Internet Engineering Task Force (IETF)
Request for Comments: 8378
Category: Experimental
ISSN: 2070-1721

V. Moreno
Cisco Systems
D. Farinacci
lispers.net

May 2018

Signal-Free Locator/ID Separation Protocol (LISP) Multicast

Abstract

When multicast sources and receivers are active at Locator/ID Separation Protocol (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 described in this document 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.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for examination, experimental implementation, and evaluation.

This document defines an Experimental Protocol for the Internet community. This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8378.

Copyright Notice

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 proposed here does not extend PIM [RFC7761] over the overlay as proposed in [RFC6831], nor does the mechanism utilize direct signaling between the Receiver-ETRs and Sender-ITRs as described in [LISP-MULTI-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 Routing Locators (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 database mapping entry in the LISP mapping system.

When the Ingress Tunnel Router (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 general procedures (Section 5) 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 Source-Specific Multicast (SSM) trees discusses the implications to the general procedures for SSM multicast trees over different topological scenarios. A section on Any-Source Multicast (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, which is 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 Re-encapsulating Tunnel Routers (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 defined here to facilitate the descriptions and discussions within this particular document.

Source: Multicast source endpoint. The host that originates multicast packets.

Receiver: Multicast group member endpoint. The host joins a 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 (RP) [RFC7761] is located. The RP site and the source site MAY be the same in some situations.

Receiver-ETR: LISP decapsulating the Tunnel Router (xTR) at the receiver site. This is a multicast ETR.

Source-ITR: LISP encapsulating 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) LISP Canonical Address Format (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 an RLE. 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 trade-off between the width of the level versus the number of levels can be selected.

ASM: Any-Source Multicast as defined in [RFC3569] where multicast distribution trees are built with a Rendezvous Point [RFC7761].

SSM: Source-Specific Multicast as defined in [RFC3569] where multicast distribution trees are built and rooted at the multicast router(s) directly connected to the multicast source.

3. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

4. 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 endpoints, respectively.

R1 is a multicast router in Source-site-1.

R2 is a multicast router at the receiver site.

Map-Servers and Map-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 distributed mapping system, such as a Delegated Database Tree (DDT).

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 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.

5. 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 endpoints 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 MUST 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.

5.1. General Receiver-Site Procedures

5.1.1. 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 to learn which multicast group the receivers are for.

5.1.2. Receiver-Site Registration

Once the Receiver-ETRs detect the presence of receivers at the receiver site, the Receiver-ETRs MUST 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 MUST use the multicast entry (Source, Group) tuple as its Endpoint ID (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 Info LCAF Type defined in [RFC8060].

The RLOC in the Map-Register message MUST be encoded using the 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 system, Map-Reply messages, as well as the map-cache at the Source-ITRs.

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

1. merge-request bit set to 1. The Map-Register messages are 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 MAY 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.

5.1.3. 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 already exists in the Map-Server's RLE-merged list, the Map-Server will replace the single RLE with the information from the Map-Register RLOC-record. The Map-Server MUST NOT merge duplicate RLOCs in the consolidated replication list.

5.2. 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.

5.2.1. Multicast Tree Building at the Source Site

When the source site receives the Map-Notify messages from the mapping system as described in Section 5.3, 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 MUST 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.

5.2.2. Multicast Destination Resolution

On reception of multicast packets, the Source-ITR obtains 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 is encoded using the Multicast Info LCAF Type, and the corresponding RLOC-records are encoded using the RLE LCAF Type. Both LCAF Types are defined in [RFC8060].

5.3. 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 [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 (when using LISP-DDT [RFC8111]), 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 across a distributed mapping transport system, per standard LISP lookup procedures, and obtain the necessary information to send the

Map-Notify messages to the source site. The Map-Notify messages are sent with an authentication length of 0 as they would not be authenticated.

When the Map-Servers are distributed, different receiver sites MAY register to different Map-Servers. However, this is not supported with the currently defined mechanisms.

6. Source-Specific Multicast Trees

The interconnection of SSM trees across sites will follow the general receiver-site procedures described in Section 5.1 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 Sections 6.1 and 6.2 .

6.1. 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 are not 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.

6.2. Source Not Directly Connected to Source-ITRs

The general LISP notification procedures described in Section 5.3 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 5.2.1.

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.

7. Multihoming Considerations

7.1. Multiple ITRs at a Source Site

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

- 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, one or more ITRs on the source LAN MUST be IGMP querier candidates. Therefore, it is required that they be configured as such.

