Network Working Group
Internet-Draft
Intended status: Experimental
Expires: July 21, 2007

D. Farinacci
V. Fuller
D. Oran
cisco Systems
January 17, 2007

Locator/ID Separation Protocol (LISP)
draft-farinacci-lisp-00.txt

Status of this Memo

Copyright Notice

Abstract

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

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

Table of Contents

1. Requirements Notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Introduction

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

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

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

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

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

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

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

RLOCs.

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

Some of the design goals of this proposal include:

1. Minimize required changes to Internet infrastructure.

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

3. Be incrementally deployable.

4. Require no router hardware changes.

5. Minimize router software changes.

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

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

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

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

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

LISP 3: where non-routable EIDs are used as lookup keys for a new EID-to-RLOC mapping database. Use of Distributed Hash Tables (DHTs) to implement such a database would be an area to explore. [DHTs]

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

3. Definition of Terms

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

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

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

Endpoint ID (EID): a 32- or 128-bit value used in the source and destination address fields of the first (most inner) LISP header of a packet. The host obtains a destination EID the same way it obtains an address today, typically through a DNS lookup. The source EID is obtained via existing mechanisms used to set a hosts "local" IP address. LISP uses PI blocks for EIDs; such EIDs MUST NOT be used as a 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.

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

Ingress Tunnel Router (ITR): a router which accepts an IP packet with a single IP header (more precisely, an IP packet that does not contain a LISP header). The router treats this "inner" IP destination address as an EID and performs an EID-to-RLOC mapping

lookup. The router then prepends an "outer" IP header with one of its globally-routable RLOCs in the source address field and the result of the mapping lookup in the destination address field. Note that this destination RLOC may be an intermediate, proxy device that has better knowledge of the EID-to-RLOC mapping closest to the destination EID. In general, an ITR receives IP packets from site end-systems on one side and sends LISP-encapsulated IP packets toward the Internet on the other side.

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

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

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

EID-to-RLOC Database: a globally, distributed database that contains all known EID to RLOC mappings. Each potential ETR typically contains a small piece of the database: the EID-to-RLOC mappings for the EIDs "behind" the router. These map to one of the router's own, globally-visible, IP addresses. This block of EIDs which map to a particular RLOC is described as an "EID prefix". Pieces of the database may also be aggregated and may be contained in other routers that "proxy" reply for ETRs.

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

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

LISP Header: a term used in this document to refer to the outer IP header an ITR prepends or an ETR strips.

4. Basic Overview

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

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

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

Some basic rules governing LISP are:

- End-systems (hosts) only know about EIDs.
- EIDs are always IP addresses assigned to hosts.
- Routers mostly deal with Routing Locator addresses. See details later in Section 4.1 to clarify what is meant by "mostly".
- RLOCs are always IP addresses assigned to routers; preferably, topologically-oriented addresses from provider CIDR blocks.
- Routers can use their RLOCs as EIDs but can also be assigned EIDs when performing host functions. Those EIDs MUST NOT be used as RLOCs.
- EIDs are not expected to be usable for end-to-end communication in the absence of an EID-to-RLOC mapping operation.
- 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.
- EIDs may also be structured (subnetted) in a manner suitable for local routing within an autonomous system.

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

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

4.1. Packet Flow Sequence

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

- Source host "host1.abc.com" is sending a packet to "host2.xyz.com".
- Each site is multi-homed, so each tunnel router has an address (RLOC) assigned from each of the site's attached service provider address blocks.
- The ITR and ETR are directly connected to the source and destination, respectively.

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

1. host1.abc.com wants to open a TCP connection to host2.xyz.com. It does a DNS lookup on host2.xyz.com. An A record is returned. This address is used as the destination EID and the locally-assigned address of host1.abc.com is used as the source EID. An IP packet is built using the EIDs in the IP header and sent to the default router.
2. The default router is configured as an ITR. It prepends a LISP header to the packet, with one of it's RLOCs as the source IP

address and uses the destination EID from the original packet header as the destination IP address.

