Workgroup:	BESS
Published:	4 July 2021
Intended Status:	Standards Track
Expires:	5 January 2022
---	---	---
Authors:	P. Thubert	A. Przygienda, Ed.
	Cisco Systems	*Juniper Networks*

Routing for eVPN Hosts

Abstract

This specification adds attributes to eVPN to carry IPv6 address metadata learned from RFC 8505 and RFC 8928 so as to maintain a synchronized copy of the 6LoWPAN ND registrar at each eVPN router and perform locally a unicast IPv6 ND service for address lookup and duplicate address detection.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

"Registration Extensions for IPv6 over 6LoWPAN Neighbor Discovery" [RFC8505] (ND) provides a routing-agnostic Host-to-Router Link-Local interface to claim an IPv6 address and obtain reachability services for that address. [RFC8505] is already used to redistribute host routes in RPL [RFC9010] and RIFT "Routing in Fat Trees" [RIFT], and to maintain a proxy-ND state in a backbone router [RFC8929].

[RFC8505] specifies a unicast address registration mechanism that enables the host called a 6LowPAN Node (6LN) to install a ND binding state in the 6LowPAN Router (6LR) that can serve as Neighbor Cache Entry (NCE), though it is not operated as a cache. The protocol provides the means to reject the registation in case of address duplication. It also enables to discriminate mobility from multihoming. [RFC8928] adds the capability to verify the ownership of the address and prevent an attacker from stealing and/or impersonating an address.

[RFC8505] defines the 6LoWPAN Border Router (6LBR) as an abstract address registrar that provides autoritative service for Address Registration and duplicate detection. The 6LBR stores address metadata that is obtained during the Address Registration, including a owner ID and a sequence counter. As part of the process of a new Address Registration, the 6LR queries the 6LBR for existing metadata related to the address being registered. This enables in particular to detect a duplication and reject the registration. This specification extends the 6LBR abstract data model to store the Link Layer Address (LLA) of the Registering Node. This enables the 6LBR to perform locally, and using unicast communication, the IPv6 ND services of address lookup and duplicate address detection.

The [RFC8505] address registrar can be centralized, but it can also be distributed and maintained synchronized using a routing protocol. This specification adds attributes to eVPN to carry the IPv6 address metadata learned from [RFC8505] so as to maintain a synchronized copy of the 6LBR abstract data at each eVPN router.

2. Terminology

2.1. Requirements Language

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

2.2. Glossary

This document uses the following acronyms:

6LN: 6LoWPAN Node (the Host)
6LR: 6LoWPAN router (the router)
6LBR: 6LoWPAN Border router (the reflector)
AMC: Address Mapping Confirmation
AMR: Address Mapping Request
EARO: Extended Address Registration Option
DAD: Duplicate Address Detection
ICMPv6: Internet Control Message Protocol for IPv6
EVPN: Ethernet VPN
LLA: Link-Layer Address (the MAC address on Ethernet)
NA: Neighbor Advertisement
NCE: Neighbor Cache Entry
ND: Neighbor Discovery
NS: Neighbor Solicitation
RA: Router Advertisement
ROVR: Registration Ownership Verifier
TID: Transaction ID (a sequence counter in the EARO)
TIO: Transit Information Option
SLLAO: Source Link-Layer Address Option
TLLAO: Target Link-Layer Address Option
ROVR MAC: MAC obtained from a host meeting requirements in Section 4
Validated ROVR MAC: ROVR MAC validated by procedures specified in [RFC8928]
ROVR Node: EVPN node capable of advertising ROVR MACs
non-ROVR Node: EVPN node not supporting extensions defined in this document.

2.3. References

This document uses the terms Clos fabric and Fat Tree interchangeably, to refer to a folded spine-and-leaf topology as defined in the terminology section of "RIFT: Routing in Fat Trees" [RIFT].

The term "leaf" represents the access switch that connects the servers to the Fat Tree. The leaf is typically a Top-of-Rack (ToR) switch.

This specification uses the terms 6LN, 6LR and 6LBR to refer specifically to nodes that implement the said roles in [RFC8505] and does not expect other functionality such as 6LoWPAN Header Compression:

- In the context of this document, the 6LN is a server that adversises an address mapping using [RFC8505], and optionally protects its ownership with [RFC8928].
- The 6LR and 6LBR function are collapsed at the leaf and its state is synchonized with that of the eVPN functional support using an internal interface that is out of scope. That interface could be "pull" meaning that the 6LBR fetches the eVPN information when it needs it, or "push", meaning that any information that eVPn distributes is immediately fed in all the 6LBRs in all the leaves. Note that this is pure control plane and is not subject to abbreviating optimization as the FIB may be.

In this document, readers will encounter terms and concepts that are discussed in the following documents:

eVPN: "BGP MPLS-Based Ethernet VPN" [RFC7432] and "Network Virtualization Overlay Solution" [RFC8365],

Classical IPv6 ND: "Neighbor Discovery for IP version 6" [RFC4861] and "IPv6 Stateless Address Autoconfiguration" [RFC4862],

6LoWPAN ND: Neighbor Discovery Optimization for Low-Power and Lossy Networks [RFC6775], "Registration Extensions for 6LoWPAN Neighbor Discovery" [RFC8505], "Address Protected Neighbor Discovery for Low-power and Lossy Networks" [RFC8928], and "IPv6 Backbone Router" [RFC8929].

3. 6LoWPAN Neighbor Discovery

This section goes through the 6LoWPAN ND mechanisms that this specification leverages, as a non-normative reference to the reader. The full normative text is to be found in [RFC6775], [RFC8505], and [RFC8928].

