Network Working Group   L. Jakab
Internet-Draft   A. Cabellos-Aparicio
Intended status: Informational   F. Coras
Expires: September 13, 2012   J. Domingo-Pascual
                                                   Technical University of
                                                                Catalonia
                                                               D. Lewis
                                                          Cisco Systems
                                                         March 12, 2012

             LISP Network Element Deployment Considerations
                    draft-ietf-lisp-deployment-03.txt

Abstract

   This document discusses the different scenarios for the deployment of
   the new network elements introduced by the Locator/Identifier
   Separation Protocol (LISP).

Table of Contents

1. Introduction

The Locator/Identifier Separation Protocol (LISP) addresses the
scaling issues of the global Internet routing system by separating
the current addressing scheme into Endpoint IDentifiers (EIDs) and
Routing LOCators (RLOCs). The main protocol specification
[I-D.ietf-lisp] describes how the separation is achieved, which new
network elements are introduced, and details the packet formats for
the data and control planes.

While the boundary between the core and edge is not strictly defined,
one widely accepted definition places it at the border routers of
stub autonomous systems, which may carry a partial or complete
default-free zone (DFZ) routing table. The initial design of LISP
took this location as a baseline for protocol development. However,
the applications of LISP go beyond of just decreasing the size of the
DFZ routing table, and include improved multihoming and ingress
traffic engineering (TE) support for edge networks, and even
individual hosts. Throughout the draft we will use the term LISP
site to refer to these networks/hosts behind a LISP Tunnel Router.
We formally define it as:

LISP site: A single host or a set of network elements in an edge
   network under the administrative control of a single organization,
   delimited from other networks by LISP Tunnel Router(s).

Since LISP is a protocol which can be used for different purposes, it
is important to identify possible deployment scenarios and the
additional requirements they may impose on the protocol specification
and other protocols. The main specification [I-D.ietf-lisp] mentions
positioning of tunnel routers, but without an in-depth discussion.
This document fills that gap, by exploring the most common cases.
While the theoretical combinations of device placements are quite
numerous, the more practical scenarios are given preference in the
following.

Additionally, this documents is intended as a guide for the
operational community for LISP deployments in their networks. It is
expected to evolve as LISP deployment progresses, and the described
scenarios are better understood or new scenarios are discovered.

Each subsection considers an element type, discussing the impact of
deployment scenarios on the protocol specification. For definition
of terms, please refer to the appropriate documents (as cited in the
respective sections).

Comments and discussions about this memo should be directed to the
LISP working group mailing list: lisp@ietf.org.

2. Tunnel Routers

   LISP is a map-and-encap protocol, with the main goal of improving
   global routing scalability. To achieve its goal, it introduces
   several new network elements, each performing specific functions
   necessary to separate the edge from the core. The device that is the
   gateway between the edge and the core is called Tunnel Router (xTR),
   performing one or both of two separate functions:

   1. Encapsulating packets originating from an end host to be
       transported over intermediary (transit) networks towards the
       other end-point of the communication

   2. Decapsulating packets entering from intermediary (transit)
       networks, originated at a remote end host.

   The first function is performed by an Ingress Tunnel Router (ITR),
   the second by an Egress Tunnel Router (ETR).

   Section 8 of the main LISP specification [I-D.ietf-lisp] has a short
   discussion of where Tunnel Routers can be deployed and some of the
   associated advantages and disadvantages. This section adds more
   detail to the scenarios presented there, and provides additional
   scenarios as well.

2.1. Customer Edge

   LISP was designed with deployment at the core-edge boundary in mind,
   which can be approximated as the set of DFZ routers belonging to non-
   transit ASes. For the purposes of this document, we will consider
   this boundary to be consisting of the routers connecting LISP sites
   to their upstreams. As such, this is the most common expected
   scenario for xTRs, and this document considers it the reference
   location, comparing the other scenarios to this one.

```
                       ISP1      ISP2
                         |         |
                         |         |
                      +----+   +----+
                   +--|xTR1|--|xTR2|--+
                   |  +----+   +----+  |
                   |                   |
                   |      LISP site    |
                   +-------------------+
```
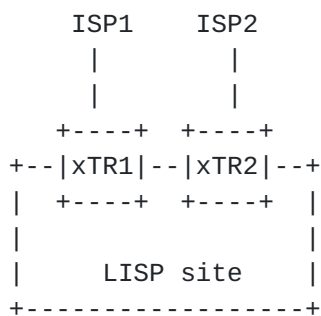
                  Figure 1: xTRs at the customer edge

   From the LISP site perspective the main advantage of this type of

deployment (compared to the one described in the next section) is
having direct control over its ingress traffic engineering. This
makes it is easy to set up and maintain active/active, active/backup,
or more complex TE policies, without involving third parties.

Being under the same administrative control, reachability information
of all ETRs is easier to synchronize, because the necessary control
traffic can be allowed between the locators of the ETRs. A correct
synchronous global view of the reachability status is thus available,
and the Loc-Status-Bits can be set correctly in the LISP data header
of outgoing packets.

By placing the tunnel router at the edge of the site, existing
internal network configuration does not need to be modified.
Firewall rules, router configurations and address assignments inside
the LISP site remain unchanged. This helps with incremental
deployment and allows a quick upgrade path to LISP. For larger sites
with many external connections, distributed in geographically diverse
PoPs, and complex internal topology, it may however make more sense
to both encapsulate and decapsulate as soon as possible, to benefit
from the information in the IGP to choose the best path (see
Section 2.3 for a discussion of this scenario).

