A Unified LISP Mapping Database for L2 and L3 Network Virtualization Overlays
draft-hertoghs-nvo3-lisp-controlplane-unified-02

Abstract

   The purpose of this draft is to document how the Locator/ID
   Separation Protocol (LISP) Control Plane can be used to offer a
   unified (offering L2 AND L3) Overlay solution that matches the
   framework and requirements of Network Virtualization over L3 (NVO3).
   This information is provided as input to the NVO3 analysis of the
   suitability of existing IETF protocols to the NVO3 requirements.

Requirements Language

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in RFC 2119 [RFC2119].

Status of this Memo

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

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

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

   This Internet-Draft will expire on January 20, 2015.

Table of Contents

[1](#). **Introduction**

The purpose of this draft is to provide a mapping between the Network
Virtualization over L3 (NVO3) framework [I-D.ietf-nvo3-framework] and
the Locator/ID Separation Protocol (LISP) [RFC6830], and in
particular how LISP components map to the reference models defined
therein. This document extends the scope of[I-D.maino-nvo3-lisp-cp]
to cover the case of a unified overlay that includes L2 and L3
services.

LISP is a flexible map and encap framework that can be used for
overlay network applications, including Data Center Network
Virtualization. The LISP framework provides two main tools for NVO3:

1. A Data Plane that specifies how Endpoint Identifiers (EIDs) are
    encapsulated in Routing Locators (RLOCs), and

2. A Control Plane that specifies the interfaces to the LISP Mapping
    System [RFC6833]. The LISP Mapping system provides the mapping
    between EIDs and RLOCs.

LISP can be leveraged to offer services to both Physical and Virtual
endpoints, and is architecturally EID address family agnostic. Some
flows will be across an L3 overlay (using IP addresses as EIDs), and
other flows will be across an L2 overlay (using MAC addresses as
EIDs).

If certain requirements are met within the architecture, the choice
of whether a flow is sent across the L2 overlay versus across the L3
overlay is not mapped one to one to the choice between intra subnet
(bridging) and inter subnet (routing) forwarding. This allows the
network administrator to offer infrastructure spanning subnets to its
tenants, while not forcing them to deploy infrastructure-wide
broadcast domains.

This document will focus on how to offer a unified L2 and L3 overlay,
where both L2 and L3 services can be offered to the tenant's traffic
simultaneously.

The draft will elaborate on achieving multi homing of Tenant Systems
(TS), as well as optimal routing and forwarding, including how VM
mobility is achieved and optimal traffic forwarding is achieved.

Although the LISP specification uses a specific data plane, its
control plane can be decoupled fairly easily from the data plane and
thus can support various data plane encapsulations. After describing
the various data plane options, this document also addresses the NVO3
data plane requirements[I-D.ietf-nvo3-dataplane-requirements].

The document continues to lay out how the NVO3 control plane
requirements [I-D.ietf-nvo3-nve-nva-cp-req] are addressed.

Finally this document will provide security considerations in Section
5

2. Definition of Terms

Flood-and-Learn: the use of dynamic (data plane) learning in VXLAN
to discover the location of a given Ethernet/IEEE 802 MAC address
in the underlay network.

For definition of NVO3 related terms, notably Tenant System (TS),
Virtual Network (VN), Virtual Network Identifier (VNI), Network
Virtualization Edge (NVE), Network Virtualization Authority (NVA),
Data Center (DC), please consult [I-D.ietf-nvo3-framework].

For definitions of LISP related terms, notably Map-Request, Map-
Reply, Ingress Tunnel Router (ITR), Egress Tunnel Router (ETR),
Endstation IDentifier (EID), Routing LOCator (RLOC), Map-Server (MS)
and Map-Resolver (MR) please consult the LISP specification
[RFC6830].

3. NVO3 Framework and LISP

3.1. NVO3 Generic Reference Model

[I-D.maino-nvo3-lisp-cp] provides a mapping of the NVO3 generic
reference model [I-D.ietf-nvo3-framework] onto the LISP architecture.
In this document we will focus on the unified L2/L3 LISP control
plane, and on how it will optimize forwarding .

3.2. NVE Reference Model

The LISP NVE Reference Model is described in [I-D.maino-nvo3-lisp-
cp]. This section will look at the different types of NVEs: L2-only,
L3-only, and unified L2/L3, as well as looking at VM Mobility, Multi-
homing, optimal forwarding and external connectivity aspects. How
TSes connect to the network is described in Section 3.4.3.

3.2.1. Types of NVE's

[RFC6830] is defined as a L3 NVE, and it can be enhanced to support
L2 NVEs.

The separation of the L2 NVE and L3 NVE functions can be based on the
nature of the packets: bridged packets are assigned to the L2 NVE
function, while routed packets are assigned to the L3 NVE function.
Therefore these discrete functions could live on discrete networking
nodes.

However, it is possible to use LISP as an unified control plane, that
combines and co-locates the function of L2/L3 NVE onto a single node.
The network administrator can choose which traffic will be forwarded
across each service type.

3.2.1.1. L2 only NVE

[I-D.smith-lisp-layer2] describes an encapsulation method for
carrying Ethernet and IEEE 802 media access control (MAC) frames
within the Locator/ID Separation Protocol (LISP). As described in
[I-D.maino-nvo3-lisp-cp] MAC addresses are used as EIDs in an L2 only
NVE. As control plane learning is used, only broadcast and multicast
traffic needs mult-destination support from the underlay.

The frame format defined in [I-D.mahalingam-dutt-dcops-vxlan], has a
header compatible with the LISP data path encapsulation header, when
MAC addresses are used as EIDs, as described in section 4.12.2 of
[I-D.ietf-lisp-lcaf].

The LISP control plane is extensible, and can support non-LISP data
path encapsulations such as NVGRE [I-D.sridharan-virtualization-
nvgre], or other encapsulations that provide support for network
virtualization.