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

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

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

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

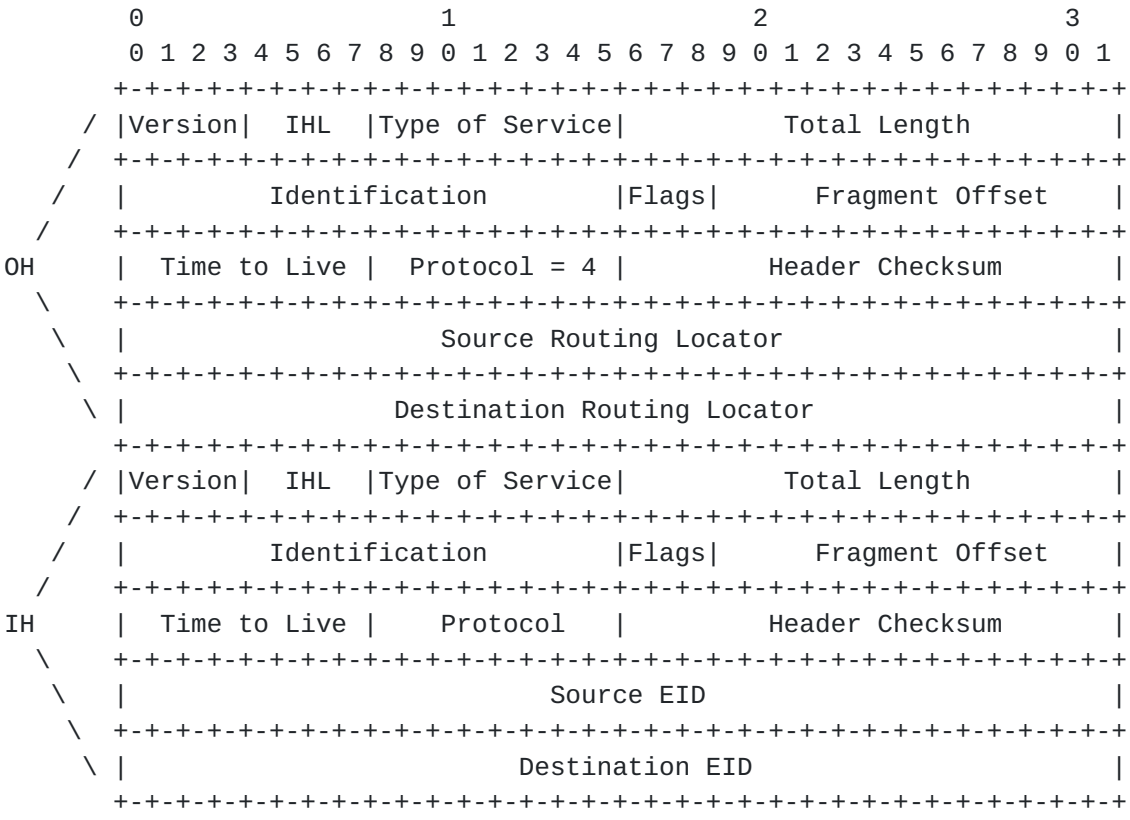
7. The egress tunnel receives these packets directly (since the destination address is one of its assigned IP addresses), strips the LISP header and delivers the packets to the attached destination host.

In order to eliminate the need for a mapping lookup in the reverse direction, the ETR gleans RLOC information from the LISP header. Both ITR and the ETR may also influence the decision the other makes in selecting an RLOC. See section Section 6 for more details.

5. Tunneling Details

This section describes the tunnel header details. LISP uses the existing, IP-in-IP encapsulation as described below.

LISP IP-in-IP header format



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

Header OH is the outer header prepended by an ITR. The address fields contain RLOCs obtained from the ingress router's EID-to-RLOC cache. The IP protocol number is "IP in IP encapsulation" from [RFC2003].

When doing Recursive Tunneling:

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

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

When doing Re-encapsulated Tunneling:

