Network Working Group
Internet-Draft
Expires: août 1, 2004

P. Thubert
M. Molteni
Cisco Systems
February 2004

IPv6 Reverse Routing Header and its application to Mobile Networks
draft-thubert-nemo-reverse-routing-header-04

Status of this Memo

Copyright Notice

Abstract

Already existing proposals enable Mobile Networks by extending Mobile IP to support Mobile Routers. In order to enable nested Mobile Networks, some involve the overhead of nested tunnels between the Mobile Routers and their Home Agents.

This proposal allows the building of a nested Mobile Network avoiding the nested tunnel overhead. This is accomplished by using a new routing header, called the reverse routing header, and by overlaying a layer 3 tree topology on the evolving Mobile Network.

Table of Contents

1. Introduction

This document assumes the reader is familiar with the Mobile Networks terminology defined in [2] and with Mobile IPv6 defined in [1].

Generally a Mobile Network may be either simple (a network with one mobile router) or nested, single or multi-homed. This proposal starts from the assumption that nested Mobile Networks will be the norm, and so presents a solution that avoids the tunnel within tunnel overhead of already existing proposals.

The solution is based on a single bi-directional tunnel between the first Mobile Router (MR) to forward a packet and its Home Agent (HA). By using IPsec ESP on that tunnel, home equivalent privacy is obtained without further encapsulation.

The solution uses a new Routing Header (RH), called the Reverse Routing Header (RRH), to provide an optimized path for the single tunnel. RRH is a variant of IPv4 Loose Source and Record Route (LSRR) [6] adapted for IPv6. RRH records the route out of the nested Mobile Network and can be trivially converted into a routing header for packets destined to the Mobile Network.

This version focuses on single-homed Mobile Networks. Hints for further optimizations and multi-homing are given in the appendixes.

Local Fixed Node (LFN) and Correspondent Node (CN) operations are left unchanged as in Mobile IPv6 [1]. Specifically the CN can also be a LFN.

Section 3 proposes an example to illustrate the operation of the proposed solution, leaving detailed specifications to the remaining chapters. The reader may refer to Section 2.1 for the specific terminology.

1.1 Recursive complexity

A number of drafts and publications suggest -or can be extended to- a model where the Home Agent and any arbitrary Correspondent would actually get individual binding from the chain of nested Mobile Routers, and form a routing header appropriately.

An intermediate MR would keep track of all the pending communications between hosts in its subtree of Mobile Networks and their CNs, and a binding message to each CN each time it changes its point of attachment.

If this was done, then each CN, by receiving all the binding messages

and processing them recursively, could infer a partial topology of the nested Mobile Network, sufficient to build a multi-hop routing header for packets sent to nodes inside the nested Mobile Network.

However, this extension has a cost:

1. Binding Update storm

   when one MR changes its point of attachment, it needs to send a BU to all the CNs of each node behind him. When the Mobile Network is nested, the number of nodes and relative CNs can be huge, leading to congestions and drops.

2. Protocol Hacks

   Also, in order to send the BUs, the MR has to keep track of all the traffic it forwards to maintain his list of CNs. In case of IPSec tunneled traffic, that CN information may not be available.

3. CN operation

   The computation burden of the CN becomes heavy, because it has to analyze each BU in a recursive fashion in order to infer nested Mobile Network topology required to build a multi hop routing header.

4. Missing BU

If a CN doesn't receive the full set of PSBU sent by the MR, it will not be able to infer the full path to a node inside the nested Mobile Network. The RH will be incomplete and the packet may or may not be delivered.

5. Obsolete BU

If the Binding messages are sent asynchronously by each MR, then, when the relative position of MRs and/or the TLMR point of attachment change rapidly, the image of Mobile Network that the CN maintains is highly unstable. If only one BU in the chain is obsolete due to the movement of an intermediate MR, the connectivity may be lost.

A conclusion is that the path information must be somehow aggregated to provide the CN with consistent snapshots of the full path across the Mobile Network. This can be achieved by an IPv6 form of loose source / record route header, that we introduce here as a Reverse Routing Header

2. Terminology and Assumptions

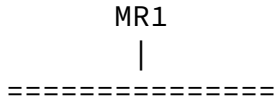
2.1 Terminology

Simple Mobile Network

One or more IP subnets attached to a MR and mobile as a unit, with respect to the rest of the Internet. A simple Mobile Network can be either single or multi-homed.

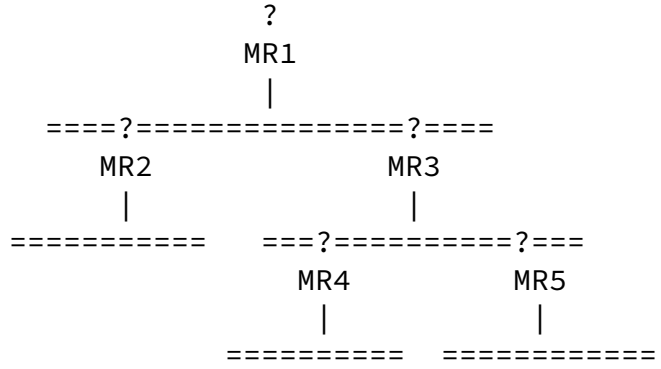
The IP subnets may have any kind of topology and may contain fixed routers. All the access points of the Mobile Network (to which further MRs may attach) are on the same layer 2 link of the MR.