3.1. RFC 6775 Address Registration

The classical "IPv6 Neighbor Discovery (IPv6 ND) Protocol" [RFC4861] [RFC4862] was defined for serial links and transit media such as Ethernet. It is a reactive protocol that relies heavily on multicast operations for Address Discovery (aka Lookup) and Duplicate Address Detection (DAD).

"Neighbor Discovery Optimizations for 6LoWPAN networks" [RFC6775] adapts IPv6 ND for operations over energy-constrained LLNs. The main functions of [RFC6775] are to proactively establish the Neighbor Cache Entry (NCE) in the 6LR and to prevent address duplication. To that effect, [RFC6775] introduces a new unicast Address Registration mechanism that contributes to reducing the use of multicast messages compared to the classical IPv6 ND protocol.

[RFC6775] defines a new Address Registration Option (ARO) that is carried in the unicast Neighbor Solicitation (NS) and Neighbor Advertisement (NA) messages between the 6LoWPAN Node (6LN) and the 6LoWPAN router (6LR). It also defines the Duplicate Address Request (DAR) and Duplicate Address Confirmation (DAC) messages between the 6LR and the 6LBR). In an LLN, the 6LBR is the central repository of all the Registered Addresses in its domain and the source of truth for uniqueness and ownership.

3.2. RFC 8505 Extended Address Registration

"Registration Extensions for 6LoWPAN Neighbor Discovery" [RFC8505] updates RFC 6775 into a generic Address Registration mechanism that can be used to access services such as routing and ND proxy. To that effect, [RFC8505] defines the Extended Address Registration Option (EARO), shown in Figure 1:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |     Type      |     Length    |    Status     |    Opaque     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Rsvd | I |R|T|     TID       |     Registration Lifetime     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
   ...             Registration Ownership Verifier                 ...
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: EARO Option Format

3.2.1. R Flag

[RFC8505] introduces the R Flag in the EARO. The Registering Node sets the R Flag to indicate whether the 6LR should ensure reachability for the Registered Address. If the R Flag is set to 0, then the Registering Node handles the reachability of the Registered Address by other means. In an eVPN network, this means that either it is a RAN that injects the route by itself or that it uses another eVPN router for reachability services.

This document specifies how the R Flag is used in the context of eVPN. An eVPN Host that implements the 6LN functionality from [RFC8505] requires reachability services for an IPv6 address if and only if it sets the R Flag in the NS(EARO) used to register the address to a 6LR acting as an eVPN border router. Upon receiving the NS(EARO), the eVPN router generates a BGP advertisement for the Registered Address if and only if the R flag is set to 1.

[RFC9010] specifies that the 'R' flags is set in the responded NA messages if and only if the route was installed. This specification echoes that behaviour.

3.2.2. TID, "I" Field and Opaque Fields

When the T Flag is set to 1, the EARO includes a sequence counter called Transaction ID (TID), that is needed to format the MAC Mobility Extended Community. This is the reason why the support of [RFC8505] by the Host, as opposed to only [RFC6775], is a prerequisite for this specification); this requirement is fully explained in Section 4.1. The EARO also transports an Opaque field and an associated "I" field that describes what the Opaque field transports and how to use it.

This document specifies the use of the "I" field and the Opaque field by a Host.

3.2.3. Status

The values of the EARO status are maintained by IANA in the Address Registration Option Status Values subregistry [IANA-EARO-STATUS] of the Internet Control Message Protocol version 6 (ICMPv6) Parameters registry.

[RFC6775] and [RFC8505] defined the original values whereas [RFC9010] reduced range to 64 values and reformated the octet field to enable to transport an external error, e.g., coming form a routing protocol.

This specification uses the format expressed in [RFC9010]. The value of 0 denotes an unqualified success, 1 indicates an address duplication, 3 a TID value that is outdated, and 4 is used in an asynchronous NA to indicate that 6LN should remove that address and possibly form new ones.

3.2.4. Route Ownership Verifier

Section 5.3 of [RFC8505] introduces the Registration Ownership Verifier (ROVR) field of variable length from 64 to 256 bits. The ROVR is a replacement of the EUI-64 in the ARO [RFC6775] that was used to identify uniquely an Address Registration with the Link-Layer address of the owner but provided no protection against spoofing.

"Address Protected Neighbor Discovery for Low-power and Lossy Networks" [RFC8928] leverages the ROVR field as a cryptographic proof of ownership to prevent a rogue third party from registering an address that is already owned. The use of ROVR field enables the 6LR to block traffic that is not sourced at an owned address.

This specification does not address how the protection by [RFC8928] could be extended for use in eVPN. On the other hand, it adds the ROVR to the BGP advertisement to share the state with the other routers via the Reflector (see Section 5.1), which means that the routers that are aware of the Host route are also aware of the ROVR associated to the Target Address, whether it is cryptographic and should be verified.

3.3. RFC 8505 Extended DAR/DAC

[RFC8505] updates the DAR/DAC messages into the Extended DAR/DAC to carry the ROVR field. The EDAR/EDAC exchange is not needed in this specification as the BGP fabric as a whole is synchronized to the same state via BGP and the reflector. Based on its local state, the router can check whether the proposed route is new and legit, and reject it otherwise.

3.4. RFC 7400 Capability Indication Option

"6LoWPAN-GHC: Generic Header Compression for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs)" [RFC7400] defines the 6LoWPAN Capability Indication Option (6CIO) that enables a node to expose its capabilities in router Advertisement (RA) messages.