3.2.1.2. L3 only NVE

LISP is defined as a virtualized IP routing and forwarding service in
[RFC6830], and as such can be used to provide L3 NVE services.

3.2.1.3. Unified L2/L3 NVE

When using a unified L2/L3 NVE, IP EIDs are registered to the LISP
mapping system with the MAC Address of the Tenant System (TS) as an
additional RLOC (next to the NVE RLOC), through the methods defined
in [I-D.ietf-lisp-lcaf], by encoding Key/Value Pairs. MAC Address
based EIDs will also be registered for TSes that are transmitting
non-IP based traffic. TSes that send out both IP and non-IP traffic
will therefore be registered twice. For the L2 overlay the Virtual
Networking Instance (VNI)/IID denotes a network-wide bridge domain,
while for the L3 overlay the VNI/IID denotes a Virtual Routing
Forwarding (VRF) instance.

Implementing an NVE with a unified L2 and L3 overlay support is
beneficial for multiple reasons:

Primarily it allows the network administrator to choose which traffic
traverses the L2 overlay versus the L3 overlay, not only on the basis
of intra-subnet (bridged) versus inter-subnet (routed) traffic flows.
As a matter of fact, it is highly beneficial to choose a policy where
all IP traffic is forwarded across the L3 overlay (i.e. both intra-
and inter-subnet traffic). Effectively this allows the 'spread' of
subnets across the Datacenter(s) without leading to network wide
broadcast and associated failure domains, while allowing free
mobility of the end-stations.

Secondarily, when all the TS IP and MAC addresses are registered with
the NVA/LISP Mapping system, optimisations in ARP and ND [RFC4861]
forwarding and handling can be achieved. ARPs and IPv6 NDs for
'unknown' destinations are by default dropped, although a policy can
allow for 'unknown' ARP/ND packets to be flooded across the underlay
if so desired by the operator (e.g. when there is a desire to
support 'silent hosts').

Finally, as all IP traffic is forwarded across a L3 overlay, and ARP/
ND operations do not need flooding services, the amount of traffic
that needs to traverse the L2 overlay is limited to non-IP traffic
only. This makes the registration of MAC-addresses as EIDs with the
LISP control plane optional. The system in this case could use
ingress replication and Flood-and-Learn to handle the non-IP traffic.
Of course, the use of the LISP control plane for MAC address based
EIDs is another option as well, but the choice is left to the network
administrator.

However, in order to achieve the benefits of this model, there are
some requirements of how TSs can connect to the unified L2/L3 NVE,
and there are also requirements on how default gateway MAC/IP
addresses are assigned to the NVE function, and how forwarding is
done on the NVE function:

o The NVE MUST always do an IP lookup for IP based traffic,
   independent of whether the destination is within the same subnet
   or not, or whether the destination TS is attached to the same VLAN
   or L2 NVI as the source TS.

o The unified L2/L3 NVE NVI instance MUST have a uniform default
   gateway MAC-address and IP address across the entire NVO3 network.
   This is to make sure that a TS can always reach its default
   gateway, irrespective of location.

o A TS can connect across a L2 switched network to a unified L2/L3
   NVE, but traffic forwarded MUST follow a simple rule, where all
   traffic from a TS MUST always be sent upstream to the unified L2/

L3 NVE, regardless of its destination MAC address, and traffic
from locally attached TS's MUST be switched through the NVE.
Directly connecting a TS to a unified L2/L3 NVE automatically
solves that requirement.

There are various options to provide unified L2 and L3 support for
LISP in the data path.

[I-D.smith-lisp-layer2] extends LISP to support MAC addresses as
EIDs, and can be used in combination with [RFC6830], using the
destination UDP port in the outer LISP header for disambiguation.

Recently extensions to both LISP and VXLAN have been proposed to
offer multiprotocol support across the same outer header format (i.e.
using a single UDP port number), as described in [I-D.lewis-lisp-
gpe], and [I-D.quinn-vxlan-gpe] respectively. It is to be noted that
some functionality offered by native LISP is no longer available when
using the [I-D.lewis-lisp-gpe]extension (namely nonce, echo-nonce,
and map-versioning). These are optional control plane optimizations
implemented in the data plane for [RFC6830], and their use is less
relevant in DC applications.

The remainder of this document assumes a unified L2/L3 NVE deployment
model.

3.2.2. Multihoming aspects

If the TSes are co-located with the xTR/NVE function, no support for
multi-homing is needed.

If the network between the L2 device connecting the TSes and the LISP
xTRs is a simple hub and spoke switched L2 topology using VLANs (this
is a common assumption in DC networks), a multi-chassis Link
Aggregation Group (LAG) solution can be used to offer redundancy,
where both xTRs will be seen by the access device as one logical
entity. xTRs connected to the same L2 switched access network are
part of the same 'LISP site', and both of them can be used to send
traffic to TSes behind them, as both xTRs are registering to the LISP
mapping system for the EIDs behind them. Registrations performed by
the individual xTR (as a result of detection of a new EID) part of
the same site are performed using the RLOCs of all xTRs connected to
that site. How the multi-chassis LAG solution is build is out of
scope of this draft.

3.2.3. External connectivity aspects

External connectivity between a LISP enabled NVO3 DC, as well as any
LISP site, and the external world can be handled through a gateway
device.

In case the gateway device handles connectivity to VPNs or the

Internet, LISP interworking will be employed at the gateway device as
per [RFC6832].

In case the gateway device is used to interconnect to another DC part
of the same administrative domain (one Mapping System governs both
DCs), the only function needed on this device is routing within the
RLOC address space.

In case separate LISP Mapping systems are used, interworking has to
be established between them, as well as providing routing between the
two administrative domain in between the RLOC address spaces of both
DCs respectively. Today there is no described way of interworking
between DDT based Mapping systems. An external controller could also
insert new RLOC locations into a DDT based Mapping system when an EID
has moved to a location not governed by this particular Mapping
system.

