Network Working Group   V. Fuller
Internet-Draft   D. Farinacci
Intended status: Experimental   D. Meyer
Expires: July 29, 2010   D. Lewis
                                                                Cisco
                                                     January 25, 2010

LISP Alternative Topology (LISP+ALT)
draft-ietf-lisp-alt-02.txt

Abstract

   This document describes a method of building an alternative, logical
   topology for managing Endpoint Identifier to Routing Locator mappings
   using the Locator/ID Separation Protocol. The logical network is
   built as an overlay on the public Internet using existing
   technologies and tools, specifically the Border Gateway Protocol and
   the Generic Routing Encapsulation. An important design goal for
   LISP+ALT is to allow for the relatively easy deployment of an
   efficient mapping system while minimizing changes to existing
   hardware and software.

Status of this Memo

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

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF), its areas, and its working groups. Note that
   other groups may also distribute working documents as Internet-
   Drafts.

   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."

   The list of current Internet-Drafts can be accessed at
   http://www.ietf.org/ietf/1id-abstracts.txt.

   The list of Internet-Draft Shadow Directories can be accessed at
   http://www.ietf.org/shadow.html.

   This Internet-Draft will expire on July 29, 2010.

Copyright Notice

Table of Contents

[1](). Requirements Notation

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

2. Introduction

   This document describes a method of building an alternative logical
   topology for managing Endpoint identifier to Routing Locator mappings
   using the Locator/ID Separation Protocol [LISP]. This logical
   topology uses existing technology and tools, specifically the Border
   Gateway Protocol [RFC4271] and its multi-protocol extension
   [RFC2858], along with the Generic Routing Encapsulation [RFC2784]
   protocol to construct an overlay network of devices that advertise
   EID-prefixes only. These Endpoint Identifier Prefix Aggregators hold
   hierarchically-assigned pieces of the Endpoint Identifier space
   (i.e., prefixes) and their next hops toward the network element which
   is authoritative for Endpoint Identifier-to-Routing Locator mapping
   for that prefix. Tunnel routers can use this overlay to make queries
   against and respond to mapping requests made against the distributed
   Endpoint Identifier-to-Routing Locator mapping database. Note the
   database is distributed (as described in [LISP]) and is stored in the
   ETRs.

   Note that an important design goal of LISP+ALT is to minimize the
   number of changes to existing hardware and/or software that are
   required to deploy the mapping system. It is envisioned that in most
   cases existing technology can be used to implement and deploy LISP+
   ALT. Since the deployment of LISP+ALT adds new devices to the
   network, existing devices not need changes or upgrades. They can
   function as they are to realize an underlying and robust physical
   topology.

   The remainder of this document is organized as follows: Section 3
   provides the definitions of terms used in this document. Section 4
   outlines the basic LISP 1.5 model. Section 5 provides a basic
   overview of the LISP Alternate Topology architecture, and Section 6
   describes how the ALT uses BGP to propagate Endpoint Identifier
   reachability over the overlay network. Section 8 describes the
   construction of the ALT aggregation hierarchy, and Section 9
   discusses how LISP+ALT elements are connected to form the overlay
   network.

3. Definition of Terms

   LISP+ALT operates on two name spaces and introduces a new network
   element, the LISP+ALT Router (see below). This section provides
   high-level definitions of the LISP+ALT name spaces, network elements,
   and message types.

   The Alternative Logical Topology (ALT): The virtual overlay network
      made up of tunnels between EID Prefix Aggregators. The Border
      Gateway Protocol (BGP) runs between ALT routers and is used to
      carry reachability information for EID prefixes.

   Legacy Internet: The portion of the Internet which does not run LISP
      and does not participate in LISP+ALT.

   LISP+ALT Router: The devices which run on the ALT. The ALT is a
      static network built using tunnels between LISP+ALT routers.
      These routers are deployed in a hierarchy in which routers at each
      level in the this hierarchy are responsible for aggregating all
      EID prefixes learned from those logically "below" them and
      advertising summary prefixes to the routers logically "above"
      them. All prefix learning and propagation between levels is done
      using BGP. A LISP+ALT router at the lowest level, or "edge" of
      the ALT, learns EID prefixes from its "client" ETRs. See
      Section 4.1 for a description of how EID prefixes are learned at
      the "edge" of the ALT. See also Section 7 for details on how BGP
      is configured between the different network elements.