Another thing to consider when placing tunnel routers are MTU issues.
Since encapsulating packets increases overhead, the MTU of the end-
to-end path may decrease, when encapsulated packets need to travel
over segments having close to minimum MTU. Some transit networks are
known to provide larger MTU than the typical value of 1500 bytes of
popular access technologies used at end hosts (e.g., IEEE 802.3 and
802.11). However, placing the LISP router connecting to such a
network at the customer edge could possibly bring up MTU issues,
depending on the link type to the provider as opposed to the
following scenario.

2.2. Provider Edge

The other location at the core-edge boundary for deploying LISP
routers is at the Internet service provider edge. The main incentive
for this case is that the customer does not have to upgrade the CE
router(s), or change the configuration of any equipment.
Encapsulation/decapsulation happens in the provider's network, which
may be able to serve several customers with a single device. For
large ISPs with many residential/business customers asking for LISP
this can lead to important savings, since there is no need to upgrade
the software (or hardware, if it's the case) at each client's
location. Instead, they can upgrade the software (or hardware) on a
few PE routers serving the customers. This scenario is depicted in
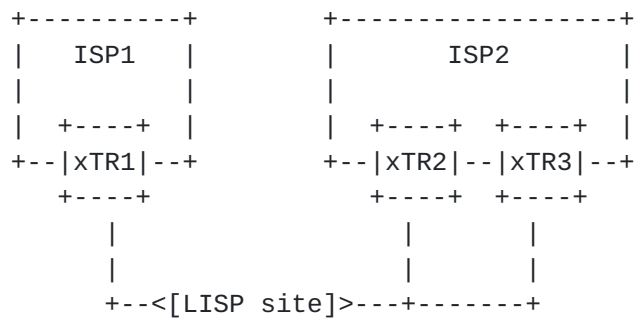Figure 2.

```
            +----------+        +------------------+
            |   ISP1   |        |      ISP2        |
            |          |        |                  |
            |  +----+  |        |  +----+  +----+  |
            +--|xTR1|--+        +--|xTR2|--|xTR3|--+
               +----+              +----+  +----+
                  |                   |       |
                  |                   |       |
               +--<[LISP site]>---+-------+
```

                   Figure 2: xTR at the PE

   While this approach can make transition easy for customers and may be
   cheaper for providers, the LISP site looses one of the main benefits
   of LISP: ingress traffic engineering. Since the provider controls
   the ETRs, additional complexity would be needed to allow customers to
   modify their mapping entries.

   The problem is aggravated when the LISP site is multihomed. Consider
   the scenario in Figure 2: whenever a change to TE policies is
   required, the customer contacts both ISP1 and ISP2 to make the
   necessary changes on the routers (if they provide this possibility).
   It is however unlikely, that both ISPs will apply changes
   simultaneously, which may lead to inconsistent state for the mappings
   of the LISP site. Since the different upstream ISPs are usually
   competing business entities, the ETRs may even be configured to
   compete, either to attract all the traffic or to get no traffic. The
   former will happen if the customer pays per volume, the latter if the
   connectivity has a fixed price. A solution could be to have the
   mappings in the Map-Server(s), and have their operator give control
   over the entries to customer, much like in today's DNS.

   Additionally, since xTR1, xTR2, and xTR3 are in different
   administrative domains, locator reachability information is unlikely
   to be exchanged among them, making it difficult to set Loc-Status-
   Bits correctly on encapsulated packets.

   Compared to the customer edge scenario, deploying LISP at the
   provider edge might have the advantage of diminishing potential MTU
   issues, because the tunnel router is closer to the core, where links
   typically have higher MTUs than edge network links.

2.3. Split ITR/ETR

   In a simple LISP deployment, xTRs are located at the border of the
   LISP site (see Section 2.1). In this scenario packets are routed
   inside the domain according to the EID. However, more complex
   networks may want to route packets according to the destination RLOC.

This would enable them to choose the best egress point.

The LISP specification separates the ITR and ETR functionality and
considers that both entities can be deployed in separated network
equipment. ITRs can be deployed closer to the host (i.e., access
routers). This way packets are encapsulated as soon as possible, and
packets exit the network through the best egress point in terms of
BGP policy. In turn, ETRs can be deployed at the border routers of
the network, and packets are decapsulated as soon as possible.
Again, once decapsulated packets are routed according to the EID, and
can follow the best path according to internal routing policy.

In the following figure we can see an example. The Source (S)
transmits packets using its EID and in this particular case packets
are encapsulated at ITR_1. The encapsulated packets are routed
inside the domain according to the destination RLOC, and can egress
the network through the best point (i.e., closer to the RLOC's AS).
On the other hand, inbound packets are received by ETR_1 which
decapsulates them. Then packets are routed towards S according to
the EID, again following the best path.

```
    +----------------------------------------+
    |                                        |
    |       +-------+              +-------+          +-------+
    |       | ITR_1 |----------+   | ETR_1 |-RLOC_A--| ISP_A |
    |       +-------+          |   +-------+          +-------+
    |  +-+           |         |       |
    |  |S|           |   IGP   |       |
    |  +-+           |         |       |
    |       +-------+          |   +-------+          +-------+
    |       | ITR_2 |----------+   | ETR_2 |-RLOC_B--| ISP_B |
    |       +-------+              +-------+          +-------+
    |                                        |
    +----------------------------------------+
```

                   Figure 3: Split ITR/ETR Scenario

This scenario has a set of implications:

o The site must carry at least partial BGP routes in order to choose
   the best egress point, increasing the complexity of the network.
   However, this is usually already the case for LISP sites that
   would benefit from this scenario.

o If the site is multihomed to different ISPs and any of the
   upstream ISPs is doing uRPF filtering, this scenario may become
   impractical. ITRs need to determine the exit ETR, for setting the
   correct source RLOC in the encapsulation header. This adds

      complexity and reliability concerns.

   o In LISP, ITRs set the reachability bits when encapsulating data
      packets. Hence, ITRs need a mechanism to be aware of the liveness
      of ETRs.

   o ITRs encapsulate packets and in order to achieve efficient
      communications, the MTU of the site must be large enough to
      accommodate this extra header.

   o In this scenario, each ITR is serving fewer hosts than in the case
      when it is deployed at the border of the network. It has been
      shown that cache hit ratio grows logarithmically with the amount
      of users [cache]. Taking this into account, when ITRs are
      deployed closer to the host the effectiveness of the mapping cache
      may be lower (i.e., the miss ratio is higher). Another
      consequence of this is that the site will transmit a higher amount
      of Map-Requests, increasing the load on the distributed mapping
      database.