3.2.4. Optimal Forwarding aspects

Implementing a co-located and unified L2 and L3 NVE, and placing that
NVE as close as possible to the TSes, leads to the most optimal
forwarding.

East-to-west traffic (from NVE to NVE) will always be mapped-and-
encapped towards the 'right' NVE, as the NVA function (the LISP
Mapping system) has knowledge about all of the TSes IP and MAC
addresses.

North to South traffic (traffic ingress into the DC) will also be
delivered to the right NVE, without traffic tromboning, as traffic is
switched based on the EID IP address, which will always point to the
correct ETR/RLOC.

Traffic going from TSes to external destinations will also not be
affected by traffic tromboning as all NVE's part of an NVI are seen
as the same default gateway, independent of location.

Traffic tromboning can occur if the last hop router is not in the
same location as the egress NVE, and if only a single L2 NVE is
deployed. In other words, the only way for a L2-only NVE based NVO3
system to avoid traffic tromboning for north-south traffic, is by
centralizing the default gateway for all VNI's in one location (that
in some cases may be suboptimal).

3.2.5. VM Mobility aspects

This section shows how the LISP control plane deals with VM mobility
when end systems are migrated from one NVE/DC to another.

We'll assume that a signaling protocol, as described in [I-D
.kompella-nvo3-server2nve], signals to the NVE operations such as
creating/terminating/migrating an end system. The signaling protocol
consists of three basic messages: "associate", "disassociate", and
"pre-associate". The signaling protocol for attach/detach is in
addition and orthogonal to the LISP control plane.

Two approaches are laid out: An approach at L2, where MAC-addresses
are used as EID, and an approach at L3, where both IP and MAC
addresses are used as EIDs.

3.2.5.1. VM Mobility at L2

VM mobility at L2 is described in [I-D.maino-nvo3-lisp-cp]

It is to be noted that the approach of solving VM mobility at L2
introduces scalability problems in terms of failure domain, NVA
scaling (as MAC addresses are a flat and non de-aggregatable address
space) and BUM containment.

3.2.5.2. VM Mobility at L3

This approach solves the scaling problems of the L2 approach by
assuming that the majority of traffic is IP based. End Systems are
therefor registered with their IP addresses as EID and xTR IP address
as an RLOC, while their MAC-address is registered as an additional
RLOC for the same EID.

```
                          ,---------.
                        ,'           `.
                       (Mapping System )
                        `.           ,'
                          `-+------+'
                        +--+--+   +-+---+
                        |MS/MR|   |MS/MR|
                        +-+---+   +-----+
                          |          |
                        .--..--. .--. ..
                       (     '          '.--.
                     .-.'         L3        '
                      (        Underlay       )
                       (                      '-'
                       ._.'--'._.'.-._.'.-._)
                  RLOC=IP_A //              \\ RLOC=IP_B
                       +---+--+              +-+--+--+
                   .--.-.|xTR A |'.-.        .| xTR B |.-.
                   (      +---+--+    )      ( +-+--+--+   )
                    (              __.        (            '.
                    ..'  LISP Site A  )       .'   LISP Site B  )
                   (            .'-'         (            .'-'
                    '--'._.'.    )\          '--'._.'.    )\
                     /      '--'  \           /      '--'  \
                  '--------'   '--------'     '--------'   '--------'
                  :  End   :   :  End   : ==> :  End   :   :  End   :
                  : Device :   : Device : ==> : Device :   : Device :
                  '--------'   '--------'     '--------'   '--------'
                     EID=          EID=<IID1,MAC_W>        EID=
                 <IID2,MAC_X>      EID=<IID1,IP_W>      <IID1,MAC_Z>
                 <IID2,IP_X>                            <IID1,IP_Z>
```

It is assumed that the LISP xTRs have a unified L2 and L3 map-en-
encap function, where IP packets, regardless of the fact they are
switched intra- or inter subnet, are mapped-and-encapped across the
L3 overlay. All other traffic (non-routable traffic, non-IP traffic)
is mapped-and-encapped by the L2 overlay. It is also assumed that
all XTRs offer the same default gateway IP and MAC address across the
network, on a per VNI instance.

A unified L2/L3 overlay will lead in the xTRs registering two sets of
EIDs for specific end systems, who deliver a mix of IP and non-IP
traffic:

o A tuple of EID=<IID, IP> to use for IP traffic across the L3
   overlay, whereby the IID maps to a VRF instance. It will register
   the EID to multiple RLOCs, one being the xTR IP address, and the
   other one being the TS MAC Address.

o A tuple EID= <IID,MAC> to use for non-routable, non-IP traffic,
   across the L2 overlay, whereby the IID maps to a network-wide
   Bridge Domain.

Assume the scenario described in Figure 1. Also assume that for the sake of this discussion, the VMs do not send out traffic that needs treatment by an L2 overlay.

As a result of the end system registration, the Mapping System contains the EID-to-RLOC mapping for end system W that associates EID=<IID1, IP_W> with the RLOC(s) associated with LISP site A (IP_A), as well as the RLOC associated with the MAC Address MAC_W of the end system W.

The process of migrating end system W from data center A to data center B is initiated.

ETR B receives a pre-associate message that includes EID=<IID1, IP_W>. ETR B sends a Map-Register to the mapping system registering RLOC=IP_B as an additional locator for end system W with priority set to 255. This means that the RLOC MUST NOT be used for unicast forwarding, but the mapping system is now aware of the new location.

During the migration process of end system W, ETR A receives a dissociate message, and sends a Map-Register with Record TTL=0 to signal the mapping system that end system W is no longer reachable at RLOC=IP_A. xTR A will also add an entry in its forwarding table that marks EID=<IID1, IP_W> as non-local.

When end system W has completed its migration, ETR B receives an associate message for end system W, and sends a Map-Register to the mapping system setting a non-255 priority for RLOC=IP_B. Now the mapping system is updated with the new EID-to-RLOC mapping for end system W with the desired priority.

