Network Working Group
Internet-Draft
Intended status: Experimental
Expires: February 2, 2018
V. Moreno
Cisco Systems
D. Farinacci
lispers.net
August 1, 2017

Signal-Free LISP Multicast
draft-ietf-lisp-signal-free-multicast-06

Abstract

When multicast sources and receivers are active at LISP sites, the core network is required to use native multicast so packets can be delivered from sources to group members. When multicast is not available to connect the multicast sites together, a signal-free mechanism can be used to allow traffic to flow between sites. The mechanism within here uses unicast replication and encapsulation over the core network for the data-plane and uses the LISP mapping database system so encapsulators at the source LISP multicast site can find decapsulators at the receiver LISP multicast sites.

Requirements Language

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].

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 February 2, 2018.

Copyright Notice

Copyright (c) 2017 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

When multicast sources and receivers are active at LISP sites, and the core network between the sites does not provide multicast support, a signal-free mechanism can be used to create an overlay that will allow multicast traffic to flow between sites and connect the multicast trees at the different sites.

The signal-free mechanism here proposed does not extend PIM over the overlay as proposed in [RFC6831], nor does the mechanism utilize direct signaling between the Receiver-ETRs and Sender-ITRs as described in [I-D.farinacci-lisp-mr-signaling]. The signal-free mechanism proposed reduces the amount of signaling required between sites to a minimum and is centered around the registration of Receiver-sites for a particular multicast-group or multicast-channel with the LISP Mapping System.

Registrations from the different receiver-sites will be merged at the Mapping System to assemble a multicast-replication-list inclusive of all RLOCs that lead to receivers for a particular multicast-group or multicast-channel. The replication-list for each specific multicast-entry is maintained as a LISP database mapping entry in the Mapping Database.

When the ITR at the source-site receives multicast traffic from sources at its site, the ITR can query the mapping system by issuing Map-Request messages for the (S,G) source and destination addresses in the packets received. The Mapping System will return the RLOC replication-list to the ITR, which the ITR will cache as per standard LISP procedure. Since the core is assumed to not support multicast, the ITR will replicate the multicast traffic for each RLOC on the replication-list and will unicast encapsulate the traffic to each RLOC. The combined function or replicating and encapsulating the traffic to the RLOCs in the replication-list is referred to as "rep-encapsulation" in this document.

The document describes the General Procedures and information encoding that are required at the Receiver-sites and Source-sites to achieve signal-free multicast interconnectivity. The General Procedures for Mapping System Notifications to different sites are also described. A section dedicated to the specific case of SSM trees discusses the implications to the General Procedures for SSM multicast trees over different topological scenarios. A section on ASM support is included to identify the constraints that come along with supporting it using LISP Signal-Free multicast.

There is a section dedicated to Replication Engineering. A mechanism to reduce the impact of head-end replication. The mapping system, via LISP Signal-Free mechanisms, can be used to build a layer of RTRs.

2. Definition of Terms

LISP related terms, notably Map-Request, Map-Reply, Ingress Tunnel Router (ITR), Egress Tunnel Router (ETR), Map-Server (MS) and Map-Resolver (MR) are defined in the LISP specification [RFC6830].

Extensions to the definitions in [RFC6830] for their application to multicast routing are documented in [RFC6831].

Terms defining interactions with the LISP Mapping System are defined in [RFC6833].

The following terms are consistent with the definitions in [RFC6830] and [RFC6831]. The terms are specific cases of the general terms and are here defined to facilitate the descriptions and discussions within this particular document.

Source: Multicast source end-point. Host originating multicast packets.

Receiver: Multicast group member end-point. Host joins multicast group as a receiver of multicast packets sent to the group.

Receiver-site: LISP site where multicast receivers are located.

Source-site: LISP site where multicast sources are located.

RP-site: LISP site where an ASM PIM Rendezvous Point is located. The RP-site and the Source-site may be the same in some situations.

Receiver-ETR: LISP xTR at the Receiver-site. This is a multicast ETR.

Source-ITR: LISP xTR at the Source-site. This is a multicast ITR.

RP-xTR: LISP xTR at the RP-site. This is typically a multicast ITR.

Replication-list: Mapping-entry containing the list of RLOCs that have registered Receivers for a particular multicast-entry.

Multicast-entry: A tuple identifying a multicast tree. Multicast-entries are in the form of (S-prefix, G-prefix).

Rep-encapsulation: The process of replicating and then encapsulating traffic to multiple RLOCs.