We like to represent a simple single-homed Mobile Network as an hanger, because it has only one uplink hook and a bar to which multiple hooks can be attached. Graphically we use the question mark "?" to show the uplink hook (interface) connected to the MR, and the "=" sign to represent the bar:



Nested Mobile Network

A group of simple Mobile Networks recursively attached together and implementing nested Mobility as defined in [2].



IPv6 Mobile Host

A IPv6 Host, with support for MIPv6 MN, and the additional Nemo capability described in this draft.

Home prefix

Network prefix, which identifies the home link within the Internet topology.

Mobile Network prefix

Network prefix, common to all IP addresses in the Mobile Network when the MR is attached to the home link. It may or may not be a subset of the Home subnet prefix.

Inbound direction:

direction from outside the Mobile Network to inside

Outbound direction:

direction from inside the Mobile Network to outside

2.2 Assumptions

We make the following assumptions:

1. A MR has one Home Agent and one Home Address -> one primary CoA.

2. A MR attaches to a single Attachment Router as default router.

3. A MR may have more than one uplink interface.

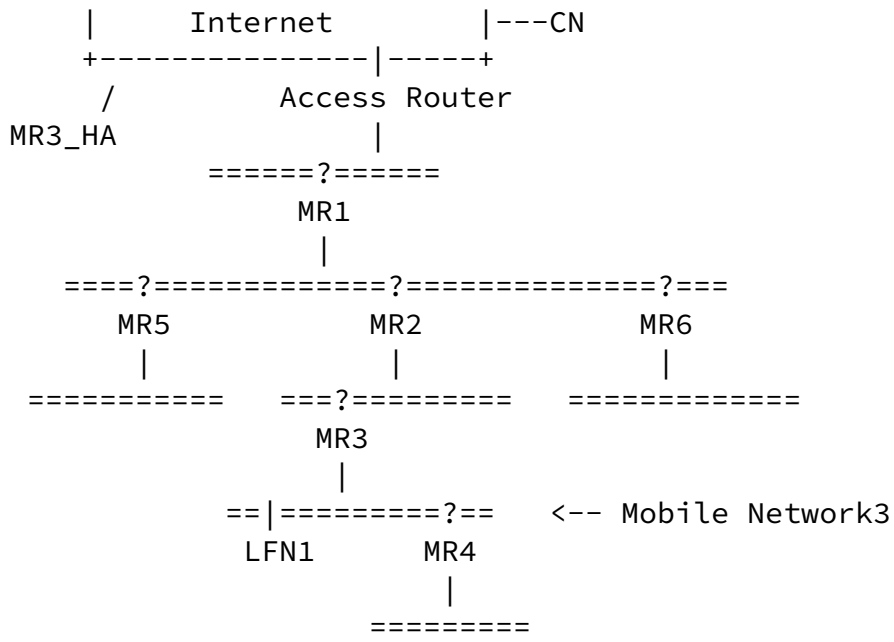
4. An interface can be either wired or wireless. The text assumes that interfaces are wireless for generality.

5. Each simple Mobile Network may have more that one L2 Access Point, all of them controlled by the same Attachment Router, which we assume to be the Mobile Router.

Since an MR has only one primary CoA, only one uplink interface can be used at a given point of time. Since the MR attaches to a single attachment router, if due care is applied to avoid loops, then the resulting topology is a tree.

3. An Example

The nested Mobile Network in the following figure has a tree topology, according to the assumptions in Section 2.2. In the tree each node is a simple Mobile Network, represented by its MR.

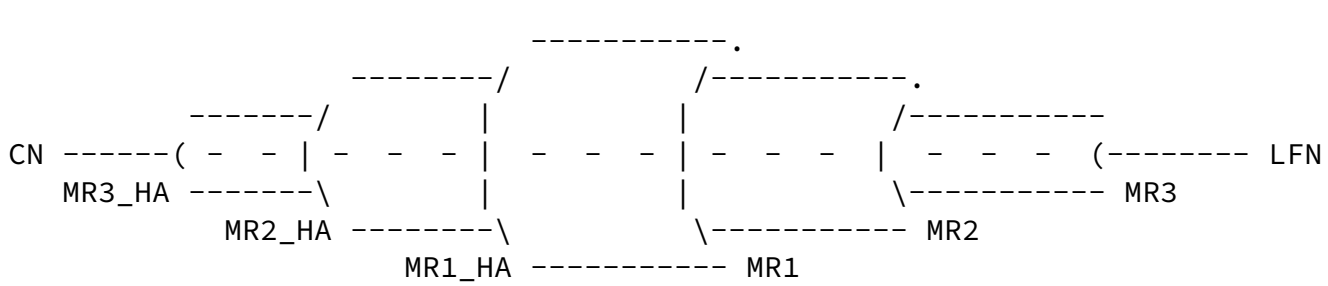


An example nested Mobile Network

This example focuses on a Mobile Network node at depth 3 (Mobile Network3) inside the tree, represented by its mobile router MR3. The path to the Top Level Mobile Router (TLMR) MR1 and then the Internet is