[RFC8505] defines a number of bits in the 6CIO, in particular:

L: Node is a 6LR.

E: Node is an IPv6 ND Registrar -- i.e., it supports registrations based on EARO.

P: Node is a Routing Registrar, -- i.e., an IPv6 ND Registrar that also provides reachability services for the Registered Address.

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     Type      |   Length = 1  |     Reserved      |D|L|B|P|E|G|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           Reserved                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: 6CIO flags

A 6LR that provides reachability services for a Host in an eVPN network as specified in this document includes a 6CIO in its RA messages and set the L, P and E flags to 1 as prescribed by [RFC8505].

3.5. Extending 6LoWPAN ND

3.5.1. Use of the R flag in NA

This document extends [RFC8928] and [RFC8505] as follows

This document also updates the behavior of a 6LR acting as eVPN router and of a 6LN acting as Host in the 6LoWPAN ND Address Registration as follows:

- The use of the R Flag is extended to the NA(EARO) to confirm whether the route was installed.

3.5.2. Distributing the 6LBR

This specification enables to distribute the 6LBR at the edge of the eVPN network and collapse the 6LBR function with that of the eVPN support. In that model, the eVPN to 6LBR interaction becomes an internal interface, where each side informs the other in case of new information concerning an IP to Link-Layer Address (LLA) mapping. Since this is an internal interface, this specification makes no assumption on whether the 6LBR stores its own representation of the full eVPN state, which means that the eVPN support informs the 6LBR in case of any change on the eVPN side (this is called the push model, see Figure 9), or if the 6LBR queries the eVPN support when it does not have a mapping to satisfy a request (pull model, see Figure 8).

This specification leverages [RFC8929] that augments the abstract data model of the 6LBR to store the LLA associated with the registered address. Based on that additional state, the 6LBR in a leaf can communicate the mapping to the colocated eVPN function and respond to unicast address mapping lookups from the server side.

In an environment where the server ranges from a classical host to a more complex platform that runs a collection of virtual hosts interconnected by a virtual switch, but where the host-to-leaf interface remains at layer 2, the 6LR and the 6LBR functions can be collapsed in the leaf. The 6LR to 6LBR interaction also becomes an internal interface, and there is no need for EDAR/EDAC messages.

In that case, the MAC address associated to the Registered Address is indicated in the Target Link-Layer Address Option (TLLAO) in the NS message used for the registration, as shown in Figure 3. In the case of a pull model, if the 6LBR does not have a local state for the mapping, it queries the eVPN support to obtain the eVPN state if any. If a mapping is known then the 6LR/6LBR evaluates the registration for address duplication and other possible issues per [RFC8505]. Else (this is for a new mapping), if the registration is accepted, then the 6LBR notifies the eVPN support to inject a route type 2 in the fabric.

```
        Server              Leaf
  +--------------+   +-------------------+
  |              |   |                   |
   6LN                6LR/6LBR       eVPN
    |                    |            |
    | <vSwitch> Ethernet | <call I/F> |
    |                    |            |
    |  NS(EARO)          |            |
    |------------------->|            |
    |                    |            | ^
    |                    |   query    | |
    |                    |----------->| | if pull
    |                    |  response  | | model
    |                    |<-----------| |
    |                                 | v
    |            Evaluation (DAD)     |
    |                                 |
    |                    |new mapping |
    |                    | indication |
    |                    |----------->|
    |                                 | Inject/maintain
    |         store a mapping state   | eVPN route type 2
    |                                 | ------------------>
    |       NA(EARO)     |            | [via BGP Signalling]
    |<-------------------|            |
    |                    |            |
```

Figure 3: Direct Registration

In another type of deployment, the 6LR may be a virtual router in the server whereas the 6LBR runs in the leaf node. To address that case, the EDAR/EDAC may be used to communicate as shown in figure 5 of [RFC8505]. This draft leverages the capability to insert IPv6 ND options in the EDAR and EDAC messages introduced in [RFC8929] to place a TLLAO that carries the MAC address associated to the Registered address in the EDAR and EDAC messages as shown in Figure 4:

```
            Server                       Leaf
      +----------------+          +----------------+
      |                |          |                |
       6LN          6LR          6LBR        eVPN
        |            |              |            |
        | <vSwitch>  |  Ethernet    | <call I/F> |
        |            |              |            |
        |  NS(EARO)  |              |            |
        |----------->|              |            |
        |            |  EDAR(TLLAO) |            |
        |            |------------->|            |
        |            |              |            | ^
        |            |              |   query    | |
        |            |              |----------->| | if pull
        |            |              |  response  | | model
        |            |              |<-----------| |
        |            |                           | v
        |            |    Evaluation (DAD)       |
        |            |                           |
        |            |              |new mapping |
        |            |              | indication |
        |            |              |----------->|
        |            |                           | Inject/maintain
        |            |   store a mapping state   | eVPN route type 2
        |            |                           | ------------------->
        |            |  EDAC(TLLAO) |            | [via BGP Signalling]
        |            |<-------------|            |
        |  NA(EARO)  |              |            |
        |<-----------|              |            |
        |            |              |            |
```