Re-encapsulating Tunnel Router (RTR): An RTR is a router that implements the re-encapsulating tunnel function detailed in Section 8 of the main LISP specification [RFC6830]. A LISP RTR performs packet re-routing by chaining ETR and ITR functions, whereby it first removes the LISP header of an ingress packet and then prepends a new LISP header to an egress packet.

RTR Level: An RTR level is encoded in a Replication-List-Entry (RLE) LCAF Type detailed in [RFC8060]. Each entry in the replication list contains an address of an xTR and a level value. Level values are used to create a replication hierarchy so that ITRs at source LISP sites replicate to the lowest (smaller value) level number RTRs in a RLE entry. And then RTRs at a given level replicate to the next higher level of RTRs. The number of RTRs at each level are engineered to control the fan-out or replication factor so a tradeoff between the width of the level versus the number of levels can be selected.

3. Reference Model

The reference model that will be used for the discussion of the Signal-Free multicast tree interconnection is illustrated in Figure 1.

```
                                 MS/MR
                                 +---+
                                 |   |
            +---+     +---+      +---+      +---+      +---+
  Src-1 ----| R1|-----|ITR|        |        |ETR|------| R2|------ Rcv-2
            +---+     +---+        |        +---+      +---+
                           \       |       /
            Source-site-1   \      |      /    Receiver-site-2
                             \     |     /
                              \    |    /
                               \   |   /
                                 Core
                               /       \
                              /         \
                             /           \
                            /             \
                           /               \
                      +---+                 +---+
  Src-3 --------------|ITR|                 |ETR|----------------- Rcv-4
                      +---+                 +---+

             Source-site-3                    Receiver-site-4


             Figure 1: LISP Multicast Generic Reference Model
```

Sites 1 and 3 are Source-sites.

Source-site-3 presents a Source (Src-3) that is directly connected to the Source-ITR

Source-site-1 presents a Source (Src-1) that is one hop or more away from the Source-ITR

Receiver-site-2 and 4 are receiver sites with not-directly connected and directly connected Receiver end-points respectively

R1 is a router in Source-site-1.

R2 is a PIM router at the Receiver-site.

The Map-Servers and Resolvers are reachable in the RLOC space in the Core, only one is shown for illustration purposes, but these can be many or even part of a DDT tree.

The procedures for interconnecting multicast Trees over an overlay can be broken down into three functional areas:

- Receiver-site procedures

- Source-site procedures

- LISP notification procedures

The receiver site procedures will be common for most tree types and topologies.

The procedures at the source site can vary depending on the type of trees being interconnected as well as based on the topological relation between sources and source-site xTRs. For ASM trees, a special case of the Source-site is the RP-site for which a variation of the Source-site procedures may be necessary if ASM trees are to be supported in future specifications of LISP Signal-Free multicast.

The LISP notification procedures between sites are normalized for the different possible scenarios. Certain scenarios may benefit from a simplified notification mechanism or no notification requirement at all.

4. General Procedures

The interconnection of multicast trees across different LISP sites involves the following procedures to build the necessary multicast distribution trees across sites.

1. The presence of multicast Receiver end-points is detected by the Receiver-ETRs at the Receiver-sites.

2. Receiver-ETRs register their RLOCs as part of the replication-list for the multicast-entry the detected Receivers subscribe to.

3. The Mapping-system merges all receiver-ETR or delivery-group RLOCs to build a comprehensive replication-list inclusive of all Receiver-sites for each multicast-entry.

4. LISP Map-Notify messages should be sent to the Source-ITR informing of any changes in the replication-list.

5. Multicast-tree building at the Source-site is initiated when the Source-ITR receives the LISP Notification.

Once the multicast distribution trees are built, the following forwarding procedures may take place:

1. The Source sends multicast packets to the multicast group destination address.

2. Multicast traffic follows the multicast tree built at the Source-site and makes its way to the Source-ITRs.

3. The Source-ITR will issue a map-request to resolve the replication-list for the multicast-entry.

4. The Mapping System responds to the Source-ITR with a map-reply containing the replication-list for the multicast group requested.

5. The Source-ITR caches the replication-list received in the map-reply for the multicast-entry.

6. Multicast traffic is rep-encapsulated. That is, the packet is replicated for each RLOC in the replication-list and then encapsulated to each one.