2.4. Inter-Service Provider Traffic Engineering

   With LISP, two LISP sites can route packets among them and control
   their ingress TE policies. Typically, LISP is seen as applicable to
   stub networks, however the LISP protocol can also be applied to
   transit networks recursively.

   Consider the scenario depicted in Figure 4. Packets originating from
   the LISP site Stub1, client of ISP_A, with destination Stub4, client
   of ISP_B, are LISP encapsulated at their entry point into the ISP_A's
   network. The external IP header now has as the source RLOC an IP
   from ISP_A's address space and destination RLOC from ISP_B's address
   space. One or more ASes separate ISP_A from ISP_B. With a single
   level of LISP encapsulation, Stub4 has control over its ingress
   traffic. However, ISP_B only has the current tools (such as BGP
   prefix deaggregation) to control on which of his own upstream or
   peering links should packets enter. This is either not feasible (if
   fine-grained per-customer control is required, the very specific
   prefixes may not be propagated) or increases DFZ table size.

```
                             _.--.
    Stub1 ...   +-------+       ,-''      `--.       +-------+   ... Stub3
          \  |   R_A1|----,'              `. ---|R_B1   |  /
           --|   R_A2|---(      Transit     )  |       |--
    Stub2 .../  |   R_A3|-----.               ,' ---|R_B2   |  \... Stub4
          +-------+      `--.      _.-'       +-------+
         ...      ISP_A           `--''            ISP_B     ...
```
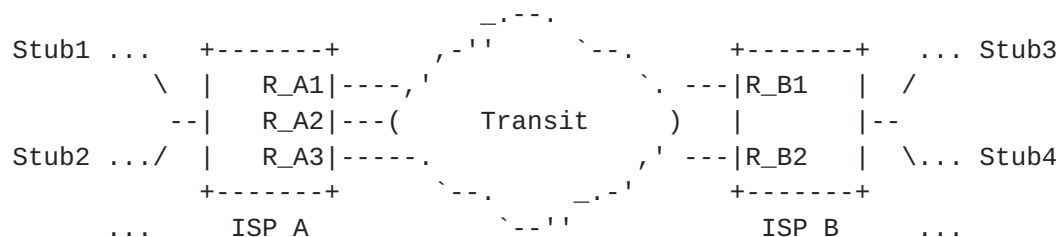
                    Figure 4: Inter-Service provider TE scenario

   A solution for this is to apply LISP recursively. ISP_A and ISP_B
   may reach a bilateral agreement to deploy their own private mapping
   system. ISP_A then encapsulates packets destined for the prefixes of
   ISP_B, which are listed in the shared mapping system. Note that in
   this case the packet is double-encapsulated (using R_A1, R_A2 or R_A3
   as source and R_B1 or R_B2 as destination in the example above).
   ISP_B's ETR removes the outer, second layer of LISP encapsulation
   from the incoming packet, and routes it towards the original RLOC,
   the ETR of Stub4, which does the final decapsulation.

   If ISP_A and ISP_B agree to share a private distributed mapping
   database, both can control their ingress TE without the need of
   disaggregating prefixes. In this scenario the private database
   contains RLOC-to-RLOC bindings. The convergence time on the TE
   policies updates is expected to be fast, since ISPs only have to
   update/query a mapping to/from the database.

   This deployment scenario includes two important recommendations.
   First, it is intended to be deployed only between two ISPs (ISP_A and
   ISP_B in Figure 4). If more than two ISPs use this approach, then
   the xTRs deployed at the participating ISPs must either query
   multiple mapping systems, or the ISPs must agree on a common shared
   mapping system. Second, the scenario is only recommended for ISPs
   providing connectivity to LISP sites, such that source RLOCs of
   packets to be reencapsulated belong to said ISP. Otherwise the
   participating ISPs must register prefixes they do not own in the
   above mentioned private mapping system. Failure to follow these
   recommendations may lead to operational and security issues when
   deploying this scenario.

   Besides these recommendations, the main disadvantages of this
   deployment case are:

   o Extra LISP header is needed. This increases the packet size and,
      for efficient communications, it requires that the MTU between
      both ISPs can accommodate double-encapsulated packets.

   o The ISP ITR must encapsulate packets and therefore must know the
      RLOC-to-RLOC binding. These bindings are stored in a mapping
      database and may be cached in the ITR's mapping cache. Cache
      misses lead to an extra lookup latency, unless NERD
      [I-D.lear-lisp-nerd] is used for the lookups.

   o The operational overhead of maintaining the shared mapping
      database.

o If an IPv6 address block is reserved for EID use, as specified in
   [I-D.ietf-lisp-eid-block], the EID-to-RLOC encapsulation (first
   level) can avoid LISP processing altogether for non-LISP
   destinations. The ISP tunnel routers however will not be able to
   take advantage of this optimization, all RLOC-to-RLOC mappings
   need a lookup in the private database (or map-cache, once results
   are cached).

2.5. Tunnel Routers Behind NAT

   NAT in this section refers to IPv4 network address and port
   translation.