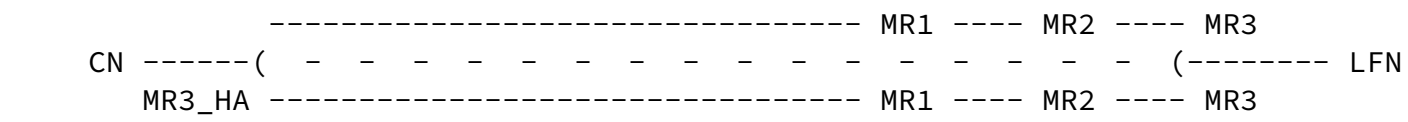
MR3 -> MR2 -> MR1 -> Internet

Consider the case where a LFN belonging to Mobile Network3 sends a packet to a CN in the Internet, and the CN replies back. With the tunnel within tunnel approach described by [3], we would have three bi-directional nested tunnels:



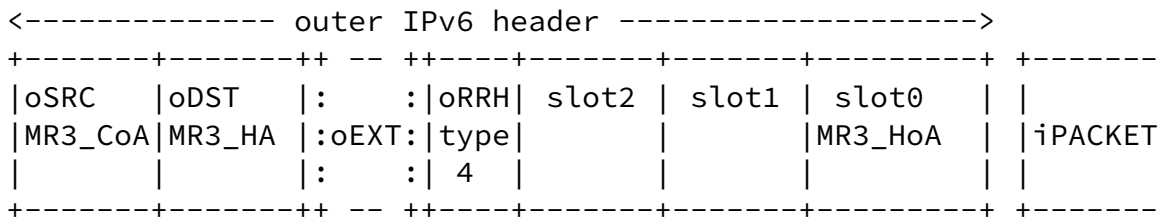
Depending on the relative location of MR1_HA, MR2_HA and MR3_HA, this may lead to a very inefficient "pinball" routing in the Infrastructure.

On the other hand, with the RRH approach we would have only one bi-directional tunnel:



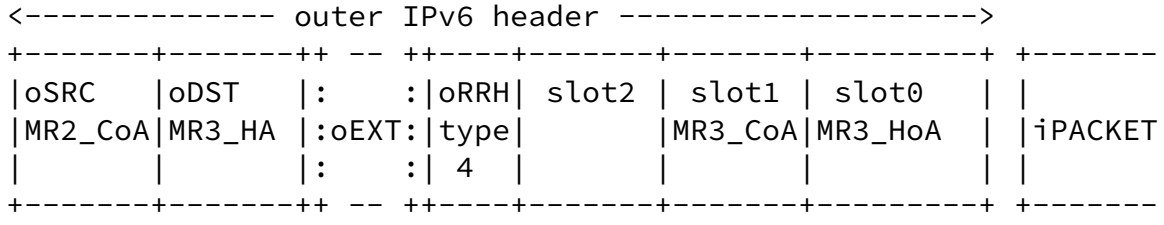
The first mobile router on the path, MR3, in addition to tunneling the packet to its HA, adds a reverse routing header with N = 3 pre-allocated slots. Choosing the right value for N is discussed in Section 6.2. The bottom slot is equivalent to the MIPv6 Home Address option. MR3 inserts its home address MR3_HoA into slot 0.

The outer packet has source MR3's Care of Address, MR3_CoA, and destination MR3's Home Agent, MR3_HA:

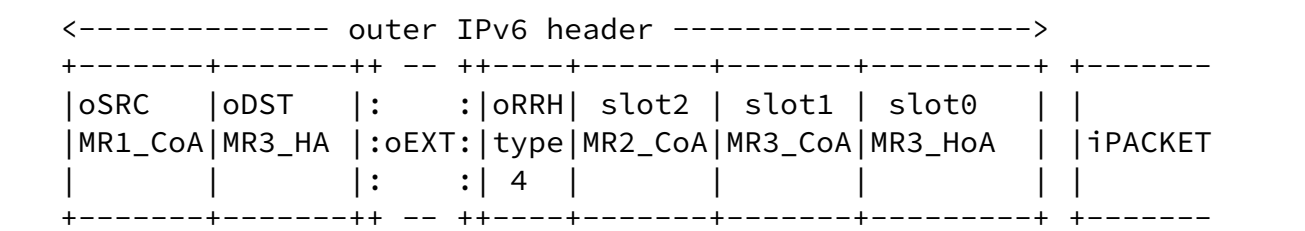


The second router on the path, MR2, notices that the packet already contains an RRH, and so it overwrites the source address of the packet with its own address, MR2_CoA, putting the old source address, MR3_CoA, in the first free slot of the RRH.