4.1. General Receiver-Site Procedures

4.1.1. Multicast Receiver Detection

When the Receiver-ETRs are directly connected to the Receivers (e.g. Receiver-site-4 in Figure 1), the Receiver-ETRs will receive IGMP Reports from the Receivers indicating which group the Receivers wish to subscribe to. Based on these IGMP Reports, the receiver-ETR is made aware of the presence of Receivers as well as which group they are interested in.

When the Receiver-ETRs are several hops away from the Receivers (e.g. Receiver-site-2 in Figure 1), the Receiver-ETRs will receive PIM join messages which will allow the Receiver-ETR to know that there are multicast Receivers at the site and also learn which multicast group the Receivers are for.

4.1.2. Receiver-Site Registration

Once the Receiver-ETRs detect the presence of Receivers at the Receiver-site, the Receiver-ETRs will issue Map-Register messages to include the Receiver-ETR RLOCs in the replication-list for the multicast-entry the Receivers joined.

The Map-Register message will use the multicast-entry (Source, Group) tuple as its EID record type with the Receiver-ETR RLOCs conforming the locator set.

The EID in the Map-Register message must be encoded using the Multicast Information LCAF type defined in [RFC8060].

The RLOC in the Map-Register message must be encoded using the Replication List Entry (RLE) LCAF type defined in [RFC8060] with the Level Value fields for all entries set to 128 (decimal).

The encoding described above must be used consistently for Map-Register messages, entries in the Mapping Database, Map-reply messages as well as the map-cache at the Source-ITRs.

The Map-Register messages [RFC6830] sent by the receiver-ETRs should have the following bits set as here specified:

1. merge-request-bit set to 1. The Map-Register messages must be sent with "Merge Semantics". The Map-Server will receive registrations from a multitude of Receiver-ETRs. The Map-Server will merge the registrations for common EIDs and maintain a consolidated replication-list for each multicast-entry.

2. want-map-notify-bit (M) set to 0. This tells the Mapping System that the receiver-ETR does not expect to receive Map-Notify messages as it does not need to be notified of all changes to the replication-list.

3. proxy-reply-bit (P) set to 1. The merged replication-list is kept in the Map-Servers. By setting the proxy-reply bit, the receiver-ETRs instruct the Mapping-system to proxy reply to map-requests issued for the multicast entries.

Map-Register messages for a particular multicast-entry should be sent for every receiver detected, even if previous receivers have been detected for the particular multicast-entry. This allows the replication-list to remain up to date.

Receiver-ETRs must be configured to know what Map-Servers Map-Register messages are sent to. The configuration is likely to be associated with an S-prefix that multiple (S,G) entries match to and are more specific for. Therefore, the S-prefix determines the Map-Server set in the least number of configuration statements.

4.1.3. Consolidation of the Replication-List

The Map-Server will receive registrations from a multitude of Receiver-ETRs. The Map-Server will merge the registrations for common EIDs and consolidate a replication-list for each multicast-entry.

When an ETR sends an RLE RLOC-record in a Map-Register and the RLE entry already exists in the Map-Server's RLE merged list, the Map-Server will replace the single RLE entry with the information from

the Map-Register RLOC-record. The Map-Server MUST never merge duplicate RLOCs in the consolidated replication-list.

4.2. General Source-Site Procedures

Source-ITRs must register the unicast EIDs of any Sources or Rendezvous Points that may be present on the Source-site. In other words, it is assumed that the Sources and RPs are LISP EIDs.

The registration of the unicast EIDs for the Sources or Rendezvous Points allows the map-server to know where to send Map-Notify messages to. Therefore, the Source-ITR must register the unicast S-prefix EID with the want-map-notify-bit set in order to receive Map-Notify messages whenever there is a change in the replication-list.

4.2.1. Multicast Tree Building at the Source-Site

When the source site receives the Map-Notify messages from the mapping system as described in Section 4.3, it will initiate the process of building a multicast distribution tree that will allow the multicast packets from the Source to reach the Source-ITR.

The Source-ITR will issue a PIM join for the multicast-entry for which it received the Map-Notify message. The join will be issued in the direction of the source or in the direction of the RP for the SSM and ASM cases respectively.

4.2.2. Multicast Destination Resolution

On reception of multicast packets, the source-ITR must obtain the replication-list for the (S,G) addresses in the packets.

In order to obtain the replication-list, the Source-ITR must issue a Map-Request message in which the EID is the (S,G) multicast tuple which is encoded using the Multicast Info LCAF type defined in [RFC8060].