2.5.1. ITR

   Packets encapsulated by an ITR are just UDP packets from a NAT
   device's point of view, and they are handled like any UDP packet,
   there are no additional requirements for LISP data packets.

   Map-Requests sent by an ITR, which create the state in the NAT table
   have a different 5-tuple in the IP header than the Map-Reply
   generated by the authoritative ETR. Since the source address of this
   packet is different from the destination address of the request
   packet, no state will be matched in the NAT table and the packet will
   be dropped. To avoid this, the NAT device has to do the following:

o Send all UDP packets with source port 4342, regardless of the
   destination port, to the RLOC of the ITR. The most simple way to
   achieve this is configuring 1:1 NAT mode from the external RLOC of
   the NAT device to the ITR's RLOC (Called "DMZ" mode in consumer
   broadband routers).

o Rewrite the ITR-AFI and "Originating ITR RLOC Address" fields in
   the payload.

   This setup supports a single ITR behind the NAT device.

2.5.2. ETR

   An ETR placed behind NAT is reachable from the outside by the
   Internet-facing locator of the NAT device. It needs to know this
   locator (and configure a loopback interface with it), so that it can
   use it in Map-Reply and Map-Register messages. Thus support for
   dynamic locators for the mapping database is needed in LISP
   equipment.

   Again, only one ETR behind the NAT device is supported.

An implication of the issues described above is that LISP sites with
xTRs can not be behind carrier based NATs, since two different sites
would collide on the port forwarding.

2.6. Summary and Feature Matrix

```
    Feature                           CE    PE    Split   Rec.
    ---------------------------------------------------------
    Control of ingress TE             x     -     x       x
    No modifications to existing
        int. network infrastructure  x     x     -       -
    Loc-Status-Bits sync              x     -     x       x
    MTU/PMTUD issues minimized        -     x     -       x
```

3. Map-Resolvers and Map-Servers

3.1. Map-Servers

The Map-Server learns EID-to-RLOC mapping entries from an
authoritative source and publishes them in the distributed mapping
database. These entries are learned through authenticated Map-
Register messages sent by authoritative ETRs. Also, upon reception
of a Map-Request, the Map-Server verifies that the destination EID
matches an EID-prefix for which it is authoritative for, and then re-
encapsulates and forwards it to a matching ETR. Map-Server
functionality is described in detail in [I-D.ietf-lisp-ms].

The Map-Server is provided by a Mapping Service Provider (MSP). A
MSP can be any of the following:

o EID registrar. Since the IPv4 address space is nearing
   exhaustion, IPv4 EIDs will come from already allocated Provider
   Independent (PI) space. The registrars in this case remain the
   current five Regional Internet Registries (RIRs). In the case of
   IPv6, the possibility of reserving a /16 block as EID space is
   currently under consideration [I-D.ietf-lisp-eid-block]. If
   granted by IANA, the community will have to determine the body
   responsible for allocations from this block, and the associated
   policies. For already allocated IPv6 prefixes the principles from
   IPv4 should be applied.

o Third parties. Participating in the LISP mapping system is
   similar to participating in global routing or DNS: as long as
   there is at least another already participating entity willing to
   forward the newcomer's traffic, there is no barrier to entry.
   Still, just like routing and DNS, LISP mappings have the issue of
   trust, with efforts underway to make the published information

verifiable. When these mechanisms will be deployed in the LISP
mapping system, the burden of providing and verifying trust should
be kept away from MSPs, which will simply host the secured
mappings. This will keep the low barrier of entry to become an
MSP for third parties.

In all cases, the MSP configures its Map-Server(s) to publish the
prefixes of its clients in the distributed mapping database and start
encapsulating and forwarding Map-Requests to the ETRs of the AS.
These ETRs register their prefix(es) with the Map-Server(s) through
periodic authenticated Map-Register messages. In this context, for
some LISP end sites, there is a need for mechanisms to:

o Automatically distribute EID prefix(es) shared keys between the
   ETRs and the EID-registrar Map-Server.

o Dynamically obtain the address of the Map-Server in the ETR of the
   AS.

The Map-Server plays a key role in the reachability of the EID-
prefixes it is serving. On the one hand it is publishing these
prefixes into the distributed mapping database and on the other hand
it is encapsulating and forwarding Map-Requests to the authoritative
ETRs of these prefixes. ITRs encapsulating towards EIDs under the
responsibility of a failed Map-Server will be unable to look up any
of their covering prefixes. The only exception are the ITRs that
already contain the mappings in their local cache. In this case ITRs
can reach ETRs until the entry expires (typically 24 hours). For
this reason, redundant Map-Server deployments are desirable. A set
of Map-Servers providing high-availability service to the same set of
prefixes is called a redundancy group. ETRs are configured to send
Map-Register messages to all Map-Servers in the redundancy group. To
achieve fail-over (or load-balancing, if desired), current known BGP
practices can be used on the LISP+ALT BGP overlay network.

Additionally, if a Map-Server has no reachability for any ETR serving
a given EID block, it should not originate that block into the
mapping system.

3.2. Map-Resolvers

A Map-Resolver a is a network infrastructure component which accepts
LISP encapsulated Map-Requests, typically from an ITR, and finds the
appropriate EID-to-RLOC mapping by either consulting its local cache
or by consulting the distributed mapping database. Map-Resolver
functionality is described in detail in [I-D.ietf-lisp-ms].