      The primary function of LISP+ALT routers is to provide a
      lightweight forwarding infrastructure for LISP control-plane
      messages (Map-Request and Map-Reply), and to transport data
      packets when the packet has the same destination address in both
      the inner (encapsulating) destination and outer destination
      addresses ((i.e., a Data Probe packet).

    Endpoint ID (EID): A 32-bit (for IPv4) or 128-bit (for ipv6) value
      used in the source and destination address fields of the first
      (most inner) LISP header of a packet. A packet that is emitted by
      a system contains EIDs in its headers and LISP headers are
      prepended only when the packet reaches an Ingress Tunnel Router
      (ITR) on the data path to the destination EID.

      In LISP+ALT, EID-prefixes MUST BE assigned in a hierarchical
      manner (in power-of-two) such that they can be aggregated by LISP+
      ALT routers. In addition, a site may have site-local structure in
      how EIDs are topologically organized (subnetting) for routing
      within the site; this structure is not visible to the global
      routing system.

   EID-Prefix Aggregate: A set of EID-prefixes said to be aggregatable
      in the [RFC4632] sense. That is, an EID-Prefix aggregate is
      defined to be a single contiguous power-of-two EID-prefix block.
      Such a block is characterized by a prefix and a length.

   Routing Locator (RLOC): An IP address of an egress tunnel router
      (ETR). It is the output of a EID-to-RLOC mapping lookup. An EID
      maps to one or more RLOCs. Typically, RLOCs are numbered from
      topologically-aggregatable blocks that are assigned to a site at
      each point to which it attaches to the global Internet; where the
      topology is defined by the connectivity of provider networks,
      RLOCs can be thought of as Provider Aggregatable (PA) addresses.
      Note that in LISP+ALT, RLOCs are not carried by LISP+ALT routers.

    EID-to-RLOC Mapping: A binding between an EID and the RLOC-set that
      can be used to reach the EID. The term "mapping" refers to an
      EID-to-RLOC mapping.

    EID Prefix Reachability: An EID prefix is said to be "reachable" if
      one or more of its locators are reachable. That is, an EID prefix
      is reachable if the ETR (or its proxy) that is authoritative for a
      given EID-to-RLOC mapping is reachable.

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

   Default Route: A Default Route in the context of LISP+ALT is a EID-
      prefix value of 0.0.0.0/0 (or 0::/0 for ipv6) which is advertised
      by BGP on top of the ALT. The Default Route is used to create a
      forwarding path for a packet to be sent into the ALT (and ALT
      datagram) on a router which does not have a full ALT forwarding
      database.

4. The LISP 1.5 model

As documented in [LISP], the LISP 1.5 model uses the same basic
query/response protocol machinery as LISP 1.0. In particular, LISP+
ALT provides two mechanisms for an ITR to obtain EID-to-RLOC mappings
(both of these techniques are described in more detail in
Section 9.2):

Data Probe: An ITR may send the first few data packets into the ALT
   to minimize packet loss and to probe for the mapping; the
   authoritative ETR will respond to the ITR with a Map-Reply message
   when it receives the data packet over the ALT. Note that in this
   case, the inner Destination Address (DA), which is an EID, is
   copied to the outer DA and is routed over the ALT.

Map-Request: An ITR may also send a Map-Request message into the ALT
   to request the mapping. As in the Data Probe case, the
   authoritative ETR will respond to the ITR with a Map-Reply
   message. Since the ALT only forwards on EID destinations, the DA
   of the Map-Request sent in to the ALT MUST be an EID. See [LISP]
   for the format of Map-Request and Map-Reply packets.

ALT datagram: A Map-Request or Data Probe to be sent into or
   forwarded on the ALT.

