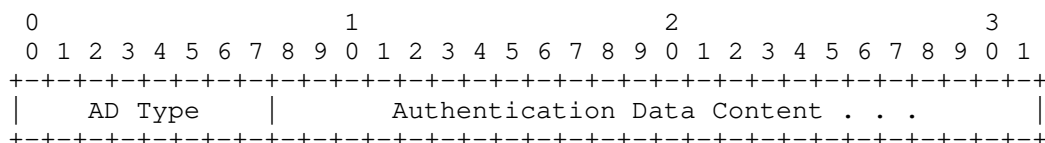
Packet field descriptions:

Type: 2 (Map-Reply)

P: This is the probe-bit which indicates that the Map-Reply is in response to a locator reachability probe Map-Request. The nonce field MUST contain a copy of the nonce value from the original Map-Request. See Section 6.3.2 for more details.

E: Indicates that the ETR which sends this Map-Reply message is advertising that the site is enabled for the Echo-Nonce locator reachability algorithm. See Section 6.3.1 for more details.

S: This is the Security bit. When set to 1 the following authentication information will be appended to the end of the Map-Reply. The detailed format of the Authentication Data Content is for further study.



Reserved: It MUST be set to 0 on transmit and MUST be ignored on receipt.

Record Count: The number of records in this reply message. A record is comprised of that portion of the packet labeled 'Record' above and occurs the number of times equal to Record count.

Nonce: A 24-bit value set in a Data-Probe packet or a 64-bit value from the Map-Request is echoed in this Nonce field of the Map-Reply. When a 24-bit value is supplied, it resides in the low-order 64 bits of the nonce field.

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'. The locator count can be 0 indicating there are no locators for the EID-prefix.

EID mask-len: Mask length for EID prefix.

ACT: This 3-bit field describes negative Map-Reply actions. In any other message type, these bits are set to 0 and ignored on receipt. These bits are used only when the 'Locator Count' field is set to 0. The action bits are encoded only in Map-Reply messages. The actions defined are used by an ITR or PITR when a destination EID matches a negative mapping cache entry. Unassigned values should cause a map-cache entry to be created and, when packets match this negative cache entry, they will be dropped. The current assigned values are:

- (0) No-Action: The map-cache is kept alive and no packet encapsulation occurs.
  - (1) Natively-Forward: The packet is not encapsulated or dropped but natively forwarded.
  - (2) Send-Map-Request: The packet invokes sending a Map-Request.
  - (3) Drop: A packet that matches this map-cache entry is dropped. An ICMP Unreachable message SHOULD be sent.
- A: The Authoritative bit, when sent is always set to 1 by an ETR. When a Map-Server is proxy Map-Replying [LISP-MS] for a LISP site, the Authoritative bit is set to 0. This indicates to requesting ITRs that the Map-Reply was not originated by a LISP node managed at the site that owns the EID-prefix.
- Map-Version Number: When this 12-bit value is non-zero the Map-Reply sender is informing the ITR what the version number is for the EID-record contained in the Map-Reply. The ETR can allocate this number internally but MUST coordinate this value with other ETRs for the site. When this value is 0, there is no versioning information conveyed. The Map-Version Number can be included in Map-Request and Map-Register messages. See Section 6.6.3 for more details.
- EID-prefix-AFI: Address family of EID-prefix according to [AFI].
- EID-prefix: 4 octets if an IPv4 address-family, 16 octets if an IPv6 address-family.
- Priority: each RLOC is assigned a unicast priority. Lower values are more preferable. When multiple RLOCs have the same priority, they MAY be used in a load-split fashion. A value of 255 means the RLOC MUST NOT be used for unicast forwarding.
- Weight: when priorities are the same for multiple RLOCs, the weight indicates how to balance unicast traffic between them. Weight is encoded as a relative weight of total unicast packets that match the mapping entry. For example if there are 4 locators in a locator set, where the weights assigned are 30, 20, 20, and 10, the first locator will get 37.5% of the traffic, the 2nd and 3rd locators will get 25% of traffic and the 4th locator will get 12.5% of the traffic. If all weights for a locator-set are equal, receiver of the Map-Reply will decide how to load-split traffic. See Section 6.5 for a suggested hash algorithm to distribute load

across locators with same priority and equal weight values.

**M Priority:** each RLOC is assigned a multicast priority used by an ETR in a receiver multicast site to select an ITR in a source multicast site for building multicast distribution trees. A value of 255 means the RLOC MUST NOT be used for joining a multicast distribution tree. For more details, see [MLISP].

**M Weight:** when priorities are the same for multiple RLOCs, the weight indicates how to balance building multicast distribution trees across multiple ITRs. The weight is encoded as a relative weight (similar to the unicast Weights) of total number of trees built to the source site identified by the EID-prefix. If all weights for a locator-set are equal, the receiver of the Map-Reply will decide how to distribute multicast state across ITRs. For more details, see [MLISP].

**Unused Flags:** set to 0 when sending and ignored on receipt.

**L:** when this bit is set, the locator is flagged as a local locator to the ETR that is sending the Map-Reply. When a Map-Server is doing proxy Map-Replying [LISP-MS] for a LISP site, the L bit is set to 0 for all locators in this locator-set.

**p:** when this bit is set, an ETR informs the RLOC-probing ITR that the locator address, for which this bit is set, is the one being RLOC-probed and MAY be different from the source address of the Map-Reply. An ITR that RLOC-probes a particular locator, MUST use this locator for retrieving the data structure used to store the fact that the locator is reachable. The "p" bit is set for a single locator in the same locator set. If an implementation sets more than one "p" bit erroneously, the receiver of the Map-Reply MUST select the first locator. The "p" bit MUST NOT be set for locator-set records sent in Map-Request and Map-Register messages.

**R:** set when the sender of a Map-Reply has a route to the locator in the locator data record. This receiver may find this useful to know if the locator is up but not necessarily reachable from the receiver's point of view. See also Section 6.4 for another way the R-bit may be used.

**Locator:** an IPv4 or IPv6 address (as encoded by the 'Loc-AFI' field) assigned to an ETR. Note that the destination RLOC address MAY be an anycast address. A source 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.

#### 6.1.5. EID-to-RLOC UDP Map-Reply Message

A Map-Reply returns an EID-prefix with a prefix length that is less than or equal to the EID being requested. The EID being requested is either from the destination field of an IP header of a Data-Probe or the EID record of a Map-Request. The RLOCs in the Map-Reply are globally-routable IP addresses of all ETRs for the LISP site. Each RLOC conveys status reachability but does not convey path reachability from a requesters perspective. Separate testing of path reachability is required, See Section 6.3 for details.

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.

When an ETR is configured with overlapping EID-prefixes, a Map-Request with an EID that longest matches any EID-prefix MUST be returned in a single Map-Reply message. For instance, if an ETR had database mapping entries for EID-prefixes:

```
10.0.0.0/8
10.1.0.0/16
10.1.1.0/24
10.1.2.0/24
```

A Map-Request for EID 10.1.1.1 would cause a Map-Reply with a record count of 1 to be returned with a mapping record EID-prefix of 10.1.1.0/24.

A Map-Request for EID 10.1.5.5, would cause a Map-Reply with a record count of 3 to be returned with mapping records for EID-prefixes 10.1.0.0/16, 10.1.1.0/24, and 10.1.2.0/24.

Note that not all overlapping EID-prefixes need to be returned, only the more specifics (note in the second example above 10.0.0.0/8 was not returned for requesting EID 10.1.5.5) entries for the matching

EID-prefix of the requesting EID. When more than one EID-prefix is returned, all SHOULD use the same Time-to-Live value so they can all time out at the same time. When a more specific EID-prefix is received later, its Time-to-Live value in the Map-Reply record can be stored even when other less specifics exist. When a less specific EID-prefix is received later, its map-cache expiration time SHOULD be set to the minimum expiration time of any more specific EID-prefix in the map-cache. This is done so the integrity of the EID-prefix set is wholly maintained so no more-specific entries are removed from the map-cache while keeping less-specific entries.

Map-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.

Map-Reply records can have an empty locator-set. A negative Map-Reply is a Map-Reply with an empty locator-set. Negative Map-Replies convey special actions by the sender to the ITR or PITR which have solicited the Map-Reply. There are two primary applications for Negative Map-Replies. The first is for a Map-Resolver to instruct an ITR or PITR when a destination is for a LISP site versus a non-LISP site. And the other is to source quench Map-Requests which are sent for non-allocated EIDs.

For each Map-Reply record, the list of locators in a locator-set MUST appear in the same order for each ETR that originates a Map-Reply message. The locator-set MUST be sorted in order of ascending IP address where an IPv4 locator address is considered numerically 'less than' an IPv6 locator address.

When sending a Map-Reply message, the destination address is copied from the one of the ITR-RLLOC fields from the Map-Request. The ETR can choose a locator address from one of the address families it supports. For Data-Probes, the destination address of the Map-Reply is copied from the source address of the Data-Probe message which is invoking the reply. The source address of the Map-Reply is one of the local IP addresses chosen to allow uRPF checks to succeed in the upstream service provider. The destination port of a Map-Reply message is copied from the source port of the Map-Request or Data-Probe and the source port of the Map-Reply message is set to the well-known UDP port 4342.

#### 6.1.5.1. Traffic Redirection with Coarse EID-Prefixes

When an ETR is misconfigured or compromised, it could return coarse EID-prefixes in Map-Reply messages it sends. The EID-prefix could

cover EID-prefixes which are allocated to other sites redirecting their traffic to the locators of the compromised site.

To solve this problem, there are two basic solutions that could be used. The first is to have Map-Servers proxy-map-reply on behalf of ETRs so their registered EID-prefixes are the ones returned in Map-Replies. Since the interaction between an ETR and Map-Server is secured with shared-keys, it is easier for an ETR to detect misbehavior. The second solution is to have ITRs and PITRs cache EID-prefixes with mask-lengths that are greater than or equal to a configured prefix length. This limits the damage to a specific width of any EID-prefix advertised, but needs to be coordinated with the allocation of site prefixes. These solutions can be used independently or at the same time.

At the time of this writing, other approaches are being considered and researched.

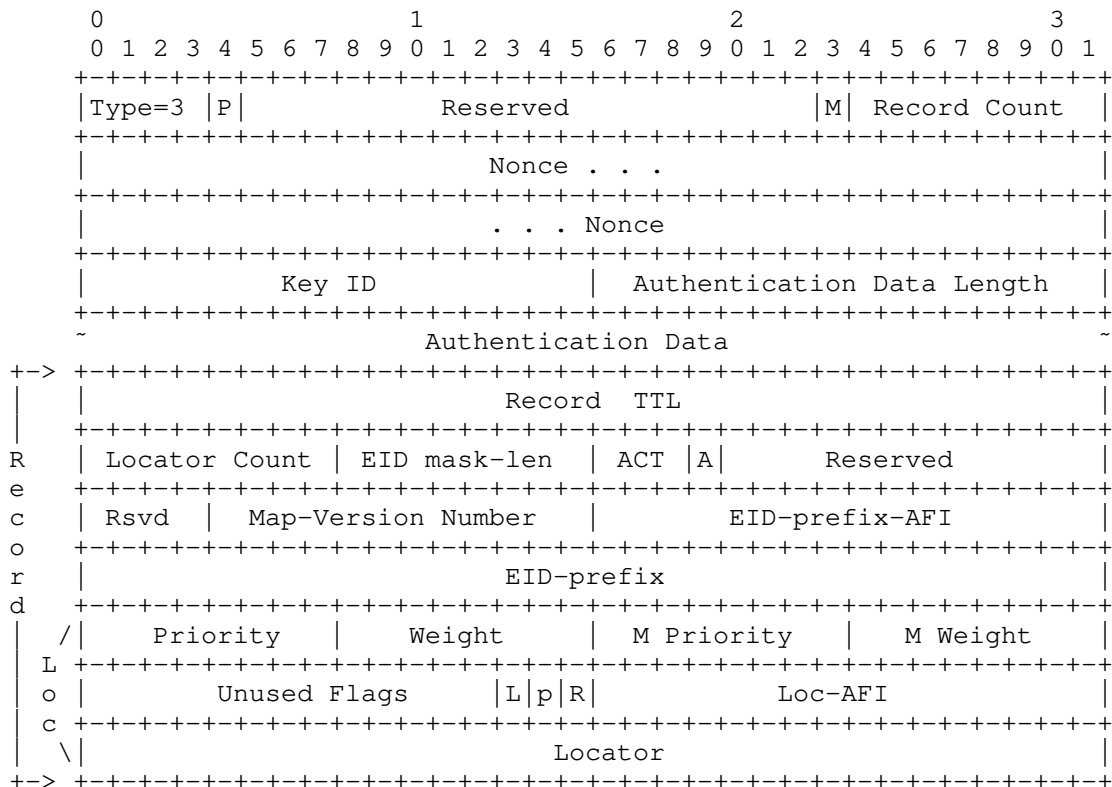
#### 6.1.6. Map-Register Message Format

The usage details of the Map-Register message can be found in specification [LISP-MS]. This section solely defines the message format.

The message is sent in UDP with a destination UDP port of 4342 and a randomly selected UDP source port number.

The Map-Register message format is:





Packet field descriptions:

Type: 3 (Map-Register)

P: This is the proxy-map-reply bit, when set to 1 an ETR sends a Map-Register message requesting for the Map-Server to proxy Map-Reply. The Map-Server will send non-authoritative Map-Replies on behalf of the ETR. Details on this usage can be found in [LISP-MS].

Reserved: It MUST be set to 0 on transmit and MUST be ignored on receipt.

M: This is the want-map-notify bit, when set to 1 an ETR is requesting for a Map-Notify message to be returned in response to sending a Map-Register message. The Map-Notify message sent by a Map-Server is used to an acknowledge receipt of a Map-Register message.

**Record Count:** The number of records in this Map-Register message. A record is comprised of that portion of the packet labeled 'Record' above and occurs the number of times equal to Record count.

**Nonce:** This 8-octet Nonce field is set to 0 in Map-Register messages. Since the Map-Register message is authenticated, the nonce field is not currently used for any security function but may be in the future as part of an anti-replay solution.

**Key ID:** A configured ID to find the configured Message Authentication Code (MAC) algorithm and key value used for the authentication function. See Section 14.4 for codepoint assignments.

**Authentication Data Length:** The length in octets of the Authentication Data field that follows this field. The length of the Authentication Data field is dependent on the Message Authentication Code (MAC) algorithm used. The length field allows a device that doesn't know the MAC algorithm to correctly parse the packet.

**Authentication Data:** The message digest used from the output of the Message Authentication Code (MAC) algorithm. The entire Map-Register payload is authenticated with this field preset to 0. After the MAC is computed, it is placed in this field. Implementations of this specification MUST include support for HMAC-SHA-1-96 [RFC2404] and support for HMAC-SHA-256-128 [RFC6234] is RECOMMENDED.

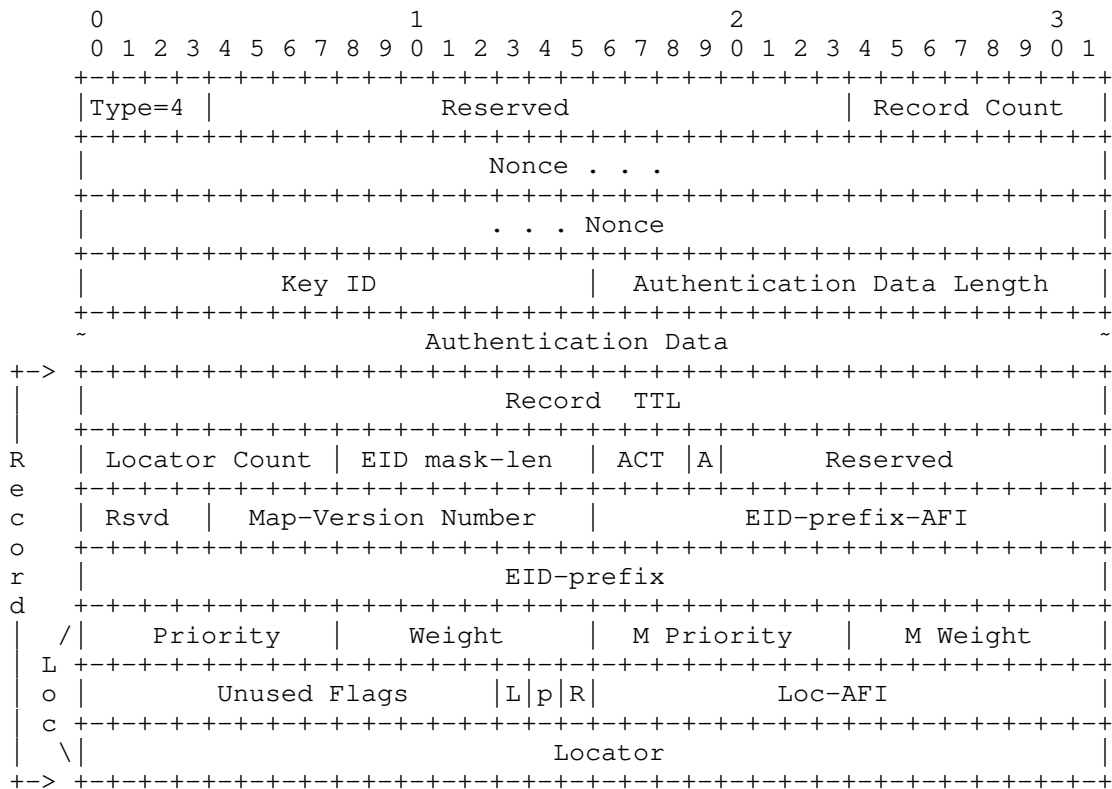
The definition of the rest of the Map-Register can be found in the Map-Reply section.

#### 6.1.7. Map-Notify Message Format

The usage details of the Map-Notify message can be found in specification [LISP-MS]. This section solely defines the message format.

The message is sent inside a UDP packet with source and destination UDP ports equal to 4342.

The Map-Notify message format is:



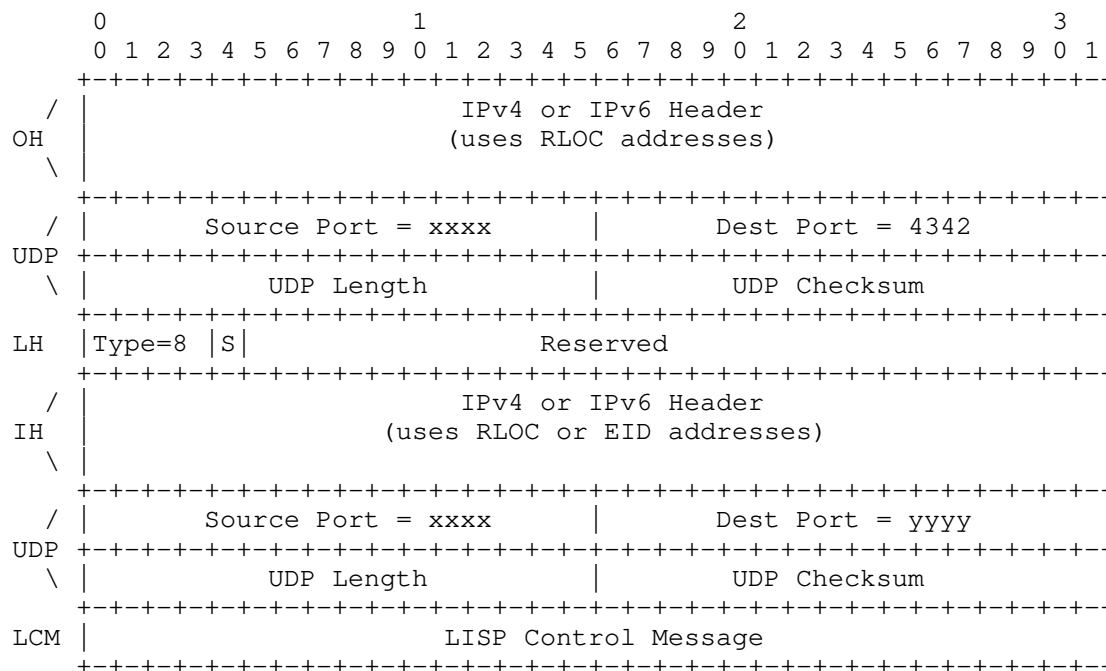
Packet field descriptions:

Type: 4 (Map-Notify)

The Map-Notify message has the same contents as a Map-Register message. See Map-Register section for field descriptions.

#### 6.1.8. Encapsulated Control Message Format

An Encapsulated Control Message (ECM) is used to encapsulate control packets sent between xTRs and the mapping database system described in [LISP-MS].



#### Packet header descriptions:

OH: The outer IPv4 or IPv6 header which uses RLOC addresses in the source and destination header address fields.

UDP: The outer UDP header with destination port 4342. The source port is randomly allocated. The checksum field MUST be non-zero.

LH: Type 8 is defined to be a "LISP Encapsulated Control Message" and what follows is either an IPv4 or IPv6 header as encoded by the first 4 bits after the reserved field.

S: This is the Security bit. When set to 1 the field following the Reserved field will have the following format. The detailed format of the Authentication Data Content is for further study.



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 a 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.
- o A "gleaned" map-cache entry, one learned from the source RLOC of a received encapsulated packet, is only stored and used for a few seconds, pending verification. Verification is performed by sending a Map-Request to the source EID (the inner header IP source address) of the received encapsulated packet. A reply to this "verifying Map-Request" is used to fully populate the map-cache entry for the "gleaned" EID and is stored and used for the time indicated from the TTL field of a received Map-Reply. When a verified map-cache entry is stored, data gleaning no longer occurs for subsequent packets which have a source EID that matches the EID-prefix of the verified entry.

RLOCs that appear in EID-to-RLOC Map-Reply messages are assumed to be reachable when the R-bit for the locator record is set to 1. When the R-bit is set to 0, an ITR or PITR MUST NOT encapsulate to the RLOC. Neither the information contained in a Map-Reply or that stored in the mapping database system provides reachability information for RLOCs. Note that reachability is not part of the mapping system and is determined using one or more of the Routing

Locator Reachability Algorithms described in the next section.

### 6.3. Routing Locator Reachability

Several mechanisms for determining RLOC reachability are currently defined:

1. An ETR may examine the Locator Status Bits in the LISP header of an encapsulated data packet received from an ITR. If the ETR is also acting as an ITR and has traffic to return to the original ITR site, it can use this status information to help select an RLOC.
2. An ITR may receive an ICMP Network or ICMP Host Unreachable message for an RLOC it is using. This indicates that the RLOC is likely down. Note, trusting ICMP messages may not be desirable but neither is ignoring them completely. Implementations are encouraged to follow current best practices in treating these conditions.
3. An ITR which participates in the global routing system can determine that an RLOC is down if no BGP RIB route exists that matches the RLOC IP address.
4. An ITR may receive an ICMP Port Unreachable message from a destination host. This occurs if an ITR attempts to use interworking [INTERWORK] and LISP-encapsulated data is sent to a non-LISP-capable site.
5. An ITR may receive a Map-Reply from an ETR in response to a previously sent Map-Request. The RLOC source of the Map-Reply is likely up since the ETR was able to send the Map-Reply to the ITR.
6. When an ETR receives an encapsulated packet from an ITR, the source RLOC from the outer header of the packet is likely up.
7. An ITR/ETR pair can use the Locator Reachability Algorithms described in this section, namely Echo-Noncing or RLOC-Probing.

When determining Locator up/down reachability by examining the Locator Status Bits from the LISP encapsulated data packet, an ETR will receive up to date status from an encapsulating ITR about reachability for all ETRs at the site. CE-based ITRs at the source site can determine reachability relative to each other using the site IGP as follows:

- o Under normal circumstances, each ITR will advertise a default route into the site IGP.
- o If an ITR fails or if the upstream link to its PE fails, its default route will either time-out or be withdrawn.

Each ITR can thus observe the presence or lack of a default route originated by the others to determine the Locator Status Bits it sets for them.

RLOCs listed in a Map-Reply are numbered with ordinals 0 to n-1. The Locator Status Bits in a LISP encapsulated packet are numbered from 0 to n-1 starting with the least significant bit. For example, if an RLOC listed in the 3rd position of the Map-Reply goes down (ordinal value 2), then all ITRs at the site will clear the 3rd least significant bit (xxxx x0xx) of the Locator Status Bits field for the packets they encapsulate.

When an ETR decapsulates a packet, it will check for any change in the Locator Status Bits field. When a bit goes from 1 to 0, the ETR if acting also as an ITR, will refrain from encapsulating packets to an RLOC that is indicated as down. It will only resume using that RLOC if the corresponding Locator Status Bit returns to a value of 1. Locator Status Bits are associated with a locator-set per EID-prefix. Therefore, when a locator becomes unreachable, the Locator Status Bit that corresponds to that locator's position in the list returned by the last Map-Reply will be set to zero for that particular EID-prefix.

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 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.

If an ITR does receive an ICMP Network or Host Unreachable message, it MAY originate its own ICMP Unreachable message destined for the host that originated the data packet the ITR encapsulated.

Also, BGP-enabled ITRs can unilaterally examine the RIB to see if a locator address from a locator-set in a mapping entry matches a prefix. If it does not find one and BGP is running in the Default Free Zone (DFZ), it can decide to not use the locator even though the



Locator Status Bits indicate the locator is up. In this case, the path from the ITR to the ETR that is assigned the locator is not available. More details are in [LOC-ID-ARCH].

Optionally, an ITR can send a Map-Request to a Locator and if a Map-Reply is returned, reachability of the Locator has been determined. 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 to be 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 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.

When an ETR decapsulates a packet, it knows that it is reachable from the encapsulating ITR because that is how the packet arrived. In most cases, the ETR can also reach the ITR but cannot assume this to be true due to the possibility of path asymmetry. In the presence of unidirectional traffic flow from an ITR to an ETR, the ITR SHOULD NOT use the lack of return traffic as an indication that the ETR is unreachable. Instead, it MUST use an alternate mechanisms to determine reachability.

#### 6.3.1. Echo Nonce Algorithm

When data flows bidirectionally between locators from different sites, a data-plane mechanism called "nonce echoing" can be used to determine reachability between an ITR and ETR. When an ITR wants to solicit a nonce echo, it sets the N and E bits and places a 24-bit nonce [RFC4086] in the LISP header of the next encapsulated data packet.

When this packet is received by the ETR, the encapsulated packet is forwarded as normal. When the ETR next sends a data packet to the ITR, it includes the nonce received earlier with the N bit set and E bit cleared. The ITR sees this "echoed nonce" and knows the path to and from the ETR is up.

The ITR will set the E-bit and N-bit for every packet it sends while in echo-nonce-request state. The time the ITR waits to process the

echoed nonce before it determines the path is unreachable is variable and a choice left for the implementation.

If the ITR is receiving packets from the ETR but does not see the nonce echoed while being in echo-nonce-request state, then the path to the ETR is unreachable. This decision may be overridden by other locator reachability algorithms. Once the ITR determines the path to the ETR is down it can switch to another locator for that EID-prefix.

Note that "ITR" and "ETR" are relative terms here. Both devices MUST be implementing both ITR and ETR functionality for the echo nonce mechanism to operate.

The ITR and ETR may both go into echo-nonce-request state at the same time. The number of packets sent or the time during which echo nonce requests are sent is an implementation specific setting. However, when an ITR is in echo-nonce-request state, it can echo the ETR's nonce in the next set of packets that it encapsulates and then subsequently, continue sending echo-nonce-request packets.

This mechanism does not completely solve the forward path reachability problem as traffic may be unidirectional. That is, the ETR receiving traffic at a site may not be the same device as an ITR which transmits traffic from that site or the site to site traffic is unidirectional so there is no ITR returning traffic.

The echo-nonce algorithm is bilateral. That is, if one side sets the E-bit and the other side is not enabled for echo-noncing, then the echoing of the nonce does not occur and the requesting side may regard the locator unreachable erroneously. An ITR SHOULD only set the E-bit in a encapsulated data packet when it knows the ETR is enabled for echo-noncing. This is conveyed by the E-bit in the Map-Reply message.

Note that other locator reachability mechanisms are being researched and can be used to compliment or even override the Echo Nonce Algorithm. See next section for an example of control-plane probing.

#### 6.3.2. RLOC Probing Algorithm

RLOC Probing is a method that an ITR or PITR can use to determine the reachability status of one or more locators that it has cached in a map-cache entry. The probe-bit of the Map-Request and Map-Reply messages are used for RLOC Probing.