Anyone with access to the distributed mapping database can set up a

Map-Resolver and provide EID-to-RLOC mapping lookup service. In the
case of the LISP+ALT mapping system, the Map-Resolver needs to become
part of the ALT overlay so that it can forward packets to the
appropriate Map-Servers. For more detail on how the ALT overlay
works, see [I-D.ietf-lisp-alt]

For performance reasons, it is recommended that LISP sites use Map-
Resolvers that are topologically close to their ITRs. ISPs
supporting LISP will provide this service to their customers,
possibly restricting access to their user base. LISP sites not in
this position can use open access Map-Resolvers, if available.
However, regardless of the availability of open access resolvers, the
MSP providing the Map-Server(s) for a LISP site should also make
available Map-Resolver(s) for the use of that site.

In medium to large-size ASes, ITRs must be configured with the RLOC
of a Map-Resolver, operation which can be done manually. However, in
Small Office Home Office (SOHO) scenarios a mechanism for
autoconfiguration should be provided.

One solution to avoid manual configuration in LISP sites of any size
is the use of anycast RLOCs for Map-Resolvers similar to the DNS root
server infrastructure. Since LISP uses UDP encapsulation, the use of
anycast would not affect reliability. LISP routers are then shipped
with a preconfigured list of well know Map-Resolver RLOCs, which can
be edited by the network administrator, if needed.

The use of anycast also helps improving mapping lookup performance.
Large MSPs can increase the number and geographical diversity of
their Map-Resolver infrastructure, using a single anycasted RLOC.
Once LISP deployment is advanced enough, very large content providers
may also be interested running this kind of setup, to ensure minimal
connection setup latency for those connecting to their network from
LISP sites.

While Map-Servers and Map-Resolvers implement different
functionalities within the LISP mapping system, they can coexist on
the same device. For example, MSPs offering both services, can
deploy a single Map-Resolver/Map-Server in each PoP where they have a
presence.

4. Proxy Tunnel Routers

4.1. P-ITR

Proxy Ingress Tunnel Routers (P-ITRs) are part of the non-LISP/LISP
transition mechanism, allowing non-LISP sites to reach LISP sites.

They announce via BGP certain EID prefixes (aggregated, whenever
possible) to attract traffic from non-LISP sites towards EIDs in the
covered range. They do the mapping system lookup, and encapsulate
received packets towards the appropriate ETR. Note that for the
reverse path LISP sites can reach non-LISP sites simply by not
encapsulating traffic. See [I-D.ietf-lisp-interworking] for a
detailed description of P-ITR functionality.

The success of new protocols depends greatly on their ability to
maintain backwards compatibility and inter-operate with the
protocol(s) they intend to enhance or replace, and on the incentives
to deploy the necessary new software or equipment. A LISP site needs
an interworking mechanism to be reachable from non-LISP sites. A
P-ITR can fulfill this role, enabling early adopters to see the
benefits of LISP, similar to tunnel brokers helping the transition
from IPv4 to IPv6. A site benefits from new LISP functionality
(proportionally with existing global LISP deployment) when going
LISP, so it has the incentives to deploy the necessary tunnel
routers. In order to be reachable from non-LISP sites it has two
options: keep announcing its prefix(es) with BGP, or have a P-ITR
announce prefix(es) covering them.

If the goal of reducing the DFZ routing table size is to be reached,
the second option is preferred. Moreover, the second option allows
LISP-based ingress traffic engineering from all sites. However, the
placement of P-ITRs significantly influences performance and
deployment incentives. Section Section 5 is dedicated to the
migration to a LISP-enabled Internet, and includes deployment
scenarios for P-ITRs.

4.2. P-ETR

In contrast to P-ITRs, P-ETRs are not required for the correct
functioning of all LISP sites. There are two cases, where they can
be of great help:

o LISP sites with unicast reverse path forwarding (uRPF)
   restrictions, and

o LISP sites without native IPv6 communicating with LISP nodes with
   IPv6-only locators.

In the first case, uRPF filtering is applied at their upstream PE
router. When forwarding traffic to non-LISP sites, an ITR does not
encapsulate packets, leaving the original IP headers intact. As a
result, packets will have EIDs in their source address. Since we are
discussing the transition period, we can assume that a prefix
covering the EIDs belonging to the LISP site is advertised to the

global routing tables by a P-ITR, and the PE router has a route
towards it. However, the next hop will not be on the interface
towards the CE router, so non-encapsulated packets will fail uRPF
checks.

To avoid this filtering, the affected ITR encapsulates packets
towards the locator of the P-ETR for non-LISP destinations. Now the
source address of the packets, as seen by the PE router is the ITR's
locator, which will not fail the uRPF check. The P-ETR then
decapsulates and forwards the packets.

The second use case is IPv4-to-IPv6 transition. Service providers
using older access network hardware, which only supports IPv4 can
still offer IPv6 to their clients, by providing a CPE device running
LISP, and P-ETR(s) for accessing IPv6-only non-LISP sites and LISP
sites, with IPv6-only locators. Packets originating from the client
LISP site for these destinations would be encapsulated towards the
P-ETR's IPv4 locator. The P-ETR is in a native IPv6 network,
decapsulating and forwarding packets. For non-LISP destination, the
packet travels natively from the P-ETR. For LISP destinations with
IPv6-only locators, the packet will go through a P-ITR, in order to
reach its destination.

For more details on P-ETRs see the [I-D.ietf-lisp-interworking]
draft.

P-ETRs can be deployed by ISPs wishing to offer value-added services
to their customers. As is the case with P-ITRs, P-ETRs too may
introduce path stretch. Because of this the ISP needs to consider
the tradeoff of using several devices, close to the customers, to
minimize it, or few devices, farther away from the customers,
minimizing cost instead.

Since the deployment incentives for P-ITRs and P-ETRs are different,
it is likely they will be deployed in separate devices, except for
the CDN case, which may deploy both in a single device.