4.1. Routeability of EIDs

As with LISP 1.0, EIDs are routable and can be used, unaltered, as
the source and destination addresses in IP datagrams. Unlike in LISP
1.0, LISP 1.5 EIDs are not routable on the public Internet; instead,
they are only routed over a separate, virtual topology referred to as
the LISP Alternative Virtual Network. This network is built as an
overlay on the public Internet using tunnels to interconnect LISP+ALT
routers. BGP is run over these tunnels to propagate the information
needed to route ALT datagrams. Importantly, while the ETRs are the
source(s) of the unaggregated EID prefix data, LISP+ALT uses existing
BGP mechanisms to aggressively aggregate this information. Note that
an ETR is not required to participate (or prevented from
participating) in LISP+ALT; an ETR may choose to communicate its
mappings to its serving LISP+ALT router(s) using subscription time
static configuration or through a dynamic mechanism such as that
described in [LISP-MS]. An ITR may similarly use a static EID
"default route" or other configuration as described in [LISP-MS] to
avoid the complexity of participating in the ALT.

4.2. Connectivity to non-LISP sites

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

4.3. Caveats on the use of Data Probes

   It is worth noting that there has been a great deal of discussion and
   controversy about whether Data Probes are a good idea. On the one
   hand, using them offers a method of avoiding the "first packet drop"
   problem when an ITR does not have a mapping for a particular EID-
   prefix. On the other hand, forwarding data packets on the ALT would
   require that it either be engineered to support relatively high
   traffic rates, which is not generally feasible for a tunneled
   network, or that it be carefully designed to aggressively rate- limit
   traffic to avoid congestion or DoS attacks. There are also other
   issues involving latency or other differences between the ALT path
   that initial a Data Probe would take and the path that subsequent
   packets on the same flow would take once a mapping were in place on
   an ITR. For these and other reasons use of Data Probes should be
   considered experimental and should be disabled by default in all ITR
   implementations.

5. **LISP+ALT: Overview**

   LISP+ALT is a hybrid push/pull architecture. Aggregated EID prefixes
   are "pushed" among the LISP+ALT routers and, optionally, out to ITRs
   (which may elect to receive the aggregated information, as opposed to
   simply using a default mapping). Specific EID-to-RLOC mappings are
   "pulled" by ITRs when they either send explicit LISP requests or data
   packets on the alternate topology that result in triggered replies
   being generated by ETRs.

   The basic idea embodied in LISP+ALT is to use BGP, running over
   tunneled overlay network, to establish reachability required to route
   ALT datagrams over an alternate logical topology (ALT). The ALT
   BGPRoute Information Base (RIB) is comprised of EID prefixes and
   associated next hops. LISP+ALT routers interconnect using eBGP and
   propagate EID prefix updates, which are learned over eBGP connections
   to authoritative ETRs, or by static configuration. ITRs may also
   eBGP peer with one or more LISP+ALT to learn the best ALT router to
   use to forward an ALT datagram for a particular prefix; in most
   cases, an ITR will have a default EID mapping pointing to one or more
   LISP+ALT routers.

   Note that while this document specifies the use of Generic Routing
   Encapsulation (GRE) as a tunneling mechanism, there is no reason that
   an ALT cannot be built using other tunneling technologies. In cases
   where GRE does not meet security, management, or other operational
   requirements, it is reasonable to use another tunneling technology
   that does. References to "GRE tunnel" in later sections of this
   document should therefore not be taken as prohibiting or precluding
   the use of other, available tunneling mechanisms. Note also that two
   LISP+ALT routers that are directly adjacent (with no layer-3 router
   hops between them) need not use a tunnel between them; in this case,
   BGP may be configured across the interfaces that connect to their
   common subnet and that subnet is considered to be part of the ALT
   topology. Use of techniques, such as "eBGP multihop", to forward ALT
   datagrams through routers that do not participate in ALT routing, is
   not recommended.