The Mapping System (most likely the Map-Server) will Map-reply with the merged replication-list maintained in the Mapping System. The Map-reply message must follow the format defined in [RFC6830], its EID must be encoded using the Multicast Info LCAF type and the corresponding RLOC-records must be encoded using the RLE LCAF type. Both LCAF types defined in [RFC8060].

4.3. General LISP Notification Procedures

The Map-Server will issue LISP Map-Notify messages to inform the Source-site of the presence of receivers for a particular multicast group over the overlay.

Updated Map-Notify messages should be issued every time a new registration is received from a Receiver-site. This guarantees that the source-sites are aware of any potential changes in the multicast-distribution-list membership.

The Map-Notify messages carry (S,G) multicast EIDs encoded using the Multicast Info LCAF type defined in [RFC8060].

Map-Notify messages will be sent by the Map-Server to the RLOCs with which the unicast S-prefix EID was registered. In the case when sources are discovered dynamically [I-D.ietf-lisp-eid-mobility], xTRs must register sources explicitly with the want-map-notify-bit set. This is so the ITR in the site the source has moved to can get the most current replication list.

When both the Receiver-sites and the Source-sites register to the same Map-Server, the Map-Server has all the necessary information to send the Map-Notify messages to the Source-site.

When the Map-Servers are distributed in a DDT, the Receiver-sites may register to one Map-Server while the Source-site registers to a different Map-Server. In this scenario, the Map-Server for the receiver sites must resolve the unicast S-prefix EID in the DDT per standard LISP lookup procedures and obtain the necessary information to send the Map-Notify messages to the Source-site. The Map-Notify messages must be sent with an authentication length of 0 as they would not be authenticated.

When the Map-Servers are distributed in a DDT, different Receiver-sites may register to different Map-Servers. This is an unsupported scenario with the currently defined mechanisms.

5. Source Specific Multicast Trees

The interconnection of Source Specific Multicast (SSM) Trees across sites will follow the General Receiver-site Procedures described in Section 4.1 on the Receiver-sites.

The Source-site Procedures will vary depending on the topological location of the Source within the Source-site as described in Section 5.1 and Section 5.2 .

5.1. Source Directly Connected to Source-ITRs

When the Source is directly connected to the source-ITR, it is not necessary to trigger signaling to build a local multicast tree at the Source-site. Therefore Map-Notify messages may not be required to initiate building of the multicast tree at the Source-site.

Map-Notify messages are still required to ensure that any changes to the replication-list are communicated to the Source-site so that the map-cache at the Source-ITRs is kept updated.

5.2. Source not Directly Connected to Source-ITRs

The General LISP Notification Procedures described in Section 4.3 must be followed when the Source is not directly connected to the source-ITR. On reception of Map-Notify messages, local multicast signaling must be initiated at the Source-site per the General Source Site Procedures for Multicast Tree building described in Section 4.2.1.

In the SSM case, the IP address of the Source is known and it is also registered with the LISP mapping system. Thus, the mapping system may resolve the mapping for the Source address in order to send Map-Notify messages to the correct source-ITR.

6. Multi-Homing Considerations

6.1. Multiple ITRs at a Source-Site

When multiple ITRs exist at a source multicast site, care should be taken that more than one ITR does not head-end replicate packets else receiver multicast sites will receive duplicate packets. The following procedures will be used for each topology scenarios:

- o When more than one ITR is directly connected to the source host, either the PIM DR or the IGMP querier (when PIM is not enabled on the ITRs) is responsible for packet replication. All other ITRs silently drop the packet. In the IGMP querier case, it is required to configure the source LAN to have one of the ITRs be the IGMP querier.

- o When more than one ITR is multiple hops away from the source host and one of the ITRs is the PIM Rendezvous Point, then the PIM RP is responsible for packet replication.

- o When more than one ITR is multiple hops away from the source host and the PIM Rendezvous Point is not one of the ITRs, then one of the ITRs must join to the RP. When a Map-Notify is received from

the Map-Server by an ITR, only the highest RLOC addressed ITR will join toward the PIM RP or toward the source.

6.2. Multiple ETRs at a Receiver-Site

When multiple ETRs exist in a receiver multicast site, and each create multicast join state, they each Map-Register their RLOC addresses to the mapping system. In this scenario, the replication happens on the overlay causing multiple ETR entry points to replicate to all receivers versus a single ETR entry point replicating to all receivers. If an ETR does not create join state, because it has not received PIM joins or IGMP reports, it will not Map-Register its RLOC addresses to the mapping system. The same procedures in Section 4.1 should be followed.