The remote ITRs that were corresponding with end system W during the migration will keep sending packets to ETR A.

ETR A will keep forwarding locally those packets until it receives a dissociate message, and the entry in the forwarding table associated with EID=<IID1, IP_W> is marked as non-local.

Subsequent packets arriving at ETR A from a remote ITR, and destined to end system W will hit the entry in the forwarding table that will generate an exception, and will generate a Solicit-Map-Request (SMR) message that is returned to the remote ITR.

Upon receiving the SMR the remote ITR will invalidate its local map-cache entry for EID=<IID1, IP_W> and send a new Map-Request for that EID. The Map-Request will generate a Map-Reply that includes the new EID-to-RLOC mapping for end system W with RLOC=IP_B.
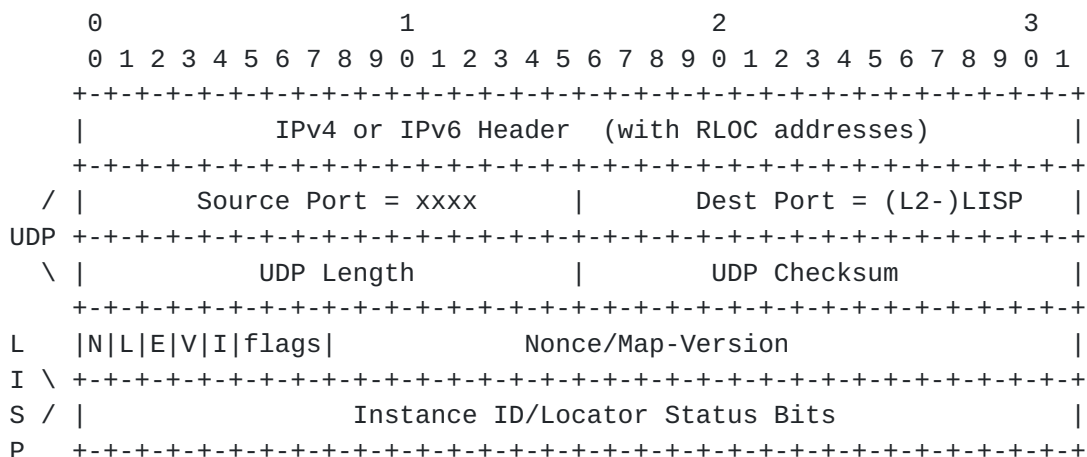
Similarly, unencapsulated packets arriving at ITR A from local end systems and destined to end system W will hit the entry in the forwarding table marked as non-local, and will generate an exception that by sending a Map-Request for EID=<IID1, IP_W> will populate the map-cache of ITR A with an EID-to-RLOC mapping for end system W with RLOC=IP_B.

3.3. LISP dataplane options and NVO3 dataplane requirements

This section maps the NVO3 data plane requirements [I-D.ietf-nvo3 -dataplane-requirements] to the various options available.

3.3.1. Native LISP Data Plane

Figure 2shows the LISP header defined in the LISP specification [RFC6830]. The UDP and LISP headers are shown below for reference. For header fields description see section 5.3 of [RFC6830].

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |              IPv4 or IPv6 Header  (with RLOC addresses)      |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     /  |         Source Port = xxxx      |       Dest Port = (L2-)LISP   |
   UDP  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     \  |           UDP Length            |        UDP Checksum         |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   L    |N|L|E|V|I|flags|               Nonce/Map-Version               |
   I \  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   S /  |                  Instance ID/Locator Status Bits             |
   P    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

When the headers are used for encapsulating IP Packets, the UDP Destination Port is set to 4341. When the headers are used for encapsulating L2 frames, the UDP Destination Port is set to 8472 [I-D .smith-lisp-layer2].

When used in private DC and Enterprise networks, the 'I'-bit (Instance bit) is set, indicating the presence of an Instance ID (IID) inside the header. A Virtual Networking Instance (VNI) is indicated by the IID, a 24 bit field, which is used as a global identifier for the tenant in LISP. When used for L3 services, the IID can be seen as a VRF, when used for L2 services, the IID can be seen as a L2 Bridge Domain instance.

Virtual Access Point (VAP) identification is naturally supported by combining LISP and Integrated Routing and Bridging (IRB). IRB allows local ports or logical ports (ports instantiated on a local port, where frames are assigned based on some fields in the header like

VLAN IDs (VIDs)), to be added to a NVE-local bridge domain. That

bridge domain instantiates the L2 Specific VNI. The bridge domain
also connects to a virtual routed port, which instantiates the L3
specific VNI.

A L2 VNI provides an emulated Ethernet Multipoint service through the
use of the LISP control plane, where it registers MAC addresses as
EIDs.

Loop-avoidance is handled by control plane learning, and control
plane enabled registration of all TS EIDs as soon as they send a
first packet. Therefore unicast traffic will never result in loops,
as there is no 'unknown' unicast. multi-destination traffic
forwarding is performed using a multicast enabled underlay and LISP
procedures laid out in [RFC6831] or through ingress replication to
the list of participating NVEs in that NVI, and therefore is loop-
free.

A L3 VNI behaves exactly as an IP VRF and therefore supports
virtualized IP routing and forwarding, through per tenant forwarding
with IP address isolation and L3 tunneling for interconnecting
instances of the same VNI on NVEs.

Note that , within this document, it is assumed that a unified L2/L3
NVE is deployed and therefore all IP traffic will be forwarded using
the L3 overlay, even intra-subnet traffic.

The LISP header performs the function of the NVO3 overlay header.

Through using the LISP control plane, the egress NVE can be looked
up.

As the outer LISP header is an IPv4 or IPv6 header, differentiated
forwarding can be supported naturally. Equally, as LISP uses IP/UDP
as a transport, multipath over LAG and ECMP in the underlay are
naturally supported, through the entropy introduced in the UDP header
by selecting per flow source UDP port numbers. A LISP based NVO3
network can work in both uniform and pipe models [RFC2983] and fully
supports ECN marking as per [RFC6040] .