Figure 4: leveraging EDAR

3.5.3. Unicast Address Lookup with the 6LBR

A classical IPv6 ND stack in the server that treats the subnet prefix as on-link (more in section 4.6.2. of [RFC4861]), will resolve an unknown LLA mapping with a multicast NS(lookup) message addressed to the solicited node multicast address (SNMA) associated with the destination address being resolved. The RECOMMENDED operation in that case is for the 6LBR that has a mapping state to forward the packet as a unicast MAC to the LLA that is stored for the IPv6 address as expected by [RFC6085]. The actual owner of the address can then answer unicast with a NA message, setting the override (O) bit to 1, as shown in Figure 5.

```
                Local              Local                        Remote
                Server             Leaf                         Server
                +-----+          +--------+                     +-----+
                |     |          |        |                     |     |
                 6LN              6LR/6LBR                       6LN
                  |                  |                            |
                  |    Ethernet      |                            |
                  |                  |      [via EVPN    ]        |
                  |    multicast     |      [Data Tunnels]        |
                  |    NS(lookup)    |                            |
                  |----------------->|                            |
                  |                  |     forward unicast        |
                  |                  |      NS(lookup)            |
                  |                  |--------------------------->|
                  |                  |                            |
                  |                  |          NA(O)             |
                  |                  |<---------------------------|
                  |      NA(O)       |                            |
                  |<-----------------|                            |
                  |                  |                            |
```

Figure 5: Forwarding legacy NS (Lookup)

Section 3.1. of [RFC8929] adds the capability to insert IPv6 ND options in the EDAR and EDAC messages. This enables the 6LBR to store the link-layer address associated with the Registered Address and to serve as a mapping server. [UNICAST-LOOKUP] leverages that state to define a new unicast address lookup operation, extending the EDAR and EDAC messages as the Address Mapping Request (AMR) and Confirmation (AMC) with a different Code Prefix [RFC8505].

In that model, the router advertises the subnet prefix as not on-link by setting the L flag to 0 in the Prefix Information Option (PIO), more in section 4.6.2. of [RFC4861]. The expected behaviour is that the host that communicates with a peer in the same subnet refrains from resolving the address mapping and passes the packets directly to the router.

In the case where the router is a virtual 6LR running in the server, and the source and destination are in the same subnet served by eVPN, the router then resolves the address mapping on behalf of the host. To that effect, the router sends a unicast AMR message to the 6LBR. The message contains the SLLAO of the router to which the 6LBR will reply. If the binding is found, the 6LBR replies with an AMC message that contains the TLLOA with the requested MAC address, as shown in Figure 6.

```
                Local                    Local             Remote
                Server                    Leaf              Server
          +-----------------+       +----------+          +-----+
          |                 |       |          |          |     |
           6LN          6LR          6LBR/eVPN             6LN
             |            |              |                  |
             |            |              |  [via EVPN    ]  |
             |  <vSwitch> |   Ethernet   |  [Data Tunnels]  |
             |            |              |                  |
             |            |              |                  |
             |   RA(PIO)  |              |                  |
             |  not onlink|              |                  |
             |<-----------|              |                  |
             |            |              |                  |
             |   Packet   |              |                  |
             |----------->|              |                  |
             |            |              |                  |
             |            | AMR (SLLAO)  |                  |
             |            |------------->|                  |
             |            |              |                  |
             |            | AMC (TLLAO)  |                  |
             |            |<-------------|                  |
             |            |              |                  |
             |     NCE in STALE state    |                  |
             |            |                                 |
             |            |            Packet               |
             |            |-------------------------------->|
             |            |                                 |
             |     NCE in DELAY state    |                  |
             |            |              |                  |
```

Figure 6: Unicast Lookup from the virtual Host

If it is not found, [UNICAST-LOOKUP] provides the capability to indicate immediately that the mapping is not known with a "not found" status in the AMC, as opposed to waiting for an NS(lookup) and retries to time out per [RFC4861].

In a fully stateful subnet where all nodes register all their addresses with [RFC8505], this means that the looked up address is not present in the network; in that case the packet is dropped and an ICMP error type 1 "Destination Unreachable" code 3 "Address unreachable" [RFC4443] is returned as shown in Figure 7.

```
              Local                 Local
              Server                 Leaf
        +----------------+       +---------+
        |                |       |         |
         6LN          6LR         6LBR/eVPN
          |            |              |
          |            |              |
          | <vSwitch>  |  Ethernet    |
          |            |              |
          |            |              |
          |   RA(PIO   |              |
          |not on-link)|              |
          |<-----------|              |
          |            |              |
          |  Packet    |              |
          |----------->|              |
          |            |              |
          |            | EDAR(lookup) |
          |            |------------->|
          |            |              |
          |            | EDAC(status= |
          |            | "Not Found") |
          |            |<-------------|
          |ICMPv6 Error|              |
          | "Address   |              |
          |Unreachable"|              |
          |<-----------|              |
          |         Drop Packet       |
          |            |              |
```

Figure 7: Unicast Lookup failure

Note that the figures above make no assumption on the pull vs. push model. In the case of pull model, the 6LBR queries the eVPN support when it does not have the mapping information to satisfy a request. Figure 8 illustrates a successful pull model lookup flow, when the route type 2 for the mapping is already known on the eVPN side.

```
            Server                          Leaf
      +-----------------+          +-----------------+
      |                 |          |                 |
       6LN          6LR             6LBR        eVPN
        |             |              |            |
        | <vSwitch>   |  Ethernet    | <call I/F> |
        |             |              |            |
        |             |              |            |
        |             |              |            | Receive eVPN
        |             |              |            | route type 2 for
        |             |              |            | remote address A'
        |             |              |            | [via BGP Signalling]
        |             |              |            |<------------------
        |             |              |     store a mapping state
        |             |              |            |
        |Packet for A'               |            |
        |----------->|               |            |
        |             |EDAR(lookup A')            |
        |             |------------->|            |
        |             |              |Query addr A'
        |             |              |----------->|
        |             |              |            |
        |             |              | return LLA |
        |             |              |<-----------|
        |             |              |            |
        |             |EDAC(A''s TLLA)            |
        |             |<-------------|            |
        |             |              |            |
        |             |            Packet for A'  |
        |             |            [via EVPN    ] |
        |             |            [Data Tunnels] |
        |             |----------------------------------->
        |             |              |            |
```