In all cases, the existence of a P-ETR involves another step in the
configuration of a LISP router. CPE routers, which are typically
configured by DHCP, stand to benefit most from P-ETRs. To enable
autoconfiguration of the P-ETR locator, a DHCP option would be
required.

As a security measure, access to P-ETRs should be limited to
legitimate users by enforcing ACLs.

5. Migration to LISP

   This section discusses a deployment architecture to support the
   migration to a LISP-enabled Internet. The loosely defined terms of
   "early transition phase", "late transition phase", and "LISP Internet
   phase" refer to time periods when LISP sites are a minority, a
   majority, or represent all edge networks respectively.

5.1. LISP+BGP

   For sites wishing to go LISP with their PI prefix the least
   disruptive way is to upgrade their border routers to support LISP,
   register the prefix into the LISP mapping system, but keep announcing
   it with BGP as well. This way LISP sites will reach them over LISP,
   while legacy sites will be unaffected by the change. The main
   disadvantage of this approach is that no decrease in the DFZ routing
   table size is achieved. Still, just increasing the number of LISP
   sites is an important gain, as an increasing LISP/non-LISP site ratio
   will slowly decrease the need for BGP-based traffic engineering that
   leads to prefix deaggregation. That, in turn, may lead to a decrease
   in the DFZ size in the late transition phase.

   This scenario is not limited to sites that already have their
   prefixes announced with BGP. Newly allocated EID blocks could follow
   this strategy as well during the early LISP deployment phase,
   depending on the cost/benefit analysis of the individual networks.
   Since this leads to an increase in the DFZ size, the following
   architecture should be preferred for new allocations.

5.2. Mapping Service Provider (MSP) P-ITR Service

   In addition to publishing their clients' registered prefixes in the
   mapping system, MSPs with enough transit capacity can offer them
   P-ITR service as a separate service. This service is especially
   useful for new PI allocations, to sites without existing BGP
   infrastructure, that wish to avoid BGP altogether. The MSP announces
   the prefix into the DFZ, and the client benefits from ingress traffic
   engineering without prefix deaggregation. The downside of this
   scenario is path stretch, which may be greater than 1.

   Routing all non-LISP ingress traffic through a third party which is
   not one of its ISPs is only feasible for sites with modest amounts of
   traffic (like those using the IPv6 tunnel broker services today),
   especially in the first stage of the transition to LISP, with a
   significant number of legacy sites. When the LISP/non-LISP site
   ratio becomes high enough, this approach can prove increasingly
   attractive.

Compared to LISP+BGP, this approach avoids DFZ bloat caused by prefix
deaggregation for traffic engineering purposes, resulting in slower
routing table increase in the case of new allocations and potential
decrease for existing ones. Moreover, MSPs serving different clients
with adjacent aggregable prefixes may lead to additional decrease,
but quantifying this decrease is subject to future research study.

5.3. Proxy-ITR Route Distribution (PITR-RD)

Instead of a LISP site, or the MSP, announcing their EIDs with BGP to
the DFZ, this function can be outsourced to a third party, a P-ITR
Service Provider (PSP). This will result in a decrease of the
operational complexity both at the site and at the MSP.

The PSP manages a set of distributed P-ITR(s) that will advertise the
corresponding EID prefixes through BGP to the DFZ. These P-ITR(s)
will then encapsulate the traffic they receive for those EIDs towards
the RLOCs of the LISP site, ensuring their reachability from non-LISP
sites.

While it is possible for a PSP to manually configure each client's
EID routes to be announced, this approach offers little flexibility
and is not scalable. This section presents a scalable architecture
that offers automatic distribution of EID routes to LISP sites and
service providers.

The architecture requires no modification to existing LISP network
elements, but it introduces a new (conceptual) network element, the
EID Route Server, defined as a router that either propagates routes
learned from other EID Route Servers, or it originates EID Routes.
The EID-Routes that it originates are those that it is authoritative
for. It propagates these routes to Proxy-ITRs within the AS of the
EID Route Server. It is worth to note that a BGP capable router can
be also considered as an EID Route Server.

Further, an EID-Route is defined as a prefix originated via the Route
Server of the mapping service provider, which should be aggregated if
the MSP has multiple customers inside a single netblock. This prefix
is propagated to other P-ITRs both within the MSP and to other P-ITR
operators it peers with. EID Route Servers are operated either by
the LISP site, MSPs or PSPs, and they may be collocated with a Map-
Server or P-ITR, but are a functionally discrete entity. They
distribute EID-Routes, using BGP, to other domains, according to
policies set by participants.

```
                          MSP (AS64500)
                          RS ---> P-ITR
                           |        /
                           |   _.--./
                         ,-''     /`--.
        LISP site   ---,' |     v      `.
                     (   |   DFZ        )----- Mapping system
     non-LISP site   ----. |     ^       ,'
                      `--. /    _.-'
                          |  `--''
                         v /
                       P-ITR
                       PSP (AS64501)
```
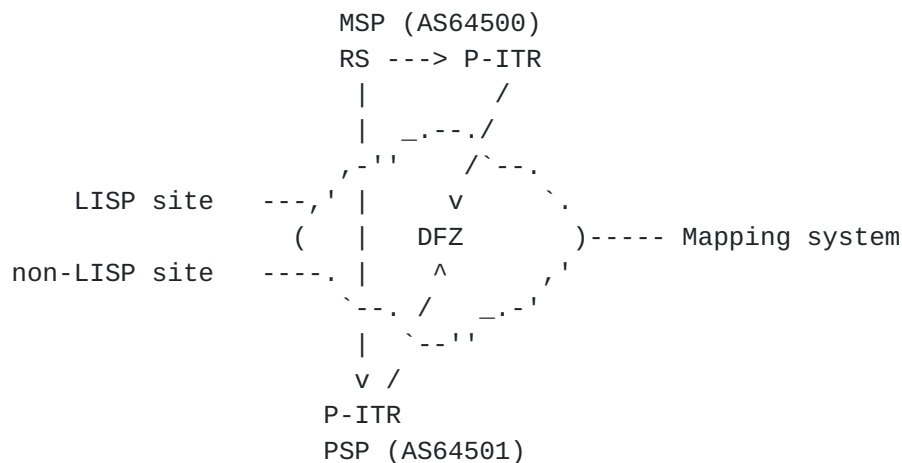
              Figure 5: The P-ITR Route Distribution architecture