The outer packet now has source MR2_CoA and destination MR3_HA; the RRH from top to bottom is MR3_CoA | MR3_HoA:



In general the process followed by the second router is repeated by all the routers on the path, including the TLMR (in this example MR1). When the packet leaves MR1 the source address is MR1_CoA and the RRH is MR2_CoA | MR3_CoA | MR3_HoA:



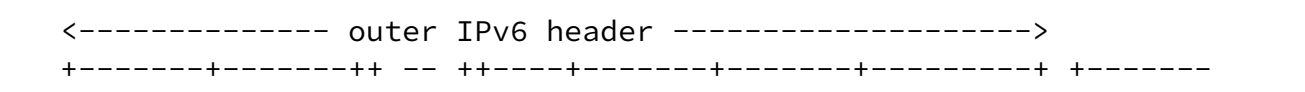
In a colloquial way we may say that while the packet travels from MR3 to MR3_HA, the Mobile Network tunnel end point "telescopes" from MR3 to MR2 to MR1.

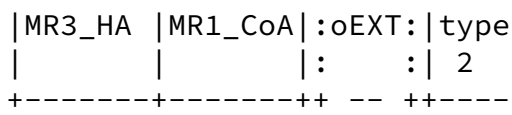
When the home agent MR3_HA receives the packet it notices that it contains a RRH and it looks at the bottom entry, MR3_HoA. This entry is used as if it were a MIPv6 Home Address destination option, i.e. as an index into the Binding Cache. When decapsulating the inner packet the home agent performs the checks described in Section 8, and if successful it forwards the inner packet to CN.

MR3_HA stores two items in the Bind Cache Entry associated with MR3: the address entries from RRH, to be used to build the RH, and the packet source address MR1_CoA, to be used as the first hop.

Further packets from the CN to the LFN are plain IPv6 packets. Destination is LFN, and so the packet reaches MR3's home network.

MR3_HA intercepts it, does a Bind Cache prefix lookup and obtains as match the MR3 entry, containing the first hop and the information required to build the RH. It then puts the packet in the tunnel MR3_HA -- MR3 as follows: source address MR3_HA and destination address the first hop, MR1_CoA. The RH is trivially built out of the previous RRH: MR2_CoA | MR3_CoA | MR3_HoA:





The packet is routed with plain IP routing up to the first destination MR1_CoA.

The RH of the outer packet is type 2 as in MIPv6 [1], but has additional semantics inherited from type 0: it contains the path information to traverse the nested Mobile Network from the TLMR to the tunnel endpoint MR3. Each intermediate destination forwards the packet to the following destination in the routing header. The security aspects of this are treated in Section 11.2.

MR1, which is the initial destination in the IP header, looks at the RH and processes it according to Section 9, updating the RH and the destination and sending it to MR2_CoA. MR2 does the same and so on until the packet reaches the tunnel endpoint, MR3.

When the packet reaches MR3, the source address in the IP header is MR3_HA, the destination is MR3_CoA and in the RH there is one segment left, MR3_HoA. As a consequence the packet belongs to the MR3_HA -- MR3 tunnel. MR3 decapsulates the inner packet, applying the rules described in Section 9 and sends it to LFN. The packet that reaches LFN is the plain IPv6 packet that was sent by CN.

4. New Routing Headers

This draft modifies the MIPv6 Routing Header type 2 and introduces two new Routing Headers, type 3 and 4. Type 3, which is an optimization of type 4 will be discussed in Appendix A.2.1. The draft presents their operation in the context of Mobile Routers although the formats are not tied to Mobile IP and could be used in other situations.

4.1 Routing Header Type 2 (MIPv6 RH with extended semantics)

Mobile IPv6 uses a Routing header to carry the Home Address for packets sent from a Correspondent Node to a Mobile Node. In [1], this Routing header (Type 2) is restricted to carry only one IPv6 address. The format proposed here extends the Routing Header type 2 to be multi-hop.