Figure 8: Pull models

In the case of push model, the eVPN support synchronizes its state upon a route type 2 with the 6LBR, and the 6LBR maintains an abstract data structure for all information known to eVPN. This way, the 6LBR already has the mapping information to satisfy any request for an existing mapping and it can answer right away. Figure 9 illustrates a successful push model lookup flow, when the 6LBR is already in possession of the mapping.

```
          Server                          Leaf
      +-----------------+          +-----------------+
      |                 |          |                 |
       6LN           6LR            6LBR         eVPN
        |             |              |             |
        | <vSwitch>   |   Ethernet   | <call I/F>  |
        |             |              |             |
        |             |              |             |
        |             |              |             |
        |             |              |             | Receive eVPN
        |             |              |             | route type 2 for
        |             |              |             | remote address A'
        |             |              |             | [via BGP Signalling]
        |             |              |             |<------------------
        |             |              |     store a mapping state
        |             |              |             |
        |             |              |indicate LLA|
        |             |              |<-----------|
        |             |    store a mapping state   |
        |             |              |             |
        |Packet for A'|              |             |
        |----------->|               |             |
        |             |EDAR(lookup A')             |
        |             |------------->|             |
        |             |              |             |
        |             |EDAC(R's TLLA)|             |
        |             |<-------------|             |
        |             |              |             |
        |             |         Packet for A'      |
        |             |         [via EVPN    ]     |
        |             |         [Data Tunnels]     |
        |             |----------------------------------->
        |             |              |             |
```

Figure 9: Push model

In a mixed environment, a lookup failure (the mapping is not found though the address is present in the network) may be caused by a legacy node that was node discovered (aka a silent node). In that case, it is an administrative decision for the 6LR to broadcast an NS(lookup) or to return an error as shown in Figure 7.

4. Requirements on the eVPN-Unaware Host

This document describes how eVPN routing can be extended to reach a Host. This section specifies the minimal eVPN-independent functionality that the Host needs to implement to obtain routing services for its addresses.

4.1. Support of 6LoWPAN ND

A host sees a prefix as not on-link (e.g., it learned that prefix in a PIO in a RA with the L flag not set) should not attempt to resolve an address within that prefix using a multicast NS(lookup). Instead, it must pass its packets to a router, preferably one that advertises that prefix in a PIO; it must register the address that it uses as source to that router to enable source address validation using [RFC8505]. It is recommended that the Host also implements [RFC8928] to prove its ownership of its addresses.

The Host is expected to request routing services from a router only if that router originates RA messages with a 6CIO that has the L, P, and E flags all set to 1 as discussed in Section 3.4, unless configured to do so. To obtain routing services for one of its addresses, the host must register the address to a router that advertises the prefix, setting the "R" and "T" flags in the EARO to 1 as discussed in Section 3.2.1 and Section 3.2.2, respectively.

This document echoes the behavior specified in [RFC9010] whereby, when the R Flag set to 1 in a NS(EARO) is not echoed in the NA(EARO), the host must understand that the route injection failed, and if the R flag is reset later in an asynchronous NA(EARO), the host must understand that routing service has failed.

The host may attach to multiple 6LRs and is expected to prefer those that provide routing services. The abstract model for this is a P2MP interface that wraps together as many P2P IP Links the host has adjacencies to 6LRs over that interface. The IPv6 address and the subnet are associated to that interface. The interface may be virtual and it may bundle multiple physical Ethernet interfaces that connect to the individual 6LRs over point to point wires, possibly via a software switch. It can also be associated to one physical interface to an external switch, either way the PI Links can be associated to sub-interface of the interface.

The Host needs to register to all the 6LRs from which it desires routing services. The multiple Address Registrations to several 6LRs should be performed in a rapid sequence, using the same EARO for the same Address. Gaps between the Address Registrations will invalidate some of the routes till the Address Registration finally shows on those routes. The routers recognize the same (ROVR, TID) as the signal of a multihomed address and maintain all the routes.

```
        Local             Local          Local
        Server            Leaf 1         Leaf 2          Reflector
       +-------+       +----------+    +----------+     +----------+
       |       |       |          |    |          |     |          |
          6LN           6LBR/eVPN      6LBR/eVPN              |
           |                |              |                  |
           |    Ethernet    |              |                  |
           |                |              |      [via BGP Signalling]
           |                |              |                  |
           |   NS(EARO)     |              |                  |
           |--------------->|              |                  |
           |      install local state      |                  |
           |                |              |                  |
           |                |----------install RT A'-------->|
           |   NA(EARO)     |              |                  |
           |<---------------|              |                  |
           |                |              |                  |
           |                |              |<------ RT  A'---|
           |                |    install remote state         |
           |                |              |                  |
           |            NS(same EARO)      |                  |
           |----------------------------->|                  |
           |                |    install local state          |
           |                |              |                  |
           |                |              |-install RT A''->|
           |            NA(same EARO)      |                  |
           |<------------------------------|                  |
           |                |              |                  |
           |                |<------------ RT A''------------|
           |     install remote state      |                  |
           |                |              |                  |
```

Figure 10: First Host Registration Flow

[RFC8505] introduces error Status values in the NA(EARO) which can be received synchronously upon an NS(EARO) or asynchronously. The Host needs to support both cases and refrain from using the address when the Status value indicates a rejection.