   In summary, LISP+ALT uses BGP to propagate EID-prefix update
   information to facilitate forwarding an ALT datagram to the ETR that
   holds the EID-to-RLOC mapping for that EID-prefix. This reachability
   is carried as IPv4 or IPv6 NLRI without modification (since an EID
   prefix has the same syntax as IPv4 or IPv6 address prefix). LISP+ALT
   routers eBGP peer with one another, forming the ALT. A LISP+ALT
   router near the edge learns EID prefixes originated by authoritative
   ETRs. This may be via eBGP with the ETRs, by static configuration,
   or through some other dynamic mechanism such as that defined in
   [LISP-MS]. A LISP+ALT router may also be configured to aggregate EID

prefixes received from ETRs or from other LISP+ALT routers that are
topologically "downstream" from it.

5.1. ITR traffic handling

When an ITR receives a packet originated by an end system within its
site (i.e. a host for which the ITR is the exit path out of the site)
and the destination for that packet is not known in the ITR's mapping
cache, the ITR encapsulates the packet in a LISP header, copying the
inner destination address (EID) to the outer destination address
(RLOC), and transmits it through a GRE tunnel to a LISP+ALT router in
the ALT (see also [LISP-MS] for non-ALT-connected ITRs, noting that
an ITR cannot send Data Probes to a Map-Server). This "first hop"
LISP+ALT router uses EID-prefix routing information learned from
other LISP+ALT routers via BGP to guide the packet to the ETR which
"owns" the prefix. Upon receipt by the ETR, normal LISP processing
occurs: the ETR responds to the ITR with a LISP Map-Reply that lists
the RLOCs (and, thus, the ETRs to use) for the EID prefix. The ETR
also de-encapsulates the packet and transmits it toward its
destination.

Upon receipt of the Map-Reply, the ITR installs the RLOC information
for a given prefix into a local mapping database. With these mapping
entries stored, additional packets destined to the given EID prefix
are routed directly to a viable ETR without use of the ALT, until
either the entry's TTL has expired, or the ITR can otherwise find no
reachable ETR. Note that a valid mapping (not timed-out) may exist
that contains no reachable RLOCs (i.e. all paths to that ETR are
down); in this case, packets destined to the EID prefix are dropped,
not routed through the ALT.

Traffic routed over the ALT therefore consists of:

o EID prefix Map-Requests, and

o data packets destined for those EID prefixes while the ITR awaits
   map replies

5.2. EID Assignment - Hierarchy and Topology

EID-prefixes will be allocated to a LISP site by Internet Registries.
Multiple allocations may not be in power-of-2 blocks. But when they
are, they will be aggregated into a single, advertised EID-prefix.
The ALT network is built in a tree-structured hierarchy to allow
aggregation at merge points in the tree. Building such a structure
should minimize the number of EID-prefixes carried by LISP+ALT nodes
near the top of the hierarchy.

Since the ALT will not need to change due to subscription or policy
reasons, the topology can remain relatively static and aggregation
can be sustained. Because routing on the ALT uses BGP, the same
rules apply for generating aggregates; in particular, a LISP+ALT
router should only be configured to generate an aggregate if it is
configured with BGP sessions to all of the originators of components
(more-specifics prefixes) of that aggregate; not all of the
components of need to be present for the aggregate to be originated
(some may be holes in the covering prefix and some may be down) but
the aggregating router must be configured to learn the state of all
of the components.

As an example, consider ETRs that are originating EID prefixes for
10.1.0.0/24, 10.1.64.0/24, 10.1.128.0/24, and 10.1.192.0/24. An ALT
router should only be configured to generate an aggregate for
10.1.0.0/16 if it has BGP sessions configured with all of these ETRs,
in other words, only if it has sufficient knowledge about the state
of those prefixes to summarize them.

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

This implies that two ALT routers that share an overlapping set of
prefixes must exchange those prefixes if either is to generate and
export a covering aggregate for those prefixes. It also implies that
an ETR which connects to the ALT using BGP must maintain BGP sessions
with all of the ALT routers that are configured to originate an
aggregate which covers that prefix. See also [LISP-MS] for an
example of other ways that prefix origin consistency and aggregation
are maintained.

