Network Working Group
Internet-Draft
Intended status: Experimental
Expires: September 5, 2011

V. Fuller
D. Farinacci
D. Meyer
D. Lewis
Cisco

March 4, 2011

LISP Alternative Topology (LISP+ALT) draft-ietf-lisp-alt-06.txt

Abstract

This document describes a simple mapping database to be used by the Locator/ID Separation Protocol (LISP) to find Endpoint Identifier (EID) to Routing Locator (RLOC) mappings. Termed the Alternative Logical Topology (ALT), the database is built as an overlay network on the public Internet using the Border Gateway Protocol (BGP) and the Generic Routing Encapsulation (GRE). Using these proven protocols, the ALT can be built and deployed relatively quickly without major changes to the existing routing infrastructure.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 5, 2011.

Copyright Notice

Table of Contents

1. Introduction

This document describes the LISP+ALT mapping database, to be used by LISP to find EID-to-RLOC mappings. The ALT network is built using the Border Gateway Protocol (BGP, [RFC4271]), the BGP multi-protocol extension [RFC4760], and the Generic Routing Encapsulation (GRE, [RFC2784]) to construct an overlay network of devices (ALT Routers) which operate on EID-prefixes and use EIDs as forwarding destinations.

ALT Routers advertise hierarchically-delegated segments of the EID namespace (i.e., prefixes) toward the rest of the ALT; they also forward traffic destined for an EID covered by one of those prefixes toward the network element that is authoritative for that EID and is the origin of the BGP advertisement for that EID-prefix. An Ingress Tunnel Router (ITR) uses this overlay to send a LISP Map-Request (see [LISP]) to the Egress Tunnel Router (ETR) that holds the EID-to-RLOC mapping for a matching EID-prefix. In most cases, an ITR does not connect directly to the overlay network but instead sends Map-Requests via a Map-Resolver (MR; see [LISP-MS]) which does. Likewise, in most cases, an ETR does not connect directly to the overlay network but instead registers its EID-prefixes with a Map-Server that advertises those EID-prefixes on to the ALT and forwards Map-Requests for them to the ETR.

It is important to note that the ALT does not distribute actual EID-to-RLOC mappings. What it does provide is a forwarding path from an ITR (or MR) which requires an EID-to-RLOC mapping to an ETR which holds that mapping. The ITR/MR uses this path to send an ALT Datagram (see Section 3) to an ETR which then responds with a Map-Reply containing the needed mapping information.

One design goal for LISP+ALT is to use existing technology wherever possible. To this end, the ALT is intended to be built using off-the-shelf routers which already implement the required protocols (BGP and GRE); little, if any, LISP-specific modifications should be needed for such devices to be deployed on the ALT. Note, though, that organizational and operational considerations suggest that ALT Routers be both logically and physically separate from the "native" Internet packet transport system; deploying this overlay on those routers which are already participating in the global routing system and actively forwarding Internet traffic is not recommended.

The remainder of this document is organized as follows: Section 2 provides the definitions of terms used in this document. Section 3 outlines the basic LISP 1.5 model. Section 4 provides a basic overview of the LISP Alternate Topology architecture, and Section 5 describes how the ALT uses BGP to propagate Endpoint Identifier

reachability over the overlay network and Section 6 describes other considerations for using BGP on the ALT. Section 7 describes the construction of the ALT aggregation hierarchy, and Section 8 discusses how LISP+ALT elements are connected to form the overlay network.

2. 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.

Alternative Logical Topology (ALT): The virtual overlay network made up of tunnels between LISP+ALT Routers. The Border Gateway Protocol (BGP) runs between ALT Routers and is used to carry reachability information for EID-prefixes. The ALT provides a way to forward Map-Requests (and, if supported, Data Probes) toward the ETR that "owns" an EID-prefix. As a tunneled overlay, its performance is expected to be quite limited so use of it to forward high-bandwidth flows of Data Probes is strongly discouraged (see Section 3.3 for additional discussion).