5. Enhancements to eVPN

This section addresses the necessary changes to EVPN formats and behavior to support address registration security per [RFC8928] and mobility per [RFC8505] while retaining interoperability with traditional nodes. With 6LR injecting not only MACs via packet sources and TTLAO options but also ROVR into mobility extended community their semantics will be somewhat extended. Specifically following issues have to be addressed:

* ROVR extends the semantics of the type-2 MAC advertisement via changes in MAC Mobility Extended Community in the sense that the MAC must be aligned with the ROVR and under normal circumstances only the validity of ROVR guarantees that the type-2 MAC can be allocated to the requester. A MAC validated by ROVR should take precedence over MAC addresses allocated without using it given it presents a much more trustworthy topological information (it will be called ROVR MAC in further text). EVPN nodes not supporting extensions introduced by this document will need to be led to believe that a MAC ROVR is to be preferred over any advertisement they see as long a ROVR MAC is

present. Nevertheless, primary key of NRLI is still the MAC address as defined in [RFC7432], Section 7.2 and 7.7. This implies that the same MAC (and consequently ROVR MAC) can be assigned multiple IP addresses and those represent independent NLRIs.

- TID field in EARO is smaller than mobility sequence number in [RFC7432]. To allow an ROVR MAC mobility to "win" over normal MACs signalling must be introduced to distinguish TID generated sequence numbers from normal sequence numbers.
- EVPN needs to deal with lifetime monitoring of the Address Registration period.

EVPN Signalling is not used to carry ROVR since without challenge per [RFC8928] they do not represent any difference over using the IP/MAC combination. Additionally, backwards compatibility could not be preserved given comparing routes based on ROVR would present a change in primary key of NLRIs which non-ROVR routes could not follow. An indication from a ROVR node that a MAC has been validated by proof of ownership is enough to convey the necessary information. Only a small hash of the ROVR is carried to allow to distinguish between MAC moves and MAC duplicate in MAC ROVR case.

5.1. ROVR MAC Mobility Extended Community

Extending MAC Mobility Extended Community allows to design a solution that, while backwards compatible, allows to introduce MAC ROVR as "more trusted" entities. Figure 11 presents the according extensions that will however necessitate some further explanation.

To introduce a "precedence" of ROVR MACs over normal EVPN MACs ROVR MACs are advertised to look like "sticky" MACs for non-ROVR nodes. As defined in the glossary, for simplicity reasons such nodes will be called non-ROVR nodes vs. ROVR nodes. The "sticky" bit will force non-ROVR nodes to disregard the sequence number and accept any NLRI route provided.

ROVR nodes MUST set the "R" flag in Mobility Extended Community to indicate that the advertisement is a ROVR MAC in case the host followed the according procedures. ROVR MACs use (instead of increasing the normal sequence number) the TID in the high bits of the sequence number field to "override" any normal MAC advertisement (further considerations will be provided in Section 5.2).

ROVR nodes MUST set the "V" flag if the address assignment passed proof of ownership per [RFC8928]. Such addresses will be preferred by ROVR nodes over non validated ROVR MACs.

In case a ROVR node configures the address as "sticky" (since the sticky bit semantics have been changed to the point a ROVR cannot tell whether address is really sticky unless advertised as such by non-ROVR node) a new "X" flag called "super sticky" is introduced.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=0x06     | Sub-Type=0x00 |       |X|V|R|S| Reserved=0    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       TID     |                  ROVR Hash                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 11: New NLRI

Flags:

S: Sticky as defined in [RFC7432].

R: ROVR Capable indicates that the advertisement is originated after processing signalling from host meeting the requirements in Section 4. This indicates a ROVR MAC.

V: ROVR Validated indicates that the MAC passed proof of ownership per [RFC8928]. Presence of this bit implies the "R" bit being set irregardless of its value.

X: Super Sticky indicates that the MAC ROVR is sticky and should follow procedures of sticky per [RFC7432].

Sequence Number Field:

TID: contains the MAC ROVR TID per [RFC8505]. This MUST NOT be zero, i.e. a ROVR

ROVR Hash: Hash of ROVR used to generate the according MAC ROVR. Hash is built by XOR'ing ROVR bytes in network order into the least significant byte and rotating the three bytes after every byte by one bit to the left.

5.2. Extended ROVR MAC Procedures

In case a non-ROVR node advertises a sticky MAC by setting the "S" bit and a ROVR node sees an ROVR address registration for the same MAC it MUST follow procedures per [RFC7432].

In case a non-ROVR node advertises a sequence number larger than the one generated by TID on a ROVR node, the ROVR node SHOULD advertise a Sequence Number consisting of all bits being set to force a "roll-over" on all nodes and then fall back to advertising the TID generated sequence number again. In case a non-ROVR node persists in increasing the sequence number after that it is indication of violation of [RFC7432] on its part.

A ROVR node advertising a ROVR MAC that has not been validated and receiving same type-2 NLRI that has been validated MUST immediately withdraw its advertisement.

A ROVR node advertising a ROVR MAC and receiving an equivalent ROVR MAC from other node with a higher TID MUST immediately withdraw its advertisement. This will allow the non-ROVR nodes to correctly interpret the sequence as MAC move despite ignoring the sequence number due to presence of "S" bit.