   The architecture described above decouples EID origination from route
   propagation, with the following benefits:

   o Can accurately represent business relationships between P-ITR
      operators

   o More mapping system agnostic (no reliance on ALT)

   o Minor changes to P-ITR implementation, no changes to other
      components

   In the example in the figure we have a MSP providing services to the
   LISP site. The LISP site does not run BGP, and gets an EID
   allocation directly from a RIR, or from the MSP, who may be a LIR.
   Existing PI allocations can be migrated as well. The MSP ensures the
   presence of the prefix in the mapping system, and runs an EID Route
   Server to distribute it to P-ITR service providers. Since the LISP
   site does not run BGP, the prefix will be originated with the AS
   number of the MSP.

   In the simple case depicted in Figure 5 the EID-Route of LISP Site
   will be originated by the Route Server, and announced to the DFZ by
   the PSP's P-ITRs with AS path 64501 64500. From that point on, the
   usual BGP dynamics apply. This way, routes announced by P-ITR are
   still originated by the authoritative Route Server. Note that the
   peering relationships between MSP/PSPs and those in the underlying
   forwarding plane may not be congruent, making the AS path to a P-ITR
   shorter than it is in reality.

   The non-LISP site will select the best path towards the EID-prefix,
   according to its local BGP policies. Since AS-path length is usually
   an important metric for selecting paths, a careful placement of P-ITR

could significantly reduce path-stretch between LISP and non-LISP
sites.

The architecture allows for flexible policies between MSP/PSPs.
Consider the EID Route Server networks as control plane overlays,
facilitating the implementation of policies necessary to reflect the
business relationships between participants. The results are then
injected to the common underlying forwarding plane. For example,
some MSP/PSPs may agree to exchange EID-Prefixes and only announce
them to each of their forwarding plane customers. Global
reachability of an EID-prefix depends on the MSP the LISP site buys
service from, and is also subject to agreement between the mentioned
parties.

In terms of impact on the DFZ, this architecture results in a slower
routing table increase for new allocations, since traffic engineering
will be done at the LISP level. For existing allocations migrating
to LISP, the DFZ may decrease since MSPs may be able to aggregate the
prefixes announced.

Compared to LISP+BGP, this approach avoids DFZ bloat caused by prefix
deaggregation for traffic engineering purposes, resulting in slower
routing table increase in the case of new allocations and potential
decrease for existing ones. Moreover, MSPs serving different clients
with adjacent aggregable prefixes may lead to additional decrease,
but quantifying this decrease is subject to future research study.

The flexibility and scalability of this architecture does not come
without a cost however: A PSP operator has to establish either
transit or peering relationships to improve their connectivity.

5.4. Migration Summary

The following table presents the expected effects of the different
transition scenarios during a certain phase on the DFZ routing table
size:

```
 Phase             | LISP+BGP     | MSP P-ITR       | PITR-RD
 ----------------+-------------+----------------+----------------
 Early transition | no change    | slower increase | slower increase
 Late transition  | may decrease | slower increase | slower increase
 LISP Internet    |             considerable decrease
```

It is expected that PITR-RD will co-exist with LISP+BGP during the
migration, with the latter being more popular in the early transition
phase. As the transition progresses and the MSP P-ITR and PITR-RD
ecosystem gets more ubiquitous, LISP+BGP should become less
attractive, slowing down the increase of the number of routes in the

DFZ.

6. Step-by-Step BGP to LISP Migration Procedure

6.1. Customer Pre-Install and Pre-Turn-up Checklist

   1. Determine how many current physical service provider connections
       the customer has and their existing bandwidth and traffic
       engineering requirements.

       This information will determine the number of routing locators,
       and the priorities and weights that should be configured on the
       xTRs.

   2. Make sure customer router has LISP capabilities.

       * Obtain output of 'show version' from the CE router.

          This information can be used to determine if the platform is
          appropriate to support LISP, in order to determine if a
          software and/or hardware upgrade is required.

       * Have customer upgrade (if necessary, software and/or hardware)
          to be LISP capable.

   3. Obtain current running configuration of CE router. A suggested
       LISP router configuration example can be customized to the
       customer's existing environment.

   4. Verify MTU Handling

       * Request increase in MTU to (1556) on service provider
          connections. Prior to MTU change verify that 1500 byte packet
          from P-xTR to RLOC with do not fragment (DF-bit) bit set.

       * Ensure they are not filtering ICMP unreachable or time-
          exceeded on their firewall or router.

       LISP, like any tunneling protocol, will increase the size of
       packets when the LISP header is appended. If increasing the MTU
       of the access links is not possible, care must be taken that ICMP
       is not being filtered in order to allow for Path MTU Discovery to
       take place.

   5. Validate member prefix allocation.

       This step is to check if the prefix used by the customer is a

      direct (Provider Independent), or if it is a prefix assigned by a
      physical service provider (Provider Allocated). If the prefixes
      are assigned by other service provivers then a Letter of
      Agreement is required to announce prefixes through the Proxy
      Service Provider.

   6. Verify the member RLOCs and their reachability.

      This step ensures that the RLOCs configured on the CE router are
      in fact reachable and working.