RLOC probing is done in the control-plane on a timer basis where an ITR or PITR will originate a Map-Request destined to a locator address from one of its own locator addresses. A Map-Request used as

an RLOC-probe is NOT encapsulated and NOT sent to a Map-Server or on the ALT like one would when soliciting mapping data. The EID record encoded in the Map-Request is the EID-prefix of the map-cache entry cached by the ITR or PITR. The ITR may include a mapping data record for its own database mapping information which contains the local EID-prefixes and RLOCs for its site. RLOC-probes are sent periodically using a jittered timer interval.

When an ETR receives a Map-Request message with the probe-bit set, it returns a Map-Reply with the probe-bit set. The source address of the Map-Reply is set according to the procedure described in Section 6.1.5. The Map-Reply SHOULD contain mapping data for the EID-prefix contained in the Map-Request. This provides the opportunity for the ITR or PITR, which sent the RLOC-probe to get mapping updates if there were changes to the ETR's database mapping entries.

There are advantages and disadvantages of RLOC Probing. The greatest benefit of RLOC Probing is that it can handle many failure scenarios allowing the ITR to determine when the path to a specific locator is reachable or has become unreachable, thus providing a robust mechanism for switching to using another locator from the cached locator. RLOC Probing can also provide rough RTT estimates between a pair of locators which can be useful for network management purposes as well as for selecting low delay paths. The major disadvantage of RLOC Probing is in the number of control messages required and the amount of bandwidth used to obtain those benefits, especially if the requirement for failure detection times are very small.

Continued research and testing will attempt to characterize the tradeoffs of failure detection times versus message overhead.

#### 6.4. EID Reachability within a LISP Site

A site may be multihomed using two or more ETRs. The hosts and infrastructure within a site will be addressed using one or more EID prefixes that are mapped to the RLOCs of the relevant ETRs in the mapping system. One possible failure mode is for an ETR to lose reachability to one or more of the EID prefixes within its own site. When this occurs when the ETR sends Map-Replies, it can clear the R-bit associated with its own locator. And when the ETR is also an ITR, it can clear its locator-status-bit in the encapsulation data header.

It is recognized there are no simple solutions to the site partitioning problem because it is hard to know which part of the EID-prefix range is partitioned. And which locators can reach any sub-ranges of the EID-prefixes. This problem is under investigation

with the expectation that experiments will tell us more. Note, this is not a new problem introduced by the LISP architecture. The problem exists today when a multi-homed site uses BGP to advertise its reachability upstream.

#### 6.5. Routing Locator Hashing

When an ETR provides an EID-to-RLOC mapping in a Map-Reply message to a requesting ITR, the locator-set for the EID-prefix may contain different priority values for each locator address. When more than one best priority locator exists, the ITR can decide how to load share traffic against the corresponding locators.

The following hash algorithm may be used by an ITR to select a locator for a packet destined to an EID for the EID-to-RLOC mapping:

1. Either a source and destination address hash can be used or the traditional 5-tuple hash which includes the source and destination addresses, source and destination TCP, UDP, or SCTP port numbers and the IP protocol number field or IPv6 next-protocol fields of a packet a host originates from within a LISP site. When a packet is not a TCP, UDP, or SCTP packet, the source and destination addresses only from the header are used to compute the hash.
2. Take the hash value and divide it by the number of locators stored in the locator-set for the EID-to-RLOC mapping.
3. The remainder will yield a value of 0 to "number of locators minus 1". Use the remainder to select the locator in the locator-set.

Note that when a packet is LISP encapsulated, the source port number in the outer UDP header needs to be set. Selecting a hashed value allows core routers which are attached to Link Aggregation Groups (LAGs) to load-split the encapsulated packets across member links of such LAGs. Otherwise, core routers would see a single flow, since packets have a source address of the ITR, for packets which are originated by different EIDs at the source site. A suggested setting for the source port number computed by an ITR is a 5-tuple hash function on the inner header, as described above.

Many core router implementations use a 5-tuple hash to decide how to balance packet load across members of a LAG. The 5-tuple hash includes the source and destination addresses of the packet and the source and destination ports when the protocol number in the packet is TCP or UDP. For this reason, UDP encoding is used for LISP encapsulation.

## 6.6. Changing the Contents of EID-to-RLOC Mappings

Since the LISP architecture uses a caching scheme to retrieve and store EID-to-RLOC mappings, the only way an ITR can get a more up-to-date mapping is to re-request the mapping. However, the ITRs do not know when the mappings change and the ETRs do not keep track of which ITRs requested its mappings. For scalability reasons, we want to maintain this approach but need to provide a way for ETRs change their mappings and inform the sites that are currently communicating with the ETR site using such mappings.

When adding a new locator record in lexicographic order to the end of a locator-set, it is easy to update mappings. We assume new mappings will maintain the same locator ordering as the old mapping but just have new locators appended to the end of the list. So some ITRs can have a new mapping while other ITRs have only an old mapping that is used until they time out. When an ITR has only an old mapping but detects bits set in the loc-status-bits that correspond to locators beyond the list it has cached, it simply ignores them. However, this can only happen for locator addresses that are lexicographically greater than the locator addresses in the existing locator-set.

When a locator record is inserted in the middle of a locator-set, to maintain lexicographic order, the SMR procedure in Section 6.6.2 is used to inform ITRs and PITRs of the new locator-status-bit mappings.

When a locator record is removed from a locator-set, ITRs that have the mapping cached will not use the removed locator because the xTRs will set the loc-status-bit to 0. So even if the locator is in the list, it will not be used. For new mapping requests, the xTRs can set the locator AFI to 0 (indicating an unspecified address), as well as setting the corresponding loc-status-bit to 0. This forces ITRs with old or new mappings to avoid using the removed locator.

If many changes occur to a mapping over a long period of time, one will find empty record slots in the middle of the locator-set and new records appended to the locator-set. At some point, it would be useful to compact the locator-set so the loc-status-bit settings can be efficiently packed.

We propose here three approaches for locator-set compaction, one operational and two protocol mechanisms. The operational approach uses a clock sweep method. The protocol approaches use the concept of Solicit-Map-Requests and Map-Versioning.

### 6.6.1. Clock Sweep

The clock sweep approach uses planning in advance and the use of count-down TTLs to time out mappings that have already been cached. The default setting for an EID-to-RLOC mapping TTL is 24 hours. So there is a 24 hour window to time out old mappings. The following clock sweep procedure is used:

1. 24 hours before a mapping change is to take effect, a network administrator configures the ETRs at a site to start the clock sweep window.
2. During the clock sweep window, ETRs continue to send Map-Reply messages with the current (unchanged) mapping records. The TTL for these mappings is set to 1 hour.
3. 24 hours later, all previous cache entries will have timed out, and any active cache entries will time out within 1 hour. During this 1 hour window the ETRs continue to send Map-Reply messages with the current (unchanged) mapping records with the TTL set to 1 minute.
4. At the end of the 1 hour window, the ETRs will send Map-Reply messages with the new (changed) mapping records. So any active caches can get the new mapping contents right away if not cached, or in 1 minute if they had the mapping cached. The new mappings are cached with a time to live equal to the TTL in the Map-Reply.

### 6.6.2. Solicit-Map-Request (SMR)

Soliciting a Map-Request is a selective way for ETRs, at the site where mappings change, to control the rate they receive requests for Map-Reply messages. SMRs are also used to tell remote ITRs to update the mappings they have cached.

Since the ETRs don't keep track of remote ITRs that have cached their mappings, they do not know which ITRs need to have their mappings updated. As a result, an ETR will solicit Map-Requests (called an SMR message) from those sites to which it has been sending encapsulated data to for the last minute. In particular, an ETR will send an SMR an ITR to which it has recently sent encapsulated data.

An SMR message is simply a bit set in a Map-Request message. An ITR or PITR will send a Map-Request when they receive an SMR message. Both the SMR sender and the Map-Request responder MUST rate-limit these messages. Rate-limiting can be implemented as a global rate-limiter or one rate-limiter per SMR destination.

The following procedure shows how a SMR exchange occurs when a site is doing locator-set compaction for an EID-to-RLOC mapping:

1. When the database mappings in an ETR change, the ETRs at the site begin to send Map-Requests with the SMR bit set for each locator in each map-cache entry the ETR caches.
2. A remote ITR which receives the SMR message will schedule sending a Map-Request message to the source locator address of the SMR message or to the mapping database system. A newly allocated random nonce is selected and the EID-prefix used is the one copied from the SMR message. If the source locator is the only locator in the cached locator-set, the remote ITR SHOULD send a Map-Request to the database mapping system just in case the single locator has changed and may no longer be reachable to accept the Map-Request.
3. The remote ITR MUST rate-limit the Map-Request until it gets a Map-Reply while continuing to use the cached mapping. When Map Versioning is used, described in Section 6.6.3, an SMR sender can detect if an ITR is using the most up to date database mapping.
4. The ETRs at the site with the changed mapping will reply to the Map-Request with a Map-Reply message that has a nonce from the SMR-invoked Map-Request. The Map-Reply messages SHOULD be rate limited. This is important to avoid Map-Reply implosion.
5. The ETRs, at the site with the changed mapping, record the fact that the site that sent the Map-Request has received the new mapping data in the mapping cache entry for the remote site so the loc-status-bits are reflective of the new mapping for packets going to the remote site. The ETR then stops sending SMR messages.

Experimentation is in progress to determine the appropriate rate-limit parameters.

For security reasons an ITR MUST NOT process unsolicited Map-Replies. To avoid map-cache entry corruption by a third-party, a sender of an SMR-based Map-Request MUST be verified. If an ITR receives an SMR-based Map-Request and the source is not in the locator-set for the stored map-cache entry, then the responding Map-Request MUST be sent with an EID destination to the mapping database system. Since the mapping database system is more secure to reach an authoritative ETR, it will deliver the Map-Request to the authoritative source of the mapping data.

When an ITR receives an SMR-based Map-Request for which it does not

have a cached mapping for the EID in the SMR message, it MAY not send a SMR-invoked Map-Request. This scenario can occur when an ETR sends SMR messages to all locators in the locator-set it has stored in its map-cache but the remote ITRs that receive the SMR may not be sending packets to the site. There is no point in updating the ITRs until they need to send, in which case, they will send Map-Requests to obtain a map-cache entry.

#### 6.6.3. Database Map Versioning

When there is unidirectional packet flow between an ITR and ETR, and the EID-to-RLOC mappings change on the ETR, it needs to inform the ITR so encapsulation can stop to a removed locator and start to a new locator in the locator-set.

An ETR, when it sends Map-Reply messages, conveys its own Map-Version number. This is known as the Destination Map-Version Number. ITRs include the Destination Map-Version Number in packets they encapsulate to the site. When an ETR decapsulates a packet and detects the Destination Map-Version Number is less than the current version for its mapping, the SMR procedure described in Section 6.6.2 occurs.

An ITR, when it encapsulates packets to ETRs, can convey its own Map-Version number. This is known as the Source Map-Version Number. When an ETR decapsulates a packet and detects the Source Map-Version Number is greater than the last Map-Version Number sent in a Map-Reply from the ITR's site, the ETR will send a Map-Request to one of the ETRs for the source site.

A Map-Version Number is used as a sequence number per EID-prefix. So values that are greater, are considered to be more recent. A value of 0 for the Source Map-Version Number or the Destination Map-Version Number conveys no versioning information and an ITR does no comparison with previously received Map-Version Numbers.

A Map-Version Number can be included in Map-Register messages as well. This is a good way for the Map-Server can assure that all ETRs for a site registering to it will be Map-Version number synchronized.

See [VERSIONING] for a more detailed analysis and description of Database Map Versioning.



## 7. Router Performance Considerations

LISP is designed to be very hardware-based forwarding friendly. A few implementation techniques can be used to incrementally implement LISP:

- o 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. There are a few proven cases where no changes to existing deployed hardware were needed to support the LISP data-plane.
- o On an ITR, prepending a new IP header consists of adding more octets to a MAC rewrite string and prepending the string as part of the outgoing encapsulation procedure. Routers that support GRE tunneling [RFC2784] or 6to4 tunneling [RFC3056] may already support this action.
- o A packet's source address or interface the packet was received on can be used to select a VRF (Virtual Routing/Forwarding). The VRF's routing table can be used to find EID-to-RLOC mappings.

For performance issues related to map-cache management, see section Section 12.

## 8. Deployment Scenarios

This section will explore how and where ITRs and ETRs can be deployed and will discuss the pros and cons of each deployment scenario. For a more detailed deployment recommendation, refer to [LISP-DEPLOY].

There are two basic deployment trade-offs 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? A centralized cache is when an ITR keeps a cache for all the EIDs it is encapsulating to. The packet takes a direct path to the destination locator. A distributed cache is when an ITR needs help from other re-encapsulating routers because it does not store all the cache entries for the EIDs it is encapsulating to. So the packet takes a path through re-encapsulating routers that have a different set of cache entries.
- 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 survey 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 Map-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 than 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 is the default behavior envisioned in the rest of this specification.

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 resilience 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. Another disadvantage of using anycast locators is the limited advertisement scope of /32 (or /128 for IPv6) routes.

### 8.3. ISP Provider-Edge (PE) Tunnel Routers

Use of ISP PE routers as tunnel endpoint routers is not the typical deployment scenario envisioned in the specification. This section attempts to capture some of reasoning behind this preference of implementing LISP on CE routers.

Use of ISP PE routers as tunnel endpoint routers gives an ISP, rather than a site, 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 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. Other disadvantages include the difficulty in synchronizing path liveness updates between CE and PE routers.

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.

### 8.4. LISP Functionality with Conventional NATs

LISP routers can be deployed behind Network Address Translator (NAT) devices to provide the same set of packet services hosts have today when they are addressed out of private address space.

It is important to note that a locator address in any LISP control message MUST be a globally routable address and therefore SHOULD NOT contain [RFC1918] addresses. If a LISP router is configured with private addresses, they MUST be used only in the outer IP header so the NAT device can translate properly. Otherwise, EID addresses MUST

be translated before encapsulation is performed. Both NAT translation and LISP encapsulation functions could be co-located in the same device.

More details on LISP address translation can be found in [INTERWORK].

#### 8.5. Packets Egressing a LISP Site

When a LISP site is using two ITRs for redundancy, the failure of one ITR will likely shift outbound traffic to the second. This second ITR's cache may not be populated with the same EID-to-RLOC mapping entries as the first. If this second ITR does not have these mappings, traffic will be dropped while the mappings are retrieved from the mapping system. The retrieval of these messages may increase the load of requests being sent into the mapping system. Deployment and experimentation will determine whether this issue requires more attention.

## 9. Traceroute Considerations

When a source host in a LISP site initiates a traceroute to a destination host in another LISP site, it is highly desirable for it to see the entire path. Since packets are encapsulated from ITR to ETR, the hop across the tunnel could be viewed as a single hop. However, LISP traceroute will provide the entire path so the user can see 3 distinct segments of the path from a source LISP host to a destination LISP host:

Segment 1 (in source LISP site based on EIDs):

source-host ---> first-hop ... next-hop ---> ITR

Segment 2 (in the core network based on RLOCs):

ITR ---> next-hop ... next-hop ---> ETR

Segment 3 (in the destination LISP site based on EIDs):

ETR ---> next-hop ... last-hop ---> destination-host

For segment 1 of the path, ICMP Time Exceeded messages are returned in the normal manner as they are today. The ITR performs a TTL decrement and test for 0 before encapsulating. So the ITR hop is seen by the traceroute source has an EID address (the address of site-facing interface).

For segment 2 of the path, ICMP Time Exceeded messages are returned to the ITR because the TTL decrement to 0 is done on the outer header, so the destination of the ICMP messages are to the ITR RLOC address, the source RLOC address of the encapsulated traceroute packet. The ITR looks inside of the ICMP payload to inspect the traceroute source so it can return the ICMP message to the address of the traceroute client as well as retaining the core router IP address in the ICMP message. This is so the traceroute client can display the core router address (the RLOC address) in the traceroute output. The ETR returns its RLOC address and responds to the TTL decrement to 0 like the previous core routers did.

For segment 3, the next-hop router downstream from the ETR will be decrementing the TTL for the packet that was encapsulated, sent into the core, decapsulated by the ETR, and forwarded because it isn't the final destination. If the TTL is decremented to 0, any router on the path to the destination of the traceroute, including the next-hop router or destination, will send an ICMP Time Exceeded message to the source EID of the traceroute client. The ICMP message will be

encapsulated by the local ITR and sent back to the ETR in the originated traceroute source site, where the packet will be delivered to the host.

### 9.1. IPv6 Traceroute

IPv6 traceroute follows the procedure described above since the entire traceroute data packet is included in ICMP Time Exceeded message payload. Therefore, only the ITR needs to pay special attention for forwarding ICMP messages back to the traceroute source.

### 9.2. IPv4 Traceroute

For IPv4 traceroute, we cannot follow the above procedure since IPv4 ICMP Time Exceeded messages only include the invoking IP header and 8 octets that follow the IP header. Therefore, when a core router sends an IPv4 Time Exceeded message to an ITR, all the ITR has in the ICMP payload is the encapsulated header it prepended followed by a UDP header. The original invoking IP header, and therefore the identity of the traceroute source is lost.

The solution we propose to solve this problem is to cache traceroute IPv4 headers in the ITR and to match them up with corresponding IPv4 Time Exceeded messages received from core routers and the ETR. The ITR will use a circular buffer for caching the IPv4 and UDP headers of traceroute packets. It will select a 16-bit number as a key to find them later when the IPv4 Time Exceeded messages are received. When an ITR encapsulates an IPv4 traceroute packet, it will use the 16-bit number as the UDP source port in the encapsulating header. When the ICMP Time Exceeded message is returned to the ITR, the UDP header of the encapsulating header is present in the ICMP payload thereby allowing the ITR to find the cached headers for the traceroute source. The ITR puts the cached headers in the payload and sends the ICMP Time Exceeded message to the traceroute source retaining the source address of the original ICMP Time Exceeded message (a core router or the ETR of the site of the traceroute destination).

The signature of a traceroute packet comes in two forms. The first form is encoded as a UDP message where the destination port is inspected for a range of values. The second form is encoded as an ICMP message where the IP identification field is inspected for a well-known value.

### 9.3. Traceroute using Mixed Locators

When either an IPv4 traceroute or IPv6 traceroute is originated and the ITR encapsulates it in the other address family header, you

cannot get all 3 segments of the traceroute. Segment 2 of the traceroute can not be conveyed to the traceroute source since it is expecting addresses from intermediate hops in the same address format for the type of traceroute it originated. Therefore, in this case, segment 2 will make the tunnel look like one hop. All the ITR has to do to make this work is to not copy the inner TTL to the outer, encapsulating header's TTL when a traceroute packet is encapsulated using an RLOC from a different address family. This will cause no TTL decrement to 0 to occur in core routers between the ITR and ETR.



## 10. Mobility Considerations

There are several kinds of mobility of which only some might be of concern to LISP. Essentially they are as follows.

### 10.1. Site Mobility

A site wishes to change its attachment points to the Internet, and its LISP Tunnel Routers will have new RLOCs when it changes upstream providers. Changes in EID-RLOC mappings for sites are expected to be handled by configuration, outside of the LISP protocol.

### 10.2. Slow Endpoint Mobility

An individual endpoint wishes to move, but is not concerned about maintaining session continuity. Renumbering is involved. LISP can help with the issues surrounding renumbering [RFC4192] [LISA96] by decoupling the address space used by a site from the address spaces used by its ISPs. [RFC4984]

### 10.3. Fast Endpoint Mobility

Fast endpoint mobility occurs when an endpoint moves relatively rapidly, changing its IP layer network attachment point. Maintenance of session continuity is a goal. This is where the Mobile IPv4 [RFC5944] and Mobile IPv6 [RFC6275] [RFC4866] mechanisms are used, and primarily where interactions with LISP need to be explored.

The problem is that as an endpoint moves, it may require changes to the mapping between its EID and a set of RLOCs for its new network location. When this is added to the overhead of mobile IP binding updates, some packets might be delayed or dropped.

In IPv4 mobility, when an endpoint is away from home, packets to it are encapsulated and forwarded via a home agent which resides in the home area the endpoint's address belongs to. The home agent will encapsulate and forward packets either directly to the endpoint or to a foreign agent which resides where the endpoint has moved to. Packets from the endpoint may be sent directly to the correspondent node, may be sent via the foreign agent, or may be reverse-tunneled back to the home agent for delivery to the mobile node. As the mobile node's EID or available RLOC changes, LISP EID-to-RLOC mappings are required for communication between the mobile node and the home agent, whether via foreign agent or not. As a mobile endpoint changes networks, up to three LISP mapping changes may be required:

- o The mobile node moves from an old location to a new visited network location and notifies its home agent that it has done so. The Mobile IPv4 control packets the mobile node sends pass through one of the new visited network's ITRs, which needs an EID-RLOC mapping for the home agent.
- o The home agent might not have the EID-RLOC mappings for the mobile node's "care-of" address or its foreign agent in the new visited network, in which case it will need to acquire them.
- o When packets are sent directly to the correspondent node, it may be that no traffic has been sent from the new visited network to the correspondent node's network, and the new visited network's ITR will need to obtain an EID-RLOC mapping for the correspondent node's site.

In addition, if the IPv4 endpoint is sending packets from the new visited network using its original EID, then LISP will need to perform a route-returnability check on the new EID-RLOC mapping for that EID.

In IPv6 mobility, packets can flow directly between the mobile node and the correspondent node in either direction. The mobile node uses its "care-of" address (EID). In this case, the route-returnability check would not be needed but one more LISP mapping lookup may be required instead:

- o As above, three mapping changes may be needed for the mobile node to communicate with its home agent and to send packets to the correspondent node.
- o In addition, another mapping will be needed in the correspondent node's ITR, in order for the correspondent node to send packets to the mobile node's "care-of" address (EID) at the new network location.

When both endpoints are mobile the number of potential mapping lookups increases accordingly.

As a mobile node moves there are not only mobility state changes in the mobile node, correspondent node, and home agent, but also state changes in the ITRs and ETRs for at least some EID-prefixes.

The goal is to support rapid adaptation, with little delay or packet loss for the entire system. Also IP mobility can be modified to require fewer mapping changes. In order to increase overall system performance, there may be a need to reduce the optimization of one area in order to place fewer demands on another.

In LISP, one possibility is to "glean" information. When a packet arrives, the ETR could examine the EID-RLOC mapping and use that mapping for all outgoing traffic to that EID. It can do this after performing a route-returnability check, to ensure that the new network location does have a internal route to that endpoint. However, this does not cover the case where an ITR (the node assigned the RLOC) at the mobile-node location has been compromised.

Mobile IP packet exchange is designed for an environment in which all routing information is disseminated before packets can be forwarded. In order to allow the Internet to grow to support expected future use, we are moving to an environment where some information may have to be obtained after packets are in flight. Modifications to IP mobility should be considered in order to optimize the behavior of the overall system. Anything which decreases the number of new EID-RLOC mappings needed when a node moves, or maintains the validity of an EID-RLOC mapping for a longer time, is useful.

#### 10.4. Fast Network Mobility

In addition to endpoints, a network can be mobile, possibly changing xTRs. A "network" can be as small as a single router and as large as a whole site. This is different from site mobility in that it is fast and possibly short-lived, but different from endpoint mobility in that a whole prefix is changing RLOCs. However, the mechanisms are the same and there is no new overhead in LISP. A map request for any endpoint will return a binding for the entire mobile prefix.

If mobile networks become a more common occurrence, it may be useful to revisit the design of the mapping service and allow for dynamic updates of the database.

The issue of interactions between mobility and LISP needs to be explored further. Specific improvements to the entire system will depend on the details of mapping mechanisms. Mapping mechanisms should be evaluated on how well they support session continuity for mobile nodes.

#### 10.5. LISP Mobile Node Mobility

A mobile device can use the LISP infrastructure to achieve mobility by implementing the LISP encapsulation and decapsulation functions and acting as a simple ITR/ETR. By doing this, such a "LISP mobile node" can use topologically-independent EID IP addresses that are not advertised into and do not impose a cost on the global routing system. These EIDs are maintained at the edges of the mapping system (in LISP Map-Servers and Map-Resolvers) and are provided on demand to only the correspondents of the LISP mobile node.

Refer to the LISP Mobility Architecture specification [LISP-MN] for more details.

## 11. 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. 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 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 [RFC5496] 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).

To avoid any EID-based multicast state in the network core, the first approach is chosen for LISP-Multicast. Details for LISP-Multicast and Interworking with non-LISP sites is described in specification [MLISP].

## 12. Security Considerations

It is believed 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-Routability (see Section 3) mechanisms evidenced by the use of a 24-bit Nonce field in the LISP encapsulation header and a 64-bit Nonce field in the LISP control message.

The nonce, coupled with the ITR accepting only solicited Map-Replies provides a basic level of security, in many ways similar to the security experienced in the current Internet routing system. It is hard for off-path attackers to launch attacks against these LISP mechanisms, as they do not have the nonce values. Sending a large number of packets to accidentally find the right nonce value is possible, but would already by itself be a denial-of-service attack. On-path attackers can perform far more serious attacks, but on-path attackers can launch serious attacks in the current Internet as well, including eavesdropping, blocking or redirecting traffic. See more discussion on this topic in Section 6.1.5.1.

LISP does not rely on a PKI 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 Map-Requests and Map-Replies to the control plane as well as rate-limiting the number of data-triggered Map-Replies.

An incorrectly implemented or malicious ITR might choose to ignore the priority and weights provided by the ETR in its Map-Reply. This traffic steering would be limited to the traffic that is sent by this ITR's site, and no more severe than if the site initiated a bandwidth DoS attack on (one of) the ETR's ingress links. The ITR's site would typically gain no benefit from not respecting the weights, and would likely to receive better service by abiding by them.

To deal with map-cache exhaustion attempts in an ITR/PITR, the implementation should consider putting a maximum cap on the number of entries stored with a reserve list for special or frequently accessed sites. This should be a configuration policy control set by the network administrator who manages ITRs and PITRs. When overlapping EID-prefixes occur across multiple map-cache entries, the integrity of the set must be wholly maintained. So if a more-specific entry cannot be added due to reaching the maximum cap, then none of the less specifics should be stored in the map-cache.

Given that the ITR/PITR maintains a cache of EID-to-RLOC mappings, cache sizing and maintenance is an issue to be kept in mind during implementation. It is a good idea to have instrumentation in place to detect thrashing of the cache. Implementation experimentation will be used to determine which cache management strategies work best. In general, it is difficult to defend against cache trashing attacks. It should be noted that an undersized cache in an ITR/PITR not only causes adverse affect on the site or region they support, but may also cause increased Map-Request load on the mapping system.

"Piggybacked" mapping data discussed in Section 6.1.3 specifies how to handle such mappings and includes the possibility for an ETR to temporarily accept such a mapping before verification when running in "trusted" environments. In such cases, there is a potential threat that a fake mapping could be inserted (even if only for a short period) into a map-cache. As noted in Section 6.1.3, an ETR MUST be specifically configured to run in such a mode and might usefully only consider some specific ITRs as also running in that same trusted environment.

There is a security risk implicit in the fact that ETRs generate the EID prefix to which they are responding. An ETR can claim a shorter prefix than it is actually responsible for. Various mechanisms to ameliorate or resolve this issue will be examined in the future, [LISP-SEC].

Spoofing of inner header addresses of LISP encapsulated packets is possible like with any tunneling mechanism. ITRs MUST verify the source address of a packet to be an EID that belongs to the site's EID-prefix range prior to encapsulation. An ETR must only decapsulate and forward datagrams with an inner header destination that matches one of its EID-prefix ranges. If, upon receipt and decapsulation, the destination EID of a datagram does not match one of the ETR's configured EID-prefixes, the ETR MUST drop the datagram. If a LISP encapsulated packet arrives at an ETR, it SHOULD compare the inner header source EID address and the outer header source RLOC address with the mapping that exists in the mapping database. Then when spoofing attacks occur, the outer header source RLOC address can be used to trace back the attack to the source site, using existing operational tools.

This experimental specification does not address automated key management (AKM). BCP 107 provides guidance in this area. In addition, at the time of this writing, substantial work is being undertaken to improve security of the routing system [KARP], [RPKI], [BGP-SEC], [LISP-SEC]. Future work on LISP should address BCP-107 as well as other open security considerations, which may require changes to this specification.

### 13. Network Management Considerations

Considerations for Network Management tools exist so the LISP protocol suite can be operationally managed. The mechanisms can be found in [LISP-MIB] and [LISP-LIG].