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

ALT Router: The devices which run on the ALT. The ALT is a static network built using tunnels between ALT Routers. These routers are deployed in a roughly-hierarchical mesh in which routers at each level in the topology are responsible for aggregating EID-prefixes learned from those logically "below" them and advertising summary prefixes to those logically "above" them. Prefix learning and propagation between ALT Routers is done using BGP. An ALT Router at the lowest level, or "edge" of the ALT, learns EID-prefixes from its "client" ETRs. See Section 3.1 for a description of how EID-prefixes are learned at the "edge" of the ALT. See also Section 6 for details on how BGP is configured between the different network elements. When an ALT Router receives an ALT Datagram, it looks up the destination EID in its forwarding table (composed of EID prefix routes it learned from neighboring ALT Routers) and forwards it to the logical next-hop on the overlay network.

Endpoint ID (EID): A 32-bit (for IPv4) or 128-bit (for ipv6) value used to identify the ultimate source or destination for a LISP-encapsulated packet. See [LISP] for details.

EID-prefix: A set of EIDs delegated in a power-of-two block. EID-prefixes are routed on the ALT (not on the global Internet) and are expected to be assigned in a hierarchical manner such that they can be aggregated by ALT Routers. Such a block is characterized by a prefix and a length. Note that while the ALT routing system considers an EID-prefix to be an opaque block of EIDs, an end site may put site-local, topologically-relevant

structure (subnetting) into an EID-prefix for intra-site routing.

Aggregated EID-prefixes: A set of individual EID-prefixes that have been aggregated in the [RFC4632] sense.

Map Server (MS): An edge ALT Router that provides a registration function for non-ALT-connected ETRs, originates EID-prefixes into the ALT on behalf of those ETRs, and forwards Map-Requests to them. See [LISP-MS] for details.

Map Resolver (MR): An edge ALT Router that accepts an Encapsulated Map-Request from a non-ALT-connected ITR, decapsulates it, and forwards it on to the ALT toward the ETR which owns the requested EID-prefix. See [LISP-MS] for details.

Ingress Tunnel Router (ITR): A router which sends LISP Map-Requests or encapsulates IP datagrams with LISP headers, as defined in [LISP]. In this document, the term refers to any device implementing ITR functionality, including a Proxy-ITR (see [LISP-IW]). Under some circumstances, a LISP Map Resolver may also originate Map-Requests (see [LISP-MS]).

Egress Tunnel Router (ETR): A router which sends LISP Map-Replies in response to LISP Map-Requests and decapsulates LISP-encapsulated IP datagrams for delivery to end systems, as defined in [LISP]. In this document, the term refers to any device implementing ETR functionality, including a Proxy-ETR (see [LISP-IW]). Under some circumstances, a LISP Map Server may also respond to Map-Requests (see [LISP-MS]).

Routing Locator (RLOC): A routable IP address for a LISP tunnel router (ITR or ETR). Interchangeably referred to as a "locator" in this document. An RLOC is also the output of an EID-to-RLOC mapping lookup; an EID-prefix maps to one or more RLOCs. Typically, RLOCs are numbered from topologically-aggregatable blocks that are assigned to a site at each point where 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. Routing for RLOCs is not carried on the ALT.

EID-to-RLOC Mapping: A binding between an EID-prefix and the set of RLOCs that can be used to reach it; sometimes referred to simply as a "mapping".

EID-prefix Reachability: An EID-prefix is said to be "reachable" if at least one of its locators is reachable. That is, an EID-prefix is reachable if the ETR 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 can be learned by configuration or from a Map-Reply message.

ALT Default Route: An EID-prefix value of 0.0.0.0/0 (or 0::/0 for ipv6) which may be learned from the ALT or statically configured on an edge ALT Router. The ALT-Default Route defines a forwarding path for a packet to be sent into the ALT on a router which does not have a full ALT forwarding database.