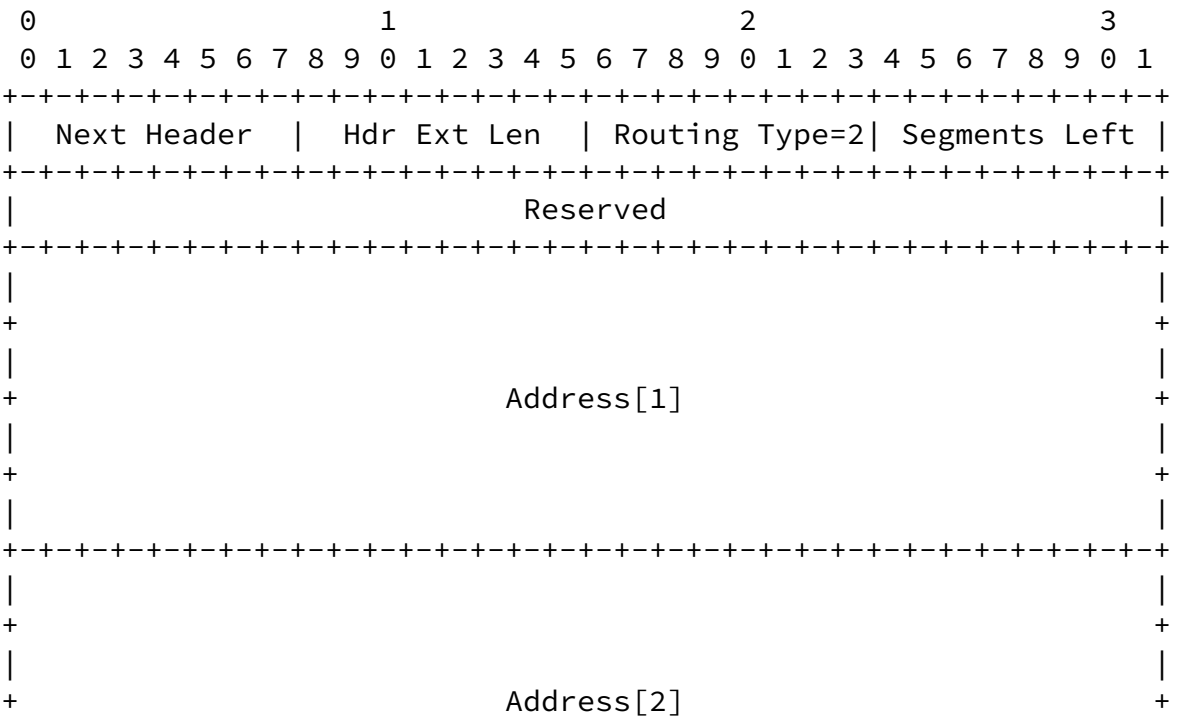
The processing of the multi-hop RH type 2 inherits from the RH type 0 described in IPv6 [10]. Specifically: the restriction on multicast addresses is the same; a RH type 2 is not examined or processed until it reaches the node identified in the Destination Address field of the IPv6 header; in that node, the RH type 0 algorithm applies, with added security checks.

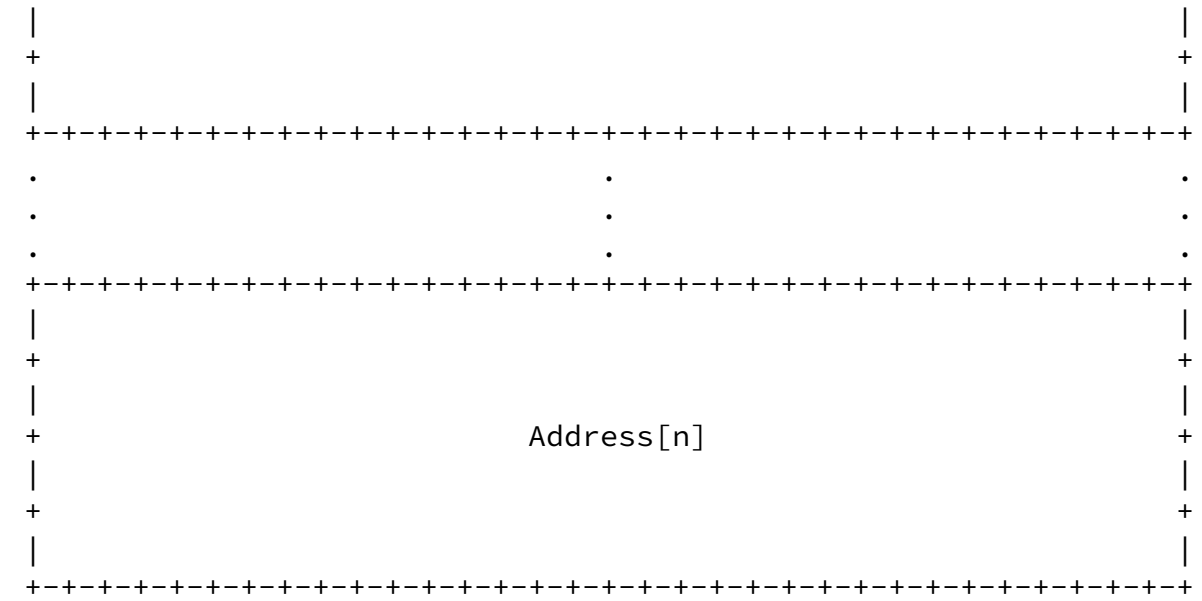
The construction of the multi-hop RH type 2 by the HA is described in Section 8; the processing by the MRs is described in Section 9.5; and the security aspects are treated in Section 11.2.

The destination node of a packet containing a RH type 2 can be a MR

or some other kind of node. If it is a MR it will perform the algorithm described in Section 9.5, otherwise it will operate as prescribed by IPv6 [10] when the routing type is unrecognized.