As it does for L3 services, the LISP control plane replaces the use
of dynamic data plane learning (Flood-and-Learn) for unicast traffic
for L2 services. Packet replication in the underlay network to
support L2 broadcast, unknown unicast (optional, as all MAC-address
are learned by the control plane) and multicast L2 and L3 overlay
services can be done by:

o Ingress replication, which reduces the need for multicast in the
   NVO3 underlay to zero.

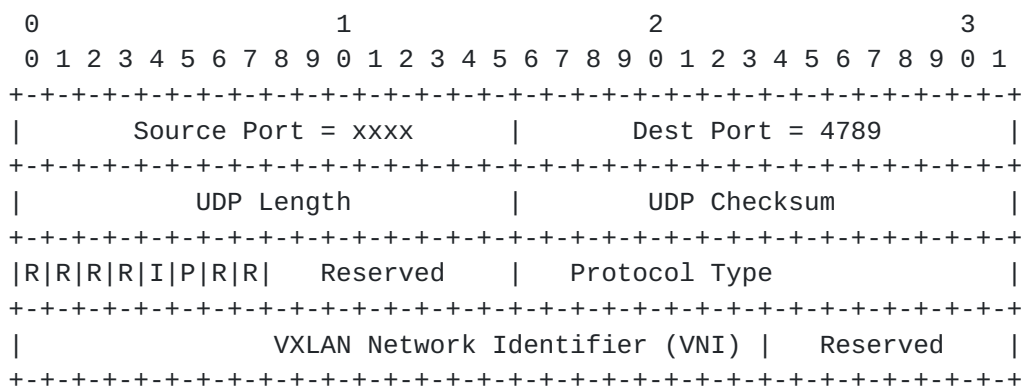o Use of underlay multicast trees. These trees can be aggregated

globally, or per tenant (rather than per VNI).

[RFC6831] and [I-D.farinacci-lisp-mr-signaling]specifies how to map a
multicast flow in the EID space during distribution tree setup and
packet delivery in the underlay network. LISP, being an IP based
map-and-encap protocol, does not require any specific data plane
functionality to make this work.

LISP interworking is described in [RFC6832] and fully supports
connectivity to the internet or VPN gateways through the use of Proxy
xTR's.

LISP, being an IP based protocol, supports ICMP-based MTU Path
Discovery [RFC1191] , [RFC1981]as well as extended MTU Path Discovery
techniques [RFC4821]. LISP also supports a stateless and stateful
way of dealing with Large Encapsulated packets, see section 5.4 of
[RFC6830].

Multi-homing is handled in the control plane, by allowing the LISP
mapping system to have multiple RLOC entries for every EID, allowing
the ITR to load balance across both ETR's. xTRs register a 'LISP
site id' to the mapping system when they come online. When the LISP
mapping system receives a registration for a given EID from a certain
xTRs, it will install that EID entry pointing to all the RLOCs/xTR
that have the same site-id. By performing LAG across multiple xTRs,
multi-homing can be provided for the access or virtual switch that
connects the TSs.

OAM can be performed across LISP in the same way as OAM is performed
over IP routed, or Ethernet L2 switched environments.

3.3.2. LISP with Generic Protocol Extension (LISP-GPE)

[I-D.lewis-lisp-gpe] introduces multiprotocol support for LISP, by
extending the LISP header, as shown in Figure 3 .

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |      Source Port = xxxx       |       Dest Port = 4341        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           UDP Length          |        UDP Checksum           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |N|L|E|V|I|P|R|R| Reserved      |Nonce/Map-Version/Protocol-Type|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 Instance ID/Locator-Status-Bits               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

A Protocol Bit (P-bit) is introduced, that when set, allows the
insertion of a 16-bit Protocol Type. The UDP destination port number
is 4341 for all protocol types.

Although the use of Nonce and Map-versioning are not allowed simultaneously with Protocol Type with this deployment, all of the solutions to the requirements described in Section 3.3.1 are exactly the same with this data plane encapsulation in an NVO3 context.

3.3.3. VxLAN-GPE

[I-D.quinn-vxlan-gpe] extends the VXLAN encapsulation with a Protocol Type, by introducing a Protocol Bit (P-bit) and a 16-bit Protocol Type in the VXLAN header, see Figure 4. Note that this data plane encapsulation is very similar to LISP-GPE, when used in an NVO3 context.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Source Port = xxxx      |       Dest Port = 4789        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           UDP Length          |         UDP Checksum          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R|R|R|R|I|P|R|R|   Reserved    |    Protocol Type              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             VXLAN Network Identifier (VNI) |   Reserved       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

All of the solutions to the requirements described in Section 3.3.1 are exactly the same with this data plane encapsulation.

3.3.4. L2 only solutions such as VxLAN and nvGRE

The LISP control plane can be leveraged to offer control plane learning for MAC Addresses for both the VXLAN [I-D.mahalingam-dutt-dcops-vxlan], as well as NVGRE [I-D.sridharan-virtualization-nvgre]. However, this solution offers sub-optimal support and hence will not be looked into further.

3.4. NVO3 control plane requirements and LISP

This section maps the NVO3 NVE to NVA control plane [I-D.ietf-nvo3 -nve-nva-cp-req]requirements to the LISP control plane.

3.4.1. Inner to Outer Address Mapping

The LISP control plane, through the use of a Mapping service,

provides inner to outer address mapping.

TS EIDs are registered to the LISP Mapping service by LISP ETRs
within the context of a LISP instance ID, (i.e An NVO3 VNI).

A LISP based NVE will check its local cache if it needs to send a
packet across the overlay. If there is a cache miss, it will request
the EID to RLOC mapping from the LISP Mapping service. If there is a
cache hit, it will use the local EID to RLOC mapping.