3. The LISP+ALT model

The LISP+ALT model uses the same basic query/response protocol that is documented in [LISP]. In particular, LISP+ALT provides two types of packet that an ITR can originate to obtain EID-to-RLOC mappings:

Map-Request: A Map-Request message is sent into the ALT to request an EID-to-RLOC mapping. The ETR which owns the mapping will respond to the ITR with a Map-Reply message. Since the ALT only forwards on EID destinations, the destination address of the Map-Request sent on the ALT must be an EID. See [LISP] for the format of Map-Request and Map-Reply packets.

Data Probe: Alternatively, an ITR may encapsulate and send the first data packet destined for an EID with no known RLOCs into the ALT as a Data Probe. This might be done minimize packet loss and to probe for the mapping. As above, the authoritative ETR for the EID-prefix will respond to the ITR with a Map-Reply message when it receives the data packet over the ALT. As a side-effect, the encapsulated data packet is delivered to the end-system at the ETR site. Note that the Data Probe's inner IP destination address, which is an EID, is copied to the outer IP destination address so that the resulting packet can be routed over the ALT. See Section 3.3 for caveats on the usability of Data Probes.

The term "ALT Datagram" is short-hand for a Map-Request or Data Probe to be sent into or forwarded on the ALT. Note that while the outer header Source Address of an ALT Datagram is currently expected to be an RLOC, there may be situations (e.g. for experimentation with caching in intermediate ALT nodes) where an EID would be used to force a Map-Reply to be routed back through the ALT.

3.1. Routeability of EIDs

A LISP EID has the same syntax as IP address and can be used, unaltered, as the source or destination of an IP datagram. In general, though, EIDs are not routable on the public Internet; LISP+ALT provides a separate, virtual network, known as the LISP Alternative Logical Topology (ALT) on which a datagram using an EID as an IP destination address may be transmitted. This network is built as an overlay on the public Internet using tunnels to interconnect ALT Routers. BGP runs over these tunnels to propagate path information needed to forward ALT Datagrams. Importantly, while the ETRs are the source(s) of the unaggregated EID-prefixes, LISP+ALT uses existing BGP mechanisms to aggregate this information.

3.1.1. Mechanisms for an ETR to originate EID-prefixes

There are three ways that an ETR may originate its mappings into the ALT:

1. By registration with a Map Server as documented in [LISP-MS]. This is the common case and is expected to be used by the majority of ETRs.

2. Using a "static route" on the ALT. Where no Map-Server is available, an edge ALT Router may be configured with a "static EID-prefix route" pointing to an ETR.

3. Edge connection to the ALT. If a site requires fine- grained control over how its EID-prefixes are advertised into the ALT, it may configure its ETR(s) with tunnel and BGP connections to edge ALT Routers.

3.1.2. Mechanisms for an ITR to forward to EID-prefixes

There are three ways that an ITR may send ALT Datagrams:

1. Through a Map Resolver as documented in [LISP-MS]. This is the common case and is expected to be used by the majority of ITRs.

2. Using a "default route". Where a Map Resolver is not available, an ITR may be configured with a static ALT Default Route pointing to an edge ALT Router.

3. Edge connection to the ALT. If a site requires fine-grained knowledge of what prefixes exist on the ALT, it may configure its ITR(s) with tunnel and BGP connections to edge ALT Routers.

3.1.3. Map Server Model preferred

The ALT-connected ITR and ETR cases are expected to be rare, as the Map Server/Map Resolver model is both simpler for an ITR/ETR operator to use, and provides a more general service interface to not only the ALT, but also to other mapping databases that may be developed in the future.

3.2. 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 [LISP-IW] for

details.

3.3. 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 may also be issues caused by different latency or other performance characteristics between the ALT path taken by an initial Data Probe and the "Internet" path taken by subsequent packets on the same flow once a mapping is in place on an ITR. For these reasons, the use of Data Probes is not recommended at this time; they should only be originated an ITR when explicitly configured to do so and such configuration should only be enabled when performing experiments intended to test the viability of using Data Probes.