When multiple ETRs exist on the same LAN as a receiver host, then the PIM DR, when PIM is enabled, or the IGMP querier is responsible for sending a Map-Register for its RLOC. In the IGMP case, it is required that the LAN is configured with one of the ETRs as IGMP querier.

6.3. Multiple RLOCs for an ETR at a Receiver-Site

It may be desirable to have multiple underlay paths to an ETR for multicast packet delivery. This can be done by having multiple RLOCs assigned to an ETR and having the ETR send Map-Registers for all its RLOCs. By doing this, an ITR can choose a specific path based on underlay performance and/or RLOC reachability.

It is suggested that an ETR sends a Map-Register with a single RLOC-record that uses the ELP LCAF type [RFC8060] that is nested inside RLE entry LCAF. For example say ETR1 has assigned RLOC1 and RLOC2 for a LISP receiver site. And there is ETR2 in another LISP receiver site, that has RLOC3. The two receiver sites have the same (S,G) being joined. Here is how the RLOC-record is encoded on each ETR:

```
ETR1: EID-record: (S,G)
      RLOC-record: RLE[ ELP{ (RLOC1,s,p), (RLOC2,s,p) } ]

ETR2: EID-record: (S,G)
      RLOC-record: RLE[ RLOC3 ]
```

And here is how the entry is merged and stored on the Map-Server since the Map-Registers have an RLE encoded RLOC-record:

```
MS: EID-record: (S,G)
    RLOC-record: RLE[ RLOC3, ELP{ (RLOC1,s,p), (RLOC2,s,p) } ]
```

When the ITR receives a packet from a multicast source S for group G, it uses the merged RLOC-record, returned from the Map-Server. The ITR replicates the packet to (RLOC3 and RLOC1) or (RLOC3 and RLOC2). Since it is required for the s-bit to be set for RLOC1, the ITR must replicate to RLOC1 if it is reachable. When the required p-bit is also set, the RLOC-reachability mechanisms from [RFC6830] are followed. If the ITR determines that RLOC1 is unreachable, it uses RLOC2, as long as RLOC2 is reachable.

6.4. Multicast RLOCs for an ETR at a Receiver-Site

This specification is focused on underlays without multicast support, but does not preclude the use of multicast RLOCs in RLE entries. ETRs MAY register multicast EID entries using multicast RLOCs. In such cases the ETRs will get joined to underlay multicast distribution trees by using IGMP as a multicast host using mechanisms in [RFC2236] and [RFC3376].

7. PIM Any Source Multicast Trees

LISP signal-free multicast can support ASM Trees in limited but acceptable topologies. It is suggested for the simplification of building ASM trees across the LISP overlay to have PIM-ASM run independently in each LISP site. What this means, is that a PIM Rendezvous Point (RP) is configured in each LISP site so PIM Register procedures and (*,G) state maintenance is contained within the LISP site.

The following procedure will be used to support ASM in each LISP site:

1. In a Receiver-site, the RP is colocated with the ETR. RPs for different groups can be spread across each ETR, but is not required.

2. When (*,G) state is created in an ETR, the procedures in Section 4.1.2 are followed. In addition, the ETR registers (S-prefix,G), where S-prefix is 0/0 (the respective unicast default route for the address-family) to the mapping system.

3. In a Source-site, the RP is colocated with the ITR. RPs for different groups can be spread across each ITR, but is not required.

4. When a multicast source sends a packet, a PIM Register message is delivered to the ITR and the procedures in Section 4.2 are followed.

5. When the the ITR sends a Map-Request for (S,G) and no Receiver-site has registered for (S,G), the mapping system will return the (0/0,G) entry to the ITR so it has a replication list of all the ETRs that have received (*,G) state.

6. The ITR stores the replication-list in its map-cache for (S,G). It replicates packets to all ETRs in the list.

7. ETRs decapsulate packets and forward based on (*,G) state in their site.

8. When last-hop PIM routers join the newly discovered (S,G), the ETR will store the state and follow the procedures in Section 4.1.2.

8. Signal-Free Multicast for Replication Engineering

The mechanisms in this draft can be applied to the LISP Replication-Engineering [I-D.coras-lisp-re] design. Rather than having the layered LISP-RE RTR hierarchy use signaling mechanisms, the RTRs can register their availability for multicast tree replication via the mapping database system.