Cache entries are aged out when no traffic is being sent to those
EIDs. The LISP control plane has ways of refreshing the local cache
after the destination EID has moved to another RLOC. For more
information, see Section 3.2.5 and [RFC6830]

3.4.2. Underlying network Multi-Destination Delivery

LISP supports delivering L2 and L3 multi-destination packets,
independent of the underlying network multicast capabilities.

[RFC6831], [I-D.farinacci-lisp-mr-signaling] , more specifically
section 6, describes how the LISP Control Plane is used by NVEs/ETRs
to join a given EID multicast group by sending LISP Map-Requests
rather than PIM Joins. Joining can be triggered by the receipt of a
PIM or IGMP join in the EID space. In the case of a L2 overlay
configuration on the NVE, the join is static.

In case the NVE/ETR is not multicast capable the ETR unicast RLOC
will be registered, and will be added to the existing RLOC set for
that given multicast EID, and the Map-Reply will contain the ITR from
which the ETR wants to replicate. LISP ITRs will retrieve this list
of ETRs from the Mapping System upon a Map-Request and will use this
as a replication list.

In case the underlying network is multicast capable the Map-Reply
will contain the multicast RLOC, which will be joined via PIM
subsequently. All this state is stored on the Mapping system, or in
the xTR caches per IID/VNI. In case ingress replication is deemed
unscaleable, [I-D.farinacci-lisp-te] can be used, allowing a Re-
encapsulating Tunnel Router (RTR) to be used as a distribution server
to replicate to all the NVEs.

It is also important to point out that, in a unified L2/L3 NVE
deployment, all IP traffic will get sent across the L3 overlay, and
that ARP and ND packets are not handled using flooding.

In case non-IP traffic needs to be supported, L2 EIDs only need to
use multicast or ingress replication for broadcast and multicast, as
unicast flows are handled by the LISP control plane. This
significantly reduces the multicast or ingress replication load.

3.4.3. VN connect/disconnect

We assume that a provisioning framework will be responsible for
provisioning end systems (e.g. VMs) in each data center. The
provisioning system configures each end system with an Ethernet/IEEE
802 MAC address and/or IP addresses and provisions the NVE with other
end system specific attributes such as VLAN information, and TS/VLAN
to VNI mapping information. LISP does not introduce new addressing
requirements for end systems.

The provisioning infrastructure is also responsible to provide a
network attach function, that notifies the NVE (the LISP site ETR)
that the end system is attached to a given virtual network
(identified by its VNI/IID) and that the end system is identified,
within that virtual network, by a given Ethernet/IEEE 802 MAC
address.

The LISP framework does not include mechanisms to provision the local
NVE with the appropriate Tenant Instance for each Tenant Systems.
Other protocols, such as VDP (in IEEE P802.1Qbg), should be used to
implement a network attach/detach function, besides using link-up
events for non-virtual end-systems. More-over it is quite common for
devices to be able to 'sense' new tenant end-systems dynamically by
tracking new mac addresses and IP addresses in case a VDP or link-up
event cant be relied upon.

The LISP control plane can take advantage of such a network attach/
detach function or the discovery of new MAC/IP addresses to trigger
the registration of a Tenant System to the Mapping System. This is
particularly helpful to handle mobility across the DC of the Tenant
System.

Upon notification of end system network attach, the ETR sends a LISP
Map-Register to the Mapping System. The Map-Register includes the
EID and RLOCs of the LISP site. The EID-to-RLOC mapping is now
available, via the Mapping System Infrastructure, to other LISP sites
that are hosting end systems that belong to the same tenant.

For more details on end system registration see [RFC6833].

3.4.4. VN name to VN ID Mapping

The LISP Control Plane uses the Instance ID to identify the NVI. The
VN Name to VNI mapping can be performed by the NVE as a result of
local provisioning. Also, using LISP LCAF , it is possible to store
ASCII Names in the Mapping Database, which can allow the system to
resolve a VN Name to an IID/VNI.

3.4.5. LISP Control Plane Characteristics in an NVO3 context

LISP is a Control Plane solution that can scale very well to the NVO3

requirements:

1. LISP ETRs register destination EIDs into the LISP Mapping System.
    LISP ITRs pull destination EIDs from the LISP Mapping System and
    cache them for as long as traffic is being sent to those
    destinations. Hence a LISP Based NVE is only holding state for
    the active TS to TS flows, and only for the NVIs that are
    configured on those NVEs.

2. The LISP Control Plane is fast to acquire the needed state for a
    given destination through issuing a single Map-Request.

3. When an ETR (lets say ETR1) detects an EID it will also register
    this EID to the Mapping system. If that EID has moved from
    another ETR (lets say ETR2), it will update the Mapping system
    with a Map-Notify saying to no longer forward packets to it, and
    will install a 'non-local' entry in the forwarding table. If an
    ITR has not updated its map-cache, and therefor sends a packet to
    ETR2, ETR will sent a Map-Notify directly to the ITR, updating
    its local cache. For further information see Section 3.2.5

4. As LISP support virtualization, the NVE running the LISP Control
    Plane will only be maintaining state for the Tenants VNIs that
    are configured on it.

5. Through leveraging the LISP DDT-based Mapping system [I-D.ietf-
    lisp-ddt], the necessary scaling can be achieved. The LISP DDT
    hierarchy can be based on address family, address family prefix,
    and IID, and scales in a very similar way as DNS.

6. The solution described in this document does not make use of
    multiple protocols, and hence is low in complexity.

7. Through the use of the LISP LCAF [I-D.ietf-lisp-lcaf] ,
    extensibility is achieved. It is possible to add new address
    families (the MAC address family is the proof point). The LCAF
    format also allows lookups on a generic Key. This Key can be an
    identifier to an ACL or policy. A combination of multiple keys
    can be achieved by doing recursive lookups, where EID attributes
    are used as keys for a subsequent lookup. LCAF allows backwards
    compatibility between systems that use different LCAF
    implementations.