4. LISP+ALT: Overview

LISP+ALT is a hybrid push/pull architecture. Aggregated EID-prefixes are advertised among the ALT Routers and to those (rare) ITRs that are directly connected via a tunnel and BGP to the ALT. Specific EID-to-RLOC mappings are requested by an ITR (and returned by an ETR) using LISP when it sends a request either via a Map Resolver or to an edge ALT Router.

The basic idea embodied in LISP+ALT is to use BGP, running on a tunneled overlay network (the ALT), to establish reachability between ALT Routers. The ALT BGP Route Information Base (RIB) is comprised of EID-prefixes and associated next hops. ALT Routers interconnect using BGP and propagate EID-prefix updates among themselves. EID-prefix information is learned from ETRs at the "edge" of the ALT either through the use of the Map Server interface (the commmon case), static configuration, or by BGP-speaking ETRs.

An ITR uses the ALT to learn the best path for forwarding an ALT Datagram destined to a particular EID-prefix. An ITR will normally use a Map Resolver to send its ALT Datagrams on to the ALT but may, in unusual circumstances, use a static ALT Default Route or connect to the ALT using BGP.

Note that while this document specifies the use of Generic Routing Encapsulation (GRE) as a tunneling mechanism, there is no reason that parts of the ALT cannot be built using other tunneling technologies, particularly in cases where GRE does not meet security, management, or other operational requirements. References to "GRE tunnel" in later sections of this document should therefore not be taken as prohibiting or precluding the use of other tunneling mechanisms. Note also that two 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 then considered to be part of the ALT topology. Use of techniques such as "eBGP multihop" to connect ALT Routers that do not share a tunnel or common subnet is not recommended as the non-ALT Routers in between the ALT Routers in such a configuration may not have information necessary to forward ALT Datagrams destined to EID-prefixes exchanged across that BGP session.

In summary, LISP+ALT uses BGP to build paths through ALT Routers so that an ALT Datagram sent into the ALT can be forwarded 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). ALT Routers establish BGP sessions with one another,

forming the ALT. An ALT Router at the "edge" of the topology learns EID-prefixes originated by authoritative ETRs. Learning may be though the Map Server interface, by static configuration, or via BGP with the ETRs. An 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.

4.1. 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 EID for that packet is not known in the ITR's mapping cache, the ITR creates either a Map-Request for the destination EID or the original packet encapsulated as a Data Probe (see Section 3.3 for caveats on the usability of Data Probes). The result, known as an ALT Datagram, is then sent to an ALT Router (see also [LISP-MS] for non-ALT-connected ITRs, noting that Data Probes cannot be sent to a Map-Resolver). This "first hop" ALT Router uses EID-prefix routing information learned from other 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. For Data Probes, the ETR also decapsulates 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 an RLOC 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 current mapping may exist that contains no reachable RLOCs; this is known as a Negative Cache Entry and it indicates that packets destined to the EID-prefix are to be dropped.

Full details on Map-Request/Map-Reply processing may be found in [LISP].

Traffic routed on to the ALT consists solely of ALT Datagrams, i.e. Map-Requests and Data Probes (if supported). Given the relatively low performance expected of a tuneled topology, ALT Routers (and Map Resolvers) should aggressively rate-limit the ingress of ALT Datagrams from ITRs and, if possible, should be configured to not accept packets that are not ALT Datagrams.

4.2. EID Assignment - Hierarchy and Topology

EID-prefixes are expected to be allocated to a LISP site by Internet Registries. Where a site has multiple allocations which are aligned

on a power-of-2 block boundary, they should be aggregated into a single EID-prefix for advertisement. The ALT network is built in a roughly hierarchical, partial mesh which is intended to allow aggregation where clearly-defined hierarchical boundaries exist. Building such a structure should minimize the number of EID-prefixes carried by LISP+ALT nodes near the top of the hierarchy.