A ROVR node that receives a ROVR MAC with "super sticky" indication and seeing the MAC locally MUST follow analogous procedures to [RFC7432].

Multi-homing a MAC on mix of ROVR and non-ROVR nodes will lead to operational notifications since per [RFC7432] the non-ROVR node will interpret the situation as a sticky MAC that has shown up on its local interface unless an implementation is somewhat clever and understands that the presence of the same ESI on all the routes indicates that this situation does not represent a sticky MAC being moved.

6. Protocol Operations

Following section illustrates several situations and resulting signalling in EVPN from the point of view of a ROVR node.

```
         Local            Local              Route           Remote
         Server           Leaf             Reflector         Leaf
        +-------+       +---------+        +-------+        +---------+
        |       |       |         |        |       |        |         |
           6LN           6LBR/eVPN            BGP            eVPN/6LBR
            |                |                 |                |
            |    Ethernet    |      [via BGP Signalling]        |
            |                |                 |                |
            |                |                 |                |
            |   NS(EARO)     |                 |                |
            |--------------->|                 |                |
            |                |                 |                |
            |      install local state         |                |
            |                |                 |                |
            |                |  MAC ROVR Route A'               |
            |                |---------------->|                |
            |   NA(EARO)     |                 |                |
            |<---------------|                 |                |
            |                |                 | MAC ROVR Route A'
            |                |                 |--------------->|
            |                |                 |                |
            |                |                 |    install local state
```

Figure 12: Host Registration

```
         Local                 Local                 Route
         Server                Leaf                Reflector
        +-------+       +------------------+       +-------+
        |       |       |                  |       |       |
           6LN           6LBR         eVPN            BGP
            |              |            |                |
            |   Ethernet   |            |                |
            |              |            |          [via BGP Signalling]
            |              |            |                |
            |  NS(EARO(    |            |                |
            |    TID+1))   |            |                |
            |------------->|            |                |
            |              |            |                |
            |   renew lifetime locally  |                |
            |              |            | MAC ROVR Route A' (TID+1)
            |              |            |---------------->|
            | NA(EARO      |            |                |
            |  status = 0))|            |                |
            |<-------------|            |                |
            |              |            |                |
```

Figure 13: Host Registration Renewal

```
         Local            Local              Route          Remote
         Server           Leaf             Reflector        Leaf
        +-------+       +---------+        +-------+        +---------+
        |       |       |         |        |       |        |         |
           6LN           6LBR/eVPN            BGP            eVPN/6LBR
             |                |                 |                |
             |   Ethernet     |                 |                |
             |                |      [via BGP Signalling]        |
             |                |                 |                |
             |  NS(EARO)      |                 |                |
             |--------------->|                 |                |
             |                |                 |                |
             |                | MAC ROVR Route A'                |
             |                |---------------->|                |
             |                |                 | MAC ROVR Route A'
             |                |                 |--------------->|
             |  NA(EARO       |                 |                |
             |  status = 0))  |               remote state with  |
             |<---------------|               same A', TID, ESI, Flags,
             |                |               ROVR Hash
             |                |                 |                |
```

Figure 14: Multihoming

```
         Local            Local              Route          Remote
         Server           Leaf             Reflector        Leaf
        +-------+       +---------+        +-------+        +---------+
        |       |       |         |        |       |        |         |
           6LN           6LBR/eVPN            BGP            eVPN/6LBR
             |                |                 |                |
             |   Ethernet     |                 |                |
             |                |      [via BGP Signalling]        |
             |                |                 |                |
             |  NS(EARO)      |                 |                |
             |--------------->|                 |                |
             |                |                 |                |
             |                | MAC ROVR Route A'                |
             |                | ESI X, TID Z    |                |
             |                | ROVR Hash F     |                |
             |                |---------------->|                |
             |                |                 |        Local A' Present
             |                |                 |        with ESI X',
             |                |                 |        ROVR Hash G
             |                |                 |                |
             |                |                 | MAC ROVR Route A'
             |                |                 | ESI X', ROVR Hash G'
             |                |                 |<---------------|
             |                |<----------------|                |
             |  NA(EARO(      |                 |                |
             |  status = 1))  |                 |                |
             |<---------------|                 |                |
             |                |                 |                |
```

Figure 15: Duplicate Addresses

```
      Local            Local              Route           Remote
      Server           Leaf             Reflector         Leaf
     +-------+       +---------+         +-------+        +---------+
     |       |       |         |         |       |        |         |
        6LN          6LBR/eVPN              BGP            eVPN/6LBR
         |                |                  |                |
         |    Ethernet    |                  |                |
         |                |        [via BGP Signalling]       |
         |                |                  |                |
         |   NS(EARO)     |                  |                |
         |--------------->|                  |                |
         |                |                  |                |
         |                |  MAC ROVR Route A'                |
         |                |  ESI X, TID Z    |                |
         |                |  ROVR Hash F     |                |
         |                |----------------->|                |
         |                |                  |         Local A' Present
         |                |                  |         with ESI X',
         |                |                  |         ROVR Hash F, TID Z+1
         |                |                  |                |
         |                |                  |  MAC ROVR Route A'
         |                |                  |  ESI X', ROVR Hash F, TID Z+1
         |                |                  |<---------------|
         |   NA(EARO(     |<-----------------|                |
         |   status = 3)) |                  |                |
         |<---------------|                  |                |
         |                |                  |                |
         |                |                  |                |
```