8. As the state is maintained in the LISP Mapping system acting as
    an NVA, adding another NVE/xTR to the network does not require
    any changes on existing NVEs.

9. LISP does not rely on Multicast in the underlay, while preserving
    the same Control Plane characteristics regardless of underlay
    multicast capability.

10. [I-D.barkai-lisp-nfv]documents one implementation of how the
     LISP Mapping System (NVA) can be programmed to create inner to
     outer address mappings. The LISP Control Plane will inform the
     xTR/NVE that hosts have moved, and if packets are attracted to
     those NVEs because of stale cache entries on other ITRs, packets
     will be routed to the right location, and the NVE will send a
     Solicited Map-Reply back to the ITR, clearing its cache, after
     which the ITR will request a new mapping. Obviously NVEs will
     be able to create inner to outer address mappings without the
     use of an orchestration solution.

11. See Section 5

3.5. NVO3 OAM Requirements and LISP

   TBD

3.6. NVO3 Management Plane Requirements and LISP

   TBD

3.7. Summary

   The LISP Control Plane, makes a very good choice to implement NVO3
   services due to the fact that it is agnostic to EID address families,
   and the fact that it provides an NVA in the form of the LISP Map
   Server with cache optimizations based on the pull-based LISP Map
   Cache on the LISP xTRs . The LISP control plane can be deployed
   across a set of different dataplane options as well. The usage of a
   unified L2 and L3 overlay , with the appropriate set of registrations
   in the LISP Mapping system, is recommended because of its optimal
   forwarding, scaling and IP centric characteristics, while supporting
   non-IP traffic as well.

4. IANA Considerations

   This document makes no request of IANA.

   Note to RFC Editor: this section may be removed on publication as an
   RFC.

5. Security Considerations

   The security requirements for a NVO3 Control Plane are defined in
   [I-D.ietf-nvo3-security-requirements] . More specifically, seven
   requirements are defined (REQ 1 to REQ 7) for NVE-NVA Control Plane
   (Network Virtualization Edge to Network Virtualization Authority
   Control Plane) and two requirements (REQ 8 and REQ 9) for NVA-NVA
   Control Plane (Network Virtualization Authority to Network

Virtualization Authority Control Plane). Table 1 provides a summary
of which document defines LISP Control Plane mechanisms that allow to
satisfy each specific requirement.

```
          +-----------+-----------------------------+
          | NVO3 Req. | LISP Control Plane Documents |
          +-----------+-----------------------------+
          | REQ 1     | [RFC6833] [I-D.ietf-lisp-sec] |
          | REQ 2     | [RFC6833] [I-D.ietf-lisp-sec] |
          | REQ 3     | Not mandatory[1]            |
          | REQ 4     | [RFC6833] [I-D.ietf-lisp-sec] |
          | REQ 5     | [RFC6833] [I-D.ietf-lisp-sec] |
          | REQ 6     | [RFC6830] [RFC6833]         |
          | REQ 7     | [RFC6830][2]                |
          | REQ 8     | [I-D.ietf-lisp-ddt]         |
          | REQ 9     | Does not apply[3]           |
          +-----------+-----------------------------+
```

[1] The requirement uses MAY as for [RFC2119] terminology. [2] Amplification attacks can be avoided by careful design of the mappings. [3] The existing LISP Control Planes do not use multicast messages.

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

[RFC6830] and [RFC6833] describe how to send control packet with limited frequencies.

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

The security of the Mapping System Infrastructure (NVA) depends on the particular mapping database used. The [I-D.ietf-lisp-ddt] specification, as an example, defines a public-key based mechanism that provides origin authentication and integrity protection to the LISP DDT protocol.

6. Acknowledgements

7. References

7.1. Normative References

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

7.2. Informative References

   [I-D.barkai-lisp-nfv]

sbarkai@gmail.com, s., Farinacci, D., Meyer, D., Maino,
F., Ermagan, V., Rodriguez-Natal, A. and A. Cabellos-
Aparicio, "LISP Based FlowMapping for Scaling NFV",
Internet-Draft draft-barkai-lisp-nfv-04, February 2014.

   [I-D.farinacci-lisp-mr-signaling]
            Farinacci, D. and M. Napierala, "LISP Control-Plane
            Multicast Signaling", Internet-Draft draft-farinacci-lisp-
            mr-signaling-04, March 2014.

   [I-D.farinacci-lisp-te]
            Farinacci, D., Kowal, M. and P. Lahiri, "LISP Traffic
            Engineering Use-Cases", Internet-Draft draft-farinacci-
            lisp-te-06, March 2014.

   [I-D.ietf-lisp-ddt]
            Fuller, V., Lewis, D., Ermagan, V. and A. Jain, "LISP
            Delegated Database Tree", Internet-Draft draft-ietf-lisp-
            ddt-01, March 2013.

   [I-D.ietf-lisp-lcaf]
            Farinacci, D., Meyer, D. and J. Snijders, "LISP Canonical
            Address Format (LCAF)", Internet-Draft draft-ietf-lisp-
            lcaf-05, May 2014.

   [I-D.ietf-lisp-sec]
            Maino, F., Ermagan, V., Cabellos-Aparicio, A. and D.
            Saucez, "LISP-Security (LISP-SEC)", Internet-Draft draft-
            ietf-lisp-sec-06, April 2014.

   [I-D.ietf-nvo3-dataplane-requirements]
            Bitar, N., Lasserre, M., Balus, F., Morin, T., Jin, L. and
            B. Khasnabish, "NVO3 Data Plane Requirements", Internet-
            Draft draft-ietf-nvo3-dataplane-requirements-03, April
            2014.

   [I-D.ietf-nvo3-framework]
            Lasserre, M., Balus, F., Morin, T., Bitar, N. and Y.
            Rekhter, "Framework for DC Network Virtualization",
            Internet-Draft draft-ietf-nvo3-framework-09, July 2014.