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

5.3. LISP+ALT Router (or ALT router for short)

A LISP+ALT Router has the following functionality:

1. It runs, at a minimum, the eBGP part of the BGP protocol.

2. It supports a separate RIB which uses next-hop GRE tunnel
    interfaces for forwarding ALT datagrams.

   3. It can act as a "proxy-ITR" to support non-LISP sites.

   4. It can act as an ETR, or as a recursive or re-encapsulating ITR
       to reduce mapping tables in site-based LISP routers.

5.4. ITR and ETR in a LISP+ALT Environment

   An ITR using LISP+ALT may have additional functionality as follows:

   1. If it is also acting as a LISP+ALT Router, it sends ALT datagrams
       on the BGP best path computed GRE tunnel for each EID prefix.

   2. When acting solely as a ITR, it sends ALT datagrams directly to a
       configured LISP+ALT router.

   An ETR using LISP+ALT may also behave slightly differently:

   1. If it is also acting as a LISP+ALT router, it advertises its
       configured EID-prefixes into BGP for distribution through the
       ALT.

   2. It receives ALT datagrams only from its "upstream" LISP+ALT
       routers over the GRE tunnel(s) configured to it/them. It
       responds with Map-Replies for the EID prefixes that it "owns".

5.5. Use of GRE and BGP between LISP+ALT Routers

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

   LISP+ALT's use of GRE and BGP reduces provider Operational Expense
   (OPEX) because no new protocols need to be either defined or used on
   the overlay topology. Also, since tunnel IP addresses are local in
   scope, no coordination is needed for their assignment; any addressing
   scheme (including private addressing) can be used for tunnel
   addressing.

6. EID Prefix Propagation and Map-Request Forwarding

   As described in Section 9.2, an ITR may send either a Map-Request or
   a data probe to find a given EID-to-RLOC mapping. The ALT provides
   the infrastructure that allows these requests to reach the
   authoritative ETR.

   Note that, under normal circumstances, Map-Replies are not sent over
   the ALT - an ETR sends a Map-Reply to the source RLOC learned from
   the original Map-Request. There may be scenarios, perhaps to
   encourage caching of EID-to-RLOC mappings by ALT routers, where Map-
   Replies could be sent over the ALT or where a "first-hop" ALT router
   might modify the originating RLOC on a Map-Request received from an
   ITR to force the Map-Reply to be sent to it; these cases will not be
   supported by initial LISP+ALT implementations but may be subject to
   future experimentation.

   LISP+ALT routers propagate mapping information for use by ITRs (when
   sending ALT datagrams) using eBGP [RFC4271]. eBGP is run on the
   inter-LISP+ALT router links, and possibly between an edge ("last
   hop") LISP+ALT router and an ETR or between an edge ("first hop")
   LISP+ALT router and an ITR. The ALT eBGP RIB consists of aggregated
   EID prefixes and their next hops toward the authoritative ETR for
   that EID prefix.

6.1. Changes to ITR behavior with LISP+ALT

   When using LISP+ALT, an ITR sends ALT datagrams to one of its
   "upstream" LISP+ALT routers; these are sent only to obtain new EID-
   to-RLOC mappings - RLOC probe and cache TTL refresh Map-Requests are
   not sent on the ALT. As in basic LISP, it should use one of its
   RLOCs as the source address of these queries; it should explicitly
   not use a tunnel interface as the source address as doing so will
   cause replies to be forwarded over the tunneled topology and may be
   problematic if the tunnel interface address is not explicitly routed
   throughout the ALT. If the ITR is running BGP with the LISP+ALT
   router(s), it selects the appropriate LISP+ALT router based on the
   BGP information received. If it is not running BGP, it uses static
   configuration to select a LISP+ALT router; in the general case, this
   will effectively be an "EID-prefix default route".

6.2. Changes to ETR behavior with LISP+ALT

   If an ETR connects using BGP to one or more LISP+ALT router(s), it
   simply announces its EID-prefix to those LISP+ALT routers. Note that
   when an ETR generates a Map-Reply message to return to a querying
   ITR, it sends it to the ITR's source-RLOC (i.e., on the underlying
   Internet topology, not on the ALT; this avoids any latency penalty

(or "stretch") that might be incurred by routing over the ALT).