#### 14. IANA Considerations

This section provides guidance to the Internet Assigned Numbers Authority (IANA) regarding registration of values related to the LISP specification, in accordance with BCP 26 and RFC 5226 [RFC5226].

There are four name spaces in LISP that require registration:

- o LISP IANA registry allocations should not be made for purposes unrelated to LISP routing or transport protocols.
- o The following policies are used here with the meanings defined in BCP 26: "Specification Required", "IETF Review", "Experimental Use", "First Come First Served".

##### 14.1. LISP ACT and Flag Fields

New ACT values (Section 6.1.4) can be allocated through IETF review or IESG approval. Four values have already been allocated by this specification (Section 6.1.4).

In addition, the LISP protocol has a number of flag and reserved fields, such as the LISP header flags field (Section 5.3). New bits for flags can be taken into use from these fields through IETF review or IESG approval, but these need not be managed by IANA.

##### 14.2. LISP Address Type Codes

LISP Address [LCAF] type codes have a range from 0 to 255. New type codes MUST be allocated consecutively starting at 0. Type Codes 0 - 127 are to be assigned by IETF review or IESG approval.

Type Codes 128 - 255 are available on a First Come First Served policy.

This registry, initially empty, is constructed for future-use experimental work of LCAF values. See [LCAF] for details for other possible unapproved address encodings. The unapproved LCAF encodings are an area for further study and experimentation.

#### 14.3. LISP UDP Port Numbers

The IANA registry has allocated UDP port numbers 4341 and 4342 for lisp-data and lisp-control operation, respectively. IANA is requested to update the description for udp ports 4341 and 4342 as follows:

lisp-data	4341 udp	LISP Data Packets
lisp-control	4342 udp	LISP Control Packets

#### 14.4. LISP Key ID Numbers

The following Key ID values are defined by this specification as used in any packet type that references a Key ID field:

Name	Number	Defined in
None	0	n/a
HMAC-SHA-1-96	1	[RFC2404]
HMAC-SHA-256-128	2	[RFC6234]

Number values are in the range of 0 to 65355. The allocation of values is on a first come first serve basis.

## 15. Known Open Issues and Areas of Future Work

As an experimental specification, this work is, by definition, incomplete. Specific areas where additional experience and work are needed include:

- o At present, only [ALT] is defined for implementing a database of EID-to-RLOC mapping information. Additional research on other mapping database systems is strongly encouraged.
- o Failure and recovery of LISP site partitioning (see Section 6.4), in the presence of redundant configuration (see Section 8.5) needs further research and experimentation.
- o The characteristics of map-cache management under exceptional conditions, such as denial-of-service attacks are not fully understood. Further experience is needed to determine whether current caching methods are practical or in need of further development. In particular, the performance, scaling and security characteristics of the map-cache will be discovered as part of this experiment. Performance metrics to be observed are packet reordering associated with the LISP data probe and loss of the first packet in a flow associated with map-caching. The impact of these upon TCP will be observed. See Section 12 for additional thoughts and considerations.
- o Preliminary work has been done to ensure that sites employing LISP can interconnect with the rest of the Internet. This work is documented in [INTERWORK], but further experimentation and experience is needed.
- o At present, no mechanism for automated key management for message authentication is defined. Addressing automated key management is necessary before this specification could be developed into a standards track RFC. See Section 12 for further details regarding security considerations.
- o In order to maintain security and stability, Internet Protocols typically isolate the control and data planes. Therefore, user activity cannot cause control plane state to be created or destroyed. LISP does not maintain this separation. The degree to which the loss of separation impacts security and stability is a topic for experimental observation.
- o LISP allows for different mapping database systems to be used. While only one [ALT] is currently well-defined, each mapping database will likely have some impact on the security of the EID-to-RLOC mappings. How each mapping database system's security

properties impact on LISP overall is for further study.

- o An examination of the implications of LISP on Internet traffic, applications, routers, and security is needed. This will help to understand the consequences for network stability, routing protocol function, routing scalability, migration and backward compatibility, and implementation scalability (as influenced by additional protocol components, additional state, and additional processing for encapsulation, decapsulation, liveness).
- o Experiments need to verify that LISP produces no significant change in the behavior of protocols run between end-systems over a LISP infrastructure versus being run directly between those same end-systems.
- o Experiments need to verify that the issues raised in the Critique section of [RFC6115] are either insignificant or have been addressed by updates to the LISP protocol.

Other LISP documents may also include open issues and areas for future work.

## 16. References

### 16.1. Normative References

- [ALT]        Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "LISP Alternative Topology (LISP-ALT)", draft-ietf-lisp-alt-10.txt (work in progress).
- [LISP-MS]   Farinacci, D. and V. Fuller, "LISP Map Server", draft-ietf-lisp-ms-16.txt (work in progress).
- [RFC0768]   Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.
- [RFC0791]   Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.
- [RFC1918]   Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.
- [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2404]   Madson, C. and R. Glenn, "The Use of HMAC-SHA-1-96 within ESP and AH", RFC 2404, November 1998.
- [RFC2460]   Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
- [RFC3168]   Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.
- [RFC3232]   Reynolds, J., "Assigned Numbers: RFC 1700 is Replaced by an On-line Database", RFC 3232, January 2002.
- [RFC4086]   Eastlake, D., Schiller, J., and S. Crocker, "Randomness Requirements for Security", BCP 106, RFC 4086, June 2005.
- [RFC4632]   Fuller, V. and T. Li, "Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan", BCP 122, RFC 4632, August 2006.
- [RFC5226]   Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

- [RFC5496]    Wijnands, IJ., Boers, A., and E. Rosen, "The Reverse Path Forwarding (RPF) Vector TLV", RFC 5496, March 2009.
- [RFC5944]    Perkins, C., "IP Mobility Support for IPv4, Revised", RFC 5944, November 2010.
- [RFC6115]    Li, T., "Recommendation for a Routing Architecture", RFC 6115, February 2011.
- [RFC6234]    Eastlake, D. and T. Hansen, "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", RFC 6234, May 2011.
- [RFC6275]    Perkins, C., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, July 2011.
- [VERSIONING]    Iannone, L., Saucez, D., and O. Bonaventure, "LISP Mapping Versioning", draft-ietf-lisp-map-versioning-09.txt (work in progress).

#### 16.2. Informative References

- [AFI]        IANA, "Address Family Indicators (AFIs)", ADDRESS FAMILY NUMBERS  
<http://www.iana.org/assignments/address-family-numbers>.
- [AFI-REGISTRY]    IANA, "Address Family Indicators (AFIs)", ADDRESS FAMILY NUMBER registry [http://www.iana.org/assignments/](http://www.iana.org/assignments/address-family-numbers/)  
[address-family-numbers/](http://www.iana.org/assignments/address-family-numbers/)  
[address-family-numbers.xml#address-family-numbers-1](http://www.iana.org/assignments/address-family-numbers/).
- [BGP-SEC]    Lepinski, M., "An Overview of BGPSEC",  
draft-lepinski-bgpsec-overview-00.txt (work in progress),  
March 2011.
- [CHIAPPA]    Chiappa, J., "Endpoints and Endpoint names: A Proposed Enhancement to the Internet Architecture", Internet-Draft <http://www.chiappa.net/~jnc/tech/endpoints.txt>.
- [CONS]        Farinacci, D., Fuller, V., and D. Meyer, "LISP-CONS: A Content distribution Overlay Network Service for LISP",  
draft-meyer-lisp-cons-04.txt (work in progress).
- [EMACS]        Brim, S., Farinacci, D., Meyer, D., and J. Curran, "EID Mappings Multicast Across Cooperating Systems for LISP",  
draft-curran-lisp-emacs-00.txt (work in progress).

- [INTERWORK]      Lewis, D., Meyer, D., Farinacci, D., and V. Fuller,  
"Interworking LISP with IPv4 and IPv6",  
draft-ietf-lisp-interworking-06.txt (work in progress).
- [KARP]      Lebovitz, G. and M. Bhatia, "Keying and Authentication for  
Routing Protocols (KARP) Design Guidelines",  
draft-ietf-karp-design-guide-06.txt (work in progress),  
October 2011.
- [LCAF]      Farinacci, D., Meyer, D., and J. Snijders, "LISP Canonical  
Address Format", draft-ietf-lisp-lcaf-00.txt (work in  
progress).
- [LISA96]      Lear, E., Katinsky, J., Coffin, J., and D. Tharp,  
"Renumbering: Threat or Menace?", Usenix .
- [LISP-DEPLOY]      Jakab, L., Coras, F., Domingo-Pascual, J., and D. Lewis,  
"LISP Network Element Deployment Considerations",  
draft-ietf-lisp-deployment-05.txt (work in progress).
- [LISP-LIG]      Farinacci, D. and D. Meyer, "LISP Internet Groper (LIG)",  
draft-ietf-lisp-lig-06.txt (work in progress).
- [LISP-MAIN]      Farinacci, D., Fuller, V., Meyer, D., and D. Lewis,  
"Locator/ID Separation Protocol (LISP)",  
draft-farinacci-lisp-12.txt (work in progress).
- [LISP-MIB]      Schudel, G., Jain, A., and V. Moreno, "LISP MIB",  
draft-ietf-lisp-mib-07.txt (work in progress).
- [LISP-MN]      Farinacci, D., Fuller, V., Lewis, D., and D. Meyer, "LISP  
Mobility Architecture", draft-meyer-lisp-mn-08.txt (work  
in progress).
- [LISP-SEC]      Maino, F., Ermagon, V., Cabellos, A., Sausez, D., and O.  
Bonaventure, "LISP-Security (LISP-SEC)",  
draft-ietf-lisp-sec-04.txt (work in progress).
- [LOC-ID-ARCH]      Meyer, D. and D. Lewis, "Architectural Implications of  
Locator/ID Separation",  
draft-meyer-loc-id-implications-02.txt (work in progress).

- [MLISP]      Farinacci, D., Meyer, D., Zwiebel, J., and S. Venaas, "LISP for Multicast Environments", draft-ietf-lisp-multicast-14.txt (work in progress).
- [NERD]      Lear, E., "NERD: A Not-so-novel EID to RLOC Database", draft-lear-lisp-nerd-08.txt (work in progress).
- [OPENLISP]      Iannone, L. and O. Bonaventure, "OpenLISP Implementation Report", draft-iannone-openlisp-implementation-01.txt (work in progress).
- [RADIR]      Narten, T., "Routing and Addressing Problem Statement", draft-narten-radir-problem-statement-05.txt (work in progress).
- [RFC1034]      Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, November 1987.
- [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.
- [RFC3261]      Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.
- [RFC4192]      Baker, F., Lear, E., and R. Droms, "Procedures for Renumbering an IPv6 Network without a Flag Day", RFC 4192, September 2005.
- [RFC4866]      Arkko, J., Vogt, C., and W. Haddad, "Enhanced Route Optimization for Mobile IPv6", RFC 4866, May 2007.
- [RFC4984]      Meyer, D., Zhang, L., and K. Fall, "Report from the IAB Workshop on Routing and Addressing", RFC 4984, September 2007.
- [RPKI]      Lepinski, M., "An Infrastructure to Support Secure Internet Routing", draft-ietf-sidr-arch-13.txt (work in progress), February 2011.
- [UDP-TUNNELS]      Eubanks, M. and P. Chimento, "UDP Checksums for Tunneled



Packets", draft-ietf-6man-udpchecksums-05.txt (work in progress), October 2012.

[UDP-ZERO]

Fairhurst, G. and M. Westerland, "IPv6 UDP Checksum Considerations", draft-ietf-6man-udpzero-07.txt (work in progress), October 2012.

## Appendix A. Acknowledgments

An initial thank you goes to Dave Oran for planting the seeds for the initial ideas for LISP. His consultation continues to provide value to the LISP authors.

A special and appreciative thank you goes to Noel Chiappa for providing architectural impetus over the past decades on separation of location and identity, as well as detailed review of the LISP architecture and documents, coupled with enthusiasm for making LISP a practical and incremental transition for the Internet.

The authors would like to gratefully acknowledge many people who have contributed discussion and ideas to the making of this proposal. They include Scott Brim, Andrew Partan, John Zwiebel, Jason Schiller, Lixia Zhang, Dorian Kim, Peter Schoenmaker, Vijay Gill, Geoff Huston, David Conrad, Mark Handley, Ron Bonica, Ted Seely, Mark Townsley, Chris Morrow, Brian Weis, Dave McGrew, Peter Lothberg, Dave Thaler, Eliot Lear, Shane Amante, Ved Kafle, Olivier Bonaventure, Luigi Iannone, Robin Whittle, Brian Carpenter, Joel Halpern, Terry Manderson, Roger Jorgensen, Ran Atkinson, Stig Venaas, Iljitsch van Beijnum, Roland Bless, Dana Blair, Bill Lynch, Marc Woolward, Damien Saucez, Damian Lezama, Attila De Groot, Parantap Lahiri, David Black, Roque Gagliano, Isidor Kouvelas, Jesper Skriver, Fred Templin, Margaret Wasserman, Sam Hartman, Michael Hofling, Pedro Marques, Jari Arkko, Gregg Schudel, Srinivas Subramanian, Amit Jain, Xu Xiaohu, Dhirendra Trivedi, Yakov Rekhter, John Scudder, John Drake, Dimitri Papadimitriou, Ross Callon, Selina Heimlich, Job Snijders, Vina Ermagan, Albert Cabellos, Fabio Maino, Victor Moreno, Chris White, Clarence Filsfils, and Alia Atlas.

This work originated in the Routing Research Group (RRG) of the IRTF. The individual submission [LISP-MAIN] was converted into this IETF LISP working group draft.

The LISP working group would like to give a special thanks to Jari Arkko, the Internet Area AD at the time the set of LISP documents were being prepared for IESG last call, for his meticulous review and detail commentary on the 7 working group last call drafts progressing toward experimental RFCs.

## Appendix B. Document Change Log

### B.1. Changes to draft-ietf-lisp-24.txt

- o Posted November 2012 for final pre-RFC version.
- o Move draft-ietf-6man-udpchecksums reference back to Informative References section.

### B.2. Changes to draft-ietf-lisp-23.txt

- o Posted May 2012 for final pre-RFC version.
- o Move only the reference draft-ietf-6man-udpzero to the Informative References section. Leave the draft-ietf-6man-udpchecksums reference in the Normative References section. After talking to many people involved with this issue at Paris IETF, all thought this would be an acceptable change.
- o Added text to IANA Considerations section 14.4 to reflect IANA comments about allocating Key-ID numbers.

### B.3. Changes to draft-ietf-lisp-22.txt

- o Posted February 2012 to reflect final DISCUSS comments from Adrian Farrel.

### B.4. Changes to draft-ietf-lisp-21.txt

- o Posted February 2012 to reflect DISCUSS comments from Adrian Farrel, Stewart Bryant, and Wesley Eddy.

### B.5. Changes to draft-ietf-lisp-20.txt

- o Posted January 2012 for resolution to Adrian Farrel's security comments as well as additions to the end of section 2, Elwyn Davies Gen-Art comments, and Ralph Droms' IANA and EID definition comments.

### B.6. Changes to draft-ietf-lisp-19.txt

- o Posted January 2012 for Stephen Farrell's comment resolution.

### B.7. Changes to draft-ietf-lisp-18.txt

- o Posted December 2011 after reflecting comments from IANA.

- o Create reference to sections 5.4.1 and 5.4.2 about DF bit setting from section 5.3.
- o Inserted two references for Route-Returnability and on-path attacks in Security Considerations section.

B.8. Changes to draft-ietf-lisp-17.txt

- o Posted December 2011 after IETF last call comments.
- o Make Map-Notify port assignment be 4342 in both source and destination ports. This change was agreed on and put in [LISP-MS] but was not updated in this spec.

B.9. Changes to draft-ietf-lisp-16.txt

- o Posted October 2011 after AD review by Jari.

B.10. Changes to draft-ietf-lisp-15.txt

- o Posted July 2011. Fixing IDnits errors.
- o Change description on how to select a source address for RLOC-probe Map-Replies to refer to the "EID-to-RLOC Map-Reply Message" section.

B.11. Changes to draft-ietf-lisp-14.txt

- o Post working group last call and pre-IESG last call review.
- o Indicate that an ICMP Unreachable message should be sent when a packet matches a drop-based negative map-cache entry.
- o Indicate how a map-cache set of overlapping EID-prefixes must maintain integrity when the map-cache maximum cap is reached.
- o Add Joel's description for the definition of an EID, that the bit string value can be an RLOC for another device in abstract but the architecture allows it to be an EID of one device and the same value as an RLOC for another device.
- o In the "Tunnel Encapsulation Details" section, indicate that 4 combinations of encapsulation are supported.
- o Add what ETR should do for a Data-Probe when received for a destination EID outside of its EID-prefix range. This was added in the Data Probe definition section.

- o Added text indicating that more-specific EID-prefixes must not be removed when less-specific entries stay in the map-cache. This is to preserve the integrity of the EID-prefix set.
- o Add clarifying text in the Security Considerations section about how an ETR must not decapsulate and forward a packet that is not for its configured EID-prefix range.

B.12. Changes to draft-ietf-lisp-13.txt

- o Posted June 2011 to complete working group last call.
- o Tracker item 87. Put Yakov suggested wording in the EID-prefix definition section to reference [INTERWORK] and [LISP-DEPLOY] about discussion on transition and access mechanisms.
- o Change "ITRs" to "ETRs" in the Locator Status Bit definition section and data packet description section per Damien's comment.
- o Remove the normative reference to [LISP-SEC] when describing the S-bit in the ECM and Map-Reply headers.
- o Tracker item 54. Added text from John Scudder in the "Packets Egressing a LISP Site" section.
- o Add sentence to the "Reencapsulating Tunnel" definition about how reencapsulation loops can occur when not coordinating among multiple mapping database systems.
- o Remove "In theory" from a sentence in the Security Considerations section.
- o Remove Security Area Statement title and reword section with Eliot's provided text. The text was agreed upon by LISP-WG chairs and Security ADs.
- o Remove word "potential" from the over-claiming paragraph of the Security Considerations section per Stephen's request.
- o Wordsmithing and other editorial comments from Alia.

B.13. Changes to draft-ietf-lisp-12.txt

- o Posted April 2011.
- o Tracker item 87. Provided rewording how an EID-prefix can be reused in the definition section of "EID-prefix".

- o Tracker item 95. Change "eliminate" to "defer" in section 4.1.
- o Tracker item 110. Added that the Mapping Protocol Data field in the Map-Reply message is only used when needed by the particular Mapping Database System.
- o Tracker item 111. Indicate that if an LSB that is associated with an anycast address, that there is at least one RLOC that is up.
- o Tracker item 108. Make clear the R-bit does not define RLOC path reachability.
- o Tracker item 107. Indicate that weights are relative to each other versus requiring an addition of up to 100%.
- o Tracker item 46. Add a sentence how LISP products should be sized for the appropriate demand so cache thrashing is avoided.
- o Change some references of RFC 5226 to [AFI] per Luigi.
- o Per Luigi, make reference to "EID-AFI" consistent to "EID-prefix-AFI".
- o Tracker item 66. Indicate that appending locators to a locator-set is done when the added locators are lexicographically greater than the previous ones in the set.
- o Tracker item 87. Once again reword the definition of the EID-prefix to reflect recent comments.
- o Tracker item 70. Added text to security section on what the implications could be if an ITR does not obey priority and weights from a Map-Reply message.
- o Tracker item 54. Added text to the new section titled "Packets Egressing a LISP Site" to describe the implications when two or more ITRs exist at a site where only one ITR is used for egress traffic and when there is a shift of traffic to the others, how the map-cache will need to be populated in those new egress ITRs.
- o Tracker item 33. Make more clear in the Routing Locator Selection section what an ITR should do when it sees an R-bit of 0 in a locator-record of a Map-Reply.
- o Tracker item 33. Add paragraph to the EID Reachability section indicating that site partitioning is under investigation.

- o Tracker item 58. Added last paragraph of Security Considerations section about how to protect inner header EID address spoofing attacks.
- o Add suggested Sam text to indicate that all security concerns need not be addressed for moving document to Experimental RFC status. Put this in a subsection of the Security Considerations section.

#### B.14. Changes to draft-ietf-lisp-11.txt

- o Posted March 30, 2011.
- o Change IANA URL. The URL we had pointed to a general protocol numbers page.
- o Added the "s" bit to the Map-Request to allow SMR-invoked Map-Requests to be sent to a MN ETR via the map-server.
- o Generalize text for the definition of Reencapsulating tunnels.
- o Add paragraph suggested by Joel to explain how implementation experimentation will be used to determine the proper cache management techniques.
- o Add Yakov provided text for the definition of "EID-to-RLOC Database".
- o Add reference in Section 8, Deployment Scenarios, to the draft-jakab-lisp-deploy-02.txt draft.
- o Clarify sentence about no hardware changes needed to support LISP encapsulation.
- o Add paragraph about what is the procedure when a locator is inserted in the middle of a locator-set.
- o Add a definition for Locator Status Bits so we can emphasize they are used as a hint for router up/down status and not path reachability.
- o Change "BGP RIB" to "RIB" per Clarence's comment.
- o Fixed complaints by IDnits.
- o Add subsection to Security Considerations section indicating how EID-prefix overclaiming in Map-Replies is for further study and add a reference to LISP-SEC.

B.15. Changes to draft-ietf-lisp-10.txt

- o Posted March 2011.
- o Add p-bit to Map-Request so there is documentary reasons to know when a PITR has sent a Map-Request to an ETR.
- o Add Map-Notify message which is used to acknowledge a Map-Register message sent to a Map-Server.
- o Add M-bit to the Map-Register message so an ETR that wants an acknowledgment for the Map-Register can request one.
- o Add S-bit to the ECM and Map-Reply messages to describe security data that can be present in each message. Then refer to [LISP-SEC] for expansive details.
- o Add Network Management Considerations section and point to the MIB and LIG drafts.
- o Remove the word "simple" per Yakov's comments.

B.16. Changes to draft-ietf-lisp-09.txt

- o Posted October 2010.
- o Add to IANA Consideration section about the use of LCAF Type values that accepted and maintained by the IANA registry and not the LCAF specification.
- o Indicate that implementations should be able to receive LISP control messages when either UDP port is 4342, so they can be robust in the face of intervening NAT boxes.
- o Add paragraph to SMR section to indicate that an ITR does not need to respond to an SMR-based Map-Request when it has no map-cache entry for the SMR source's EID-prefix.

B.17. Changes to draft-ietf-lisp-08.txt

- o Posted August 2010.
- o In section 6.1.6, remove statement about setting TTL to 0 in Map-Register messages.
- o Clarify language in section 6.1.5 about Map-Replying to Data-Probes or Map-Requests.



- o Indicate that outer TTL should only be copied to inner TTL when it is less than inner TTL.
- o Indicate a source-EID for RLOC-probes are encoded with an AFI value of 0.
- o Indicate that SMRs can have a global or per SMR destination rate-limiter.
- o Add clarifications to the SMR procedures.
- o Add definitions for "client-side" and "server-side" terms used in this specification.
- o Clear up language in section 6.4, last paragraph.
- o Change ACT of value 0 to "no-action". This is so we can RLOC-probe a PETR and have it return a Map-Reply with a locator-set of size 0. The way it is spec'ed the map-cache entry has action "dropped". Drop-action is set to 3.
- o Add statement about normalizing locator weights.
- o Clarify R-bit definition in the Map-Reply locator record.
- o Add section on EID Reachability within a LISP site.
- o Clarify another disadvantage of using anycast locators.
- o Reworded Abstract.
- o Change section 2.0 Introduction to remove obsolete information such as the LISP variant definitions.
- o Change section 5 title from "Tunneling Details" to "LISP Encapsulation Details".
- o Changes to section 5 to include results of network deployment experience with MTU. Recommend that implementations use either the stateful or stateless handling.
- o Make clarification wordsmithing to Section 7 and 8.
- o Identify that if there is one locator in the locator-set of a map-cache entry, that an SMR from that locator should be responded to by sending the the SMR-invoked Map-Request to the database mapping system rather than to the RLOC itself (which may be unreachable).

- o When describing Unicast and Multicast Weights indicate the the values are relative weights rather than percentages. So it doesn't imply the sum of all locator weights in the locator-set need to be 100.
- o Do some wordsmithing on copying TTL and TOS fields.
- o Numerous wordsmithing changes from Dave Meyer. He fine toothed combed the spec.
- o Removed Section 14 "Prototype Plans and Status". We felt this type of section is no longer appropriate for a protocol specification.
- o Add clarification text for the IRC description per Damien's commentary.
- o Remove text on copying nonce from SMR to SMR-invoked Map- Request per Vina's comment about a possible DoS vector.
- o Clarify (S/2 + H) in the stateless MTU section.
- o Add text to reflect Damien's comment about the description of the "ITR-RLOC Address" field in the Map-Request. that the list of RLOC addresses are local addresses of the Map-Requester.

B.18. Changes to draft-ietf-lisp-07.txt

- o Posted April 2010.
- o Added I-bit to data header so LSB field can also be used as an Instance ID field. When this occurs, the LSB field is reduced to 8-bits (from 32-bits).
- o Added V-bit to the data header so the 24-bit nonce field can also be used for source and destination version numbers.
- o Added Map-Version 12-bit value to the EID-record to be used in all of Map-Request, Map-Reply, and Map-Register messages.
- o Added multiple ITR-RLOC fields to the Map-Request packet so an ETR can decide what address to select for the destination of a Map-Reply.
- o Added L-bit (Local RLOC bit) and p-bit (Probe-Reply RLOC bit) to the Locator-Set record of an EID-record for a Map-Reply message. The L-bit indicates which RLOCs in the locator-set are local to the sender of the message. The P-bit indicates which RLOC is the

source of a RLOC-probe Reply (Map-Reply) message.

- o Add reference to the LISP Canonical Address Format [LCAF] draft.
- o Made editorial and clarification changes based on comments from Dhirendra Trivedi.
- o Added wordsmithing comments from Joel Halpern on DF=1 setting.
- o Add John Zwiebel clarification to Echo Nonce Algorithm section 6.3.1.
- o Add John Zwiebel comment about expanding on proxy-map-reply bit for Map-Register messages.
- o Add NAT section per Ron Bonica comments.
- o Fix IDnits issues per Ron Bonica.
- o Added section on Virtualization and Segmentation to explain the use if the Instance ID field in the data header.
- o There are too many P-bits, keep their scope to the packet format description and refer to them by name every where else in the spec.
- o Scanned all occurrences of "should", "should not", "must" and "must not" and uppercased them.
- o John Zwiebel offered text for section 4.1 to modernize the example. Thanks Z!
- o Make it more clear in the definition of "EID-to-RLOC Database" that all ETRs need to have the same database mapping. This reflects a comment from John Scudder.
- o Add a definition "Route-returnability" to the Definition of Terms section.
- o In section 9.2, add text to describe what the signature of traceroute packets can look like.
- o Removed references to Data Probe for introductory example. Data-probes are still part of the LISP design but not encouraged.
- o Added the definition for "LISP site" to the Definition of Terms" section.

B.19. Changes to draft-ietf-lisp-06.txt

Editorial based changes:

- o Posted December 2009.
- o Fix typo for flags in LISP data header. Changed from "4" to "5".
- o Add text to indicate that Map-Register messages must contain a computed UDP checksum.
- o Add definitions for PITR and PETR.
- o Indicate an AFI value of 0 is an unspecified address.
- o Indicate that the TTL field of a Map-Register is not used and set to 0 by the sender. This change makes this spec consistent with [LISP-MS].
- o Change "... yield a packet size of L octets" to "... yield a packet size greater than L octets".
- o Clarify section 6.1.5 on what addresses and ports are used in Map-Reply messages.
- o Clarify that LSBs that go beyond the number of locators do not to be SMRed when the locator addresses are greater lexicographically than the locator in the existing locator-set.
- o Add Gregg, Srini, and Amit to acknowledgment section.
- o Clarify in the definition of a LISP header what is following the UDP header.
- o Clarify "verifying Map-Request" text in section 6.1.3.
- o Add Xu Xiaohu to the acknowledgment section for introducing the problem of overlapping EID-prefixes among multiple sites in an RRG email message.

Design based changes:

- o Use stronger language to have the outer IPv4 header set DF=1 so we can avoid fragment reassembly in an ETR or PETR. This will also make IPv4 and IPv6 encapsulation have consistent behavior.
- o Map-Requests should not be sent in ECM with the Probe bit is set. These type of Map-Requests are used as RLOC-probes and are sent

directly to locator addresses in the underlying network.

- o Add text in section 6.1.5 about returning all EID-prefixes in a Map-Reply sent by an ETR when there are overlapping EID-prefixes configure.
- o Add text in a new subsection of section 6.1.5 about dealing with Map-Replies with coarse EID-prefixes.