- 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.

7.2. Multiple ETRs at a Receiver Site

When multiple ETRs exist in a receiver multicast site and each one creates a multicast join state, each Map-Registers its RLOC address to the mapping system. In this scenario, the replication happens on the overlay causing multiple ETR entry points to replicate to all receivers instead of 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 5.1 are 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, one or more ETRs on a LAN MUST be IGMP querier candidates. Therefore, it is required that they are configured as such.

7.3. 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 recommended that an ETR send a Map-Register with a single RLOC-record that uses the Explicit Locator Path (ELP) LCAF Type [RFC8060] that is nested inside the RLE LCAF. For example, say ETR1 has assigned RLOC1 and RLOC2 for a LISP receiver site. Also, 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.

7.4. Multicast RLOCs for an ETR at a Receiver Site

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

8. 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 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 co-located 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 5.1.2 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 co-located 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 5.2 are followed.

5. When 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 5.1.2.

9. Signal-Free Multicast for Replication Engineering

The mechanisms in this specification can be applied to the "LISP Replication Engineering" [LISP-RE] design. Rather than have 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 [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 of an (S,G) mapping database entry, it could replicate to one or more layer 0 RTRs in the LISP-RE hierarchy.

This document specifies how the RTR hierarchy is determined but not the optimal layers of RTRs to be used. Methods for determining optimal paths or RTR topological closeness are out of scope for this 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 that 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 will be 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 keeping 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 in which they were registered and are stored in a Map-Server. When a Map-Server uses 'complete-format', the 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 a priority greater than 1.

When a Map_server uses 'filtered-format', the 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 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 instead of 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
```

10. Security Considerations

[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 the 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. As an example, [RFC8111] 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 legitimate source.

Data-plane encryption can be used when doing unicast rep-encapsulation as described in [RFC8061].

11. IANA Considerations

This document has no IANA actions.

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, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2236] Fenner, W., "Internet Group Management Protocol, Version 2", RFC 2236, DOI 10.17487/RFC2236, November 1997, <https://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, <https://www.rfc-editor.org/info/rfc3376>.

[RFC3569] Bhattacharyya, S., Ed., "An Overview of Source-Specific Multicast (SSM)", RFC 3569, DOI 10.17487/RFC3569, July 2003, <https://www.rfc-editor.org/info/rfc3569>.

[RFC6830] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "The Locator/ID Separation Protocol (LISP)", RFC 6830, DOI 10.17487/RFC6830, January 2013, <https://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, <https://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, <https://www.rfc-editor.org/info/rfc6833>.

[RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March 2016, <https://www.rfc-editor.org/info/rfc7761>.

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

[RFC8111] Fuller, V., Lewis, D., Ermagan, V., Jain, A., and A. Smirnov, "Locator/ID Separation Protocol Delegated Database Tree (LISP-DDT)", RFC 8111, DOI 10.17487/RFC8111, May 2017, <https://www.rfc-editor.org/info/rfc8111>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

12.2. Informative References

[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", Work in Progress, draft-ietf-lisp-eid-mobility-01, November 2017.

[LISP-MULTI-SIGNALING] Farinacci, D. and M. Napierala, "LISP Control-Plane Multicast Signaling", Work in Progress, draft-farinacci-lisp-mr-signaling-06, February 2015.

[LISP-RE] Coras, F., Cabellos-Aparicio, A., Domingo-Pascual, J., Maino, F., and D. Farinacci, "LISP Replication Engineering", Work in Progress, draft-coras-lisp-re-08, November 2015.

[LISP-SEC] Maino, F., Ermagan, V., Cabellos-Aparicio, A., and D. Saucez, "LISP-Security (LISP-SEC)", Work in Progress, draft-ietf-lisp-sec-15, April 2018.

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

Acknowledgements

The authors want to thank Greg Shepherd, Joel Halpern, and Sharon Barkai for their insightful contribution to shaping the ideas in this document. A special thanks to Luigi Iannone, LISP WG co-chair, for shepherding this working group document. Thanks also goes to Jimmy Kyriannis, Paul Vinciguerra, Florin Coras, and Yan Filyurin for testing an implementation of this document.

Authors' Addresses

Victor Moreno
Cisco Systems
170 Tasman Drive
San Jose, California 95134
United States of America

Email: vimoreno@cisco.com

Dino Farinacci
lispers.net
San Jose, CA 95120
United States of America

Email: farinacci@gmail.com