Routes on the ALT do not need to respond to changes in policy, subscription, or underlying physical connectivity, so 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 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-specific 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.

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.

An ALT Router must not generate an aggregate that includes a non-LISP-speaking hole unless it can be configured to return a Negative Map-Reply with action="Natively-Forward" (see [LISP]) if it receives an ALT Datagram that matches that hole. If it receives an ALT Datagram that matches a LISP-speaking hole that is currently not reachable, it should return a Negative Map-Reply with action="drop". Negative Map-Replies should be returned with a short TTL, as specified in [LISP-MS]. Note that an off-the-shelf, non-LISP-speaking router configured as an aggregating ALT Router cannot send Negative Map-Replies, so such a router must never originate an aggregate that includes a non-LISP-speaking hole.

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 and that each of those ALT Routers must be explicitly configured to know the set of EID-prefixes that make up any aggregate that it originates. See also [LISP-MS] for an example of other ways that prefix origin consistency and aggregation can be maintained.

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. If the Router originating 10.1.0.0/16 receives an ALT Datagram destined for 10.1.77.88, a non-LISP destination covered by the aggregate, it returns a Negative Map-Reply with action "Natively-Forward". If it receives an ALT Datagram destined for 10.1.128.199 but the configured LISP prefix 10.1.128.0/24 is unreachable, it returns a Negative Map-Reply with action "drop".

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.

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

The ALT network is built using GRE tunnels between ALT Routers. BGP sessions are configured over those tunnels, with each 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 facilities deployment and operation of LISP because no new protocols need to be defined, implemented, or used on the overlay topology; existing BGP/GRE tools and operational expertise are also re-used. Tunnel address assignment is also easy: since the addresses on an ALT tunnel are only used by the pair of routers connected to the tunnel, the only requirement of the IP addresses used to establish that tunnel is that the attached routers be reachable by each other; any addressing plan, including private addressing, can therefore be used for ALT tunnels.

5. EID-prefix Propagation and Map-Request Forwarding

As described in Section 8.2, an ITR sends an ALT Datagram to 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 one of the ITR RLOCs 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 returned to the "first-hop" ALT Router. These cases will not be supported by initial LISP+ALT implementations but may be subject to future experimentation.

ALT Routers propagate path information via BGP ([RFC4271]) that is used by ITRs to send ALT Datagrams toward the appropriate ETR for each EID-prefix. BGP is run on the inter-ALT Router links, and possibly between an edge ("last hop") ALT Router and an ETR or between an edge ("first hop") ALT Router and an ITR. The ALT BGP RIB consists of aggregated EID-prefixes and their next hops toward the authoritative ETR for that EID-prefix.

5.1. Changes to ITR behavior with LISP+ALT

As previously described, an ITR will usually use the Map Resolver interface and will send its Map Requests to a Map Resolver. When an ITR instead connects via tunnels and BGP to the ALT, it sends ALT Datagrams to one of its "upstream" 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 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 routed throughout the ALT. If the ITR is running BGP with the LISP+ALT router(s), it selects the appropriate ALT Router based on the BGP information received. If it is not running BGP, it uses a statically-configued ALT Default Route to select an ALT Router.

5.2. Changes to ETR behavior with LISP+ALT

As previously described, an ETR will usually use the Map Server interface (see [LISP-MS]) and will register its EID-prefixes with its configured Map Servers. When an ETR instead connects using BGP to one or more ALT Routers, it announces its EID-prefix(es) to those ALT Routers.

As documented in [LISP], when an ETR generates a Map-Reply message to return to a querying ITR, it sets the outer header IP destination address to one of the requesting ITR's RLOCs so that the Map-Reply will be sent on the underlying Internet topology, not on the ALT; this avoids any latency penalty (or "stretch") that might be incurred by sending the Map-Reply via the ALT, reduces load on the ALT, and ensures that the Map-Reply can be routed even if the original ITR does not have an ALT-routed EID. For details on how an ETR selects which ITR RLOC to use, see section 6.1.5 of [LISP].