#### B.20. Changes to draft-ietf-lisp-05.txt

- o Posted September 2009.
- o Added this Document Change Log appendix.
- o Added section indicating that encapsulated Map-Requests must use destination UDP port 4342.
- o Don't use AH in Map-Registers. Put key-id, auth-length, and auth-data in Map-Register payload.
- o Added Jari to acknowledgment section.
- o State the source-EID is set to 0 when using Map-Requests to refresh or RLOC-probe.
- o Make more clear what source-RLOC should be for a Map-Request.
- o The LISP-CONS authors thought that the Type definitions for CONS should be removed from this specification.
- o Removed nonce from Map-Register message, it wasn't used so no need for it.
- o Clarify what to do for unspecified Action bits for negative Map-Replies. Since No Action is a drop, make value 0 Drop.

#### B.21. Changes to draft-ietf-lisp-04.txt

- o Posted September 2009.
- o How do deal with record count greater than 1 for a Map-Request. Damien and Joel comment. Joel suggests: 1) Specify that senders compliant with the current document will always set the count to 1, and note that the count is included for future extensibility. 2) Specify what a receiver compliant with the draft should do if it receives a request with a count greater than 1. Presumably, it should send some error back?

- o Add Fred Templin in acknowledgment section.
- o Add Margaret and Sam to the acknowledgment section for their great comments.
- o Say more about LAGs in the UDP section per Sam Hartman's comment.
- o Sam wants to use MAY instead of SHOULD for ignoring checksums on ETR. From the mailing list: "You'd need to word it as an ITR MAY send a zero checksum, an ETR MUST accept a 0 checksum and MAY ignore the checksum completely. And of course we'd need to confirm that can actually be implemented. In particular, hardware that verifies UDP checksums on receive needs to be checked to make sure it permits 0 checksums."
- o Margaret wants a reference to <http://www.ietf.org/id/draft-eubanks-chimento-6man-00.txt>.
- o Fix description in Map-Request section. Where we describe Map-Reply Record, change "R-bit" to "M-bit".
- o Add the mobility bit to Map-Replies. So PITRs don't probe so often for MNs but often enough to get mapping updates.
- o Indicate SHA1 can be used as well for Map-Registers.
- o More Fred comments on MTU handling.
- o Isidor comment about spec'ing better periodic Map-Registers. Will be fixed in draft-ietf-lisp-ms-02.txt.
- o Margaret's comment on gleaning: "The current specification does not make it clear how long gleaned map entries should be retained in the cache, nor does it make it clear how/ when they will be validated. The LISP spec should, at the very least, include a (short) default lifetime for gleaned entries, require that they be validated within a short period of time, and state that a new gleaned entry should never overwrite an entry that was obtained from the mapping system. The security implications of storing "gleaned" entries should also be explored in detail."
- o Add section on RLOC-probing per working group feedback.
- o Change "loc-reach-bits" to "loc-status-bits" per comment from Noel.
- o Remove SMR-bit from data-plane. Dino prefers to have it in the control plane only.

- o Change LISP header to allow a "Research Bit" so the Nonce and LSB fields can be turned off and used for another future purpose. For Luigi et al versioning convergence.
- o Add a N-bit to the data header suggested by Noel. Then the nonce field could be used when N is not 1.
- o Clarify that when E-bit is 0, the nonce field can be an echoed nonce or a random nonce. Comment from Jesper.
- o Indicate when doing data-gleaning that a verifying Map-Request is sent to the source-EID of the gleaned data packet so we can avoid map-cache corruption by a 3rd party. Comment from Pedro.
- o Indicate that a verifying Map-Request, for accepting mapping data, should be sent over the ALT (or to the EID).
- o Reference IPsec RFC 4302. Comment from Sam and Brian Weis.
- o Put E-bit in Map-Reply to tell ITRs that the ETR supports echo-nouncing. Comment by Pedro and Dino.
- o Jesper made a comment to loosen the language about requiring the copy of inner TTL to outer TTL since the text to get mixed-AF traceroute to work would violate the "MUST" clause. Changed from MUST to SHOULD in section 5.3.

#### B.22. Changes to draft-ietf-lisp-03.txt

- o Posted July 2009.
- o Removed loc-reach-bits longword from control packets per Damien comment.
- o Clarifications in MTU text from Roque.
- o Added text to indicate that the locator-set be sorted by locator address from Isidor.
- o Clarification text from John Zwiebel in Echo-Nonce section.

#### B.23. Changes to draft-ietf-lisp-02.txt

- o Posted July 2009.
- o Encapsulation packet format change to add E-bit and make loc-reach-bits 32-bits in length.

- o Added Echo-Nonce Algorithm section.
- o Clarification how ECN bits are copied.
- o Moved S-bit in Map-Request.
- o Added P-bit in Map-Request and Map-Reply messages to anticipate RLOC-Probe Algorithm.
- o Added to Mobility section to reference [LISP-MN].

B.24. Changes to draft-ietf-lisp-01.txt

- o Posted 2 days after draft-ietf-lisp-00.txt in May 2009.
- o Defined LEID to be a "LISP EID".
- o Indicate encapsulation use IPv4 DF=0.
- o Added negative Map-Reply messages with drop, native-forward, and send-map-request actions.
- o Added Proxy-Map-Reply bit to Map-Register.

B.25. Changes to draft-ietf-lisp-00.txt

- o Posted May 2009.
- o Rename of draft-farinacci-lisp-12.txt.
- o Acknowledgment to RRG.



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

Email: dmm@cisco.com

Darrel Lewis  
cisco Systems  
170 Tasman Drive  
San Jose, CA  
USA

Email: darlewis@cisco.com



Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: June 8, 2012

V. Fuller  
D. Farinacci  
D. Meyer  
D. Lewis  
Cisco  
December 6, 2011

LISP Alternative Topology (LISP+ALT)  
draft-ietf-lisp-alt-10.txt

## Abstract

This document describes a simple distributed index system to be used by a Locator/ID Separation Protocol (LISP) Ingress Tunnel Router (ITR) or Map Resolver (MR) to find the Egress Tunnel Router (ETR) which holds the mapping information for a particular Endpoint Identifier (EID). The MR can then query that ETR to obtain the actual mapping information, which consists of a list of Routing Locators (RLOCs) for the EID. Termed the Alternative Logical Topology (ALT), the index is built as an overlay network on the public Internet using the Border Gateway Protocol (BGP) and the Generic Routing Encapsulation (GRE).

## Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on June 8, 2012.

## Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	4
2. Definition of Terms . . . . .	6
3. The LISP+ALT model . . . . .	9
3.1. Routeability of EIDs . . . . .	9
3.1.1. Mechanisms for an ETR to originate EID-prefixes . . . . .	10
3.1.2. Mechanisms for an ITR to forward to EID-prefixes . . . . .	10
3.1.3. Map Server Model preferred . . . . .	10
3.2. Connectivity to non-LISP sites . . . . .	10
3.3. Caveats on the use of Data Probes . . . . .	11
4. LISP+ALT: Overview . . . . .	12
4.1. ITR traffic handling . . . . .	13
4.2. EID Assignment - Hierarchy and Topology . . . . .	14
4.3. Use of GRE and BGP between LISP+ALT Routers . . . . .	15
5. EID-prefix Propagation and Map-Request Forwarding . . . . .	16
5.1. Changes to ITR behavior with LISP+ALT . . . . .	16
5.2. Changes to ETR behavior with LISP+ALT . . . . .	17
5.3. ALT Datagram forwarding failure . . . . .	17
6. BGP configuration and protocol considerations . . . . .	19
6.1. Autonomous System Numbers (ASNs) in LISP+ALT . . . . .	19
6.2. Sub-Address Family Identifier (SAFI) for LISP+ALT . . . . .	19
7. EID-prefix Aggregation . . . . .	20
7.1. Stability of the ALT . . . . .	20
7.2. Traffic engineering using LISP . . . . .	20
7.3. Edge aggregation and dampening . . . . .	21
7.4. EID assignment flexibility vs. ALT scaling . . . . .	21
8. Connecting sites to the ALT network . . . . .	23
8.1. ETRs originating information into the ALT . . . . .	23
8.2. ITRs Using the ALT . . . . .	23
9. IANA Considerations . . . . .	25
10. Security Considerations . . . . .	26
10.1. Apparent LISP+ALT Vulnerabilities . . . . .	26
10.2. Survey of LISP+ALT Security Mechanisms . . . . .	27
10.3. Use of new IETF standard BGP Security mechanisms . . . . .	27
11. Acknowledgments . . . . .	28
12. References . . . . .	29
12.1. Normative References . . . . .	29
12.2. Informative References . . . . .	29

Authors' Addresses . . . . . 30

## 1. Introduction

This document describes the LISP+ALT system, used by a [LISP] ITR or MR to find the ETR that holds the RLOC mapping information for a particular EID. The ALT network is built using the Border Gateway Protocol (BGP, [RFC4271]), the BGP multi-protocol extension [RFC4760], and the Generic Routing Encapsulation (GRE, [RFC2784]) to construct an overlay network of devices (ALT Routers) which operate on EID-prefixes and use EIDs as forwarding destinations.

ALT Routers advertise hierarchically-delegated segments of the EID namespace (i.e., prefixes) toward the rest of the ALT; they also forward traffic destined for an EID covered by one of those prefixes toward the network element that is authoritative for that EID and is the origin of the BGP advertisement for that EID-prefix. An Ingress Tunnel Router (ITR) uses this overlay to send a LISP Map-Request (defined in [LISP]) to the Egress Tunnel Router (ETR) that holds the EID-to-RLOC mapping for a matching EID-prefix. In most cases, an ITR does not connect directly to the overlay network but instead sends Map-Requests via a Map-Resolver (described in [LISP-MS]) which does. Likewise, in most cases, an ETR does not connect directly to the overlay network but instead registers its EID-prefixes with a Map-Server that advertises those EID-prefixes on to the ALT and forwards Map-Requests for them to the ETR.

It is important to note that the ALT does not distribute actual EID-to-RLOC mappings. What it does provide is a forwarding path from an ITR (or MR) which requires an EID-to-RLOC mapping to an ETR which holds that mapping. The ITR/MR uses this path to send an ALT Datagram (see Section 3) to an ETR which then responds with a Map-Reply containing the needed mapping information.

One design goal for LISP+ALT is to use existing technology wherever possible. To this end, the ALT is intended to be built using off-the-shelf routers which already implement the required protocols (BGP and GRE); little, if any, LISP-specific modifications should be needed for such devices to be deployed on the ALT (see Section 7 for aggregation requirements). Note, though, that organizational and operational considerations suggest that ALT Routers be both logically and physically separate from the "native" Internet packet transport system; deploying this overlay on those routers which are already participating in the global routing system and actively forwarding Internet traffic is not recommended.

This specification is experimental, and there are areas where further experience is needed to understand the best implementation strategy, operational model, and effects on Internet operations. These areas include:

- o application effects of on-demand route map discovery
- o tradeoff in connection setup time vs. ALT design and performance when using a Map Request instead of carrying initial user data in a Data Probe
- o best practical ways to build ALT hierarchies
- o effects of route leakage from ALT to the current Internet, particularly for LISP-to-non-LISP interworking
- o effects of exceptional situations, such as denial-of-service attacks

Experimentation, measurements, and deployment experience on these aspects is appreciated. While these issues are conceptually well-understood (e.g. an ALT lookup causes potential delay for the first packet destined to a given network), the real-world operational effects are much less clear.

The remainder of this document is organized as follows: Section 2 provides the definitions of terms used in this document. Section 3 outlines the LISP ALT model, where EID prefixes are routed across an overlay network. Section 4 provides a basic overview of the LISP Alternate Topology architecture, and Section 5 describes how the ALT uses BGP to propagate Endpoint Identifier reachability over the overlay network and Section 6 describes other considerations for using BGP on the ALT. Section 7 describes the construction of the ALT aggregation hierarchy, and Section 8 discusses how LISP+ALT elements are connected to form the overlay network.

## 2. Definition of Terms

This section provides high-level definitions of LISP concepts and components involved with and affected by LISP+ALT.

**Alternative Logical Topology (ALT):** The virtual overlay network made up of tunnels between LISP+ALT Routers. The Border Gateway Protocol (BGP) runs between ALT Routers and is used to carry reachability information for EID-prefixes. The ALT provides a way to forward Map-Requests (and, if supported, Data Probes) toward the ETR that "owns" an EID-prefix. As a tunneled overlay, its performance is expected to be quite limited so use of it to forward high-bandwidth flows of Data Probes is strongly discouraged (see Section 3.3 for additional discussion).

**ALT Router:** The devices which run on the ALT. The ALT is a static network built using tunnels between ALT Routers. These routers are deployed in a roughly-hierarchical mesh in which routers at each level in the topology are responsible for aggregating EID-prefixes learned from those logically "below" them and advertising summary prefixes to those logically "above" them. Prefix learning and propagation between ALT Routers is done using BGP. An ALT Router at the lowest level, or "edge" of the ALT, learns EID-prefixes from its "client" ETRs. See Section 3.1 for a description of how EID-prefixes are learned at the "edge" of the ALT. See also Section 6 for details on how BGP is configured between the different network elements. When an ALT Router receives an ALT Datagram, it looks up the destination EID in its forwarding table (composed of EID prefix routes it learned from neighboring ALT Routers) and forwards it to the logical next-hop on the overlay network.

**Endpoint ID (EID):** A 32-bit (for IPv4) or 128-bit (for ipv6) value used to identify the ultimate source or destination for a LISP-encapsulated packet. See [LISP] for details.

**EID-prefix:** A set of EIDs delegated in a power-of-two block. EID-prefixes are routed on the ALT (not on the global Internet) and are expected to be assigned in a hierarchical manner such that they can be aggregated by ALT Routers. Such a block is characterized by a prefix and a length. Note that while the ALT routing system considers an EID-prefix to be an opaque block of EIDs, an end site may put site-local, topologically-relevant structure (subnetting) into an EID-prefix for intra-site routing.



**Aggregated EID-prefixes:** A set of individual EID-prefixes that have been aggregated in the [RFC4632] sense.

**Map Server (MS):** An edge ALT Router that provides a registration function for non-ALT-connected ETRs, originates EID-prefixes into the ALT on behalf of those ETRs, and forwards Map-Requests to them. See [LISP-MS] for details.

**Map Resolver (MR):** An edge ALT Router that accepts an Encapsulated Map-Request from a non-ALT-connected ITR, decapsulates it, and forwards it on to the ALT toward the ETR which owns the requested EID-prefix. See [LISP-MS] for details.

**Ingress Tunnel Router (ITR):** A router which sends LISP Map-Requests or encapsulates IP datagrams with LISP headers, as defined in [LISP]. In this document, the term refers to any device implementing ITR functionality, including a Proxy-ITR (see [LISP-IW]). Under some circumstances, a LISP Map Resolver may also originate Map-Requests (see [LISP-MS]).

**Egress Tunnel Router (ETR):** A router which sends LISP Map-Replies in response to LISP Map-Requests and decapsulates LISP-encapsulated IP datagrams for delivery to end systems, as defined in [LISP]. In this document, the term refers to any device implementing ETR functionality, including a Proxy-ETR (see [LISP-IW]). Under some circumstances, a LISP Map Server may also respond to Map-Requests (see [LISP-MS]).

**Routing Locator (RLOC):** A routable IP address for a LISP tunnel router (ITR or ETR). Interchangeably referred to as a "locator" in this document. An RLOC is also the output of an EID-to-RLOC mapping lookup; an EID-prefix maps to one or more RLOCs. Typically, RLOCs are numbered from topologically-aggregatable blocks that are assigned to a site at each point where it attaches to the global Internet; where the topology is defined by the connectivity of provider networks, RLOCs can be thought of as Provider Aggregatable (PA) addresses. Routing for RLOCs is not carried on the ALT.

**EID-to-RLOC Mapping:** A binding between an EID-prefix and the set of RLOCs that can be used to reach it; sometimes referred to simply as a "mapping".

**EID-prefix Reachability:** An EID-prefix is said to be "reachable" if at least one of its locators is reachable. That is, an EID-prefix is reachable if the ETR that is authoritative for a given EID-to-RLOC mapping is reachable.

**Default Mapping:** A Default Mapping is a mapping entry for EID-prefix 0.0.0.0/0 (::/0 for ipv6). It maps to a locator-set used for all EIDs in the Internet. If there is a more specific EID-prefix in the mapping cache it overrides the Default Mapping entry. The Default Mapping can be learned by configuration or from a Map-Reply message.

**ALT Default Route:** An EID-prefix value of 0.0.0.0/0 (or ::/0 for ipv6) which may be learned from the ALT or statically configured on an edge ALT Router. The ALT-Default Route defines a forwarding path for a packet to be sent into the ALT on a router which does not have a full ALT forwarding database.

### 3. The LISP+ALT model

The LISP+ALT model uses the same basic query/response protocol that is documented in [LISP]. In particular, LISP+ALT provides two types of packet that an ITR can originate to obtain EID-to-RLOC mappings:

**Map-Request:** A Map-Request message is sent into the ALT to request an EID-to-RLOC mapping. The ETR which owns the mapping will respond to the ITR with a Map-Reply message. Since the ALT only forwards on EID destinations, the destination address of the Map-Request sent on the ALT must be an EID.

**Data Probe:** Alternatively, an ITR may encapsulate and send the first data packet destined for an EID with no known RLOCs into the ALT as a Data Probe. This might be done to minimize packet loss and to probe for the mapping. As above, the authoritative ETR for the EID-prefix will respond to the ITR with a Map-Reply message when it receives the data packet over the ALT. As a side-effect, the encapsulated data packet is delivered to the end-system at the ETR site. Note that the Data Probe's inner IP destination address, which is an EID, is copied to the outer IP destination address so that the resulting packet can be routed over the ALT. See Section 3.3 for caveats on the usability of Data Probes.

The term "ALT Datagram" is short-hand for a Map-Request or Data Probe to be sent into or forwarded on the ALT. Note that such packets use an RLOC as the outer header source IP address and an EID as the outer header destination IP address.

Detailed descriptions of the LISP packet types referenced by this document may be found in [LISP].

#### 3.1. Routeability of EIDs

A LISP EID has the same syntax as IP address and can be used, unaltered, as the source or destination of an IP datagram. In general, though, EIDs are not routable on the public Internet; LISP+ALT provides a separate, virtual network, known as the LISP Alternative Logical Topology (ALT) on which a datagram using an EID as an IP destination address may be transmitted. This network is built as an overlay on the public Internet using tunnels to interconnect ALT Routers. BGP runs over these tunnels to propagate path information needed to forward ALT Datagrams. Importantly, while the ETRs are the source(s) of the unaggregated EID-prefixes, LISP+ALT uses existing BGP mechanisms to aggregate this information.

### 3.1.1. Mechanisms for an ETR to originate EID-prefixes

There are three ways that an ETR may originate its mappings into the ALT:

1. By registration with a Map Server as documented in [LISP-MS]. This is the common case and is expected to be used by the majority of ETRs.
2. Using a "static route" on the ALT. Where no Map-Server is available, an edge ALT Router may be configured with a "static EID-prefix route" pointing to an ETR.
3. Edge connection to the ALT. If a site requires fine-grained control over how its EID-prefixes are advertised into the ALT, it may configure its ETR(s) with tunnel and BGP connections to edge ALT Routers.

### 3.1.2. Mechanisms for an ITR to forward to EID-prefixes

There are three ways that an ITR may send ALT Datagrams:

1. Through a Map Resolver as documented in [LISP-MS]. This is the common case and is expected to be used by the majority of ITRs.
2. Using a "default route". Where a Map Resolver is not available, an ITR may be configured with a static ALT Default Route pointing to an edge ALT Router.
3. Edge connection to the ALT. If a site requires fine-grained knowledge of what prefixes exist on the ALT, it may configure its ITR(s) with tunnel and BGP connections to edge ALT Routers.

### 3.1.3. Map Server Model preferred

The ALT-connected ITR and ETR cases are expected to be rare, as the Map Server/Map Resolver model is both simpler for an ITR/ETR operator to use, and provides a more general service interface to not only the ALT, but also to other mapping databases that may be developed in the future.

## 3.2. Connectivity to non-LISP sites

As stated above, EIDs used as IP addresses by LISP sites are not routable on the public Internet. This implies that, absent a mechanism for communication between LISP and non-LISP sites, connectivity between them is not possible. To resolve this problem, an "interworking" technology has been defined; see [LISP-IW] for

details.

### 3.3. Caveats on the use of Data Probes

It is worth noting that there has been a great deal of discussion and controversy about whether Data Probes are a good idea. On the one hand, using them offers a method of avoiding the "first packet drop" problem when an ITR does not have a mapping for a particular EID-prefix. On the other hand, forwarding data packets on the ALT would require that it either be engineered to support relatively high traffic rates, which is not generally feasible for a tunneled network, or that it be carefully designed to aggressively rate-limit traffic to avoid congestion or DoS attacks. There may also be issues caused by different latency or other performance characteristics between the ALT path taken by an initial Data Probe and the "Internet" path taken by subsequent packets on the same flow once a mapping is in place on an ITR. For these reasons, the use of Data Probes is not recommended at this time; they should only be originated on an ITR when explicitly configured to do so and such configuration should only be enabled when performing experiments intended to test the viability of using Data Probes.

#### 4. LISP+ALT: Overview

LISP+ALT is a hybrid push/pull architecture. Aggregated EID-prefixes are advertised among the ALT Routers and to those (rare) ITRs that are directly connected via a tunnel and BGP to the ALT. Specific EID-to-RLOC mappings are requested by an ITR (and returned by an ETR) using LISP when it sends a request either via a Map Resolver or to an edge ALT Router.

The basic idea embodied in LISP+ALT is to use BGP, running on a tunneled overlay network (the ALT), to establish reachability between ALT Routers. The ALT BGP Route Information Base (RIB) is comprised of EID-prefixes and associated next hops. ALT Routers interconnect using BGP and propagate EID-prefix updates among themselves. EID-prefix information is learned from ETRs at the "edge" of the ALT either through the use of the Map Server interface (the common case), static configuration, or by BGP-speaking ETRs.

Map Resolvers learn paths through the ALT to Map Servers for EID-prefixes. An ITR will normally use a Map Resolver to send its ALT Datagrams on to the ALT but may, in unusual cases (see Section 3.1.2), use a static ALT Default Route or connect to the ALT using BGP. Likewise, an ETR will normally register its prefixes in the mapping database using a Map Server but can sometimes (see Section 3.1.1) connect directly to the ALT using BGP. See [LISP-MS] for details on Map Servers and Map Resolvers.

Note that while this document specifies the use of Generic Routing Encapsulation (GRE) as a tunneling mechanism, there is no reason that parts of the ALT cannot be built using other tunneling technologies, particularly in cases where GRE does not meet security, management, or other operational requirements. References to "GRE tunnel" in later sections of this document should therefore not be taken as prohibiting or precluding the use of other tunneling mechanisms. Note also that two ALT Routers that are directly adjacent (with no layer-3 router hops between them) need not use a tunnel between them; in this case, BGP may be configured across the interfaces that connect to their common subnet and that subnet is then considered to be part of the ALT topology. Use of techniques such as "eBGP multihop" to connect ALT Routers that do not share a tunnel or common subnet is not recommended as the non-ALT Routers in between the ALT Routers in such a configuration may not have information necessary to forward ALT Datagrams destined to EID-prefixes exchanged across that BGP session.

In summary, LISP+ALT uses BGP to build paths through ALT Routers so that an ALT Datagram sent into the ALT can be forwarded to the ETR that holds the EID-to-RLOC mapping for that EID-prefix. This

reachability is carried as IPv4 or ipv6 NLRI without modification (since an EID-prefix has the same syntax as IPv4 or ipv6 address prefix). ALT Routers establish BGP sessions with one another, forming the ALT. An ALT Router at the "edge" of the topology learns EID-prefixes originated by authoritative ETRs. Learning may be through the Map Server interface, by static configuration, or via BGP with the ETRs. An ALT Router may also be configured to aggregate EID-prefixes received from ETRs or from other LISP+ALT Routers that are topologically "downstream" from it.

#### 4.1. ITR traffic handling

When an ITR receives a packet originated by an end system within its site (i.e. a host for which the ITR is the exit path out of the site) and the destination EID for that packet is not known in the ITR's mapping cache, the ITR creates either a Map-Request for the destination EID or the original packet encapsulated as a Data Probe (see Section 3.3 for caveats on the usability of Data Probes). The result, known as an ALT Datagram, is then sent to an ALT Router (see also [LISP-MS] for non-ALT-connected ITRs, noting that Data Probes cannot be sent to a Map-Resolver). This "first hop" ALT Router uses EID-prefix routing information learned from other ALT Routers via BGP to guide the packet to the ETR which "owns" the prefix. Upon receipt by the ETR, normal LISP processing occurs: the ETR responds to the ITR with a LISP Map-Reply that lists the RLOCs (and, thus, the ETRs to use) for the EID-prefix. For Data Probes, the ETR also decapsulates the packet and transmits it toward its destination.

Upon receipt of the Map-Reply, the ITR installs the RLOC information for a given prefix into a local mapping database. With these mapping entries stored, additional packets destined to the given EID-prefix are routed directly to an RLOC without use of the ALT, until either the entry's TTL has expired, or the ITR can otherwise find no reachable ETR. Note that a current mapping may exist that contains no reachable RLOCs; this is known as a Negative Cache Entry and it indicates that packets destined to the EID-prefix are to be dropped.

Full details on Map-Request/Map-Reply processing may be found in [LISP].

Traffic routed on to the ALT consists solely of ALT Datagrams, i.e. Map-Requests and Data Probes (if supported). Given the relatively low performance expected of a tunneled topology, ALT Routers (and Map Resolvers) should aggressively rate-limit the ingress of ALT Datagrams from ITRs and, if possible, should be configured to not accept packets that are not ALT Datagrams.

#### 4.2. EID Assignment - Hierarchy and Topology

The ALT database is organized in a hierarchical manner with EID-prefixes aggregated on power-of-2 block boundaries. Where a LISP site has multiple EID-prefixes that are aligned on a power-of-2 block boundary, they should be aggregated into a single EID-prefix for advertisement. The ALT network is built in a roughly hierarchical, partial mesh which is intended to allow aggregation where clearly-defined hierarchical boundaries exist. Building such a structure should minimize the number of EID-prefixes carried by LISP+ALT nodes near the top of the hierarchy.

Routes on the ALT do not need to respond to changes in policy, subscription, or underlying physical connectivity, so the topology can remain relatively static and aggregation can be sustained. Because routing on the ALT uses BGP, the same rules apply for generating aggregates; in particular, a ALT Router should only be configured to generate an aggregate if it is configured with BGP sessions to all of the originators of components (more-specific prefixes) of that aggregate. Not all of the components need to be present for the aggregate to be originated (some may be holes in the covering prefix and some may be down) but the aggregating router must be configured to learn the state of all of the components.

Under what circumstances the ALT Router actually generates the aggregate is a matter of local policy: in some cases, it will be statically configured to do so at all times with a "static discard" route. In other cases, it may be configured to only generate the aggregate prefix if at least one of the components of the aggregate is learned via BGP.

An ALT Router must not generate an aggregate that includes a non-LISP-speaking hole unless it can be configured to return a Negative Map-Reply with action="Natively-Forward" (see [LISP]) if it receives an ALT Datagram that matches that hole. If it receives an ALT Datagram that matches a LISP-speaking hole that is currently not reachable, it should return a Negative Map-Reply with action="drop". Negative Map-Replies should be returned with a short TTL, as specified in [LISP-MS]. Note that an off-the-shelf, non-LISP-speaking router configured as an aggregating ALT Router cannot send Negative Map-Replies, so such a router must never originate an aggregate that includes a non-LISP-speaking hole.

This implies that two ALT Routers that share an overlapping set of prefixes must exchange those prefixes if either is to generate and export a covering aggregate for those prefixes. It also implies that an ETR which connects to the ALT using BGP must maintain BGP sessions with all of the ALT Routers that are configured to originate an



aggregate which covers that prefix and that each of those ALT Routers must be explicitly configured to know the set of EID-prefixes that make up any aggregate that it originates. See also [LISP-MS] for an example of other ways that prefix origin consistency and aggregation can be maintained.