As stated in [I-D.coras-lisp-re], the RTR layered hierarchy is used to avoid head-end replication in replicating nodes closest to a multicast source. Rather than have multicast ITRs replicate to each ETR in an RLE entry of a (S,G) mapping database entry, it could replicate to one or more layer-0 RTRs in the LISP-RE hierarchy.

This draft documents how the RTR hierarchy is determined but not what are the optimal layers of RTRs to use. Methods for determining optimal paths or RTR topological closeness are out of scope for his document.

There are two formats an (S,G) mapping database entry could have. One format is a 'complete-format' and the other is a 'filtered-format'. A 'complete-format' entails an (S,G) entry having multiple RLOC records which contain both ETRs that have registered as well as the RTRs at the first level of the LISP-RE hierarchy for the ITR to replicate to. When using 'complete-format', the ITR has the ability to select if it replicates to RTRs or to the registered ETRs at the receiver sites. A 'filtered-format' (S,G) entry is one where the Map-Server returns the RLOC-records that it decides the ITR should use. So replication policy is shifted from the ITRs to the mapping system. The Map-Servers can also decide for a given ITR, if it uses a different set of replication targets per (S,G) entry for which the ITR is replicating for.

The procedure for the LISP-RE RTRs to make themselves available for replication can occur before or after any receivers join an (S,G) entry or any sources send for a particular (S,G) entry. Therefore, newly configured RTR state will be used to create new (S,G) state and inherited into existing (S,G) state. A set of RTRs can register themselves to the mapping system or a third-party can do so on their behalf. When RTR registration occurs, it is done with an (S-prefix, G-prefix) entry so it can advertise its replication services for a wide-range of source/group combinations.

When a Map-Server receives (S,G) registrations from ETRs and (S-prefix, G-prefix) registrations from RTRs, it has the option of merging the RTR RLOC-records for each (S,G) that is more-specific for the (S-prefix, G-prefix) entry or keep them separate. When merging, a Map-Server is ready to return a 'complete-format' Map-Reply. When keeping the entries separate, the Map-Server can decide what to include in a Map-Reply when a Map-Request is received. It can include a combination of RLOC-records from each entry or decide to use one or the other depending on policy configured.

```
                      +---+                 +----+
  Src-1 --------------|ITR|                 |ETR1|----------------- Rcv-1
                      +---+                 +----+
                          \                 /
           Source-site-1   \               /    Receiver-site-1
                            \             /
                             \           /
                  +----+      \         /     +----+
                  |RTR1|       \       /      |RTR2|     Level-0
                  +----+        \     /       +----+
                        \  <^^^^^^^^^^^^^^>  /
                         \ <              > /
                           < Core-Network >
                           <              >
                           <vvvvvvvvvvvvvv>
                           /     /   \    \
                          /     /     \    \
                  +----+ /     /       \    \ +----+
                  |RTR3|      /         \     |RTR4|     Level-1
                  +----+     /           \    +----+
                            /             \
                           /               \
                      +----+                +----+
  Rcv-2 --------------|ETR2|                |ETR3|----------------- Rcv-3
                      +----+                +----+

             Receiver-site-2                   Receiver-site-3
```

Figure 2: LISP-RE Reference Model

Here is a specific example, illustrated in Figure 2, of (S,G) and (S-prefix, G-prefix) mapping database entries when a source S is behind an ITR and there are receiver sites joined to (S,G) via ETR1, ETR2, and ETR3. And there exists a LISP-RE hierarchy of RTR1 and RTR2 at level-0 and RTR3 and RTR4 at level-1:

```
    EID-record: (S,G)
        RLOC-record: RLE: (ETR1, ETR2, ETR3), p1
    EID-record: (S-prefix, G-prefix)
        RLOC-record: RLE: (RTR1(L0), RTR2(L0), RTR3(L1), RTR4(L1)), p1
```

The above entries are in the form of how they were registered and stored in a Map-Server. When a Map-Server uses 'complete-format', a Map-Reply it originates has the mapping record encoded as:

```
EID-record: (S,G)
    RLOC-record: RLE: (RTR1(L0), RTR3(L1)), p1
    RLOC-record: RLE: (ETR1, ETR2, ETR3), p1
```

The above Map-Reply allows the ITR to decide if it replicates to the ETRs or if it should replicate only to level-0 RTR1. This decision is left to the ITR since both RLOC-records have priority 1. If the Map-Server wanted to force the ITR to replicate to RTR1, it would set the ETRs RLOC-record to priority greater than 1.