   [I-D.ietf-nvo3-nve-nva-cp-req]
            Kreeger, L., Dutt, D., Narten, T. and D. Black, "Network
            Virtualization NVE to NVA Control Protocol Requirements",
            Internet-Draft draft-ietf-nvo3-nve-nva-cp-req-02, April
            2014.

   [I-D.ietf-nvo3-overlay-problem-statement]
            Narten, T., Gray, E., Black, D., Fang, L., Kreeger, L. and
            M. Napierala, "Problem Statement: Overlays for Network
            Virtualization", Internet-Draft draft-ietf-nvo3-overlay-
            problem-statement-04, July 2013.

   [I-D.ietf-nvo3-security-requirements]

Hartman, S., Zhang, D. and M. Wasserman, "Security
Requirements of NVO3", Internet-Draft draft-ietf-nvo3
-security-requirements-02, January 2014.

   [I-D.kompella-nvo3-server2nve]
            Kompella, K., Rekhter, Y., Morin, T. and D. Black,
            "Signaling Virtual Machine Activity to the Network
            Virtualization Edge", Internet-Draft draft-kompella-
            nvo3-server2nve-02, April 2013.

   [I-D.lewis-lisp-gpe]
            Lewis, D., Agarwal, P., Kreeger, L., Maino, F., Quinn, P.,
            Smith, M. and N. Yadav, "LISP Generic Protocol Extension",
            Internet-Draft draft-lewis-lisp-gpe-02, July 2014.

   [I-D.mahalingam-dutt-dcops-vxlan]
            Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger,
            L., Sridhar, T., Bursell, M. and C. Wright, "VXLAN: A
            Framework for Overlaying Virtualized Layer 2 Networks over
            Layer 3 Networks", Internet-Draft draft-mahalingam-dutt-
            dcops-vxlan-09, April 2014.

   [I-D.maino-nvo3-lisp-cp]
            Maino, F., Ermagan, V., Hertoghs, Y., Farinacci, D. and M.
            Smith, "LISP Control Plane for Network Virtualization
            Overlays", Internet-Draft draft-maino-nvo3-lisp-cp-03,
            October 2013.

   [I-D.quinn-vxlan-gpe]
            Quinn, P., Agarwal, P., Fernando, R., Lewis, D., Kreeger,
            L., Smith, M., Yadav, N., Yong, L., Xu, X., Elzur, U. and
            P. Garg, "Generic Protocol Extension for VXLAN", Internet-
            Draft draft-quinn-vxlan-gpe-03, July 2014.

   [I-D.smith-lisp-layer2]
            Smith, M., Dutt, D., Farinacci, D. and F. Maino, "Layer 2
            (L2) LISP Encapsulation Format", Internet-Draft draft-
            smith-lisp-layer2-03, September 2013.

   [I-D.sridharan-virtualization-nvgre]
            Sridharan, M., Greenberg, A., Wang, Y., Garg, P.,
            Venkataramiah, N., Duda, K., Ganga, I., Lin, G., Pearson,
            M., Thaler, P. and C. Tumuluri, "NVGRE: Network
            Virtualization using Generic Routing Encapsulation",
            Internet-Draft draft-sridharan-virtualization-nvgre-04,
            February 2014.

   [RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191,
            November 1990.

   [RFC1981] McCann, J., Deering, S. and J. Mogul, "Path MTU Discovery
            for IP version 6", RFC 1981, August 1996.

   [RFC2983] Black, D., "Differentiated Services and Tunnels", RFC
              2983, October 2000.

   [RFC3971]   Arkko, J., Kempf, J., Zill, B. and P. Nikander, "SEcure
               Neighbor Discovery (SEND)", RFC 3971, March 2005.

   [RFC4821]   Mathis, M. and J. Heffner, "Packetization Layer Path MTU
               Discovery", RFC 4821, March 2007.

   [RFC4861]   Narten, T., Nordmark, E., Simpson, W. and H. Soliman,
               "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861,
               September 2007.

   [RFC6040]   Briscoe, B., "Tunnelling of Explicit Congestion
               Notification", RFC 6040, November 2010.

   [RFC6830]   Farinacci, D., Fuller, V., Meyer, D. and D. Lewis, "The
               Locator/ID Separation Protocol (LISP)", RFC 6830, January
               2013.

   [RFC6831]   Farinacci, D., Meyer, D., Zwiebel, J. and S. Venaas, "The
               Locator/ID Separation Protocol (LISP) for Multicast
               Environments", RFC 6831, January 2013.

   [RFC6832]   Lewis, D., Meyer, D., Farinacci, D. and V. Fuller,
               "Interworking between Locator/ID Separation Protocol
               (LISP) and Non-LISP Sites", RFC 6832, January 2013.

   [RFC6833]   Fuller, V. and D. Farinacci, "Locator/ID Separation
               Protocol (LISP) Map-Server Interface", RFC 6833, January
               2013.

   [RFC6836]   Fuller, V., Farinacci, D., Meyer, D. and D. Lewis,
               "Locator/ID Separation Protocol Alternative Logical
               Topology (LISP+ALT)", RFC 6836, January 2013.

Authors' Addresses

   Yves Hertoghs
   Cisco Systems
   6a De Kleetlaan
   Diegem, 1831
   Belgium2

   Phone: +32-2778-435
   Email: yves@cisco.com

   Fabio Maino
   Cisco Systems
   170 Tasman Drive
   San Jose, California 95134

USA

   Email: fmaino@cisco.com

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

   Email: vimoreno@cisco.com

   Michael Smith
   Cisco Systems
   170 Tasman Drive
   San Jose, California 95134
   USA

   Email: michsmit@cisco.com

   Dino Farinacci
   lispers.net

   Email: farinacci@gmail.com

   Luigi Iannone
   Telecom ParisTech

   Email: luigi.iannone@telecom-paristech.fr