The multi-hop Routing Header type 2, as extended by this draft, has the following format:





Next Header

8-bit selector. Identifies the type of header immediately following the Routing header. Uses the same values as the IPv4 Protocol field [13].

Hdr Ext Len

8-bit unsigned integer. Length of the Routing header in 8-octet

units, not including the first 8 octets. For the Type 2 Routing header, Hdr Ext Len is equal to two times the number of addresses in the header.

Routing Type

8-bit unsigned integer. Set to 2.

Segments Left

8-bit unsigned integer. Number of route segments remaining, i.e., number of explicitly listed intermediate nodes still to be visited before reaching the final destination.

Reserved

32-bit reserved field. Initialized to zero for transmission; ignored on reception.

Address[1..n]

Vector of 128-bit addresses, numbered 1 to n.

4.2 Routing Header Type 4 (The Reverse Routing Header)

The Routing Header type 4, or Reverse Routing Header (RRH), is a variant of IPv4 loose source and record route (LSRR) [6] adapted for IPv6.

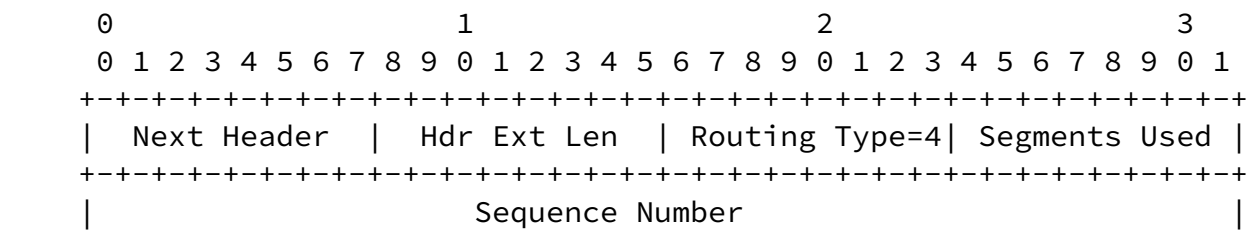
Addresses are added from bottom to top (0 to n-1 in the picture). The RRH is designed to help the destination build an RH for the return path.

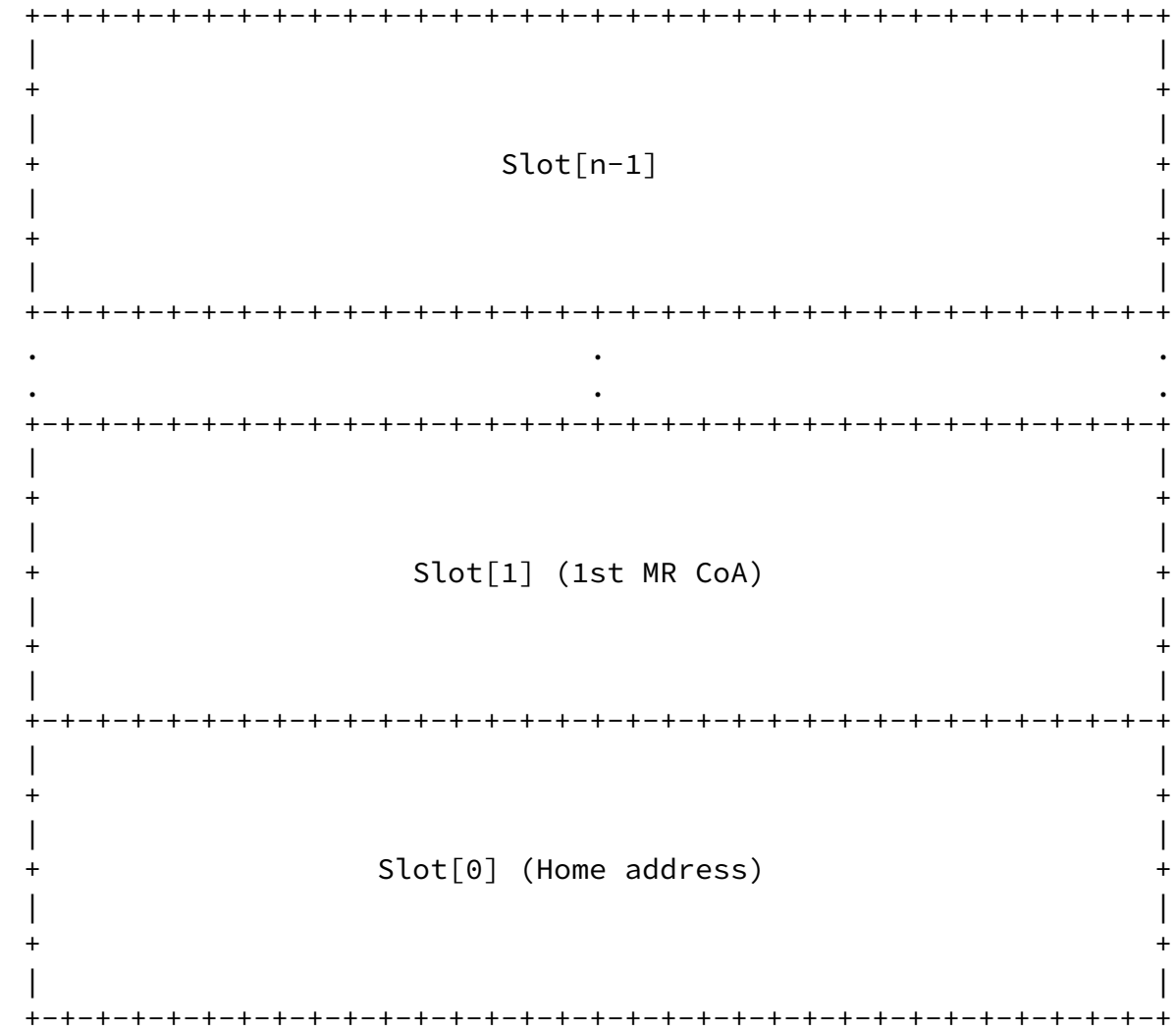
When a RRH is present in a packet, the rule for upper-layer checksum computing is that the source address used in the pseudo-header is that of the original source, located in the slot 0 of the RRH, unless the RRH slot 0 is empty, in which case the source in the IP header of the packet is used.