As an example, consider ETRs that are originating EID-prefixes for 10.1.0.0/24, 10.1.64.0/24, 10.1.128.0/24, and 10.1.192.0/24. An ALT Router should only be configured to generate an aggregate for 10.1.0.0/16 if it has BGP sessions configured with all of these ETRs, in other words, only if it has sufficient knowledge about the state of those prefixes to summarize them. If the Router originating 10.1.0.0/16 receives an ALT Datagram destined for 10.1.77.88, a non-LISP destination covered by the aggregate, it returns a Negative Map-Reply with action "Natively-Forward". If it receives an ALT Datagram destined for 10.1.128.199 but the configured LISP prefix 10.1.128.0/24 is unreachable, it returns a Negative Map-Reply with action "drop".

Note: much is currently uncertain about the best way to build the ALT network; as testing and prototype deployment proceeds, a guide to how to best build the ALT network will be developed.

#### 4.3. Use of GRE and BGP between LISP+ALT Routers

The ALT network is built using GRE tunnels between ALT Routers. BGP sessions are configured over those tunnels, with each ALT Router acting as a separate AS "hop" in a Path Vector for BGP. For the purposes of LISP+ALT, the AS-path is used solely as a shortest-path determination and loop-avoidance mechanism. Because all next-hops are on tunnel interfaces, no IGP is required to resolve those next-hops to exit interfaces.

LISP+ALT's use of GRE and BGP facilitates deployment and operation of LISP because no new protocols need to be defined, implemented, or used on the overlay topology; existing BGP/GRE tools and operational expertise are also re-used. Tunnel address assignment is also easy: since the addresses on an ALT tunnel are only used by the pair of routers connected to the tunnel, the only requirement of the IP addresses used to establish that tunnel is that the attached routers be reachable by each other; any addressing plan, including private addressing, can therefore be used for ALT tunnels.

## 5. EID-prefix Propagation and Map-Request Forwarding

As described in Section 8.2, an ITR sends an ALT Datagram to a given EID-to-RLOC mapping. The ALT provides the infrastructure that allows these requests to reach the authoritative ETR.

Note that under normal circumstances Map-Replies are not sent over the ALT; an ETR sends a Map-Reply to one of the ITR RLOCs learned from the original Map-Request. See sections 6.1.2 and 6.2 of [LISP] for more information on the use of the Map-Request ITR RLOC field. Keep in mind that the ITR RLOC field supports multiple RLOCs in multiple address families, so a Map-Reply sent in response to a Map-Request is not necessarily sent back to the Map-Request RLOC source.

There may be scenarios, perhaps to encourage caching of EID-to-RLOC mappings by ALT Routers, where Map-Replies could be sent over the ALT or where a "first-hop" ALT Router might modify the originating RLOC on a Map-Request received from an ITR to force the Map-Reply to be returned to the "first-hop" ALT Router. These cases will not be supported by initial LISP+ALT implementations but may be subject to future experimentation.

ALT Routers propagate path information via BGP ([RFC4271]) that is used by ITRs to send ALT Datagrams toward the appropriate ETR for each EID-prefix. BGP is run on the inter-ALT Router links, and possibly between an edge ("last hop") ALT Router and an ETR or between an edge ("first hop") ALT Router and an ITR. The ALT BGP RIB consists of aggregated EID-prefixes and their next hops toward the authoritative ETR for that EID-prefix.

### 5.1. Changes to ITR behavior with LISP+ALT

As previously described, an ITR will usually use the Map Resolver interface and will send its Map Requests to a Map Resolver. When an ITR instead connects via tunnels and BGP to the ALT, it sends ALT Datagrams to one of its "upstream" ALT Routers; these are sent only to obtain new EID-to-RLOC mappings - RLOC probe and cache TTL refresh Map-Requests are not sent on the ALT. As in basic LISP, it should use one of its RLOCs as the source address of these queries; it should not use a tunnel interface as the source address as doing so will cause replies to be forwarded over the tunneled topology and may be problematic if the tunnel interface address is not routed throughout the ALT. If the ITR is running BGP with the LISP+ALT router(s), it selects the appropriate ALT Router based on the BGP information received. If it is not running BGP, it uses a statically-configured ALT Default Route to select an ALT Router.

## 5.2. Changes to ETR behavior with LISP+ALT

As previously described, an ETR will usually use the Map Server interface (see [LISP-MS]) and will register its EID-prefixes with its configured Map Servers. When an ETR instead connects using BGP to one or more ALT Routers, it announces its EID-prefix(es) to those ALT Routers.

As documented in [LISP], when an ETR generates a Map-Reply message to return to a querying ITR, it sets the outer header IP destination address to one of the requesting ITR's RLOCs so that the Map-Reply will be sent on the underlying Internet topology, not on the ALT; this avoids any latency penalty (or "stretch") that might be incurred by sending the Map-Reply via the ALT, reduces load on the ALT, and ensures that the Map-Reply can be routed even if the original ITR does not have an ALT-routed EID. For details on how an ETR selects which ITR RLOC to use, see section 6.1.5 of [LISP].

## 5.3. ALT Datagram forwarding failure

Intermediate ALT Routers, forward ALT Datagrams using normal, hop-by-hop routing on the ALT overlay network. Should an ALT router not be able to forward an ALT Datagram, whether due to an unreachable next-hop, TTL exceeded, or other problem, it has several choices:

- o If the ALT Router understands the LISP protocol, as is the case for a Map Resolver or Map Server, it may respond to a forwarding failure by returning a negative Map-Reply, as described in Section 4.2 and [LISP-MS].
- o If the ALT Router does not understand LISP, it may attempt to return an ICMP message to the source IP address of the packet that cannot be forwarded. Since the source address is an RLOC, an ALT Router would send this ICMP message using "native" Internet connectivity, not via the ALT overlay.
- o A non-LISP-capable ALT Router may also choose to silently drop the non-forwardable ALT Datagram.

[LISP] and [LISP-MS] define how the source of an ALT Datagram should handle each of these cases. The last case, where an ALT Datagram is silently discarded, will generally result in several retransmissions by the source, followed by treating the destination as unreachable via LISP when no Map-Reply is received. If a problem on the ALT is severe enough to prevent ALT Datagrams from being delivered to a specific EID, this is probably the only sensible way to handle this case.

Note that the use of GRE tunnels should prevent MTU problems from ever occurring on the ALT; an ALT Datagram that exceeds an intermediate MTU will be fragmented at that point and will be reassembled by the target of the GRE tunnel.

## 6. BGP configuration and protocol considerations

### 6.1. Autonomous System Numbers (ASNs) in LISP+ALT

The primary use of BGP today is to define the global Internet routing topology in terms of its participants, known as Autonomous Systems. LISP+ALT specifies the use of BGP to create a global overlay network (the ALT) for finding EID-to-RLOC mappings. While related to the global routing database, the ALT serves a very different purpose and is organized into a very different hierarchy. Because LISP+ALT does use BGP, however, it uses ASNs in the paths that are propagated among ALT Routers. To avoid confusion, LISP+ALT should use newly-assigned AS numbers that are unrelated to the ASNs used by the global routing system. Exactly how this new space will be assigned and managed will be determined during the deployment of LISP+ALT.

Note that the ALT Routers that make up the "core" of the ALT will not be associated with any existing core-Internet ASN because the ALT topology is completely separate from, and independent of, the global Internet routing system.

### 6.2. Sub-Address Family Identifier (SAFI) for LISP+ALT

As defined by this document, LISP+ALT may be implemented using BGP without modification. Given the fundamental operational difference between propagating global Internet routing information (the current dominant use of BGP) and creating an overlay network for finding EID-to-RLOC mappings (the use of BGP proposed by this document), it may be desirable to assign a new SAFI [RFC4760] to prevent operational confusion and difficulties, including the inadvertent leaking of information from one domain to the other. Use of a separate SAFI would make it easier to debug many operational problems but would come at a significant cost: unmodified, off-the-shelf routers which do not understand the new SAFI could not be used to build any part of the ALT network. At present, this document does not request the assignment of a new SAFI; additional experimentation may suggest the need for one in the future.

## 7. EID-prefix Aggregation

The ALT BGP peering topology should be arranged in a tree-like fashion (with some meshiness), with redundancy to deal with node and link failures. A basic assumption is that as long as the routers are up and running, the underlying Internet will provide alternative routes to maintain BGP connectivity among ALT Routers.

Note that, as mentioned in Section 4.2, the use of BGP by LISP+ALT requires that information only be aggregated where all active more-specific prefixes of a generated aggregate prefix are known. This is no different than the way that BGP route aggregation works in the existing global routing system: a service provider only generates an aggregate route if it is configured to learn to all prefixes that make up that aggregate.

### 7.1. Stability of the ALT

It is worth noting that LISP+ALT does not directly propagate EID-to-RLOC mappings. What it does is provide a mechanism for an ITR to communicate with the ETR that holds the mapping for a particular EID-prefix. This distinction is important when considering the stability of BGP on the ALT network as compared to the global routing system. It also has implications for how site-specific EID-prefix information may be used by LISP but not propagated by LISP+ALT (see Section 7.2 below).

RLOC prefixes are not propagated through the ALT so their reachability is not determined through use of LISP+ALT. Instead, reachability of RLOCs is learned through the LISP ITR-ETR exchange. This means that link failures or other service disruptions that may cause the reachability of an RLOC to change are not known to the ALT. Changes to the presence of an EID-prefix on the ALT occur much less frequently: only at subscription time or in the event of a failure of the ALT infrastructure itself. This means that "flapping" (frequent BGP updates and withdrawals due to prefix state changes) is not likely and mapping information cannot become "stale" due to slow propagation through the ALT BGP mesh.

### 7.2. Traffic engineering using LISP

Since an ITR learns an EID-to-RLOC mapping directly from the ETR that owns it, it is possible to perform site-to-site traffic engineering by setting the preference and/or weight fields, and by including more-specific EID-to-RLOC information in Map-Reply messages.

This is a powerful mechanism that can conceivably replace the traditional practice of routing prefix deaggregation for traffic

engineering purposes. Rather than propagating more-specific information into the global routing system for local- or regional- optimization of traffic flows, such more-specific information can be exchanged, through LISP (not LISP+ALT), on an as-needed basis between only those ITRs/ETRs (and, thus, site pairs) that need it. Such an exchange of "more-specifics" between sites facilitates traffic engineering, by allowing richer and more fine-grained policies to be applied without advertising additional prefixes into either the ALT or the global routing system.

Note that these new traffic engineering capabilities are an attribute of LISP and are not specific to LISP+ALT; discussion is included here because the BGP-based global routing system has traditionally used propagation of more-specific routes as a crude form of traffic engineering.

### 7.3. Edge aggregation and dampening

Normal BGP best common practices apply to the ALT network. In particular, first-hop ALT Routers will aggregate EID prefixes and dampen changes to them in the face of excessive updates. Since EID-prefix assignments are not expected to change as frequently as global routing BGP prefix reachability, such dampening should be very rare, and might be worthy of logging as an exceptional event. It is again worth noting that the ALT carries only EID-prefixes, used to a construct BGP path to each ETR (or Map-Server) that originates each prefix; the ALT does not carry reachability about RLOCs. In addition, EID-prefix information may be aggregated as the topology and address assignment hierarchy allow. Since the topology is all tunneled and can be modified as needed, reasonably good aggregation should be possible. In addition, since most ETRs are expected to connect to the ALT using the Map Server interface, Map Servers will implement a natural "edge" for the ALT where dampening and aggregation can be applied. For these reasons, the set of prefix information on the ALT can be expected to be both better aggregated and considerably less volatile than the actual EID-to-RLOC mappings.

### 7.4. EID assignment flexibility vs. ALT scaling

There are major open questions regarding how the ALT will be deployed and what organization(s) will operate it. In a simple, non-distributed world, centralized administration of EID prefix assignment and ALT network design would facilitate a well- aggregated ALT routing system. Business and other realities will likely result in a more complex, distributed system involving multiple levels of prefix delegation, multiple operators of parts of the ALT infrastructure, and a combination of competition and cooperation among the participants. In addition, re-use of existing IP address

assignments, both provider-independent ("PI") and provider-assigned ("PA"), to avoid renumbering when sites transition to LISP will further complicate the processes of building and operating the ALT.

A number of conflicting considerations need to be kept in mind when designing and building the ALT. Among them are:

1. Target ALT routing state size and level of aggregation. As described in Section 7.1, the ALT should not suffer from some of the performance constraints or stability issues as the Internet global routing system, so some reasonable level of deaggregation and increased number of EID prefixes beyond what might be considered ideal should be acceptable. That said, measures, such as tunnel rehomeing to preserve aggregation when sites move from one mapping provider to another and implementing aggregation at multiple levels in the hierarchy to collapse de-aggregation at lower levels, should be taken to reduce unnecessary explosion of ALT routing state.
2. Number of operators of parts of the ALT and how they will be organized (hierarchical delegation vs. shared administration). This will determine not only how EID prefixes are assigned but also how tunnels are configured and how EID prefixes can be aggregated between different parts of the ALT.
3. Number of connections between different parts of the ALT. Trade-offs will need to be made among resilience, performance, and placement of aggregation boundaries.
4. EID prefix portability between competing operators of the ALT infrastructure. A significant benefit for an end-site to adopt LISP is the availability of EID space that is not tied to a specific connectivity provider; it is important to ensure that an end site doesn't trade lock-in to a connectivity provider for lock-in to a provider of its EID assignment, ALT connectivity, or Map Server facilities.

This is, by no means, an exhaustive list.

While resolving these issues is beyond the scope of this document, the authors recommend that existing distributed resource structures, such as the IANA/Regional Internet Registries and the ICANN/Domain Registrar, be carefully considered when designing and deploying the ALT infrastructure.



## 8. Connecting sites to the ALT network

### 8.1. ETRs originating information into the ALT

EID-prefix information is originated into the ALT by three different mechanisms:

**Map Server:** In most cases, a site will configure its ETR(s) to register with one or more Map Servers (see [LISP-MS]), and does not participate directly in the ALT.

**BGP:** For a site requiring complex control over their EID-prefix origination into the ALT, an ETR may connect to the LISP+ALT overlay network by running BGP to one or more ALT Router(s) over tunnel(s). The ETR advertises reachability for its EID-prefixes over these BGP connection(s). The edge ALT Router(s) that receive(s) these prefixes then propagate(s) them into the ALT. Here the ETR is simply an BGP peer of ALT Router(s) at the edge of the ALT. Where possible, an ALT Router that receives EID-prefixes from an ETR via BGP should aggregate that information.

**Configuration:** One or more ALT Router(s) may be configured to originate an EID-prefix on behalf of the non-BGP-speaking ETR that is authoritative for a prefix. As in the case above, the ETR is connected to ALT Router(s) using GRE tunnel(s) but rather than BGP being used, the ALT Router(s) are configured with what are in effect "static routes" for the EID-prefixes "owned" by the ETR. The GRE tunnel is used to route Map-Requests to the ETR.

**Note:** in all cases, an ETR may register to multiple Map Servers or connect to multiple ALT Routers for the following reasons:

- \* redundancy, so that a particular ETR is still reachable even if one path or tunnel is unavailable.
- \* to connect to different parts of the ALT hierarchy if the ETR "owns" multiple EID-to-RLOC mappings for EID-prefixes that cannot be aggregated by the same ALT Router (i.e. are not topologically "close" to each other in the ALT).

### 8.2. ITRs Using the ALT

In the common configuration, an ITR does not need to know anything about the ALT, since it sends Map-Requests to one of its configured Map-Resolvers (see [LISP-MS]). There are two exceptional cases:

Static default: If a Map Resolver is not available but an ITR is adjacent to an ALT Router (either over a common subnet or through the use of a tunnel), it can use an ALT Default Route route to cause all ALT Datagrams to be sent that ALT Router. This case is expected to be rare.

Connection to ALT: A site with complex Internet connectivity needs may need more fine-grained distinction between traffic to LISP-capable and non-LISP-capable sites. Such a site may configure each of its ITRs to connect directly to the ALT, using a tunnel and BGP connection. In this case, the ITR will receive EID-prefix routes from its BGP connection to the ALT Router and will LISP-encapsulate and send ALT Datagrams through the tunnel to the ALT Router. Traffic to other destinations may be forwarded (without LISP encapsulation) to non-LISP next-hop routers that the ITR knows.

In general, an ITR that connects to the ALT does so only to ALT Routers at the "edge" of the ALT (typically two for redundancy). There may, though, be situations where an ITR would connect to other ALT Routers to receive additional, shorter path information about a portion of the ALT of interest to it. This can be accomplished by establishing GRE tunnels between the ITR and the set of ALT Routers with the additional information. This is a purely local policy issue between the ITR and the ALT Routers in question.

As described in [LISP-MS], Map-Resolvers do not accept or forward Data Probes; in the rare scenario that an ITR does support and originate Data Probes, it must do so using one of the exceptional configurations described above. Note that the use of Data Probes is discouraged at this time (see Section 3.3).

## 9. IANA Considerations

This document makes no request of the IANA.

## 10. Security Considerations

LISP+ALT shares many of the security characteristics of BGP. Its security mechanisms are comprised of existing technologies in wide operational use today, so securing the ALT should be mostly a matter of applying the same technology that is used to secure the BGP-based global routing system (see Section 10.3 below).

### 10.1. Apparent LISP+ALT Vulnerabilities

This section briefly lists the known potential vulnerabilities of LISP+ALT.

**Mapping Integrity:** Potential for an attacker to insert bogus mappings to black-hole (create Denial-of-Service, or DoS attack) or intercept LISP data-plane packets.

**ALT Router Availability:** Can an attacker DoS the ALT Routers connected to a given ETR? If a site's ETR cannot advertise its EID-to-RLLOC mappings, the site is essentially unavailable.

**ITR Mapping/Resources:** Can an attacker force an ITR or ALT Router to drop legitimate mapping requests by flooding it with random destinations for which it will generate large numbers of Map-Requests and fill its mapping cache? Further study is required to see the impact of admission control on the overlay network.

**EID Map-Request Exploits for Reconnaissance:** Can an attacker learn about a LISP site's TE policy by sending legitimate mapping requests and then observing the RLLOC mapping replies? Is this information useful in attacking or subverting peer relationships? Note that any public LISP mapping database will have similar data-plane reconnaissance issue.

**Scaling of ALT Router Resources:** Paths through the ALT may be of lesser bandwidth than more "direct" paths; this may make them more prone to high-volume denial-of-service attacks. For this reason, all components of the ALT (ETRs and ALT Routers) should be prepared to rate-limit traffic (ALT Datagrams) that could be received across the ALT.

**UDP Map-Reply from ETR:** Since Map-Replies are sent directly from the ETR to the ITR's RLLOC, the ITR's RLLOC may be vulnerable to various types of DoS attacks (this is a general property of LISP, not an LISP+ALT vulnerability).

More-specific prefix leakage: Because EID-prefixes on the ALT are expected to be fairly well-aggregated and EID-prefixes propagated out to the global Internet (see [LISP-IW]) much more so, accidental leaking or malicious advertisement of an EID-prefix into the global routing system could cause traffic redirection away from a LISP site. This is not really a new problem, though, and its solution can only be achieved by much more strict prefix filtering and authentication on the global routing system. Section 10.3 describes an existing approach to solving this problem.

#### 10.2. Survey of LISP+ALT Security Mechanisms

Explicit peering: The devices themselves can both prioritize incoming packets, as well as potentially do key checks in hardware to protect the control plane.

Use of TCP to connect elements: This makes it difficult for third parties to inject packets.

Use of HMAC to protect BGP/TCP connections: HMAC [RFC5925] is used to verify the integrity and authenticity of TCP connections used to exchange BGP messages, making it nearly impossible for third party devices to either insert or modify messages.

Message sequence numbers and nonce values in messages: This allows an ITR to verify that the Map-Reply from an ETR is in response to a Map-Request originated by that ITR (this is a general property of LISP; LISP+ALT does not change this behavior).

#### 10.3. Use of new IETF standard BGP Security mechanisms

LISP+ALT's use of BGP allows it to take advantage of BGP security features designed for existing Internet BGP use. This means that LISP+ALT can and should use technology developed for adding security to BGP (in the IETF SIDR working group or elsewhere) to provide authentication of EID-prefix origination and EID-to-RLOC mappings.

## 11. Acknowledgments

The authors would like to specially thank J. Noel Chiappa who was a key contributor to the design of the LISP-CONS mapping database (many ideas from which made their way into LISP+ALT) and who has continued to provide invaluable insight as the LISP effort has evolved. Others who have provided valuable contributions include John Zwiebel, Hannu Flinck, Amit Jain, John Scudder, Scott Brim, and Jari Arkko.

## 12. References

### 12.1. Normative References

- [LISP]      Farinacci, D., Fuller, V., Meyer, D., and D. Lewis,  
"Locator/ID Separation Protocol (LISP)",  
draft-ietf-lisp-15.txt (work in progress), July 2011.
- [LISP-MS]   Fuller, V. and D. Farinacci, "LISP Map Server",  
draft-ietf-lisp-ms-12.txt (work in progress),  
October 2011.
- [RFC2784]   Farinacci, D., Li, T., Hanks, S., Meyer, D., and P.  
Traina, "Generic Routing Encapsulation (GRE)", RFC 2784,  
March 2000.
- [RFC4271]   Rekhter, Y., Li, T., and S. Hares, "A Border Gateway  
Protocol 4 (BGP-4)", RFC 4271, January 2006.
- [RFC4632]   Fuller, V. and T. Li, "Classless Inter-domain Routing  
(CIDR): The Internet Address Assignment and Aggregation  
Plan", BCP 122, RFC 4632, August 2006.
- [RFC4760]   Bates, T., Chandra, R., Katz, D., and Y. Rekhter,  
"Multiprotocol Extensions for BGP-4", RFC 4760,  
January 2007.

### 12.2. Informative References

- [LISP-IW]   Lewis, D., Meyer, D., Farinacci, D., and V. Fuller,  
"Interworking LISP with IPv4 and ipv6",  
draft-ietf-lisp-interworking-02.txt (work in progress),  
March 2011.
- [RFC5925]   Touch, J., Mankin, A., and R. Bonica, "The TCP  
Authentication Option", RFC 5925, June 2010.

Authors' Addresses

Vince Fuller  
Cisco  
Tasman Drive  
San Jose, CA 95134  
USA

Email: vaf@cisco.com

Dino Farinacci  
Cisco  
Tasman Drive  
San Jose, CA 95134  
USA

Email: dino@cisco.com

Dave Meyer  
Cisco  
Tasman Drive  
San Jose, CA 95134  
USA

Email: dmm@cisco.com

Darrel Lewis  
Cisco  
Tasman Drive  
San Jose, CA 95134  
USA

Email: darlewis@cisco.com





Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: March 12, 2012

D. Farinacci  
D. Meyer  
cisco Systems  
September 9, 2011

LISP Internet Groper (LIG)  
draft-ietf-lisp-lig-06

Abstract

A simple tool called the LISP Internet Groper or 'lig' can be used to query the LISP mapping database. This draft describes how it works.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on March 12, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Definition of Terms . . . . .	4
3. Basic Overview . . . . .	7
4. Implementation Details . . . . .	9
4.1. LIISP Router Implementation . . . . .	9
4.2. Public Domain Host Implementation . . . . .	10
5. Testing the ALT . . . . .	12
6. Future Enhancements . . . . .	13
7. Deployed Network Diagnostic Tools . . . . .	14
8. Security Considerations . . . . .	15
9. IANA Considerations . . . . .	16
10. References . . . . .	17
10.1. Normative References . . . . .	17
10.2. Informative References . . . . .	17
Appendix A. Acknowledgments . . . . .	18
Authors' Addresses . . . . .	19

## 1. Introduction

LISP [LISP] specifies an architecture and mechanism for replacing the addresses currently used by IP with two separate name spaces: Endpoint IDs (EIDs), used within sites, and Routing Locators (RLOCs), used on the transit networks that make up the Internet infrastructure. To achieve this separation, the Locator/ID Separation Protocol (LISP) defines protocol mechanisms for mapping from EIDs to RLOCs. In addition, LISP assumes the existence of a database to store and propagate those mappings globally. Several such databases have been proposed, among them: LISP-CONS [CONS], LISP-NERD [NERD], and LISP+ALT [ALT], with LISP+ALT being the system that is currently being implemented and deployed on the pilot LISP network.

In conjunction with the various mapping systems, there exists a network based API called LISP Map-Server [LISP-MS]. Using Map-Resolvers and Map-Servers allows LISP sites to query and register into the database in a uniform way independent of the mapping system used. Sending Map-Requests to Map-Resolvers provides a secure mechanism to obtain a Map-Reply containing the authoritative EID-to-RLOC mapping for a destination LISP site.

The 'lig' is a manual management tool to query the mapping database. It can be run by all devices which implement LISP, including ITRs, ETRs, PITRs, PETRs, Map-Resolvers, Map-Servers, and LISP-ALT routers, as well as by a host system at either a LISP-capable or non-LISP-capable site.

The mapping database system is typically a public database used for wide-range connectivity across Internet sites. The information in the public database is purposely not kept private so it can be generally accessible for public use.

## 2. Definition of Terms

**Map-Server:** a network infrastructure component which learns EID-to-RLOC mapping entries from an authoritative source (typically, an ETR, though static configuration or another out-of-band mechanism may be used). A Map-Server advertises these mappings in the distributed mapping database.

**Map-Resolver:** a network infrastructure component which accepts LISP Encapsulated Map-Requests, typically from an ITR, quickly determines whether or not the destination IP address is part of the EID namespace; if it is not, a Negative Map-Reply is immediately returned. Otherwise, the Map-Resolver finds the appropriate EID-to-RLOC mapping by consulting the distributed mapping database system.

**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. Thus, the topology is defined by the connectivity of provider networks and RLOCs can be thought of as PA addresses. Multiple RLOCs can be assigned to the same ETR device or to multiple ETR devices at a site.

**Endpoint ID (EID):** a 32-bit (for IPv4) or 128-bit (for IPv6) 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 a destination address today, for example through a DNS lookup. The source EID is obtained via existing mechanisms used to set a host's "local" IP address. An EID is allocated to a host from an EID-prefix block associated with the site where the host is located. An EID can be used by a host to refer to other hosts. 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-to-RLOC Cache:** a short-lived, on-demand table 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 full "database" of EID-to-RLOC mappings, it is dynamic, local to the ITR(s), and relatively small while the database is distributed, relatively static, and much more global in scope.

**EID-to-RLOC Database:** a global 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.

**Encapsulated Map-Request (EMR):** an EMR is a Map-Request message which is encapsulated with another LISP header using UDP destination port number 4341. It is used so an ITR, PITR, or a system initiating a 'lig' command can get the Map-Request to a Map-Resolver by using locator addresses. When the Map-Request is decapsulated by the Map-Resolver it will be forwarded on the ALT network to the Map-Server that has injected the EID-prefix for a registered site. The Map-Server will then encapsulate the Map-Request in a LISP packet and send it to an ETR at the site. The ETR will then return an authoritative reply to the system that initiated the request. See [LISP] for packet format details.

**Ingress Tunnel Router (ITR):** An ITR is 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 closer 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.

**Egress Tunnel Router (ETR):** An ETR is a router that accepts an IP packet where the 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. ETR functionality does not have to be limited to a router device. A server host can be the endpoint of a LISP tunnel as well.

**Proxy ITR (PITR):** A PITR is also known as a PTR is defined and described in [INTERWORK], a PITR acts like an ITR but does so on behalf of non-LISP sites which send packets to destinations at LISP sites.

Proxy ETR (PETR): A PETR is defined and described in [INTERWORK], a PETR acts like an ETR but does so on behalf of LISP sites which send packets to destinations at non-LISP sites.

xTR: A xTR is a reference to an ITR or ETR when direction of data flow is not part of the context description. xTR refers to the router that is the tunnel endpoint. Used synonymously with the term "Tunnel Router". For example, "An xTR can be located at the Customer Edge (CE) router", meaning both ITR and ETR functionality is at the CE router.

Provider Assigned (PA) Addresses: PA addresses are an address block assigned to a site by each service provider to which a site connects. Typically, each block is sub-block of a service provider Classless Inter-Domain Routing (CIDR) [RFC4632] 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.

### 3. Basic Overview

When the lig command is run, a Map-Request is sent for a destination EID. When a Map-Reply is returned, the contents are displayed to the user. The information displayed includes:

- o The EID-prefix for the site the queried destination EID matches.
- o The locator address of the Map Replier.
- o The locator-set for the mapping entry which includes the locator address, up/down status, priority, and weight of each locator.
- o An round-trip-time estimate for the Map-Request/Map-Reply exchange.