7. BGP configuration and protocol considerations

7.1. Autonomous System Numbers (ASNs) in LISP+ALT

   The primary use of BGP today is to define the global Internet routing
   topology in terms of its participants, known as Autonomous Systems.
   LISP+ALT specifies the use of BGP to create a global EID-to-RLOC
   mapping database which, while related to the global routing database,
   serves a very different purpose and is organized into a very
   different hierarchy. Because LISP+ALT does use BGP, however, it uses
   ASNs in the paths that are propagated among LISP+ALT routers. To
   avoid confusion, it needs to be stressed that that these LISP+ALT
   ASNs use a new numbering space that is unrelated to the ASNs used by
   the global routing system. Exactly how this new space will be
   assigned and managed will be determined during experimental
   deployment of LISP+ALT.

   Note that the LISP+ALT routers that make up the "core" of the ALT
   will not be associated with any existing core-Internet ASN because
   topology, hierarchy, and aggregation boundaries are completely
   separate from and independent of the global Internet routing system.

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

   As defined by this document, LISP+ALT may be implemented using BGP
   without modification. Given the fundamental operational difference
   between propagating global Internet routing information (the current,
   dominant use of BGP) and managing the global EID-to-RLOC database
   (the use of BGP proposed by this document), it may be desirable to
   assign a new SAFI [RFC2858] to prevent operational confusion and
   difficulties, including the inadvertent leaking of information from
   one domain to the other. At present, this document does not require
   the assignment of a new SAFI but the authors anticipate that
   experimentation may suggest the need for one in the future.

8. EID-Prefix Aggregation

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

Note that, as mentioned in Section 5.2, the use of BGP by LISP+ALT
requires that information can only be aggregated where all active
more-specific prefixes of a generated aggregate prefix are known.
This implies, for example, that if a given set of prefixes is used by
multiple, ALT networks, those networks must interconnect and share
information about all of the prefixes if either were to generate an
aggregate prefix that covered all of them. This is no different than
the way that BGP route aggregation works in the existing global
routing system: a service provider only generates an aggregate route
if it is configured to learn to all prefixes that make up that
aggregate.

8.1. Traffic engineering with LISP and LISP+ALT

It is worth noting that LISP+ALT does not directly propagate EID-to-
RLOC mappings. What it does is provide a mechanism for a LISP ITR to
find the ETR that holds the mapping for a particular EID prefix.
This distinction is important for several reasons. First, it means
that the reachability of RLOCs is learned through the LISP ITR-ETR
exchange so "flapping" of state information through BGP is not likely
nor can mapping information become "stale" by slow propagation
through the ALT BGP mesh. Second, by deferring EID-to-RLOC mapping
to an ITR-ETR exchange, it is possible to perform site-to-site
traffic engineering through a combination of setting the preference
and weight fields and by returning more-specific EID-to-RLOC
information in LISP Map-Reply messages. This is a powerful mechanism
that can conceivably replace the traditional practice of routing
prefix deaggregation for traffic engineering purposes. Rather than
propagating more-specific information into the global routing system
for local- or regional-optimization of traffic flows, such more-
specific information can be exchanged, through LISP (not LISP+ALT),
on an as-needed basis between only those ITRs/ETRs (and, thus, site
pairs) that need it; should a receiving ITR decide that it does not
wish to store such more-specific information, it has the option of
discarding it as long as a shorter, covering EID prefix exists. Not
only does this greatly improve the scalability of the global routing
system but it also allows improved traffic engineering techniques by
allowing richer and more fine-grained policies to be applied.

8.2. Edge aggregation and dampening

   Note also that normal BGP best common practices apply to the ALT
   network. In particular, first-hop ALT routers will aggregate EID
   prefixes and dampen changes to them in the face of excessive updates.
   Since EID prefix assignments are not expected to change with anywhere
   as frequently BGP prefix reachability on the Internet, such dampening
   should be very rare and might be worthy of logging as an exceptional
   event. It is again worth noting that the ALT carries only EID
   prefixes, along with BGP-generated paths to the ETRs that source
   those prefixes as advertisements travel over the logical topology;
   this set of information is considerablly less volatile than the
   actual EID-to-RLOC mappings.