As the 'segment left' field of the generic RH is reassigned to the number of segments used, an IPv6 node that does not support RRH will discard the packet, unless the RRH is empty.

The RRH contains n pre-allocated address slots, to be filled by each MR in the path. It is possible to optimize the number of slots using the Tree Information Option described in Section 6.2.

The Type 4 Routing Header has the following format:





Next Header

8-bit selector. Identifies the type of header immediately following the Routing header. Uses the same values as the IPv4 Protocol field [13].

Hdr Ext Len

8-bit unsigned integer. Length of the Routing header in 8-octet units, not including the first 8 octets. For the Type 4 Routing header, Hdr Ext Len is equal to two times the number of addresses

in the header.

Routing Type

8-bit unsigned integer. Set to 4.

Segments Used

8-bit unsigned integer. Number of slots used. Initially set to 1 by the MR when only the Home Address is there. Incremented by the MRs on the way as they add the packets source addresses to the RRH.

Sequence Number

32-bit unsigned integer. The Sequence Number starts at 0, and is incremented by the source upon each individual packet. Using the Radia Perlman's lollipop algorithm, values between 0 and 255 are 'negative', left to indicate a reboot or the loss of HA connectivity, and are skipped when wrapping and upon positive Binding Ack. The sequence number is used to check the freshness of the RRH; anti-replay protection is left to IPsec AH.

Slot[n-1..0]

Vector of 128-bit addresses, numbered n-1 to 0.

When applied to the Nemo problem, the RRH can be used to update the HA on the actual location of the MR. Only MRs forwarding packets on an egress interface while not at home update it on the fly.

A RRH is inserted by the first MR on the Mobile Network outbound path, as part of the reverse tunnel encapsulation; it is removed by the associated HA when the tunneled packet is decapsulated.

4.3 Extension Header order

The RH type 2 is to be placed as any RH as described in [10] section 4.1. If a RH type 0 is present in the packet, then the RH type 2 is placed immediately after the RH type 0, and the RH type 0 MUST be consumed before the RH type 2.

RH type 3 and 4 are mutually exclusive. They are to be placed right after the Hop-by-Hop Options header if any, or else right after the IPv6 header.

As a result, the order prescribed in section 4.1 of RFC 2460 becomes:

IPv6 header

Hop-by-Hop Options header

Routing header type 3 or 4

Destination Options header (note 1)

Routing header type 0

Routing header type 2

Fragment header

Authentication header (note 2)

Encapsulating Security Payload header (note 2)

Destination Options header (note 3)

upper-layer header

5. ICMP

The RRH could have fewer slots than the number of MRs in the path because either the nested Mobile Network topology is changing too quickly or the MR that inserted the RRH could have a wrong representation of the topology.

To solve this problem a new ICMP message is introduced, "RRH Warning", type 64. Note that this ICMP message creates a new class of warning messages besides the error messages and the control messages of ICMP.