A possible syntax for a lig command could be:

```
lig <destination> [source <source>] [to <map-resolver>]
```

Parameter description:

<destination>: is either a Fully Qualified Domain Name or a destination EID for a remote LIISP site.

source <source>: is an optional source EID to be inserted in the "Source EID" field of the Map-Request.

to <map-resolver>: is an optional Fully Qualified Domain Name or RLOC address for a Map-Resolver.

The lig utility has two use cases. The first being a way to query the mapping database for a particular EID. And the other to verify if a site has registered successfully with a Map-Server.

The first usage has already been described. Verifying registration is called "ligging yourself". What occurs is in the lig initiator, a Map-Request is sent for one of the EIDs for the lig initiator's site. The Map-Request is then returned to one of the ETRs for the lig initiating site. In response to the Map-Request, a Map-Reply is sent back to the locator address of the lig initiator (note the Map-Reply could be sent by the lig initiator). That Map-Reply is processed and the mapping data for the lig initiating site is displayed for the user. Refer to the syntax in section Section 4.1 for an implementation of "ligging yourself". However, for host-based implementations within a LIISP site, "lig self" is less useful since the host may not have an RLOC to receive a Map-Reply with. But, lig



can be used in a non-LIISP site as well as from infrastructure hosts to get mapping information.

## 4. Implementation Details

### 4.1. LISP Router Implementation

The cisco LISP prototype implementation has support for lig for IPv4 and IPv6. The command line description is:

```
lig <dest-eid> [source <source-eid>] [to <mr>] [count <1-5>]
```

This command initiates the LISP Internet Groper. It is similar to the DNS analogue 'dig' but works on the LISP mapping database. When this command is invoked, the local system will send a Map-Request to the configured Map-Resolver. When a Map-Reply is returned, its contents will be displayed to the user. By default, up to 3 Map-Requests are sent if no Map-Reply is returned but once a Map-Reply is returned no other Map-Requests are sent. The destination can take a DNS name, or an IPv4 or IPv6 EID address. The <source-eid> can be one of the EID addresses assigned to the site in the default VRF. When <mr> is specified, then the Map-Request is sent to the address. Otherwise, the Map-Request is sent to a configured Map-Resolver. When a Map-Resolver is not configured then the Map-Request is sent on the ALT network if the local router is attached to the ALT. When "count <1-5>" is specified, 1, 2, 3, 4, or 5 Map-Requests are sent.

Some sample output:

```
router# lig abc.example.com
Send map-request to 10.0.0.1 for 192.168.1.1 ...
Received map-reply from 10.0.0.2 with rtt 0.081468 secs

Map-cache entry for abc.example.com EID 192.168.1.1:
192.168.1.0/24, uptime: 13:59:59, expires: 23:59:58,
via map-reply, auth
  Locator      Uptime      State      Priority/Weight  Packets In/Out
  10.0.0.2      13:59:59    up         1/100            0/14
```

Using lig to "lig yourself" is accomplished with the following syntax:

```
lig {self | self6} [source <source-eid>] [to <mr>] [count <1-5>]
```

Use this command for a simple way to see if the site is registered with the mapping database system. The destination-EID address for the Map-Request will be the first configured EID-prefix for the site (with the host-bits set to 0). For example, if the site's EID-prefix

is 192.168.1.0/24, the destination-EID for the Map-Request is 192.168.1.0. The source-EID address for the Map-Request will also be 192.168.1.0 (in this example) and the Map-Request is sent to the configured Map-Resolver. If the Map-Resolver and Map-Server are the same LISP system, then the "lig self" is testing if the Map-Resolver can "turn back a Map-Request to the site". If another Map-Resolver is used, it can test that the site's EID-prefix has been injected into the ALT infrastructure in which case the lig Map-Request is processed by the Map-Resolver, propagated through each ALT router hop to the site's registered Map-Server. Then the Map-Server returns the Map-Request to the originating site. In which case, an xTR at the originating site sends a Map-Reply to the source of the Map-Request (could be itself or another xTR for the site). All other command parameters are described above. Using "lig self6" tests for registering of IPv6 EID- prefixes.

Some sample output for ligging yourself:

```
router# lig self
Send loopback map-request to 10.0.0.1 for 192.168.2.0 ...
Received map-reply from 10.0.0.3 with rtt 0.001592 secs

Map-cache entry for EID 192.168.2.0:
192.168.2.0/24, uptime: 00:00:02, expires: 23:59:57
via map-reply, self
  Locator      Uptime      State  Priority/Weight  Packets In/Out
  10.0.0.3      00:00:02   up     1/100            0/0

router# lig self6
Send loopback map-request to 10.0.0.1 for 2001:db8:1:: ...
Received map-reply from 10::1 with rtt 0.044372 secs

Map-cache entry for EID 192:168:1:::
2001:db8:1::/48, uptime: 00:00:01, expires: 23:59:58
via map-reply, self
  Locator      Uptime      State  Priority/Weight  Packets In/Out
  10.0.0.3      00:00:01   up     1/100            0/0
  2001:db8:ffff::1 00:00:01   up     2/0              0/0
```

#### 4.2. Public Domain Host Implementation

There is a public domain implementation that can run on any x86 based system. The only requirement is that the system that initiates lig must have an address assigned from the locator namespace.

```
lig [-d] <eid> -m <map-resolver> [-c <count>] [-t <timeout>]
```

Parameter description:

-d: prints additional protocol debug output.

<eid>: is the destination EID or FQDN of a LISP host.

-m <map-resolver>: is the RLOC address or FQDN of a Map-Resolver.

-c <count>: the number of Map-Requests to send before the first Map-Reply is returned. The default value is 3. The range is from 1 to 5.

-t <timeout>: the amount of time, in seconds, before another Map-Request is sent when no Map-Reply is returned. The default value is 2 seconds. The range is from 1 to 5.

Some sample output:

```
% lig xyz.example.com -m 10.0.0.1
Send map-request to 10.0.0.1 for 192.168.1.1 ...
Received map-reply from 10.0.0.2 with rtt 0.04000 sec
```

```
Mapping entry for EID 192.168.1.1:
192.168.1.0/24, record ttl: 60
Locator      State      Priority/Weight
10.0.0.1     up         1/25
10.0.0.2     up         1/25
10.0.0.3     up         1/25
10.0.0.4     up         2/25
```

The public domain implementation of lig is available at  
<http://github.com/davidmeyer/lig>.

## 5. Testing the ALT

There are cases where a Map-Reply is returned from a lig request but the user doesn't really know how much of the mapping infrastructure was tested. There are two cases to consider, avoiding the ALT and traversing the ALT.

When an ITR sends a lig request to its Map-Resolver for a destination-EID, the Map-Resolver could also be configured as a Map-Server. And if the destination-EID is for a site that registers with this Map-Server, the Map-Request is sent to the site directly without testing the ALT. This occurs because the Map-Server is the source of the advertisement for the site's EID-prefix. So if the map-reply is returned to the lig requesting site, you cannot be sure that other sites can reach the same destination-EID.

If a Map-Resolver is used that is not a Map-Server for the EID-prefix being sought, then the ALT infrastructure can be tested. This test case is testing the functionality of the Map-Resolver, traversal of the ALT (testing BGP-over-GRE), and the Map-Server.

It is recommended that users issue 2 lig requests, each of which send Map-Requests to different Map-Resolvers.

The network can have a LISP-ALT router deployed as a "ALT looking-glass" node. This type of router has BGP peering sessions with other ALT routers where it does not inject any EID-prefixes into the ALT but just learns ones advertised by other ALT routers and Map-Servers. This router is configured as a Map-Resolver. Lig users can point to the ALT looking-glass router for Map-Resolver services via the "to <map-resolver>" parameter on the lig command. The ALT looking-glass node can be used to lig other sites as well as your own site. When the ALT looking-glass is used as a Map-Resolver, you can be assured the ALT network is being tested.

## 6. Future Enhancements

When negative Map-Replies have been further developed and implemented, lig should be modified appropriately to process and clearly indicate how and why a negative Map-Reply was received. Negative Map-Replies could be sent in the following cases, the lig request was initiated for a non-EID address or the Map-Request initiated by lig request is being rejected due to rate-limiting on the replier.

## 7. Deployed Network Diagnostic Tools

There is an web-based interface to do auto-polling with lig on the back-end for most of the LISP sites on the LISP test network. The web-page can be accessed at <http://www.lisp4.net/status>.

There is a LISP site monitoring web-based interface that can be found at <http://www.lisp4.net/lisp-site>.

At <http://baldomar.ccaba.upc.edu/lispmon>, written by the folks at UPC, shows a geographical map indicating where each LISP site resides.

## 8. Security Considerations

The use of lig does not affect the security of the LISP infrastructure as it is simply a tool that facilitates diagnostic querying. See [LISP], [ALT], and [LISP-MS] for descriptions of the security properties of the LISP infrastructure.

Lig provides easy access to the information in the public mapping database. Therefore, it is important to protect the mapping information for private use. This can be provided by disallowing access to specific mapping entries or to place such entries in a private mapping database system.



## 9. IANA Considerations

This document makes no request of the IANA.

## 10. References

### 10.1. Normative References

[INTERWORK]

Lewis, D., Meyer, D., Farinacci, D., and V. Fuller,  
"Interworking LISP with IPv4 and IPv6",  
draft-ietf-lisp-interworking-02.txt (work in progress).

[LISP]

Farinacci, D., Fuller, V., Meyer, D., and D. Lewis,  
"Locator/ID Separation Protocol (LISP)",  
draft-ietf-lisp-15.txt (work in progress).

[LISP-MS]

Farinacci, D. and V. Fuller, "LISP Map Server",  
draft-ietf-lisp-ms-11.txt (work in progress).

[RFC4632]

Fuller, V. and T. Li, "Classless Inter-domain Routing  
(CIDR): The Internet Address Assignment and Aggregation  
Plan", BCP 122, RFC 4632, August 2006.

### 10.2. Informative References

[ALT]

Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "LISP  
Alternative Topology (LISP-ALT)",  
draft-ietf-lisp-alt-08.txt (work in progress).

[CONS]

Farinacci, D., Fuller, V., and D. Meyer, "LISP-CONS: A  
Content distribution Overlay Network Service for LISP",  
draft-meyer-lisp-cons-04.txt (work in progress).

[LISP-LIG]

Farinacci, D. and D. Meyer, "LISP Internet Groper (LIG)",  
draft-farinacci-lisp-lig-02.txt (work in progress).

[NERD]

Lear, E., "NERD: A Not-so-novel EID to RLOC Database",  
draft-lear-lisp-nerd-08.txt (work in progress).

## Appendix A. Acknowledgments

Thanks and kudos to John Zwiebel, Andrew Partan, Darrel Lewis, and Vince Fuller for providing critical feedback on the lig design and prototype implementations. These folks as well as all the people on `lisp-beta@external.cisco.com` who tested lig functionality and continue to do so, we extend our sincere thanks.

This working group draft is based on individual contribution `draft-farinacci-lisp-lig-02.txt` [LISP-LIG].

Authors' Addresses

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

Email: [dino@cisco.com](mailto:dino@cisco.com)

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

Email: [dmm@cisco.com](mailto:dmm@cisco.com)



Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: September 5, 2012

V. Fuller  
D. Farinacci  
cisco Systems  
March 4, 2012

LISP Map Server Interface  
draft-ietf-lisp-ms-16.txt

Abstract

This draft describes the Mapping Service for the Locator Identifier Separation Protocol (LISP), implemented by two new types of LISP-speaking devices, the LISP Map Resolver and LISP Map Server, that provides a simplified "front end" to for one or more Endpoint ID to Routing Locator mapping databases.

By using this service interface and communicating with Map Resolvers and Map Servers, LISP Ingress Tunnel Routers and Egress Tunnel Routers, are not dependent on the details of mapping database systems, which facilitates experimentation with different database designs. Since these devices implement the "edge" of the LISP infrastructure, connect directly to LISP-capable Internet end sites, and comprise the bulk of LISP-speaking devices, reducing their implementation and operational complexity should also reduce the overall cost and effort of deploying LISP.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on September 5, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Definition of Terms . . . . .	4
3. Basic Overview . . . . .	5
4. Interactions With Other LIISP Components . . . . .	6
4.1. ITR EID-to-RLLOC Mapping Resolution . . . . .	6
4.2. EID Prefix Configuration and ETR Registration . . . . .	7
4.3. Map Server Processing . . . . .	8
4.4. Map Resolver Processing . . . . .	9
4.4.1. Anycast Map Resolver Operation . . . . .	10
5. Open Issues and Considerations . . . . .	11
6. IANA Considerations . . . . .	12
7. Security Considerations . . . . .	13
8. References . . . . .	14
8.1. Normative References . . . . .	14
8.2. Informative References . . . . .	14
Appendix A. Acknowledgments . . . . .	15
Authors' Addresses . . . . .	16

## 1. Introduction

[LISP], the Locator Identifier Separation Protocol, specifies an architecture and mechanism for replacing the addresses currently used by IP with two separate name spaces: Endpoint IDs (EIDs), used within sites, and Routing Locators (RLOCs), used on the transit networks that make up the Internet infrastructure. To achieve this separation, LISP defines protocol mechanisms for mapping from EIDs to RLOCs. In addition, LISP assumes the existence of a database to store and propagate those mappings globally. Several such databases have been proposed, among them: LISP-CONS [CONS], LISP-NERD, [NERD] and LISP+ALT [ALT].

The LISP Mapping Service defines two new types of LISP-speaking devices: the Map Resolver, which accepts Map-Requests from an Ingress Tunnel Router (ITR) and "resolves" the EID-to-RLOC mapping using a mapping database, and the Map Server, which learns authoritative EID-to-RLOC mappings from an Egress Tunnel Router (ETR) and publishes them in a database.

Conceptually, LISP Map Servers share some of the same basic configuration and maintenance properties as Domain Name System (DNS) [RFC1035] servers; likewise, Map Resolvers are conceptually similar to DNS caching resolvers. With this in mind, this specification borrows familiar terminology (resolver and server) from the DNS specifications.

Note that while this document assumes a LISP+ALT database mapping infrastructure to illustrate certain aspects of Map Server and Map Resolver operation, the Mapping Service interface can (and likely will) be used by ITRs and ETRs to access other mapping database systems as the LISP infrastructure evolves.

Section 5 of this document notes a number of issues with the Map Server and Map Resolver design that are not yet completely understood and are subjects of further experimentation.

The LISP Mapping Service is an important component of the LISP toolset. Issues and concerns about the deployment of LISP for Internet traffic are discussed in [LISP].



## 2. Definition of Terms

**Map Server:** a network infrastructure component which learns of EID-prefix mapping entries from an ETR, via the registration mechanism described below, or some other authoritative source if one exists. A Map Server publishes these EID-prefixes in a mapping database.

**Map Resolver:** a network infrastructure component which accepts LISP Encapsulated Map-Requests, typically from an ITR, determines whether or not the destination IP address is part of the EID namespace; if it is not, a Negative Map-Reply is returned. Otherwise, the Map Resolver finds the appropriate EID-to-RLOC mapping by consulting a mapping database system.

**Encapsulated Map-Request:** a LISP Map-Request carried within an Encapsulated Control Message, which has an additional LISP header prepended. Sent to UDP destination port 4342. The "outer" addresses are globally-routeable IP addresses, also known as RLOCs. Used by an ITR when sending to a Map Resolver and by a Map Server when forwarding a Map-Request to an ETR.

**Negative Map-Reply:** a LISP Map-Reply that contains an empty locator-set. Returned in response to a Map-Request if the destination EID does not exist in the mapping database. Typically, this means that the "EID" being requested is an IP address connected to a non-LISP site.

**Map-Register message:** a LISP message sent by an ETR to a Map Server to register its associated EID-prefixes. In addition to the set of EID-prefixes to register, the message includes one or more RLOCs to be used by the Map Server when forwarding Map-Requests (re-formatted as Encapsulated Map-Requests) received through the database mapping system. An ETR may request that the Map Server answer Map-Requests on its behalf by setting the "proxy-map-reply" flag (P-bit) in the message.

**Map-Notify message:** a LISP message sent by a Map Server to an ETR to confirm that a Map-Register has been received and processed. An ETR requests that a Map-Notify be returned by setting the "want-map-notify" or "M" bit in the Map-Register message. Unlike a Map-Reply, a Map-Notify uses UDP port 4342 for both source and destination.

For definitions of other terms, notably Map-Request, Map-Reply, Ingress Tunnel Router (ITR), and Egress Tunnel Router (ETR), please consult the LISP specification [LISP].

### 3. Basic Overview

A Map Server is a device which publishes EID-prefixes in a LISP mapping database on behalf of a set of ETRs. When it receives a Map Request (typically from an ITR) it consults the mapping database to find an ETR that can answer with the set of RLOCs for an EID-prefix. To publish its EID-prefixes, an ETR periodically sends Map-Register messages to the Map Server. A Map-Register message contains a list of EID-prefixes plus a set of RLOCs that can be used to reach the ETR when a Map Server needs to forward a Map-Request to it.

When LISP+ALT is used as the mapping database, a Map Server connects to ALT network and acts as a "last-hop" ALT router. Intermediate ALT routers forward Map-Requests to the Map Server that advertises a particular EID-prefix and the Map Server forwards them to the owning ETR, which responds with Map-Reply messages.

A Map Resolver receives Encapsulated Map-Requests from its client ITRs and uses a mapping database system to find the appropriate ETR to answer those requests. On a LISP+ALT network, a Map Resolver acts as a "first-hop" ALT router. It has GRE tunnels configured to other ALT routers and uses BGP to learn paths to ETRs for different prefixes in the LISP+ALT database. The Map Resolver uses this path information to forward Map-Requests over the ALT to the correct ETRs.

Note that while it is conceivable that a Map Resolver could cache responses to improve performance, issues surrounding cache management will need to be resolved for doing so to be reliable and practical. As initially deployed, Map Resolvers will operate only in a non-caching mode, de-decapsulating and forwarding Encapsulated Map Requests received from ITRs. Any specification of caching functionality is left for future work.

Note that a single device can implement the functions of both a Map Server and a Map Resolver and, in many cases, the functions will be co-located in that way.

Detailed descriptions of the LISP packet types referenced by this document may be found in [LISP].

## 4. Interactions With Other LISP Components

### 4.1. ITR EID-to-RLOC Mapping Resolution

An ITR is configured with one or more Map Resolver addresses. These addresses are "locators" (or RLOCs) and must be routeable on the underlying core network; they must not need to be resolved through LISP EID-to-RLOC mapping as that would introduce a circular dependency. When using a Map Resolver, an ITR does not need to connect to any other database mapping system. In particular, the ITR need not connect to the LISP+ALT infrastructure or implement the BGP and GRE protocols that it uses.

An ITR sends an Encapsulated Map-Request to a configured Map Resolver when it needs an EID-to-RLOC mapping that is not found in its local map-cache. Using the Map Resolver greatly reduces both the complexity of the ITR implementation and the costs associated with its operation.

In response to an Encapsulated Map-Request, the ITR can expect one of the following:

- o An immediate Negative Map-Reply (with action code of "forward-native", 15-minute TTL) from the Map Resolver if the Map Resolver can determine that the requested EID does not exist. The ITR saves the EID-prefix returned in the Map-Reply in its cache, marking it as non-LISP-capable and knows not to attempt LISP encapsulation for destinations matching it.
- o A Negative Map-Reply (with action code of "forward-native") from the Map Server that has an aggregate EID-covering the EID in the Map-Request but where the EID matches a "hole" in the aggregate. If the "hole" is for a LISP EID-prefix that is defined in the Map Server configuration but for which no ETRs are currently registered, a 1-minute TTL is returned. If the "hole" is for an unassigned part of the aggregate, then it is not a LISP EID and a 15-minute TTL is returned. See Section 4.2 for discussion of aggregate EID-prefixes and details of Map Server EID-prefix matching.
- o A LISP Map-Reply from the ETR that owns the EID-to-RLOC mapping or possibly from a Map Server answering on behalf of the ETR. See (Section 4.4) for more details on Map Resolver message processing.

Note that an ITR may be configured to both use a Map Resolver and to participate in a LISP+ALT logical network. In such a situation, the ITR should send Map-Requests through the ALT network for any EID-prefix learned via ALT BGP. Such a configuration is expected to be

very rare, since there is little benefit to using a Map Resolver if an ITR is already using LISP+ALT. There would be, for example, no need for such an ITR to send a Map-Request to a possibly non-existent EID (and rely on Negative Map-Replies) if it can consult the ALT database to verify that an EID-prefix is present before sending that Map-Request.

#### 4.2. EID Prefix Configuration and ETR Registration

An ETR publishes its EID-prefixes on a Map Server by sending LISP Map-Register messages. A Map-Register message includes authentication data, so prior to sending a Map-Register message, the ETR and Map Server must be configured with a shared secret or other relevant authentication information. A Map Server's configuration must also include a list of the EID-prefixes for which each ETR is authoritative. Upon receipt of a Map-Register from an ETR, a Map Server accepts only EID-prefixes that are configured for that ETR. Failure to implement such a check would leave the mapping system vulnerable to trivial EID-prefix hijacking attacks. As developers and operators gain experience with the mapping system, additional, stronger security measures may be added to the registration process.

In addition to the set of EID-prefixes defined for each ETR that may register, a Map Server is typically also configured with one or more aggregate prefixes that define the part of the EID numbering space assigned to it. When LISP+ALT is the database in use, aggregate EID-prefixes are implemented as discard routes and advertised into ALT BGP. The existence of aggregate EID-prefixes in a Map Server's database means that it may receive Map Requests for EID-prefixes that match an aggregate but do not match a registered prefix; Section 4.3 describes how this is handled.

Map-Register messages are sent periodically from an ETR to a Map Server with a suggested interval between messages of one minute. A Map Server should time-out and remove an ETR's registration if it has not received a valid Map-Register message within the past three minutes. When first contacting a Map Server after restart or changes to its EID-to-RLOC database mappings, an ETR may initially send Map-Register messages at an increased frequency, up to one every 20 seconds. This "quick registration" period is limited to five minutes in duration.

An ETR may request that a Map Server explicitly acknowledge receipt and processing of a Map-Register message by setting the "want-map-notify" ("M" bit) flag. A Map Server that receives a Map-Register with this flag set will respond with a Map-Notify message. Typical use of this flag by an ETR would be to set it for Map-Register messages sent during the initial "quick registration" with a Map

Server but then set it only occasionally during steady-state maintenance of its association with that Map Server. Note that the Map-Notify message is sent to UDP destination port 4342, not to the source port specified in the original Map-Register message.

Note that a one-minute minimum registration interval during maintenance of an ETR-MS association places a lower-bound on how quickly and how frequently a mapping database entry can be updated. This may have implications for what sorts of mobility can be supported directly by the mapping system; shorter registration intervals or other mechanisms might be needed to support faster mobility in some cases. For a discussion on one way that faster mobility may be implemented for individual devices, please see [LISP-MN].

An ETR may also request, by setting the "proxy-map-reply" flag (P-bit) in the Map-Register message, that a Map Server answer Map-Requests instead of forwarding them to the ETR. See [LISP] for details on how the Map Server sets certain flags (such as those indicating whether the message is authoritative and how returned locators should be treated) when sending a Map-Reply on behalf of an ETR. When an ETR requests proxy reply service, it should include all RLOCs for all ETRs for the EID-prefix being registered, along with the routable flag ("R-bit") setting for each RLOC. The Map Server includes all of this information in Map Reply messages that it sends on behalf of the ETR. This differs from a non-proxy registration since the latter need only provide one or more RLOCs for a Map Server to use for forwarding Map-Requests; the registration information is not used in Map-Replies so it being incomplete is not incorrect.

An ETR which uses a Map Server to publish its EID-to-RLOC mappings does not need to participate further in the mapping database protocol(s). When using a LISP+ALT mapping database, for example, this means that the ETR does not need to implement GRE or BGP, which greatly simplifies its configuration and reduces its cost of operation.

Note that use of a Map Server does not preclude an ETR from also connecting to the mapping database (i.e. it could also connect to the LISP+ALT network) but doing so doesn't seem particularly useful as the whole purpose of using a Map Server is to avoid the complexity of the mapping database protocols.

#### 4.3. Map Server Processing

Once a Map Server has EID-prefixes registered by its client ETRs, it can accept and process Map-Requests for them.

In response to a Map-Request (received over the ALT if LISP+ALT is in use), the Map Server first checks to see if the destination EID matches a configured EID-prefix. If there is no match, the Map Server returns a negative Map-Reply with action code "forward-native" and a 15-minute TTL. This may occur if a Map Request is received for a configured aggregate EID-prefix for which no more-specific EID-prefix exists; it indicates the presence of a non-LISP "hole" in the aggregate EID-prefix.

Next, the Map Server checks to see if any ETRs have registered the matching EID-prefix. If none are found, then the Map Server returns a negative Map-Reply with action code "forward-native" and a 1-minute TTL.

If any of the registered ETRs for the EID-prefix have requested proxy reply service, then the Map Server answers the request instead of forwarding it. It returns a Map-Reply with the EID-prefix, RLOCs, and other information learned through the registration process.

If none of the ETRs have requested proxy reply service, then the Map Server re-encapsulates and forwards the resulting Encapsulated Map-Request to one of the registered ETRs. It does not otherwise alter the Map-Request so any Map-Reply sent by the ETR is returned to the RLOC in the Map-Request, not to the Map Server. Unless also acting as a Map Resolver, a Map Server should never receive Map-Replies; any such messages should be discarded without response, perhaps accompanied by logging of a diagnostic message if the rate of Map-Replies is suggestive of malicious traffic.

#### 4.4. Map Resolver Processing

Upon receipt of an Encapsulated Map-Request, a Map Resolver de-encapsulates the enclosed message then searches for the requested EID in its local database of mapping entries (statically configured or learned from associated ETRs if the Map Resolver is also a Map Server offering proxy reply service). If it finds a matching entry, it returns a LISP Map-Reply with the known mapping.

If the Map Resolver does not have the mapping entry and if it can determine that the EID is not in the mapping database (for example, if LISP+ALT is used, the Map Resolver will have an ALT forwarding table that covers the full EID space) it immediately returns a negative LISP Map-Reply, with action code "forward-native" and a 15-minute TTL. To minimize the number of negative cache entries needed by an ITR, the Map Resolver should return the least-specific prefix which both matches the original query and does not match any EID-prefix known to exist in the LISP-capable infrastructure.

If the Map Resolver does not have sufficient information to know whether the EID exists, it needs to forward the Map-Request to another device which has more information about the EID being requested. To do this, it forwards the unencapsulated Map-Request, with the original ITR RLOC as the source, to the mapping database system. Using LISP+ALT, the Map Resolver is connected to the ALT network and sends the Map-Request to the next ALT hop learned from its ALT BGP neighbors. The Map Resolver does not send any response to the ITR; since the source RLOC is that of the ITR, the ETR or Map Server which receives the Map-Request over the ALT and responds will do so directly to the ITR.

#### 4.4.1. Anycast Map Resolver Operation

A Map Resolver can be set up to use "anycast", where the same address is assigned to multiple Map Resolvers and is propagated through IGP routing, to facilitate the use of a topologically-close Map Resolver each ITR.

Note that Map Server associations with ETRs should not use anycast addresses as registrations need to be established between an ETR and a specific set of Map Servers, each identified by a specific registration association.

## 5. Open Issues and Considerations

There are a number of issues with the Map Server and Map Resolver design that are not yet completely understood. Among these are:

- o Constants, such as those used for Map-Register frequency, retransmission timeouts, retransmission limits, negative Map-Reply TTLs, et al are subject to further refinement as more experience with prototype deployment is gained.
- o Convergence time when an EID-to-RLOC mapping changes and mechanisms for detecting and refreshing or removing stale, cached information
- o Deployability and complexity trade-offs of implementing stronger security measures in both EID-prefix registration and Map-Request/Map-Reply processing
- o Requirements for additional state in the registration process between Map Servers and ETRs

A discussion of other issues surrounding LISP deployment may also be found in Section 15 of [LISP].

The authors expect that experimentation on the LISP pilot network will help answer open questions surrounding these and other issues.



## 6. IANA Considerations

This document makes no request of the IANA.

## 7. Security Considerations

The 2-way LISP header nonce exchange documented in [LISP] can be used to avoid ITR spoofing attacks.

To publish an authoritative EID-to-RLOC mapping with a Map Server, an ETR includes authentication data that is a hash of the message using pair-wise shared key. An implementation must support use of HMAC-SHA-1-96 [RFC2104] and should support use of HMAC-SHA-256-128 [RFC6234] (SHA-256 truncated to 128 bits).