   7. Prepare for cut-over.

      * If possible, have a host outside of all security and filtering
         policies connected to the console port of the edge router or
         switch.

      * Make sure customer has access to the router in order to
         configure it.

[6.2](). Customer Activating LISP Service

   1. Customer configures LISP on CE router(s) from service provider
      recommended configuration.

      The LISP configuration consists of the EID prefix, the locators,
      and the weights and priorities of the mapping between the two
      values. In addition, the xTR must be configured with Map-
      Resolver(s), Map-Server(s) and the shared key for registering to
      Map-Server(s). If required, Proxy-ETR(s) may be configured as
      well.

      In addition to the LISP configuration, the following:

      * Ensure default route(s) to next-hop external neighbors are
         included and RLOCs are present in configuration.

      * If two or more routers are used, ensure all RLOCs are included
         in the LISP configuration on all routers.

      * It will be necessary to redistribute default route via IGP
         between the external routers.

   2. When transition is ready perform a soft shutdown on existing eBGP
      peer session(s)

      * From CE router, use LIG to ensure registration is successful.

     * To verify LISP connectivity, ping LISP connected sites. See
        http://www.lisp4.net/ and/or http://www.lisp6.net/ for
        potential candidates.

     * To verify connectivity to non-LISP sites, try accessing major
        Internet sites via a web browser.

6.3. Cut-Over Provider Preparation and Changes

   1. Verify site configuration and then active registration on Map-
       Server(s)

     * Authentication key

     * EID prefix

   2. Add EID space to map-cache on proxies

   3. Add networks to BGP advertisement on proxies

     * Modify route-maps/policies on P-xTRs

     * Modify route policies on core routers (if non-connected
        member)

     * Modify ingress policers on core routers

     * Ensure route announcement in looking glass servers, RouteViews

   4. Perform traffic verification test

     * Ensure MTU handling is as expected (PMTUD working)

     * Ensure proxy-ITR map-cache population

     * Ensure access from traceroute/ping servers around Internet

     * Use a looking glass, to check for external visibility of
        registration via several Map-Resolvers (e.g.,
        http://lispmon.net/).

7. Security Considerations

   Security implications of LISP deployments are to be discussed in
   separate documents. [I-D.saucez-lisp-security] gives an overview of
   LISP threat models, while securing mapping lookups is discussed in
   [I-D.ietf-lisp-sec].

8. IANA Considerations

   This memo includes no request to IANA.

9. Acknowledgements

   Many thanks to Margaret Wasserman for her contribution to the IETF76
   presentation that kickstarted this work. The authors would also like
   to thank Damien Saucez, Luigi Iannone, Joel Halpern, Vince Fuller,
   Dino Farinacci, Terry Manderson, Noel Chiappa, Hannu Flinck, and
   everyone else who provided input.

10. References

10.1. Normative References

   [I-D.ietf-lisp]
              Farinacci, D., Fuller, V., Meyer, D., and D. Lewis,
              "Locator/ID Separation Protocol (LISP)",
              draft-ietf-lisp-15 (work in progress), July 2011.

   [I-D.ietf-lisp-alt]
              Fuller, V., Farinacci, D., Meyer, D., and D. Lewis, "LISP
              Alternative Topology (LISP+ALT)", draft-ietf-lisp-alt-09
              (work in progress), September 2011.

   [I-D.ietf-lisp-interworking]
              Lewis, D., Meyer, D., Farinacci, D., and V. Fuller,
              "Interworking LISP with IPv4 and IPv6",
              draft-ietf-lisp-interworking-02 (work in progress),
              June 2011.

   [I-D.ietf-lisp-ms]
              Fuller, V. and D. Farinacci, "LISP Map Server Interface",
              draft-ietf-lisp-ms-12 (work in progress), October 2011.

   [I-D.ietf-lisp-sec]
              Maino, F., Ermagan, V., Cabellos-Aparicio, A., Saucez, D.,
              and O. Bonaventure, "LISP-Security (LISP-SEC)",
              draft-ietf-lisp-sec-00 (work in progress), July 2011.

   [I-D.saucez-lisp-security]
              Saucez, D., Iannone, L., and O. Bonaventure, "LISP
              Security Threats", draft-saucez-lisp-security-03 (work in
              progress), March 2011.

10.2. Informative References

   [I-D.ietf-lisp-eid-block]
              Lewis, D., Meyer, D., Iannone, L., and V. Fuller, "LISP
              EID Block", draft-ietf-lisp-eid-block-01 (work in
              progress), October 2011.

   [I-D.lear-lisp-nerd]
              Lear, E., "NERD: A Not-so-novel EID to RLOC Database",
              draft-lear-lisp-nerd-08 (work in progress), March 2010.

   [cache]   Jung, J., Sit, E., Balakrishnan, H., and R. Morris, "DNS
              performance and the effectiveness of caching", 2002.

Authors' Addresses

   Lorand Jakab
   Technical University of Catalonia
   C/Jordi Girona, s/n
   BARCELONA 08034
   Spain

   Email: ljakab@ac.upc.edu

   Albert Cabellos-Aparicio
   Technical University of Catalonia
   C/Jordi Girona, s/n
   BARCELONA 08034
   Spain

   Email: acabello@ac.upc.edu

   Florin Coras
   Technical University of Catalonia
   C/Jordi Girona, s/n
   BARCELONA 08034
   Spain

   Email: fcoras@ac.upc.edu

Jordi Domingo-Pascual
Technical University of Catalonia
C/Jordi Girona, s/n
BARCELONA 08034
Spain

Email: jordi.domingo@ac.upc.edu

Darrel Lewis
Cisco Systems
170 Tasman Drive
San Jose, CA 95134
USA

Email: darlewis@cisco.com