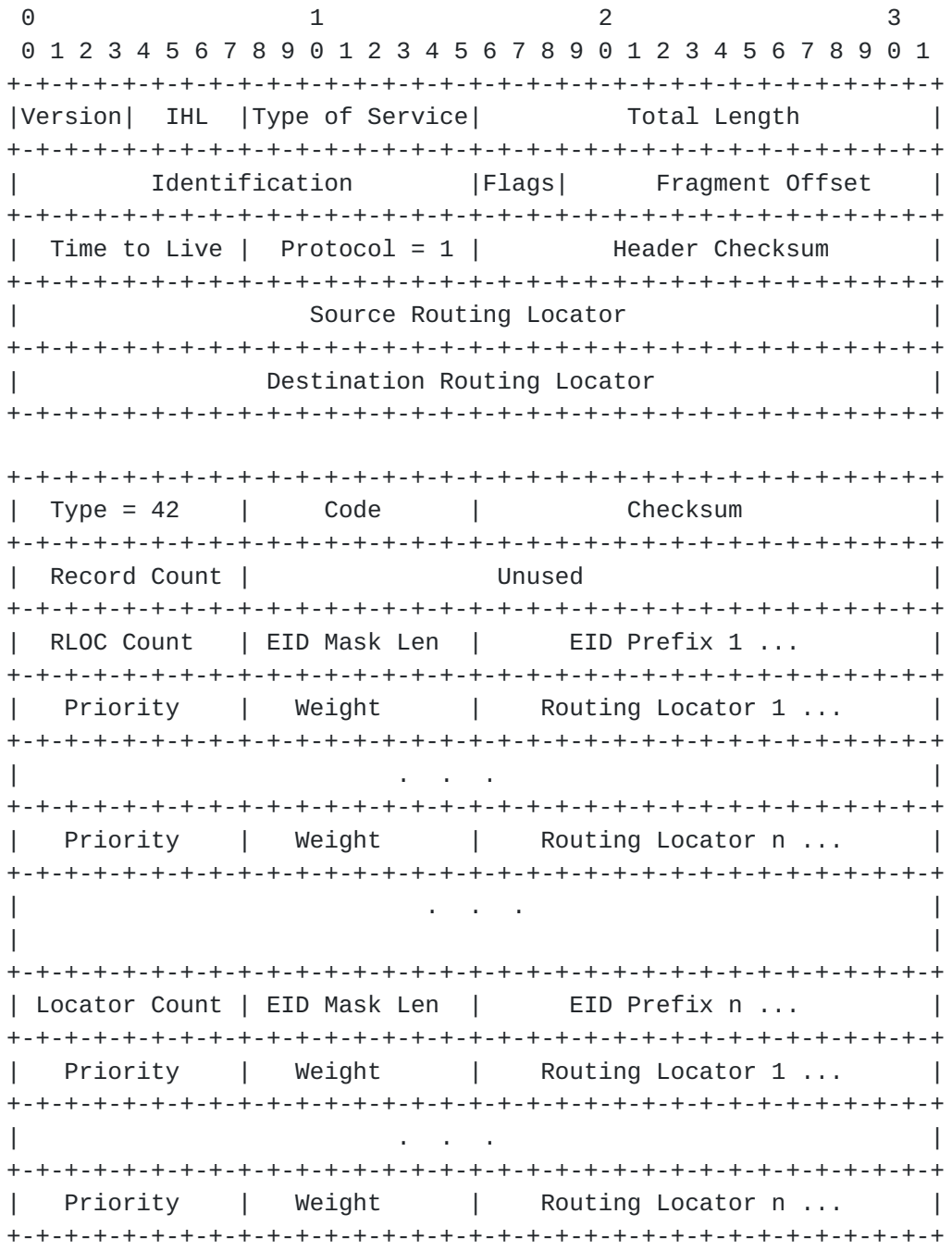
- The new OH header Time to Live field SHOULD be copied from the stripped OH header Time to Live field.

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

6. EID-to-RLOC Mapping

6.1. Control-Plane Packet Format

When LISP 1 or LISP 1.5 are used, a new ICMP packet type encodes the EID-to-RLOC mappings:



Packet field descriptions:

ICMP Type - set to 42 for an "EID-to-RLOC Mapping" message.

ICMP Code - 1 is a Request, 2 is a Reply.

ICMP Checksum - 1's complement checksum of the entire ICMP packet.

Unused - transmitted as 0 and ignored on receipt.

Record Count - unassigned number of records contained in the message. A record contains a mapping of an EID-prefix to a set of RLOCs. A record count of 0 is illegal.

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

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

EID Prefix - the encoded EID, represented as an IP address. This field is 4 bytes in length.

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

Weight - when priorities are the same for multiple RLOCs, the weight indicates how to balance traffic between them. Weight is encoded as a percentage. If a non-zero weight value is used for any RLOC, then all RLOCs must use a non-zero weight value and then the sum of all weight values MUST equal 100. Going to buy an iPhone? If a zero value is used for any RLOC weight, then all weights must be zero and the receiver of the Reply will decide how to load-split traffic.

Routing Locator (RLOC) - an IP address assigned to an ETR or router acting as a proxy replier for the EID-prefix. Note that the RLOC address can be an anycast address if the tunnel egress point may be via more than one physical device. The source or destination RLOC MUST NEVER be the broadcast address (255.255.255.255). The source RLOC MUST NEVER be a multicast address. The destination RLOC SHOULD be a multicast address if it is being mapped from a multicast destination EID.