During experimental and prototype deployment, all authentication key configuration will be manual. Should LISP and its components be considered for IETF standardization, further work will be required to follow the BCP 107 [RFC4107] recommendations on automated key management.

As noted in Section 4.2, a Map Server should verify that all EID-prefixes registered by an ETR match configuration stored on the Map Server.

The currently-defined authentication mechanism for Map-Register messages does not provide protection against "replay" attacks by a "man-in-the-middle". Additional work is needed in this area.

[LISP-SEC] defines a proposed mechanism for providing origin authentication, integrity, anti-replay protection, and prevention of man-in-the-middle and "overclaiming" attacks on the Map-Request/Map-Reply exchange. Work is ongoing on this and other proposals for resolving these open security issues

While beyond the scope of securing an individual Map Server or Map Resolver, it should be noted that a BGP-based LISP+ALT network (if ALT is used as the mapping database infrastructure) can take advantage standards work on adding security to BGP.

## 8. References

### 8.1. Normative References

- [ALT] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "LISP Alternative Topology (LISP-ALT)", draft-ietf-lisp-alt-10.txt (work in progress), December 2011.
- [LISP] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "Locator/ID Separation Protocol (LISP)", draft-ietf-lisp-22.txt (work in progress), February 2012.
- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.
- [RFC2104] Krawczyk, H., Bellare, M., and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, February 1997.
- [RFC6234] Eastlake, D. and T. Hansen, "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", RFC 6234, May 2011.

### 8.2. Informative References

- [CONS] Farinacci, D., Fuller, V., and D. Meyer, "LISP-CONS: A Content distribution Overlay Network Service for LISP", draft-meyer-lisp-cons-04.txt (work in progress), April 2008.
- [LISP-MN] Farinacci, D., Lewis, D., Meyer, D., and C. White, "LISP Mobile Node Architecture", draft-meyer-lisp-mn-06.txt (work in progress), October 2011.
- [LISP-SEC] Maino, F., Ermagan, V., Cabellos, A., Sanchez, D., and O. Bonaventure, "LISP-Security", draft-ietf-lisp-sec-01.txt (work in progress), January 2012.
- [NERD] Lear, E., "NERD: A Not-so-novel EID to RLOC Database", draft-lear-lisp-nerd-08.txt (work in progress), March 2010.
- [RFC4107] Bellovin, S. and R. Housley, "Guidelines for Cryptographic Key Management", BCP 107, RFC 4107, June 2005.

## Appendix A. Acknowledgments

The authors would like to thank Greg Schudel, Darrel Lewis, John Zwiebel, Andrew Partan, Dave Meyer, Isidor Kouvelas, Jesper Skriver, Fabio Maino, and members of the `lisp@ietf.org` mailing list for their feedback and helpful suggestions.

Special thanks are due to Noel Chiappa for his extensive work on caching with LISP-CONS, some of which may be used by Map Resolvers.

Authors' Addresses

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

Email: vaf@cisco.com

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

Email: dino@cisco.com



Network Working Group  
Internet-Draft  
Intended status: Experimental  
Expires: August 11, 2012

D. Farinacci  
D. Meyer  
J. Zwiebel  
S. Venaas  
cisco Systems  
February 8, 2012

LISP for Multicast Environments  
draft-ietf-lisp-multicast-14

Abstract

This draft describes how inter-domain multicast routing will function in an environment where Locator/ID Separation is deployed using the LISP architecture.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on August 11, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Requirements Notation . . . . .	4
2. Introduction . . . . .	5
3. Definition of Terms . . . . .	7
4. Basic Overview . . . . .	10
5. Source Addresses versus Group Addresses . . . . .	13
6. Locator Reachability Implications on LISP-Multicast . . . . .	14
7. Multicast Protocol Changes . . . . .	15
8. LISP-Multicast Data-Plane Architecture . . . . .	18
8.1. ITR Forwarding Procedure . . . . .	18
8.1.1. Multiple RLOCs for an ITR . . . . .	18
8.1.2. Multiple ITRs for a LISP Source Site . . . . .	19
8.2. ETR Forwarding Procedure . . . . .	19
8.3. Replication Locations . . . . .	20
9. LISP-Multicast Interworking . . . . .	21
9.1. LISP and non-LISP Mixed Sites . . . . .	21
9.1.1. LISP Source Site to non-LISP Receiver Sites . . . . .	22
9.1.2. Non-LISP Source Site to non-LISP Receiver Sites . . . . .	23
9.1.3. Non-LISP Source Site to Any Receiver Site . . . . .	24
9.1.4. Unicast LISP Source Site to Any Receiver Sites . . . . .	25
9.1.5. LISP Source Site to Any Receiver Sites . . . . .	25
9.2. LISP Sites with Mixed Address Families . . . . .	26
9.3. Making a Multicast Interworking Decision . . . . .	28
10. Considerations when RP Addresses are Embedded in Group Addresses . . . . .	29
11. Taking Advantage of Upgrades in the Core . . . . .	30
12. Mtrace Considerations . . . . .	31
13. Security Considerations . . . . .	32
14. Acknowledgments . . . . .	33
15. IANA Considerations . . . . .	34
16. References . . . . .	35
16.1. Normative References . . . . .	35
16.2. Informative References . . . . .	36
Appendix A. Document Change Log . . . . .	37
A.1. Changes to draft-ietf-lisp-multicast-14.txt . . . . .	37
A.2. Changes to draft-ietf-lisp-multicast-13.txt . . . . .	37
A.3. Changes to draft-ietf-lisp-multicast-12.txt . . . . .	37
A.4. Changes to draft-ietf-lisp-multicast-11.txt . . . . .	37
A.5. Changes to draft-ietf-lisp-multicast-10.txt . . . . .	37
A.6. Changes to draft-ietf-lisp-multicast-09.txt . . . . .	37
A.7. Changes to draft-ietf-lisp-multicast-08.txt . . . . .	37
A.8. Changes to draft-ietf-lisp-multicast-07.txt . . . . .	38
A.9. Changes to draft-ietf-lisp-multicast-06.txt . . . . .	38
A.10. Changes to draft-ietf-lisp-multicast-05.txt . . . . .	38
A.11. Changes to draft-ietf-lisp-multicast-04.txt . . . . .	38
A.12. Changes to draft-ietf-lisp-multicast-03.txt . . . . .	38
A.13. Changes to draft-ietf-lisp-multicast-02.txt . . . . .	39



A.14. Changes to draft-ietf-lisp-multicast-01.txt . . . . .	39
A.15. Changes to draft-ietf-lisp-multicast-00.txt . . . . .	39
Authors' Addresses . . . . .	40

## 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

The Locator/ID Separation Architecture [LISP] provides a mechanism to separate out Identification and Location semantics from the current definition of an IP address. By creating two namespaces, an Endpoint ID (EID) namespace used by sites and a Routing Locator (RLOC) namespace used by core routing, the core routing infrastructure can scale by doing topological aggregation of routing information.

Since LISP creates a new namespace, a mapping function must exist to map a site's EID prefixes to its associated locators. For unicast packets, both the source address and destination address must be mapped. For multicast packets, only the source address needs to be mapped. The destination group address doesn't need to be mapped because the semantics of an IPv4 or IPv6 group address are logical in nature and not topology-dependent. Therefore, this specification focuses on to map a source EID address of a multicast flow during distribution tree setup and packet delivery.

This specification will address the following scenarios:

1. How a multicast source host in a LISP site sends multicast packets to receivers inside of its site as well as to receivers in other sites that are LISP enabled.
2. How inter-domain (or between LISP sites) multicast distribution trees are built and how forwarding of multicast packets leaving a source site toward receivers sites is performed.
3. What protocols are affected and what changes are required to such multicast protocols.
4. How ASM-mode (Any Source Multicast), SSM-mode (Single Source Multicast), and Bidir-mode (Bidirectional Shared Trees) service models will operate.
5. How multicast packet flow will occur for multiple combinations of LISP and non-LISP capable source and receiver sites, for example:
  - A. How multicast packets from a source host in a LISP site are sent to receivers in other sites when they are all non-LISP sites.
  - B. How multicast packets from a source host in a LISP site are sent to receivers in both LISP-enabled sites and non-LISP sites.

- C. How multicast packets from a source host in a non-LISP site are sent to receivers in other sites when they are all LISP-enabled sites.
- D. How multicast packets from a source host in a non-LISP site are sent to receivers in both LISP-enabled sites and non-LISP sites.

This specification focuses on what changes are needed to the multicast routing protocols to support LISP-Multicast as well as other protocols used for inter-domain multicast, such as Multi-protocol BGP (MBGP) [RFC4760]. The approach proposed in this specification requires no packet format changes to the protocols and no operational procedural changes to the multicast infrastructure inside of a site when all sources and receivers reside in that site, even when the site is LISP enabled. That is, internal operation of multicast is unchanged regardless of whether or not the site is LISP enabled or whether or not receivers exist in other sites which are LISP-enabled.

Therefore, we see only operational (and not protocol) changes for PIM-ASM [RFC4601], MSDP [RFC3618], and PIM-SSM [RFC4607]. Bidir-PIM [RFC5015], which typically does not run in an inter-domain environment is not addressed in depth in this version of the specification.

Also, the current version of this specification does not describe multicast-based Traffic Engineering relative to the TE-ITR (Traffic Engineering based Ingress Tunnel Router) and TE-ETR (Traffic Engineering based Egress Tunnel Router) descriptions in [LISP]. Further work is also needed to determine the detailed behavior for multicast proxy ITRs (mPITRs) (Section 9.1.3), mtrace (Section 12), and locator reachability (Section 6). Finally, further deployment and experimentation would be useful to understand the real-life performance of the LISP-Multicast solution. For instance, the design optimizes for minimal state and control traffic in the core, but can in some cases cause extra multicast traffic to be sent Section 8.1.2.

Issues and concerns about the deployment of LISP for Internet traffic are discussed in [LISP]. Section 12 provides additional issues and concerns raised by this document.

### 3. Definition of Terms

The terminology in this section is consistent with the definitions in [LISP] but is extended specifically to deal with the application of the terminology to multicast routing.

**LISP-Multicast:** a reference to the design in this specification. That is, when any site that is participating in multicast communication has been upgraded to be a LISP site, the operation of control-plane and data-plane protocols is considered part of the LISP-Multicast architecture.

**Endpoint ID (EID):** a 32-bit (for IPv4) or 128-bit (for IPv6) value used in the source address field of the first (most inner) LISP header of a multicast packet. The host obtains a destination group address the same way it obtains one today, as it would when it is a non-LISP site. The source EID is obtained via existing mechanisms used to set a host's "local" IP address. An EID is allocated to a host from an EID prefix block associated with the site the host is located in. An EID can be used by a host to refer to another host, as when it joins an SSM (S-EID,G) route using IGMP version 3 [RFC4604]. LISP uses Provider Independent (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.

**Routing Locator (RLOC):** the IPv4 or IPv6 address of an ingress tunnel router (ITR), the router in the multicast source host's site that encapsulates multicast packets. 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 Provider Assigned (PA) addresses. Multiple RLOCs can be assigned to the same ITR device or to multiple ITR devices at a site.

**Ingress Tunnel Router (ITR):** a router which accepts an IP multicast 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 multicast address opaquely so it doesn't need to perform a map lookup on the group address because it is topologically insignificant. The router then prepends an "outer" IP header with one of its globally-routable RLOCs as the source address field. This RLOC is known to other multicast receiver

sites which have used the mapping database to join a multicast tree for which the ITR is the root. In general, an ITR receives IP packets from site end systems on one side and sends LISP-encapsulated multicast IP packets out all external interfaces which have been joined.

An ITR would receive a multicast packet from a source inside of its site when 1) it is on the path from the multicast source to internally joined receivers, or 2) when it is on the path from the multicast source to externally joined receivers.

**Egress Tunnel Router (ETR):** a router that is on the path from a multicast source host in another site to a multicast receiver in its own site. An ETR accepts a PIM Join/Prune message from a site internal PIM router destined for the source's EID in the multicast source site. The ETR maps the source EID in the Join/Prune message to an RLOC address based on the EID-to-RLOC mapping. This sets up the ETR to accept multicast encapsulated packets from the ITR in the source multicast site. A multicast ETR decapsulates multicast encapsulated packets and replicates them on interfaces leading to internal receivers.

**xTR:** is a reference to an ITR or ETR when direction of data flow is not part of the context description. xTR refers to the router that is the tunnel endpoint. Used synonymously with the term "Tunnel Router". For example, "An xTR can be located at the Customer Edge (CE) router", meaning both ITR and ETR functionality can be at the CE router.

**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 headers and an ETR strips headers. A LISP encapsulated multicast packet will have an "inner" header with the source EID in the source field; an "outer" header with the source RLOC in the source field; and the same globally unique group address in the destination field of both the inner and outer header.

**(S,G) State:** the formal definition is in the PIM Sparse Mode [RFC4601] specification. For this specification, the term is used generally to refer to multicast state. Based on its topological location, the (S,G) state resides in routers can be either (S-EID,G) state (at a location where the (S,G) state resides) or (S-RLOC,G) state (in the Internet core).

**(S-EID,G) State:** refers to multicast state in multicast source and receiver sites where S-EID is the IP address of the multicast source host (its EID). An S-EID can appear in an IGMPv3 report, an MSDP SA message or a PIM Join/Prune message that travels inside

of a site.

(S-RLOC,G) State: refers to multicast state in the core where S is a source locator (the IP address of a multicast ITR) of a site with a multicast source. The (S-RLOC,G) is mapped from (S-EID,G) entry by doing a mapping database lookup for the EID prefix that S-EID maps to. An S-RLOC can appear in a PIM Join/Prune message when it travels from an ETR to an ITR over the Internet core.

uLISP Site: a unicast only LISP site according to [LISP] which has not deployed the procedures of this specification and therefore, for multicast purposes, follows the procedures from Section 9. A uLISP site can be a traditional multicast site.

LISP Site: a unicast LISP site (uLISP Site) that is also multicast capable according to the procedures in this specification.

mPETR: this is a multicast proxy-ETR that is responsible for advertising a very coarse EID prefix which non-LISP and uLISP sites can target their (S-EID,G) PIM Join/Prune message to. mPETRs are used so LISP source multicast sites can send multicast packets using source addresses from the EID namespace. mPETRs act as Proxy ETRs for supporting multicast routing in a LISP infrastructure. It is likely an uPITR [INTWORK] and a mPETR will be co-located since the single device advertises a coarse EID-prefix in the underlying unicast routing system.

Mixed Locator-Sets: this is a locator-set for a LISP database mapping entry where the RLOC addresses in the locator-set are in both IPv4 and IPv6 format.

Unicast Encapsulated PIM Join/Prune Message: this is a standard PIM Join/Prune message (LISP encapsulated with destination UDP port 4341) which is sent by ETRs at multicast receiver sites to an ITR at a multicast source site. This message is sent periodically as long as there are interfaces in the OIF-list for the (S-EID,G) entry the ETR is joining for.

OIF-list: this is notation to describe the outgoing interface list a multicast router stores per multicast routing table entry so it knows what interfaces to replicate multicast packets on.

RPF: Reverse Path Forwarding is a procedure used by multicast routers. A router will accept a multicast packet for forwarding if the packet was received on the path that the router would use to forward unicast packets to the multicast packet's source.

#### 4. Basic Overview

LISP, when used for unicast routing, increases the site's ability to control ingress traffic flows. Egress traffic flows are controlled by the IGP in the source site. For multicast, the IGP coupled with PIM can decide which path multicast packets ingress. By using the traffic engineering features of LISP [LISP], a multicast source site can control the egress of its multicast traffic. By controlling the priorities of locators from a mapping database entry, a source multicast site can control which way multicast receiver sites join to the source site.

At this point in time, there is no requirement for different locator-sets, priority, and weight policies for multicast than there is for unicast. However, when traffic engineering policies are different for unicast versus multicast flows, it will be desirable to use multicast-based priority and weight values in Map-Reply messages.

The fundamental multicast forwarding model is to encapsulate a multicast packet into another multicast packet. An ITR will encapsulate multicast packets received from sources that it serves in a LISP multicast header. The destination group address from the inner header is copied to the destination address of the outer header. The inner source address is the EID of the multicast source host and the outer source address is the RLOC of the encapsulating ITR.

The LISP-Multicast architecture will follow this high-level protocol and operational sequence:

1. Receiver hosts in multicast sites will join multicast content the way they do today, they use IGMP. When they use IGMPv3 where they specify source addresses, they use source EIDs, that is they join (S-EID,G). If the multicast source is external to this receiver site, the PIM Join/Prune message flows toward the ETRs, finding the shortest exit (that is the closest exit for the Join/Prune message and the closest entrance for the multicast packet to the receiver).
2. The ETR does a mapping database lookup for S-EID. If the mapping is cached from a previous lookup (from either a previous Join/Prune for the source multicast site or a unicast packet that went to the site), it will use the RLOC information from the mapping. The ETR will use the same priority and weighting mechanism as for unicast. So the source site can decide which way multicast packets egress.



3. The ETR will build two PIM Join/Prune messages, one that contains a (S-EID,G) entry that is unicast to the ITR that matches the RLOC the ETR selects, and the other which contains a (S-RLOC,G) entry so the core network can create multicast state from this ETR to the ITR.
4. When the ITR gets the unicast Join/Prune message (see Section 3 for formal definition), it will process (S-EID,G) entries in the message and propagate them inside of the site where it has explicit routing information for EIDs via the IGP. When the ITR receives the (S-RLOC,G) PIM Join/Prune message it will process it like any other join it would get in today's Internet. The S-RLOC address is the IP address of this ITR.
5. At this point there is (S-EID,G) state from the joining host in the receiver multicast site to the ETR of the receiver multicast site. There is (S-RLOC,G) state across the core network from the ETR of the multicast receiver site to the ITR in the multicast source site and (S-EID,G) state in the source multicast site. Note, the (S-EID,G) state is the same S-EID in each multicast site. As other ETRs join the same multicast tree, they can join through the same ITR (in which case the packet replication is done in the core) or a different ITR (in which case the packet replication is done at the source site).
6. When a packet is originated by the multicast host in the source site, the packet will flow to one or more ITRs which will prepend a LISP header. By copying the group address to the outer destination address field, the ITR insert its own locator address in the outer source address field. The ITR will look at its (S-RLOC,G) state, where S-RLOC is its own locator address, and replicate the packet on each interface a (S-RLOC,G) joined was received on. The core has (S-RLOC,G) so where fanout occurs to multiple sites, a core router will do packet replication.
7. When either the source site or the core replicates the packet, the ETR will receive a LISP packet with a destination group address. It will decapsulate packets because it has receivers for the group. Otherwise, it would have not received the packets because it would not have joined. The ETR decapsulates and does a (S-EID,G) lookup in its multicast FIB to forward packets out one or more interfaces to forward the packet to internal receivers.

This architecture is consistent and scalable with the architecture presented in [LISP] where multicast state in the core operates on locators and multicast state at the sites operates on EIDs.

Alternatively, [LISP] also has a mechanism where (S-EID,G) state can reside in the core through the use of RPF-vectors [RFC5496] in PIM Join/Prune messages. However, few PIM implementations support RPF vectors and LISP should avoid S-EID state in the core. See Section 5 for details.

However, some observations can be made on the algorithm above. The control plane can scale but at the expense of sending data to sites which may have not joined the distribution tree where the encapsulated data is being delivered. For example, one site joins (S-EID1,G) and another site joins (S-EID2,G). Both EIDs are in the same multicast source site. Both multicast receiver sites join to the same ITR with state (S-RLOC,G) where S-RLOC is the RLOC for the ITR. The ITR joins both (S-EID1,G) and (S-EID2,G) inside of the site. The ITR receives (S-RLOC,G) joins and populates the OIF-list state for it. Since both (S-EID1,G) and (S-EID2,G) map to the one (S-RLOC,G) packets will be delivered by the core to both multicast receiver sites even though each have joined a single source-based distribution tree. This behavior is a consequence of the many-to-one mapping between S-EIDs and a S-RLOC.

There is a possible solution to this problem which reduces the number of many-to-one occurrences of (S-EID,G) entries aggregating into a single (S-RLOC,G) entry. If a physical ITR can be assigned multiple RLOC addresses and these addresses are advertised in mapping database entries, then ETRs at receiver sites have more RLOC address options and therefore can join different (RLOC,G) entries for each (S-EID,G) entry joined at the receiver site. It would not scale to have a one-to-one relationship between the number of S-EID sources at a source site and the number of RLOCs assigned to all ITRs at the site, but "n" can reduce to a smaller number in the "n-to-1" relationship. And in turn, reduce the opportunity for data packets to be delivered to sites for groups not joined.

## 5. Source Addresses versus Group Addresses

Multicast group addresses don't have to be associated with either the EID or RLOC namespace. They actually are a namespace of their own that can be treated as logical with relatively opaque allocation. So, by their nature, they don't detract from an incremental deployment of LISP-Multicast.

As for source addresses, as in the unicast LISP scenario, there is a decoupling of identification from location. In a LISP site, packets are originated from hosts using their allocated EIDs. EID addresses are used to identify the host as well as where in the site's topology the host resides but not how and where it is attached to the Internet.

Therefore, when multicast distribution tree state is created anywhere in the network on the path from any multicast receiver to a multicast source, EID state is maintained at the source and receiver multicast sites, and RLOC state is maintained in the core. That is, a multicast distribution tree will be represented as a 3-tuple of  $\{(S-EID, G) (S-RLOC, G) (S-EID, G)\}$  where the first element of the 3-tuple is the state stored in routers from the source to one or more ITRs in the source multicast site, the second element of the 3-tuple is the state stored in routers downstream of the ITR, in the core, to all LISP receiver multicast sites, and the third element in the 3-tuple is the state stored in the routers downstream of each ETR, in each receiver multicast site, reaching each receiver. Note that  $(S-EID, G)$  is the same in both the source and receiver multicast sites.

The concatenation/mapping from the first element to the second element of the 3-tuples is done by the ITR and from the second element to the third element is done at the ETRs.

## 6. Locator Reachability Implications on LISP-Multicast

Multicast state as it is stored in the core is always (S,G) state as it exists today or (S-RLOC,G) state as it will exist when LISP sites are deployed. The core routers cannot distinguish one from the other. They don't need to because it is state that RPFs against the core routing tables in the RLOC namespace. The difference is where the root of the distribution tree for a particular source is. In the traditional multicast core, the source S is the source host's IP address. For LISP-Multicast the source S is a single ITR of the multicast source site.

An ITR is selected based on the LISP EID-to-RLOC mapping used when an ETR propagates a PIM Join/Prune message out of a receiver multicast site. The selection is based on the same algorithm an ITR would use to select an ETR when sending a unicast packet to the site. In the unicast case, the ITR can change on a per-packet basis depending on the reachability of the ETR. So an ITR can change relatively easily using local reachability state. However, in the multicast case, when an ITR goes unreachable, new distribution tree state must be built because the encapsulating root has changed. This is more significant than an RPF-change event, where any router would typically locally change its RPF-interface for its existing tree state. But when an encapsulating LISP-Multicast ITR goes unreachable, new distribution state must be rebuilt and reflect the new encapsulator. Therefore, when an ITR goes unreachable, all ETRs that are currently joined to that ITR will have to trigger a new Join/Prune message for (S-RLOC,G) to the new ITR as well as send a unicast encapsulated Join/Prune message telling the new ITR which (S-EID,G) is being joined.

This issue can be mitigated by using anycast addressing for the ITRs so the problem does reduce to an RPF change in the core, but still requires a unicast encapsulated Join/Prune message to tell the new ITR about (S-EID,G). The problem with this approach is that the ETR really doesn't know when the ITR has changed so the new anycast ITR will get the (S-EID,G) state only when the ETR sends it the next time during its periodic sending procedures.

## 7. Multicast Protocol Changes

A number of protocols are used today for inter-domain multicast routing:

IGMPv1-v3, MLDv1-v2: These protocols [RFC4604] do not require any changes for LISP-Multicast for two reasons. One being that they are link-local and not used over site boundaries and second, they advertise group addresses that don't need translation. Where source addresses are supplied in IGMPv3 and MLDv2 messages, they are semantically regarded as EIDs and don't need to be converted to RLOCs until the multicast tree-building protocol, such as PIM, is received by the ETR at the site boundary. Addresses used for IGMP and MLD come out of the source site's allocated addresses which are therefore from the EID namespace.

MBGP: Even though MBGP [RFC4760] is not a multicast routing protocol, it is used to find multicast sources when the unicast BGP peering topology and the multicast MBGP peering topology are not congruent. When MBGP is used in a LISP-Multicast environment, the prefixes which are advertised are from the RLOC namespace. This allows receiver multicast sites to find a path to the source multicast site's ITRs. MBGP peering addresses will be from the RLOC namespace. There are no MBGP protocol changes required to support LISP-Multicast.

MSDP: MSDP [RFC3618] is used to announce active multicast sources to other routing domains (or LISP sites). The announcements come from the PIM Rendezvous Points (RPs) from sites where there are active multicast sources sending to various groups. In the context of LISP-Multicast, the source addresses advertised in MSDP will semantically be from the EID namespace since they describe the identity of a source multicast host. It will be true that the state stored in MSDP caches from core routers will be from the EID namespace. An RP address inside of site will be from the EID namespace so it can be advertised and reached by internal unicast routing mechanism. However, for MSDP peer-RPF checking to work properly across sites, the RP addresses must be converted or mapped into a routable address that is advertised and maintained in the BGP routing tables in the core. MSDP peering addresses can come out of either the EID or a routable address namespace. And the choice can be made unilaterally because the ITR at the site will determine which namespace the destination peer address is out of by looking in the mapping database service. There are no MSDP protocol changes required to support LISP-Multicast.

**PIM-SSM:** In the simplest form of distribution tree building, when PIM operates in SSM mode [RFC4607], a source distribution tree is built and maintained across site boundaries. In this case, there is a small modification to how PIM Join/Prune messages are sent by the LISP-Multicast component. No modifications to any message format, but to support taking a Join/Prune message originated inside of a LISP site with embedded addresses from the EID namespace and converting them to addresses from the RLOC namespace when the Join/Prune message crosses a site boundary. This is similar to the requirements documented in [RFC5135].

**PIM-Bidir:** Bidirectional PIM [RFC5015] is typically run inside of a routing domain, but if deployed in an inter-domain environment, one would have to decide if the RP address of the shared-tree would be from the EID namespace or the RLOC namespace. If the RP resides in a site-based router, then the RP address is from the EID namespace. If the RP resides in the core where RLOC addresses are routed, then the RP address is from the RLOC namespace. This could be easily distinguishable if the EID address were well-known address allocation block from the RLOC namespace. Also, when using Embedded-RP for RP determination [RFC3956], the format of the group address could indicate the namespace the RP address is from. However, refer to Section 10 for considerations core routers need to make when using Embedded-RP IPv6 group addresses. When using Bidir-PIM for inter-domain multicast routing, it is recommended to use statically configured RPs. Allowing core routers to associate a Bidir group's RP address with an ITR's RLOC address. And site routers to associate the Bidir group's RP address as an EID address. With respect to DF-election in Bidir PIM, no changes are required since all messaging and addressing is link-local.

**PIM-ASM:** The ASM mode of PIM [RFC4601], the most popular form of PIM, is deployed in the Internet today is by having shared-trees within a site and using source-trees across sites. By the use of MSDP and PIM-SSM techniques described above, multicast connectivity can occur across LISP sites. Having said that, that means there are no special actions required for processing (\*,G) or (S,G,R) Join/Prune messages since they all operate against the shared-tree which is site resident. Just like with ASM, there is no (\*,G) in the core when LISP-Multicast is in use. This is also true for the RP-mapping mechanisms Auto-RP and BSR.

Based on the protocol description above, the conclusion is that there are no protocol message format changes, just a translation function performed at the control-plane. This will make for an easier and faster transition for LISP since fewer components in the network have to change.

It should also be stated just like it is in [LISP] that no host changes, whatsoever, are required to have a multicast source host send multicast packets and for a multicast receiver host to receive multicast packets.

## 8. LISP-Multicast Data-Plane Architecture

The LISP-Multicast data-plane operation conforms to the operation and packet formats specified in [LISP]. However, encapsulating a multicast packet from an ITR is a much simpler process. The process is simply to copy the inner group address to the outer destination address. And to have the ITR use its own IP address (its RLOC) as the source address. The process is simpler for multicast because there is no EID-to-RLOC mapping lookup performed during packet forwarding.

In the decapsulation case, the ETR simply removes the outer header and performs a multicast routing table lookup on the inner header (S-EID,G) addresses. Then the OIF-list for the (S-EID,G) entry is used to replicate the packet on site-facing interfaces leading to multicast receiver hosts.