6. BGP configuration and protocol considerations

6.1. 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 overlay network (the ALT) for finding EID-to-RLOC mappings. While related to the global routing database, the ALT 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 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 the deployment of LISP+ALT.

Note that the ALT Routers that make up the "core" of the ALT will not be associated with any existing core-Internet ASN because the ALT topology is completely separate from, and independent of, the global Internet routing system.

6.2. 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 creating an overlay network for finding EID-to-RLOC mappings (the use of BGP proposed by this document), it may be desirable to assign a new SAFI [RFC4760] to prevent operational confusion and difficulties, including the inadvertent leaking of information from one domain to the other. Use of a separate SAFI would make it easier to debug many operational problems but would come at a significant cost: unmodified, off-the-shelf routers which do not understand the new SAFI could not be used to build any part of the ALT network. At present, this document does not request the assignment of a new SAFI; additional experimentation may suggest the need for one in the future.

7. 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 Internet will provide alternative routes to maintain BGP connectivity among ALT Routers.

Note that, as mentioned in Section 4.2, the use of BGP by LISP+ALT requires that information only be aggregated where all active more-specific prefixes of a generated aggregate prefix are known. 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.

7.1. Stability of the ALT

It is worth noting that LISP+ALT does not directly propagate EID-to-RLOC mappings. What it does is provide a mechanism for an ITR to commonicate with the ETR that holds the mapping for a particular EID-prefix. This distinction is important when considering the stability of BGP on the ALT network as compared to the global routing system. It also has implications for how site-specific EID-prefix information may be used by LISP but not propagated by LISP+ALT (see Section 7.2 below).

RLOC prefixes are not propagated through the ALT so their reachability is not determined through use of LISP+ALT. Instead, reachability of RLOCs is learned through the LISP ITR-ETR exchange. This means that link failures or other service disruptions that may cause the reachability of an RLOC to change are not known to the ALT. Changes to the presence of an EID-prefix on the ALT occur much less frequently: only at subscription time or in the event of a failure of the ALT infrastructure itself. This means that "flapping" (frequent BGP updates and withdrawals due to prefix state changes) is not likely and mapping information cannot become "stale" due to slow propagation through the ALT BGP mesh.

7.2. Traffic engineering using LISP

Since an ITR learns an EID-to-RLOC mapping directly from the ETR that owns it, it is possible to perform site-to-site traffic engineering by setting the preference and/or weight fields, and by including more-specific EID-to-RLOC information in 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. Such an exchange of "more-specifics" between sites facilitates traffic engineering, by allowing richer and more fine-grained policies to be applied without advertising additional prefixes into either the ALT or the global routing system.

Note that these new traffic engineering capabilities are an attribute of LISP and are not specific to LISP+ALT; discussion is included here because the BGP-based global routing system has traditionally used propagation of more-specific routes as a crude form of traffic engineering.

7.3. Edge aggregation and dampening

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 as frequently as global routing BGP prefix reachability, 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, used to a construct BGP path to each ETR (or Map-Server) that originates each prefix; the ALT does not carry reachability about RLOCs. In addition, EID-prefix information may be aggregated as the topology and address assignment hierarchy allow. Since the topology is all tunneled and can be modified as needed, reasonably good aggregation should be possible. In addition, since most ETRs are expected to connect to the ALT using the Map Server interface, Map Servers will implement a natural "edge" for the ALT where dampening and aggregation can be applied. For these reasons, the set of prefix information on the ALT can be expected to be both better aggregated and considerably less volatile than the actual EID-to-RLOC mappings.