When a Map_server uses "filtered-format', a Map-Reply it originates has the mapping record encoded as:

```
EID-record: (S,G)
    RLOC-record: RLE: (RTR1(L0), RTR3(L1)), p1
```

An (S,G) entry can contain alternate RTRs. So rather than replicating to multiple RTRs, one of a RTR set may be used based on the RTR reachability status. An ITR can test reachability status to any layer-0 RTR using RLOC-probing so it can choose one RTR from a set to replicate to. When this is done the RTRs are encoded in different RLOC-records versus together in one RLE RLOC-record. This moves the replication load off the ITRs at the source site to the RTRs inside the network infrastructure. This mechanism can also be used by level-n RTRs to level-n+1 RTRs.

The following mapping would be encoded in a Map-Reply sent by a Map-Server and stored in the ITR. The ITR would use RTR1 until it went unreachable and then switch to use RTR2:

```
EID-record: (S,G)
    RLOC-record: RTR1, p1
    RLOC-record: RTR2, p2
```

9. Security Considerations

[I-D.ietf-lisp-sec] defines a set of security mechanisms that provide origin authentication, integrity and anti-replay protection to LISP's EID-to-RLOC mapping data conveyed via mapping lookup process. LISP-SEC also enables verification of authorization on EID-prefix claims in Map-Reply messages.

Additional security mechanisms to protect the LISP Map-Register messages are defined in [RFC6833].

The security of the Mapping System Infrastructure depends on the particular mapping database used. The [I-D.ietf-lisp-ddt] specification, as an example, defines a public-key based mechanism

that provides origin authentication and integrity protection to the LISP DDT protocol.

Map-Replies received by the source-ITR can be signed (by the Map-Server) so the ITR knows the replication-list is from a legit source.

Data-plane encryption can be used when doing unicast rep-encapsulation as described in [RFC8061]. For further study we will look how to do multicast rep-encapsulation.

10. IANA Considerations

This document has no IANA implications

11. Acknowledgements

The authors want to thank Greg Shepherd, Joel Halpern and Sharon Barkai for their insightful contribution to shaping the ideas in this document. Thanks also goes to Jimmy Kyriannis, Paul Vinciguerra, Florin Coras, and Yan Filyurin for testing an implementation of this draft.

12. References

12.1. Normative References

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

[RFC2236] Fenner, W., "Internet Group Management Protocol, Version 2", RFC 2236, DOI 10.17487/RFC2236, November 1997, <http://www.rfc-editor.org/info/rfc2236>.

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, DOI 10.17487/RFC3376, October 2002, <http://www.rfc-editor.org/info/rfc3376>.

[RFC3618] Fenner, B., Ed. and D. Meyer, Ed., "Multicast Source Discovery Protocol (MSDP)", RFC 3618, DOI 10.17487/RFC3618, October 2003, <http://www.rfc-editor.org/info/rfc3618>.

[RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, DOI 10.17487/RFC4601, August 2006, <http://www.rfc-editor.org/info/rfc4601>.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, DOI 10.17487/RFC4607, August 2006, <http://www.rfc-editor.org/info/rfc4607>.

12.2. Informative References

[I-D.coras-lisp-re] Coras, F., Cabellos-Aparicio, A., Domingo-Pascual, J., Maino, F., and D. Farinacci, "LISP Replication Engineering", draft-coras-lisp-re-08 (work in progress), November 2015.

[I-D.farinacci-lisp-mr-signaling] Farinacci, D. and M. Napierala, "LISP Control-Plane Multicast Signaling", draft-farinacci-lisp-mr-signaling-06 (work in progress), February 2015.

[I-D.ietf-lisp-ddt] Fuller, V., Lewis, D., Ermagan, V., Jain, A., and A. Smirnov, "LISP Delegated Database Tree", draft-ietf-lisp-ddt-09 (work in progress), January 2017.

[I-D.ietf-lisp-eid-mobility] Portoles-Comeras, M., Ashtaputre, V., Moreno, V., Maino, F., and D. Farinacci, "LISP L2/L3 EID Mobility Using a Unified Control Plane", draft-ietf-lisp-eid-mobility-00 (work in progress), May 2017.

[I-D.ietf-lisp-sec] Maino, F., Ermagan, V., Cabellos-Aparicio, A., and D. Saucez, "LISP-Security (LISP-SEC)", draft-ietf-lisp-sec-12 (work in progress), November 2016.