Figure 16: Address Move on Registration

```
     Local             Local                Route            Remote
     Server            Leaf               Reflector          Leaf
    +-------+        +---------+         +-------+         +---------+
    |       |        |         |         |       |         |         |
       6LN            6LBR/eVPN              BGP            eVPN/6LBR
        |                |                  |                |
        |    Ethernet    |                  |                |
        |                |       [via BGP Signalling]        |
        |                |                  |                |
        |   NS(EARO)     |                  |                |
        |--------------->|                  |                |
        |                |                  |                |
        |          Local State:             |                |
        |          MAC ROVR Route A'        |                |
        |          ESI X, TID Z             |                |
        |          ROVR Hash F              |                |
        |                |                  |        Local A' Present
        |                |                  |        with ESI X',
        |                |                  |        ROVR Hash F, TID Z+1
        |                |                  |                |
        |                |                  |  MAC ROVR Route A'
        |                |                  |  ESI X', Hash F, TID Z+1
        |                |                  |<---------------|
        |   NA(EARO(     |<-----------------|                |
        |   status = 4)) |                  |                |
        |<---------------|                  |                |
        |                | Withdraw MAC ROVR Route A'        |
        |                |----------------->|                |
        |                |                  |--------------->|
```

Figure 17: Address Move

```
     Local             Local                Route            Remote
     Server            Leaf               Reflector          Leaf
    +-------+        +---------+         +-------+         +---------+
    |       |        |         |         |       |         |         |
       6LN            6LBR/eVPN              BGP            eVPN/6LBR
        |                |                  |                |
        |    Ethernet    |                  |                |
        |                |       [via BGP Signalling]        |
        |                |                  |                |
    Lifetime Expired     |                  |                |
        |                | Withdraw MAC ROVR Route A'        |
        |                |----------------->|                |
        |                |                  |--------------->|
```

Figure 18: Lifetime Elapse

7. Security Considerations

TBD

8. IANA Considerations

9. Acknowledgments

The authors wish to thank you for reading that far.

10. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <https://www.rfc-editor.org/info/rfc4861>.

[RFC6775] Shelby, Z., Ed., Chakrabarti, S., Nordmark, E., and C. Bormann, "Neighbor Discovery Optimization for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs)", RFC 6775, DOI 10.17487/RFC6775, November 2012, <https://www.rfc-editor.org/info/rfc6775>.

[RFC6085] Gundavelli, S., Townsley, M., Troan, O., and W. Dec, "Address Mapping of IPv6 Multicast Packets on Ethernet", RFC 6085, DOI 10.17487/RFC6085, January 2011, <https://www.rfc-editor.org/info/rfc6085>.

[RFC7400] Bormann, C., "6LoWPAN-GHC: Generic Header Compression for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs)", RFC 7400, DOI 10.17487/RFC7400, November 2014, <https://www.rfc-editor.org/info/rfc7400>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

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

[RFC8505] Thubert, P., Ed., Nordmark, E., Chakrabarti, S., and C. Perkins, "Registration Extensions for IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN) Neighbor Discovery", RFC 8505, DOI 10.17487/RFC8505, November 2018, <https://www.rfc-editor.org/info/rfc8505>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <https://www.rfc-editor.org/info/rfc8365>.

[RFC8928] Thubert, P., Ed., Sarikaya, B., Sethi, M., and R. Struik, "Address-Protected Neighbor Discovery for Low-Power and Lossy Networks", RFC 8928, DOI 10.17487/RFC8928, November 2020, <https://www.rfc-editor.org/info/rfc8928>.

[RFC9010] Thubert, P., Ed. and M. Richardson, "Routing for RPL (Routing Protocol for Low-Power and Lossy Networks) Leaves", RFC 9010, DOI 10.17487/RFC9010, April 2021, <https://www.rfc-editor.org/info/rfc9010>.

[UNICAST-LOOKUP] Thubert, P. and E. Levy-Abegnoli, "IPv6 Neighbor Discovery Unicast Lookup", Work in Progress, Internet-Draft, draft-thubert-6lo-unicast-lookup-00, 25 January 2019, <https://tools.ietf.org/html/draft-thubert-6lo-unicast-lookup-00>.

11. Informative References

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, DOI 10.17487/RFC4862, September 2007, <https://www.rfc-editor.org/info/rfc4862>.

[RFC8929] Thubert, P., Ed., Perkins, C.E., and E. Levy-Abegnoli, "IPv6 Backbone Router", RFC 8929, DOI 10.17487/RFC8929, November 2020, <https://www.rfc-editor.org/info/rfc8929>.

[RIFT] Przygienda, T., Sharma, A., Thubert, P., Rijsman, B., and D. Afanasiev, "RIFT: Routing in Fat Trees", Work in Progress, Internet-Draft, draft-ietf-rift-rift-12, 26 May 2020, <https://tools.ietf.org/html/draft-ietf-rift-rift-12>.

[IANA-EARO-STATUS] IANA, "Address Registration Option Status Values", <https://www.iana.org/assignments/icmpv6-parameters/icmpv6-parameters.xhtml#address-registration>.

Authors' Addresses

Pascal Thubert
Cisco Systems, Inc
Building D
45 Allee des Ormes - BP1200
06254 Mougins - Sophia Antipolis
France
Phone: +33 497 23 26 34
Email: pthubert@cisco.com

Tony Przygienda (editor)
Juniper Networks
1137 Innovation Way
Sunnyvale, CA
United States of America
Email: prz@juniper.net