7.4. EID assignment flexibility vs. ALT scaling

There are major open questions regarding how the ALT will be deployed and what organization(s) will operate it. In a simple, non-distributed world, centralized administration of EID prefix assignment and ALT network design would facilitate a well- aggregated ALT routing system. Business and other realities will likely result in a more complex, distributed system involving multiple levels of

prefix delegation, multiple operators of parts of the ALT infrastructure, and a combination of competition and cooperation among the participants. In addition, re-use of existing IP address assignments, both "PI" and "PA", to avoid renumbering when sites transition to LISP will further complicate the processes of building and operating the ALT.

A number of conflicting considerations need to be kept in mind when designing and building the ALT. Among them are:

1. Target ALT routing state size and level of aggregation. As described in Section 7.1, the ALT should not suffer from some of the performance constraints or stability issues as the Internet global routing system, so some reasonable level of deaggregation and increased number of EID prefixes beyond what might be considered ideal should be acceptable. That said, measures, such as tunnel rehoming to preserve aggregation when sites move from one mapping provider to another and implementing aggregation at multiple levels in the hierarchy to collapse de-aggregation at lower levels, should be taken to reduce unnecessary explosion of ALT routing state.

2. Number of operators of parts of the ALT and how they will be organized (hierarchical delegation vs. shared administration). This will determine not only how EID prefixes are assigned but also how tunnels are configured and how EID prefixes can be aggregated between different parts of the ALT.

3. Number of connections between different parts of the ALT. Trade-offs will need to be made among resilience, performance, and placement of aggregation boundaries.

4. EID prefix portability between competing operators of the ALT infrastructure. A significant benefit for an end-site to adopt LISP is the availability of EID space that is not tied to a specific connectivity provider; it is important to ensure that an end site doesn't trade lock-in to a connectivity provider for lock-in to a provider of its EID assignment, ALT connectivity, or Map Server facilities.

This is, by no means, and exhaustive list.

While resolving these issues is beyond the scope of this document, the authors recommend that existing distributed resource structures, such as the IANA/Regional Internet Registries and the ICANN/Domain Registrar, be carefully considered when designing and deploying the ALT infrastructure.

8. Connecting sites to the ALT network

8.1. ETRs originating information into the ALT

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

Map Server: In most cases, a site will configure its ETR(s) to register with one or more Map Servers (see [LISP-MS]), and does not participate directly in the ALT.

BGP: For a site requiring complex control over their EID-prefix origination into the ALT, an ETR may connect to the LISP+ALT overlay network by running BGP to one or more ALT Router(s) over tunnel(s). The ETR advertises reachability for its EID-prefixes over these BGP connection(s). The edge ALT Router(s) that receive(s) these prefixes then propagate(s) them into the ALT. Here the ETR is simply an BGP peer of ALT Router(s) at the edge of the ALT. Where possible, an ALT Router that receives EID-prefixes from an ETR via BGP should aggregate that information.

Configuration: One or more 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 ALT Router(s) using GRE tunnel(s) but rather than BGP being used, the 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: in all cases, an ETR may register to multiple Map Servers or connect to multiple ALT Routers for the following reasons:

* redundancy, so that a particular ETR is still reachable 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 ALT Router (i.e. are not topologically "close" to each other in the ALT).

8.2. ITRs Using the ALT

In the common configuration, an ITR does not need to know anything about the ALT, since it sends Map-Requests to one of its configured Map-Resolvers (see [LISP-MS]). There are two exceptional cases:

Static default: If a Map Resolver is not available but an ITR is adjacent to an ALT Router (either over a common subnet or through the use of a tunnel), it can use an ALT Default Route route to cause all ALT Datagrams to be sent that ALT Router. This case is expected to be rare.

Connection to ALT: A site with complex Internet connectivity needs may need more fine-grained distinction between traffic to LISP-capable and non-LISP-capable sites. Such a site may configure each of its ITRs to connect directly to the ALT, using a tunnel and BGP connection. In this case, the ITR will receive EID-prefix routes from its BGP connection to the ALT Router and will LISP-encapsulate and send ALT Datagrams through the tunnel to the ALT Router. Traffic to other destinations may be forwarded (without LISP encapsulation) to non-LISP next-hop routers that the ITR knows.