9. Connecting sites to the ALT network

9.1. ETRs originating information into the ALT

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

   eBGP: An ETR usually participates in the LISP+ALT overlay network by
      running eBGP to one or more LISP+ALT router(s) over tunnel(s).
      The ETR advertises reachability for its EID prefixes over these
      eBGP connection(s). The LISP+ALT router(s) that receive(s) these
      prefixes then propagate(s) them into the ALT. Here the ETR is
      simply an eBGP peer of LISP+ALT router(s) at the edge of the ALT.
      Where possible, a LISP+ALT router that receives EID prefixes from
      an ETR via eBGP should aggregate that information.

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

   Note: in both cases, an ETR may have connections to to multiple
      LISP+ALT routers for the following reasons:

      * redundancy, so that a particular ETR is still reachable through
         the ALT even if one path or tunnel is unavailable.

      * to connect to different parts of the ALT hierarchy if the ETR
         "owns" multiple EID-to-RLOC mappings for EID prefixes that
         cannot be aggregated by the same LISP+ALT router (i.e. are not
         topologically "close" to each other in the ALT).

9.2. ITRs Using the ALT

   In order to source Map-Requests to the ALT or to route a Data Probe
   packet over the ALT, each ITR participating in the ALT establishes a
   connection to one or more LISP+ALT routers. These connections can be
   either eBGP or TCP (as described above).

   In the case in which the ITR is running eBGP, the peer LISP+ALT
   routers use these connections to advertise highly aggregated EID-
   prefixes to the peer ITRs. The ITR then installs the received
   prefixes into a forwarding table that is used to to send LISP Map-

Requests to the appropriate LISP+ALT router. In most cases, a LISP+
ALT router will send a default mapping to its client ITRs so that
they can send request for any EID prefix into the ALT.

In the case in which the ITR is connected to some set of LISP+ALT
routers without eBGP, the ITR sends Map-Requests to any of its
connected LISP+ALT routers.

An ITR may also choose to send the first few data packets over the
ALT to minimize packet loss and reduce mapping latency. In this
case, the data packet serves as a mapping probe (Data Probe) and the
ETR which receives the data packet (over the ALT) responds with a
Map-Reply is sent to the ITR's source-RLOC using the underlying
topology. Note that the use of Data Probes is discouraged at this
time (see Section 4.3).

In general, an ITR will establish connections only to LISP+ALT
routers at the "edge" of the ALT (typically two for redundancy) but
there may also be situations where an ITR would connect to other
LISP+ALT routers to receive additional, shorter path information
about a portion of the ALT of interest to it. This can be
accomplished by establishing GRE tunnels between the ITR and the set
of LISP+ALT routers with the additional information. This is a
purely local policy issue between the ITR and the LISP+ALT routers in
question.

10. IANA Considerations

   This document makes no request of the IANA.

11. Security Considerations

   LISP+ALT shares many of the security characteristics of BGP. Its
   security mechanisms are comprised of existing technologies in wide
   operational use today. Securing LISP+ALT is much simpler than
   securing BGP.

   Compared to BGP, LISP+ALT routers are not topologically bound,
   allowing them to be put in locations away from the vulnerable AS
   border (unlike eBGP speakers).

11.1. Apparent LISP+ALT Vulnerabilities

   This section briefly lists of the apparent vulnerabilities of LISP+
   ALT.

   Mapping Integrity: Can an attacker insert bogus mappings to black-
      hole (create a DoS) or intercept LISP data-plane packets?

   LISP+ALT router Availability: Can an attacker DoS the LISP+ALT
      routers connected to a given ETR? without access to its mappings,
      a site is essentially unavailable.