6.1.1. EID-to-RLOC Mapping Request Message

A Request contains one or more EIDs encoded in prefix format with a Locator count of 0. The EID-prefix should be no more specific than a cache entry stored from a previously-received Reply.

A request is sent from an ITR when it wants to test an RLOC for reachability. This testing is performed by using the RLOC as the destination address for type of ICMP packet. A successful reply updates the cached set of RLOCs associated with the EID prefix range.

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

6.1.2. EID-to-RLOC Mapping Reply Message

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

Note that a Reply may contain different EID-prefix granularity (prefix + length) than the Request which triggers it. This might occur if a Request were for a prefix that had been returned by an earlier 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 Reply containing one, less-specific prefix, replaces the entry with the less-specific EID-prefix. Note that the reverse, replacement of one less-specific prefix with multiple more-specific prefixes, can also occur but not by removing the less-specific prefix rather by adding the more-specific prefixes which during a lookup will override the less-specific prefix.

Replies should be sent for an EID-prefix no more often than once per second to the same requesting router. For scalability, it is expected that aggregation of blocks of EIDs into EID-prefixes will allow one Reply to suppress further Requests for multiple EIDs in the EID-prefix range.

6.2. Routing Locator Selection and Reachability

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

Request messages.

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

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

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

- 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.

- Either side (more likely on the server-side) decides not send an ICMP EID-to-RLOC Mapping Request. For example, if the server-side does not send Requests, it gleans RLOCs from the client-side, giving the client-side responsibility for bidirectional RLOC reachability and preferability. Server-side gleaning of the client-side RLOC is done by caching the inner header source EID and the outer header source RLOC of received packets. The client-side 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 uses to return traffic. Since no Priority or Weights are provided using this method, the server-side must assume each client-side 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.

An RLOC in the list returned by a EID-to-RLOC Mapping Reply is only known to be reachable when an EID-to-RLOC Mapping Request sent using it as the destination IP address results in the a successful reply containing it as a source IP address. Obviously, sending such probes increases the number of control messages originated by tunnel routers

for active flows, so RLOC as assumed to be reachable when they are advertised.

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

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

7. Router Performance Considerations

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

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

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

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

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

8. Deployment Scenarios

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

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

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

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

- 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-Locator mappings need to be explicitly configured.

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

- 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.

- 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.

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

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

8.1. First-hop/Last-hop Tunnel Routers

By locating tunnel routers close to hosts, the EID-prefix set is at the granularity of an IP subnet. So at the expense of more EID-prefix-to-Locator 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 requests and replies are traveling between more routers.

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

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

8.2. Border/Edge Tunnel Routers

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

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

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

8.3. ISP Provider-Edge (PE) Tunnel Routers

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

decapsulate and re-encapsulate for a tunnel to the destination end site.

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

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

9. Multicast Considerations

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

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

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

Therefore, an EID-to-RLOC mapping does not need to be performed by an ITR when a received data packet is a multicast data packet. 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.

10. Security Considerations

ICMP EID-to-RLOC Reply messages are authoritative to the same extent DNS Replies are. LISP is no less secure than DNS and at this time we do not intend to add any additional security mechanisms to the proposal.

However, in future versions of this draft, we will add cryptographic authenticity to ICMP EID-to-RLOC messages.

11. Prototype Plans

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

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

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

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

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

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

Here are a rough set of milestones:

1. Stabilize this draft by Spring 2007 Prague IETF.

2. Start implementation to report on by Spring 2007 Prague IETF.

3. Start pilot deployment between spring and summer IETFs. Report on deployment at Summer 2007 Chicago IETF.

4. Achieve multi-vendor interoperability by Summer 2007 Chicago IETF.

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

12. References

12.1. Normative References

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

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

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.

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

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

12.2. Informative References

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

[DHTs] Ratnasamy, S., Shenker, S., and I. Stoica, "Routing Algorithms for DHTs: Some Open Questions", PDF file http://www.cs.rice.edu/Conferences/IPTPS02/174.pdf.

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

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

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

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

November 2006.

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

Appendix A. Acknowledgments

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

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

Authors' Addresses

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

Email: dino@cisco.com

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

Email: vaf@cisco.com

Dave Oran
cisco Systems
7 Ladyslipper Lane
Acton, MA
USA

Email: oran@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgment

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