[RFC6830] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "The Locator/ID Separation Protocol (LISP)", RFC 6830, DOI 10.17487/RFC6830, January 2013, <http://www.rfc-editor.org/info/rfc6830>.

[RFC6831] Farinacci, D., Meyer, D., Zwiebel, J., and S. Venaas, "The Locator/ID Separation Protocol (LISP) for Multicast Environments", RFC 6831, DOI 10.17487/RFC6831, January 2013, <http://www.rfc-editor.org/info/rfc6831>.

[RFC6833] Fuller, V. and D. Farinacci, "Locator/ID Separation Protocol (LISP) Map-Server Interface", RFC 6833, DOI 10.17487/RFC6833, January 2013, <http://www.rfc-editor.org/info/rfc6833>.

[RFC8060] Farinacci, D., Meyer, D., and J. Snijders, "LISP Canonical Address Format (LCAF)", RFC 8060, DOI 10.17487/RFC8060, February 2017, <http://www.rfc-editor.org/info/rfc8060>.

[RFC8061] Farinacci, D. and B. Weis, "Locator/ID Separation Protocol (LISP) Data-Plane Confidentiality", RFC 8061, DOI 10.17487/RFC8061, February 2017, <http://www.rfc-editor.org/info/rfc8061>.

Appendix A. Document Change Log

A.1. Changes to draft-ietf-lisp-signal-free-multicast-06

- Posted July 2017.

- Stig made a comment about referencing RFC6831 when an RLOC is a multicast address. It opens up a lot of assumptions on what parts of RFC6831 is performed and which parts should not be performed. In the case of signal-free-multicast, join the underlay trees as a multicast host by using IGMP.

A.2. Changes to draft-ietf-lisp-signal-free-multicast-05

- Posted July 2017.

- Make it clear that when a RLE is sent by an ETR and it is already in the merged RLE list on the Map-Server, that the Map-Server replaces the RLE entry (versus adding a duplicate RLE entry to the list).

- Make it clear that an RLOC can be a unicast or multicast address. Then make a reference to RFC6831 about mechanisms to support multicast RLOCs.

- Fix some typos.

A.3. Changes to draft-ietf-lisp-signal-free-multicast-04

- Posted May 2017.

- Make it clear that recieiver-ETRs need configuraiton information for what Map-Servers (S,G) entries are registered to.

- Make it clear this document indicates what RTR layered hierarchy to use and not if its the best hierarchy to use.

A.4. Changes to draft-ietf-lisp-signal-free-multicast-03

- Posted April 2017.

- Add "Multi-Homing Considerations" section to describe the case where a source LISP site has multiple ITRs and the multicast distribution tree at the source site branches to more than one ITR. And at receiver sites where there are multiple ETRs and multiple RLOCs per ETR.

A.5. Changes to draft-ietf-lisp-signal-free-multicast-02

- Posted October 2016.

- Updated document expiration timer.

A.6. Changes to draft-ietf-lisp-signal-free-multicast-01

- Posted April 2016.

- Add text to define RTRs and indicate how RTR level number is used for LISP-RE.

- Draw figure 2 that shows a LISP-RE topology.

- Indicate that PIM-ASM or (*,G) trees can be supported in LISP Signal-Free Multicast.

A.7. Changes to draft-ietf-lisp-signal-free-multicast-00

- Posted late December 2015.

- Converted draft-farinacci-lisp-signal-free-multicast-04 into LISP working group draft.

A.8. Changes to draft-farinacci-lisp-signal-free-multicast-04

- Posted early December 2015.

- Update references and document timer.

A.9. Changes to draft-farinacci-lisp-signal-free-multicast-03

- Posted June 2015.

- Update references and document timer.

A.10. Changes to draft-farinacci-lisp-signal-free-multicast-02

- Posted December 2014.

- Added section about how LISP-RE can use the mechanisms from signal-free-multicast so we can avoid head-end replication and avoid signalling across a layered RE topology.

A.11. Changes to draft-farinacci-lisp-signal-free-multicast-01

- Posted June 2014.

- Changes based on implementation experience of this draft.

A.12. Changes to draft-farinacci-lisp-signal-free-multicast-00

- Posted initial draft February 2014.

Authors' Addresses

Victor Moreno
Cisco Systems
170 Tasman Drive
San Jose, California 95134
USA

Email: vimoreno@cisco.com

Dino Farinacci
lispers.net
San Jose, CA 95120
USA

Email: farinacci@gmail.com