In general, an ITR that connects to the ALT does so only to to ALT Routers at the "edge" of the ALT (typically two for redundancy). There may, though, be situations where an ITR would connect to other 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 ALT Routers with the additional information. This is a purely local policy issue between the ITR and the ALT Routers in question.

As described in [LISP-MS], Map-Resolvers do not accept or forward Data Probes; in the rare scenario that an ITR does support and originate Data Probes, it must do so using one of the exceptional configurations described above. Note that the use of Data Probes is discouraged at this time (see Section 3.3).

9. IANA Considerations

This document makes no request of the IANA.

10. 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, so securing the ALT should be mostly a matter of applying the same technology that is used to secure the BGP-based global routing system (see Section 10.3 below).

10.1. Apparent LISP+ALT Vulnerabilities

This section briefly lists the known potential vulnerabilities of LISP+ALT.

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

ALT Router Availability: Can an attacker DoS the ALT Routers connected to a given ETR? If a site's ETR cannot advertise its EID-to-RLOC mappings, the site is essentially unavailable.

ITR Mapping/Resources: Can an attacker force an ITR or ALT Router to drop legitimate mapping requests by flooding it with random destinations for which it will generate large numbers of Map-Requests and fill its mapping cache? 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 site's TE policy by sending legitimate mapping requests and then observing the RLOC mapping replies? Is this information useful in attacking or subverting peer relationships? Note that any public LISP mapping database will have similar data-plane reconnaissance issue.

Scaling of 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 are sent directly from the ETR to the ITR's RLOC, the ITR's RLOC may be vulnerable to various types of DoS attacks (this is a general property of LISP, not an LISP+ALT vulnerability).

More-specific prefix leakage: Because EID-prefixes on the ALT are expected to be fairly well-aggregated and EID-prefixes propagated out to the global Internet (see [LISP-IW] much more so, accidental leaking or malicious advertisement of an EID-prefix into the global routing system could cause traffic redirection away from a LISP site. This is not really a new problem, though, and its solution can only be achieved by much more strict prefix filtering and authentication on the global routing system.

10.2. 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 BGP/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 an ITR to verify that the Map-Reply from an ETR is in response to a Map-Request originated by that ITR (this is a general property of LISP; LISP+ALT does not change this behavior).

10.3. Use of new IETF standard BGP Security mechanisms

LISP+ALT's use of BGP allows the ALT to take advantage of BGP security features designed for existing Internet BGP use. Should the Internet community converge on the work currently being done in the IETF SIDR working group or should either S-BGP [I-D.murphy-bgp-secr] or soBGP [I-D.white-sobgparchitecture] be implemented and widely-deployed, LISP+ALT can readily use these mechanisms to provide authentication of EID-prefix origination and EID-to-RLOC mappings.

11. Acknowledgments

The authors would like to specially thank J. Noel Chiappa who was a key contributer to the design of the LISP-CONS mapping database (many ideas from which made their way into LISP+ALT) and who has continued to provide invaluable insight as the LISP effort has evolved. Others who have provided valuable contributions include John Zwiebel, Hannu Flinck, Amit Jain, John Scudder, and Scott Brim.

12. References

12.1. Normative References

[LISP] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "Locator/ID Separation Protocol (LISP)", draft-ietf-lisp-10.txt (work in progress), March 2011.

[LISP-MS] Fuller, V. and D. Farinacci, "LISP Map Server", draft-ietf-lisp-ms-07.txt (work in progress), March 2011.

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

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

12.2. 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.

[LISP-IW] Lewis, D., Meyer, D., Farinacci, D., and V. Fuller, "Interworking LISP with IPv4 and ipv6", draft-ietf-lisp-interworking-02.txt (work in progress), March 2011.

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