There is no Data-Probe logic for ETRs as there can be in the unicast forwarding case.

### 8.1. ITR Forwarding Procedure

The following procedure is used by an ITR, when it receives a multicast packet from a source inside of its site:

1. A multicast data packet sent by a host in a LISP site will have the source address equal to the host's EID and the destination address equal to the group address of the multicast group. It is assumed the group information is obtained by current methods. The same is true for a multicast receiver to obtain the source and group address of a multicast flow.
2. When the ITR receives a multicast packet, it will have both S-EID state and S-RLOC state stored. Since the packet was received on a site-facing interface, the RPF lookup is based on the S-EID state. If the RPF check succeeds, then the OIF-list contains interfaces that are site-facing and external-facing. For the site-facing interfaces, no LISP header is prepended. For the external-facing interfaces a LISP header is prepended. When the ITR prepends a LISP header, it uses its own RLOC address as the source address and copies the group address supplied by the IP header the host built as the outer destination address.

#### 8.1.1. Multiple RLOCs for an ITR

Typically, an ITR will have a single RLOC address but in some cases there could be multiple RLOC addresses assigned from either the same or different service providers. In this case when (S-RLOC,G) Join/



Prune messages are received for each RLOC, there is a OIF-list merging action that must take place. Therefore, when a packet is received from a site-facing interface that matches on a (S-EID,G) entry, the interfaces of the OIF-list from all (RLOC,G) entries joined to the ITR as well as the site-facing OIF-list joined for (S-EID,G) must be part be included in packet replication. In addition to replicating for all types of OIF-lists, each oif entry must be tagged with the RLOC address, so encapsulation uses the outer source address for the RLOC joined.

#### 8.1.2. Multiple ITRs for a LISP Source Site

Note when ETRs from different multicast receiver sites receive (S-EID,G) joins, they may select a different S-RLOC for a multicast source site due to policy (the multicast ITR can return different multicast priority and weight values per ETR Map-Request). In this case, the same (S-EID,G) is being realized by different (S-RLOC,G) state in the core. This will not result in duplicate packets because each ITR in the multicast source site will choose their own RLOC for the source address for encapsulated multicast traffic. The RLOC addresses are the ones joined by remote multicast ETRs.

When different (S-EID,G) traffic is combined into a single (RLOC,G) core distribution tree, this may cause traffic to go to a receiver multicast site when it does not need to. This happens when one receiver multicast site joins (S1-EID,Gi) through a core distribution tree of (RLOC1,Gi) and another multicast receiver site joins (S2-EID,Gi) through the same core distribution tree of (RLOC1,Gi). When ETRs decapsulate such traffic, they should know from their local (S-EID,G) state if the packet should be forwarded. If there is no (S-EID,G) state that matches the inner packet header, the packet is discarded.

#### 8.2. ETR Forwarding Procedure

The following procedure is used by an ETR, when it receives a multicast packet from a source outside of its site:

1. When a multicast data packet is received by an ETR on an external-facing interface, it will do an RPF lookup on the S-RLOC state it has stored. If the RPF check succeeds, the interfaces from the OIF-list are used for replication to interfaces that are site-facing as well as interfaces that are external-facing (this ETR can also be a transit multicast router for receivers outside of its site). When the packet is to be replicated for an external-facing interface, the LISP encapsulation header are not stripped. When the packet is replicated for a site-facing interface, the encapsulation header is stripped.

2. The packet without a LISP header is now forwarded down the (S-EID,G) distribution tree in the receiver multicast site.

### 8.3. Replication Locations

Multicast packet replication can happen in the following topological locations:

- o In an IGP multicast router inside a site which operates on S-EIDs.
- o In a transit multicast router inside of the core which operates on S-RLOCs.
- o At one or more ETR routers depending on the path a Join/Prune message exits a receiver multicast site.
- o At one or more ITR routers in a source multicast site depending on what priorities are returned in a Map-Reply to receiver multicast sites.

In the last case the source multicast site can do replication rather than having a single exit from the site. But this only can occur when the priorities in the Map-Reply are modified for different receiver multicast site so that the PIM Join/Prune messages arrive at different ITRs.

This policy technique, also used in [ALT] for unicast, is useful for multicast to mitigate the problems of changing distribution tree state as discussed in Section 6.

## 9. LISP-Multicast Interworking

This section will describe the multicast corollary to [INTWORK] which describes the interworking of multicast routing among LISP and non-LISP sites.

### 9.1. LISP and non-LISP Mixed Sites

Since multicast communication can involve more than two entities to communicate together, the combinations of interworking scenarios are more involved. However, the state maintained for distribution trees at the sites is the same regardless of whether or not the site is LISP enabled or not. So most of the implications are in the core with respect to storing routable EID prefixes from either PA or PI blocks.

Before enumerating the multicast interworking scenarios, let's define 3 deployment states of a site:

- o A non-LISP site which will run PIM-SSM or PIM-ASM with MSDP as it does today. The addresses for the site are globally routable.
- o A site that deploys LISP for unicast routing. The addresses for the site are not globally routable. Let's define the name for this type of site as a uLISP site.
- o A site that deploys LISP for both unicast and multicast routing. The addresses for the site are not globally routable. Let's define the name for this type of site as a LISP-Multicast site.

What will not be considered is a LISP site enabled for multicast purposes only but do consider a uLISP site as documented in [INTWORK]. In this section there is no discussion how a LISP site sends multicast packets when all receiver sites are LISP-Multicast enabled; that has been discussed in previous sections.

The following scenarios exist to make LISP-Multicast sites interwork with non-LISP-Multicast sites:

1. A LISP site must be able to send multicast packets to receiver sites which are a mix of non-LISP sites and uLISP sites.
2. A non-LISP site must be able to send multicast packets to receiver sites which are a mix of non-LISP sites and uLISP sites.
3. A non-LISP site must be able to send multicast packets to receiver sites which are a mix of LISP sites, uLISP sites, and non-LISP sites.

4. A uLISP site must be able to send multicast packets to receiver sites which are a mix of LISP sites, uLISP sites, and non-LISP sites.
5. A LISP site must be able to send multicast packets to receiver sites which are a mix of LISP sites, uLISP sites, and non-LISP sites.

#### 9.1.1. LISP Source Site to non-LISP Receiver Sites

In the first scenario, a site is LISP capable for both unicast and multicast traffic and as such operates on EIDs. Therefore there is a possibility that the EID prefix block is not routable in the core. For LISP receiver multicast sites this isn't a problem but for non-LISP or uLISP receiver multicast sites, when a PIM Join/Prune message is received by the edge router, it has no route to propagate the Join/Prune message out of the site. This is no different than the unicast case that LISP-NAT in [INTWORK] solves.

LISP-NAT allows a unicast packet that exits a LISP site to get its source address mapped to a globally routable address before the ITR realizes that it should not encapsulate the packet destined to a non-LISP site. For a multicast packet to leave a LISP site, distribution tree state needs to be built so the ITR can know where to send the packet. So the receiver multicast sites need to know about the multicast source host by its routable address and not its EID address. When this is the case, the routable address is the (S-RLOC,G) state that is stored and maintained in the core routers. It is important to note that the routable address for the host cannot be the same as an RLOC for the site because it is desirable for ITRs to process a received PIM Join/Prune message from an external-facing interface to be propagated inside of the site so the site-part of the distribution tree is built.

Using a globally routable source address allows non-LISP and uLISP multicast receiver to join, create, and maintain a multicast distribution tree. However, the LISP multicast receiver site will want to perform an EID-to-RLOC mapping table lookup when a PIM Join/Prune message is received on a site-facing interface. It does this because it wants to find a (S-RLOC,G) entry to Join in the core. So there is a conflict of behavior between the two types of sites.

The solution to this problem is the same as when an ITR wants to send a unicast packet to a destination site but needs determine if the site is LISP capable or not. When it is not LISP capable, the ITR does not encapsulate the packet. So for the multicast case, when ETR receives a PIM Join/Prune message for (S-EID,G) state, it will do a mapping table lookup on S-EID. In this case, S-EID is not in the

mapping database because the source multicast site is using a routable address and not an EID prefix address. So the ETR knows to simply propagate the PIM Join/Prune message to a external-facing interface without converting the (S-EID,G) because it is an (S,G) where S is routable and reachable via core routing tables.

Now that the multicast distribution tree is built and maintained from any non-LISP or uLISP receiver multicast site, the way packet forwarding model is performed can be explained.

Since the ITR in the source multicast site has never received a unicast encapsulated PIM Join/Prune message from any ETR in a receiver multicast site, it knows there are no LISP-Multicast receiver sites. Therefore, there is no need for the ITR to encapsulate data. Since it will know a priori (via configuration) that its site's EIDs are not routable (and not registered to the mapping database system), it assumes that the multicast packets from the source host are sent by a routable address. That is, it is the responsibility of the multicast source host's system administrator to ensure that the source host sends multicast traffic using a routable source address. When this happens, the ITR acts simply as a router and forwards the multicast packet like an ordinary multicast router.

There is an alternative to using a LISP-NAT scheme just like there is for unicast [INTWORK] forwarding by using Proxy Tunnel Routers (PxTRs). This can work the same way for multicast routing as well, but the difference is that non-LISP and uLISP sites will send PIM Join/Prune messages for (S-EID,G) which make their way in the core to multicast PxTRs. Let's call this use of a PxTR as a "Multicast Proxy-ETR" (or mPETR). Since the mPETRs advertise very coarse EID prefixes, they draw the PIM Join/Prune control traffic making them the target of the distribution tree. To get multicast packets from the LISP source multicast sites, the tree needs to be built on the path from the mPETR to the LISP source multicast site. To make this happen the mPETR acts as a "Proxy ETR" (where in unicast it acts as a "Proxy ITR", or an uPITR [INTWORK]).

The existence of mPETRs in the core allows source multicast site ITRs to encapsulate multicast packets according to (S-RLOC,G) state. The (S-RLOC,G) state is built from the mPETRs to the multicast ITRs. The encapsulated multicast packets are decapsulated by mPETRs and then forwarded according to (S-EID,G) state. The (S-EID,G) state is built from the non-LISP and uLISP receiver multicast sites to the mPETRs.

#### 9.1.2. Non-LISP Source Site to non-LISP Receiver Sites

Clearly non-LISP multicast sites can send multicast packets to non-LISP receiver multicast sites. That is what they do today. However,

discussion is required to show how non-LISP multicast sites send multicast packets to uLISP receiver multicast sites.

Since uLISP receiver multicast sites are not targets of any (S,G) state, they simply send (S,G) PIM Join/Prune messages toward the non-LISP source multicast site. Since the source multicast site, in this case has not been upgraded to LISP, all multicast source host addresses are routable. So this case is simplified to where a uLISP receiver multicast site looks to the source multicast site as a non-LISP receiver multicast site.

#### 9.1.3. Non-LISP Source Site to Any Receiver Site

When a non-LISP source multicast site has receivers in either a non-LISP/uLISP site or a LISP site, one needs to decide how the LISP receiver multicast site will attach to the distribution tree. It is known from Section 9.1.2 that non-LISP and uLISP receiver multicast sites can join the distribution tree, but a LISP receiver multicast site ETR will need to know if the source address of the multicast source host is routable or not. It has been shown in Section 9.1.1 that an ETR, before it sends a PIM Join/Prune message on an external-facing interface, does a EID-to-RLOC mapping lookup to determine if it should convert the (S,G) state from a PIM Join/Prune message received on a site-facing interface to a (S-RLOC,G). If the lookup fails, the ETR can conclude the source multicast site is a non-LISP site so it simply forwards the Join/Prune message (it also doesn't need to send a unicast encapsulated Join/Prune message because there is no ITR in a non-LISP site and there is namespace continuity between the ETR and source).

For a non-LISP source multicast site, (S-EID,G) state could be limited to the edges of the network with the use of multicast proxy-ITRs (mPITRs). The mPITRs can take native, unencapsulated multicast packets from non-LISP source multicast and uLISP sites and encapsulate them to ETRs in receiver multicast sites or to mPETRs that can decapsulate for non-LISP receiver multicast or uLISP sites. The mPITRs are responsible for sending (S-EID,G) joins to the non-LISP source multicast site. To connect the distribution trees together, multicast ETRs will need to be configured with the mPITR's RLOC addresses so they can send both (S-RLOC,G) joins to build a distribution tree to the mPITR as well as for sending unicast joins to mPITRs so they can propagate (S-EID,G) joins into source multicast sites. The use of mPITRs is undergoing more study and is work in progress.

#### 9.1.4. Unicast LISP Source Site to Any Receiver Sites

In the last section, it was explained how an ETR in a multicast receiver site can determine if a source multicast site is LISP-enabled by looking into the mapping database. When the source multicast site is a uLISP site, it is LISP enabled but the ITR, by definition is not capable of doing multicast encapsulation. So for the purposes of multicast routing, the uLISP source multicast site is treated as non-LISP source multicast site.

Non-LISP receiver multicast sites can join distribution trees to a uLISP source multicast site since the source site behaves, from a forwarding perspective, as a non-LISP source site. This is also the case for a uLISP receiver multicast site since the ETR does not have multicast functionality built-in or enabled.

Special considerations are required for LISP receiver multicast sites since they think the source multicast site is LISP capable, the ETR cannot know if ITR is LISP-Multicast capable. To solve this problem, each mapping database entry will have a multicast 2-tuple (Mpriority, Mweight) per RLOC [LISP]. When the Mpriority is set to 255, the site is considered not multicast capable. So an ETR in a LISP receiver multicast site can distinguish whether a LISP source multicast site is LISP-Multicast site from a uLISP site.

#### 9.1.5. LISP Source Site to Any Receiver Sites

When a LISP source multicast site has receivers in LISP, non-LISP, and uLISP receiver multicast sites, it has a conflict about how it sends multicast packets. The ITR can either encapsulate or natively forward multicast packets. Since the receiver multicast sites are heterogeneous in their behavior, one packet forwarding mechanism cannot satisfy both. However, if a LISP receiver multicast site acts like a uLISP site then it could receive packets like a non-LISP receiver multicast site making all receiver multicast sites have homogeneous behavior. However, this poses the following issues:

- o LISP-NAT techniques with routable addresses would be required in all cases.
- o Or alternatively, mPETR deployment would be required forcing coarse EID prefix advertisement in the core.
- o But what is most disturbing is that when all sites that participate are LISP-Multicast sites but then a non-LISP or uLISP site joins the distribution tree, then the existing joined LISP receiver multicast sites would have to change their behavior. This would create too much dynamic tree-building churn to be a

viable alternative.

So the solution space options are:

1. Make the LISP ITR in the source multicast site send two packets, one that is encapsulated with (S-RLOC,G) to reach LISP receiver multicast sites and another that is not encapsulated with (S-EID,G) to reach non-LISP and uLISP receiver multicast sites.
2. Make the LISP ITR always encapsulate packets with (S-RLOC,G) to reach LISP-Multicast sites and to reach mPETRs that can decapsulate and forward (S-EID,G) packets to non-LISP and uLISP receiver multicast sites.

#### 9.2. LISP Sites with Mixed Address Families

A LISP database mapping entry that describes the locator-set, Mpriority and Mweight per locator address (RLOC), for an EID prefix associated with a site could have RLOC addresses in either IPv4 or IPv6 format. When a mapping entry has a mix of RLOC formatted addresses, it is an implicit advertisement by the site that it is a dual-stack site. That is, the site can receive IPv4 or IPv6 unicast packets.

To distinguish if the site can receive dual-stack unicast packets as well as dual-stack multicast packets, the Mpriority value setting will be relative to an IPv4 or IPv6 RLOC See [LISP] for packet format details.

If one considers the combinations of LISP, non-LISP, and uLISP sites sharing the same distribution tree and considering the capabilities of supporting IPv4, IPv6, or dual-stack, the number of total combinations grows beyond comprehension.

Using some combinatorial math, the following profiles of a site and the combinations that can occur:

1. LISP-Multicast IPv4 Site
2. LISP-Multicast IPv6 Site
3. LISP-Multicast Dual-Stack Site
4. uLISP IPv4 Site
5. uLISP IPv6 Site



6. uLISP Dual-Stack Site
7. non-LISP IPv4 Site
8. non-LISP IPv6 Site
9. non-LISP Dual-Stack Site

Lets define  $(m\ n) = m!/(n!(m-n)!)$ , pronounced "m choose n" to illustrate some combinatorial math below.

When 1 site talks to another site, the combinatorial is  $(9\ 2)$ , when 1 site talks to another 2 sites, the combinatorial is  $(9\ 3)$ . If sum this up to  $(9\ 9)$ , then:

$$(9\ 2) + (9\ 3) + (9\ 4) + (9\ 5) + (9\ 6) + (9\ 7) + (9\ 8) + (9\ 9) =$$

$$36 + 84 + 126 + 126 + 84 + 36 + 9 + 1$$

Which results in the total number of cases to be considered at 502.

This combinatorial gets even worse when one considers a site using one address family inside of the site and the xTRs use the other address family (as in using IPv4 EIDs with IPv6 RLOCs or IPv6 EIDs with IPv4 RLOCs).

To rationalize this combinatorial nightmare, there are some guidelines which need to be put in place:

- o Each distribution tree shared between sites will either be an IPv4 distribution tree or an IPv6 distribution tree. Therefore, head-end replication can be avoided by building and sending packets on each address family based distribution tree. Even though there might be an urge to do multicast packet translation from one address family format to the other, it is a non-viable over-complicated urge. Multicast ITRs will only encapsulate packets where the inner and outer headers are from the same address family.
- o All LISP sites on a multicast distribution tree must share a common address family which is determined by the source site's locator-set in its LISP database mapping entry. All receiver multicast sites will use the best RLOC priority controlled by the source multicast site. This is true when the source site is either LISP-Multicast or uLISP capable. This means that priority-based policy modification is prohibited. When a receiver multicast site ETR receives a (S-EID,G) join, it must select a

S-RLOC for the same address family as S-EID.

- o When a multicast locator-set has more than one locator, only locators from the same address-family MUST be set to the same best priority value. A mixed locator-set can exist (for unicast use), but the multicast priorities MUST be the set for the same address family locators.
- o When the source site is not LISP capable, it is up to how receivers find the source and group information for a multicast flow. That mechanism decides the address family for the flow.

### 9.3. Making a Multicast Interworking Decision

This Multicast Interworking section has shown all combinations of multicast connectivity that could occur. As already concluded, this can be quite complicated and if the design is too ambitious, the dynamics of the protocol could cause a lot of instability.

The trade-off decisions are hard to make and so the same single solution is desirable to work for both IPv4 and IPv6 multicast. It is imperative to have an incrementally deployable solution for all of IPv4 unicast and multicast and IPv6 unicast and multicast while minimizing (or eliminating) both unicast and multicast EID namespace state.

Therefore the design decision to go with uPITRs [INTWORK] for unicast routing and mPETRs for multicast routing seems to be the sweet spot in the solution space so state requirements can be optimized and avoid head-end data replication at ITRs.

#### 10. Considerations when RP Addresses are Embedded in Group Addresses

When ASM and PIM-Bidir is used in an IPv6 inter-domain environment, a technique exists to embed the unicast address of an RP in a IPv6 group address [RFC3956]. When routers in end sites process a PIM Join/Prune message which contain an embedded-RP group address, they extract the RP address from the group address and treat it from the EID namespace. However, core routers do not have state for the EID namespace, and need to extract an RP address from the RLOC namespace.

Therefore, it is the responsibility of ETRs in multicast receiver sites to map the group address into a group address where the embedded-RP address is from the RLOC namespace. The mapped RP-address is obtained from a EID-to-RLOC mapping database lookup. The ETR will also send a unicast (\*,G) Join/Prune message to the ITR so the branch of the distribution tree from the source site resident RP to the ITR is created.

This technique is no different than the techniques described in this specification for translating (S,G) state and propagating Join/Prune messages into the core. The only difference is that the (\*,G) state in Join/Prune messages are mapped because they contain unicast addresses encoded in an Embedded-RP group address.

## 11. Taking Advantage of Upgrades in the Core

If the core routers are upgraded to support [RFC5496], then the EID specific data can be passed through the core without, possibly, having to store the state in the core.

By doing this one can eliminate the ETR from unicast encapsulating PIM Join/Prune messages to the source site's ITR.

However, this solution is restricted to a small set of workable cases which would not be good for general use of LISP-Multicast. In addition due to slow convergence properties, it is not being recommended for LISP-Multicast.

## 12. Mtrace Considerations

Mtrace functionality MUST be consistent with unicast traceroute functionality where all hops from multicast receiver to multicast source are visible.

The design for mtrace for use in LISP-Multicast environments is to be determined but should build upon the mtrace version 2 specified in [MTRACE].

### 13. Security Considerations

The security concerns for LISP multicast are mainly the same as for the base LISP specification [LISP] and for multicast in general, including PIM-ASM [RFC4601].

There may be a security concern with respect to unicast PIM messages. When multiple receiver sites are joining a (S-EID1,G) distribution tree that maps to a (RLOC1,G) core distribution tree, and a malicious receiver site joins a (S-EID2,G) distribution tree that also maps to the (RLOC1,G) core distribution tree, the legitimate sites will receive data from S-EID2 when they did not ask for it.

Other than as noted above there are currently no known security differences between multicast with LISP and multicast without LISP. However this has not been a topic that has been investigated deeply so far therefore additional issues might arise in future.

#### 14. Acknowledgments

The authors would like to gratefully acknowledge the people who have contributed discussion, ideas, and commentary to the making of this proposal and specification. People who provided expert review were Scott Brim, Greg Shepherd, and Dave Oran. Other commentary from discussions at Summer 2008 Dublin IETF were Toerless Eckert and Ijsbrand Wijnands.

The authors would also like to thank the MBONED working group for constructive and civil verbal feedback when this draft was presented at the Fall 2008 IETF in Minneapolis. In particular, good commentary came from Tom Pusateri, Steve Casner, Marshall Eubanks, Dimitri Papadimitriou, Ron Bonica, Lenny Guardino, Alia Atlas, Jesus Arango, and Jari Arkko.

An expert review of this specification was done by Yiqun Cai and Liming Wei. The authors thank them for their detailed comments.

This work originated in the Routing Research Group (RRG) of the IRTF. The individual submission [MLISP] was converted into this IETF LISP working group draft.

## 15. IANA Considerations

This document makes no request of the IANA.



## 16. References

### 16.1. Normative References

- [INTWORK] Lewis, D., Meyer, D., and D. Farinacci, "Interworking LISP with IPv4 and IPv6", draft-ietf-lisp-interworking-02.txt (work in progress).
- [LISP] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "Locator/ID Separation Protocol (LISP)", draft-ietf-lisp-16.txt (work in progress).
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3618] Fenner, B. and D. Meyer, "Multicast Source Discovery Protocol (MSDP)", RFC 3618, October 2003.
- [RFC3956] Savola, P. and B. Haberman, "Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address", RFC 3956, November 2004.
- [RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, August 2006.
- [RFC4604] Holbrook, H., Cain, B., and B. Haberman, "Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast", RFC 4604, August 2006.
- [RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, August 2006.
- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.
- [RFC5015] Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", RFC 5015, October 2007.
- [RFC5135] Wing, D. and T. Eckert, "IP Multicast Requirements for a Network Address Translator (NAT) and a Network Address Port Translator (NAPT)", BCP 135, RFC 5135, February 2008.
- [RFC5496] Wijnands, IJ., Boers, A., and E. Rosen, "The Reverse Path Forwarding (RPF) Vector TLV", RFC 5496, March 2009.

## 16.2. Informative References

- [ALT] Farinacci, D., Fuller, V., and D. Meyer, "LISP Alternative Topology (LISP-ALT)", draft-ietf-lisp-alt-09.txt (work in progress).
- [MLISP] Farinacci, D., Meyer, D., Zwiebel, J., and S. Venaas, "LISP for Multicast Environments", draft-farinacci-lisp-multicast-01.txt (work in progress).
- [MTRACE] Asaeda, H., Jinmei, T., Fenner, W., and S. Casner, "Mtrace Version 2: Traceroute Facility for IP Multicast", draft-ietf-mboned-mtrace-v2-08.txt (work in progress).

## Appendix A. Document Change Log

## A.1. Changes to draft-ietf-lisp-multicast-14.txt

- o Posted February 2012.
- o Resolve Adrian Farrel's final DISCUSS comment.

## A.2. Changes to draft-ietf-lisp-multicast-13.txt

- o Posted February 2012.
- o Resolution to Stewart Bryant's and Adrian Farrel's comments.

## A.3. Changes to draft-ietf-lisp-multicast-12.txt

- o Posted January 2012.
- o Added more security disclaimers to the Security Considerations section.

## A.4. Changes to draft-ietf-lisp-multicast-11.txt

- o Posted November 2011.
- o Added Stig text to Security Considerations section to reflect comments from IESG review comment from Stephen Farrell.
- o Changed how an unicast PIM join gets sent. Do not use an ECM or else an instance-ID cannot be included in the join. So go back to what we had where the unicast PIM join is encapsulated in a 4341 UDP packet.

## A.5. Changes to draft-ietf-lisp-multicast-10.txt

- o Posted second half of October 2011. Changes to reflect IESG review comments from Stephen Farrell.

## A.6. Changes to draft-ietf-lisp-multicast-09.txt

- o Posted October 2011. Changes to reflect IESG review comments from Ralph Droms and Kathleen Moriarty.

## A.7. Changes to draft-ietf-lisp-multicast-08.txt

- o Posted September 2011. Minor editorial changes from Jari's commentary.

## A.8. Changes to draft-ietf-lisp-multicast-07.txt

- o Posted July 2011. Fixing IDnits errors.

## A.9. Changes to draft-ietf-lisp-multicast-06.txt

- o Posted June 2011 to complete working group last call.
- o Added paragraph to section 8.1.2 based on Jesus comment about making it more clear what happens when two (S-EID,G) trees use the same (RLOC,G) tree.
- o Make more references to [INTWORK] when mentioning uPITRs and uPETRs.
- o Made many changes based on editorial and wordsmithing comments from Alia.

## A.10. Changes to draft-ietf-lisp-multicast-05.txt

- o Posted April 2011 to reset expiration timer.
- o Updated references.

## A.11. Changes to draft-ietf-lisp-multicast-04.txt

- o Posted October 2010 to reset expiration timer.
- o Updated references.

## A.12. Changes to draft-ietf-lisp-multicast-03.txt

- o Posted April 2010.
- o Added section 8.1.2 to address Joel Halpern's comment about receiver sites joining the same source site via 2 different RLOCs, each being a separate ITR.
- o Change all occurrences of "mPTR" to "mPETR" to become more consistent with uPITRs and uPETRs described in [INTWORK]. That is, an mPETR is a LISP multicast router that decapsulates multicast packets that are encapsulated to it by ITRs in multicast source sites.
- o Add clarifications in section 9 about how homogeneous multicast encapsulation should occur. As well as describing in this section, how to deal with mixed-locator sets to avoid heterogeneous encapsulation.

- o Introduce concept of mPITRs to help reduce (S-EID,G) to the edges of LISP global multicast network.

A.13. Changes to draft-ietf-lisp-multicast-02.txt

- o Posted September 2009.
- o Added Document Change Log appendix.
- o Specify that the LISP Encapsulated Control Message be used for unicasting PIM Join/Prune messages from ETRs to ITRs.

A.14. Changes to draft-ietf-lisp-multicast-01.txt

- o Posted November 2008.
- o Specified that PIM Join/Prune unicast messages that get sent from ETRs to ITRs of a source multicast site get LISP encapsulated in destination UDP port 4342.
- o Add multiple RLOCs per ITR per Yiqun's comments.
- o Indicate how static RPs can be used when LISP is run using Bidir-PIM in the core.
- o Editorial changes per Liming comments.
- o Add Mtrace Considerations section.

A.15. Changes to draft-ietf-lisp-multicast-00.txt

- o Posted April 2008.
- o Renamed from draft-farinacci-lisp-multicast-01.txt.

Authors' Addresses

Dino Farinacci  
cisco Systems  
Tasman Drive  
San Jose, CA  
USA

Email: dino@cisco.com

Dave Meyer  
cisco Systems  
Tasman Drive  
San Jose, CA  
USA

Email: dmm@cisco.com

John Zwiebel  
cisco Systems  
Tasman Drive  
San Jose, CA  
USA

Email: jzwiebel@cisco.com

Stig Venaas  
cisco Systems  
Tasman Drive  
San Jose, CA  
USA

Email: stig@cisco.com