   ITR Mapping/Resources: Can an attacker force an ITR or LISP+ALT
      router to drop legitimate mapping requests by flooding it with
      random destinations that it will have to query for. Further study
      is required to see the impact of admission control on the overlay
      network.

   EID Map-Request Exploits for Reconnaissance: Can an attacker learn
      about a LISP destination sites' TE policy by sending legitimate
      mapping requests messages and then observing the RLOC mapping
      replies? Is this information useful in attacking or subverting
      peer relationships? Note that LISP 1.0 has a similar data-plane
      reconnaissance issue.

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

   UDP Map-Reply from ETR: Since Map-Replies packets are sent directly
      from the ETR to the ITR's RLOC, the ITR's RLOC may be vulnerable
      to various types of DoS attacks.

11.2. **Survey of LISP+ALT Security Mechanisms**

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

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

   Use of HMAC Protected TCP Connections: HMAC is used to verify
      message integrity and authenticity, making it nearly impossible
      for third party devices to either insert or modify messages.

   Message Sequence Numbers and Nonce Values in Messages: This allows
      for devices to verify that the mapping-reply packet was in
      response to the mapping-request that they sent.

11.3. **Using existing BGP Security mechanisms**

   LISP+ALT's use of BGP allows for the ALT to take advantage of BGP
   security features designed for existing Internet BGP use.

   For example, should either sBGP [I-D.murphy-bgp-secr] or soBGP
   [I-D.white-sobgparchitecture] become widely deployed it expected that
   LISP+ALT could use these mechanisms to provide authentication of EID-
   to-RLOC mappings, and EID origination.

12. Acknowledgments

   Many of the ideas described in this document were developed during
   detailed discussions with Scott Brim and Darrel Lewis, who made many
   insightful comments on earlier versions of this document. Additional
   thanks are due to Hannu Flinck and Amit Jain who offered many helpful
   suggestions for the -02 version.

13. References

13.1. Normative References

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

   [RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P.
              Traina, "Generic Routing Encapsulation (GRE)", RFC 2784,
              March 2000.

   [RFC2858] Bates, T., Rekhter, Y., Chandra, R., and D. Katz,
              "Multiprotocol Extensions for BGP-4", RFC 2858, June 2000.

   [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway
              Protocol 4 (BGP-4)", RFC 4271, January 2006.

   [RFC4632] Fuller, V. and T. Li, "Classless Inter-domain Routing
              (CIDR): The Internet Address Assignment and Aggregation
              Plan", BCP 122, RFC 4632, August 2006.

13.2. Informative References

   [I-D.murphy-bgp-secr]
              Murphy, S., "BGP Security Analysis",
              draft-murphy-bgp-secr-04 (work in progress),
              November 2001.

   [I-D.white-sobgparchitecture]
              White, R., "Architecture and Deployment Considerations for
              Secure Origin BGP (soBGP)",
              draft-white-sobgparchitecture-00 (work in progress),
              May 2004.

   [Interworking]
              Lewis, D., Meyer, D., Farinacci, D., and V. Fuller,
              "Interworking LISP with IPv4 and ipv6",
              draft-ietf-lisp-interworking-01.txt (work in progress),
              January 2010.

   [LISP]   Farinacci, D., Fuller, V., Meyer, D., and D. Lewis,
              "Locator/ID Separation Protocol (LISP)",
              draft-ietf-lisp-06.txt (work in progress), January 2010.

   [LISP-MS] Fuller, V. and D. Farinacci, "LISP Map Server",
              draft-ietf-lisp-ms-04.txt (work in progress),
              October 2009.

Authors' Addresses

   Vince Fuller
   Cisco
   Tasman Drive
   San Jose, CA 95134
   USA

   Email: vaf@cisco.com

   Dino Farinacci
   Cisco
   Tasman Drive
   San Jose, CA 95134
   USA

   Email: dino@cisco.com

   Dave Meyer
   Cisco
   Tasman Drive
   San Jose, CA 95134
   USA

   Email: dmm@cisco.com

   Darrel Lewis
   Cisco
   Tasman Drive
   San Jose, CA 95134
   USA

   Email: darlewis@cisco.com