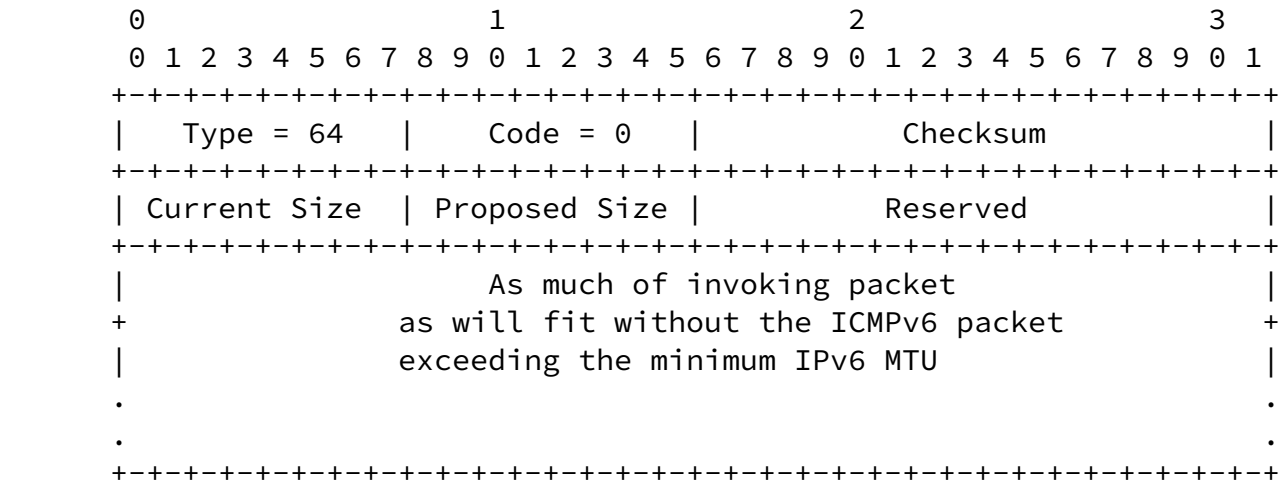
This message allows a MR on the path to propose a larger number of slots to the MR that creates the RRH. The Proposed Size MUST be larger than the current size and MUST NOT be larger than 8.

The originating MR must rate-limit the ICMP messages to avoid excessive ICMP traffic in the case of the source failing to operate as requested.

The originating MR must insert an RH type 2 based on the RRH in the associated IP header, in order to route the ICMP message back to the source of the reverse tunnel. A MR that receives this ICMP message is the actual destination and it MUST NOT forward it to the (LFN) source of the tunneled packet.

A MR on the path that finds no more space in the RRH SHOULD send an ICMP "RRH warning" back to the MR that inserted the RRH. On the other hand, a MR should always be able, by receiving TI option with up to date tree depth (see Section Section 6.2). to correctly size the RRH to insert in an outgoing packet.

The type 64 ICMP has the following format:



Type

64 [To Be Assigned]

Code 0: RRH too small

The originating MR requires the source to set the RRH size to a larger value. The packet that triggered the ICMP will still be forwarded by the MR, but the path cannot be totally optimized (see Section 9.3).

Checksum

The ICMP checksum [12].

Current Size

RRH size of the invoking packet, as a reference.

Proposed Size

The new value, expressed as a number of IPv6 addresses that can fit in the RRH.

Reserved

16-bit reserved field. Initialized to zero for transmission; ignored on reception.

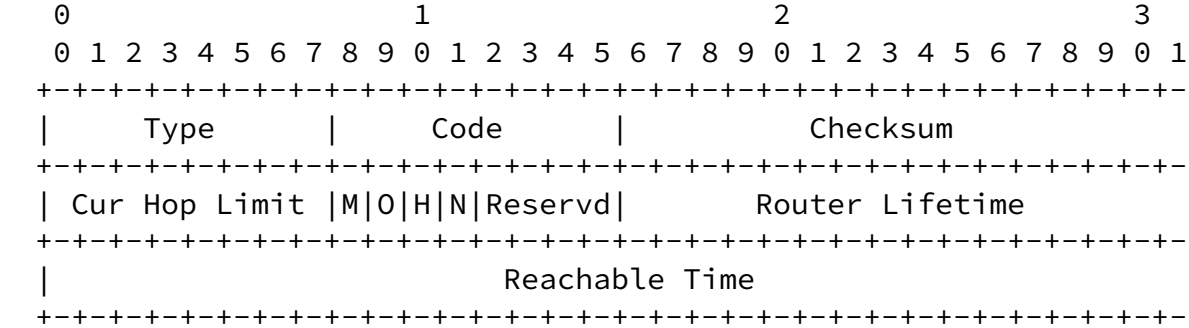
6. Modifications to IPv6 Neighbor Discovery

6.1 Modified Router Advertisement Message Format

Mobile IPv6 [1] modifies the format of the Router Advertisement message [11] by the addition of a single flag bit (H) to indicate that the router sending the Advertisement message is serving as a home agent on this link.

This draft adds another single flag bit (N) to indicate that the router sending the advertisement message is a MR. This means that the link on which the message is sent is a Mobile Network, which may or may not be at home.

The Router Advertisement message has the following format:



```
|                          Retrans Timer                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Options ...
+-+-+-+-+-+-+-+-+-+-+-
```

This format represents the following changes over that originally specified for Neighbor Discovery [11]:

Home Agent (H)

The Home Agent (H) bit is set in a Router Advertisement to indicate that the router sending this Router Advertisement is also functioning as a Mobile IP home agent on this link.

NEMO Capable (N)

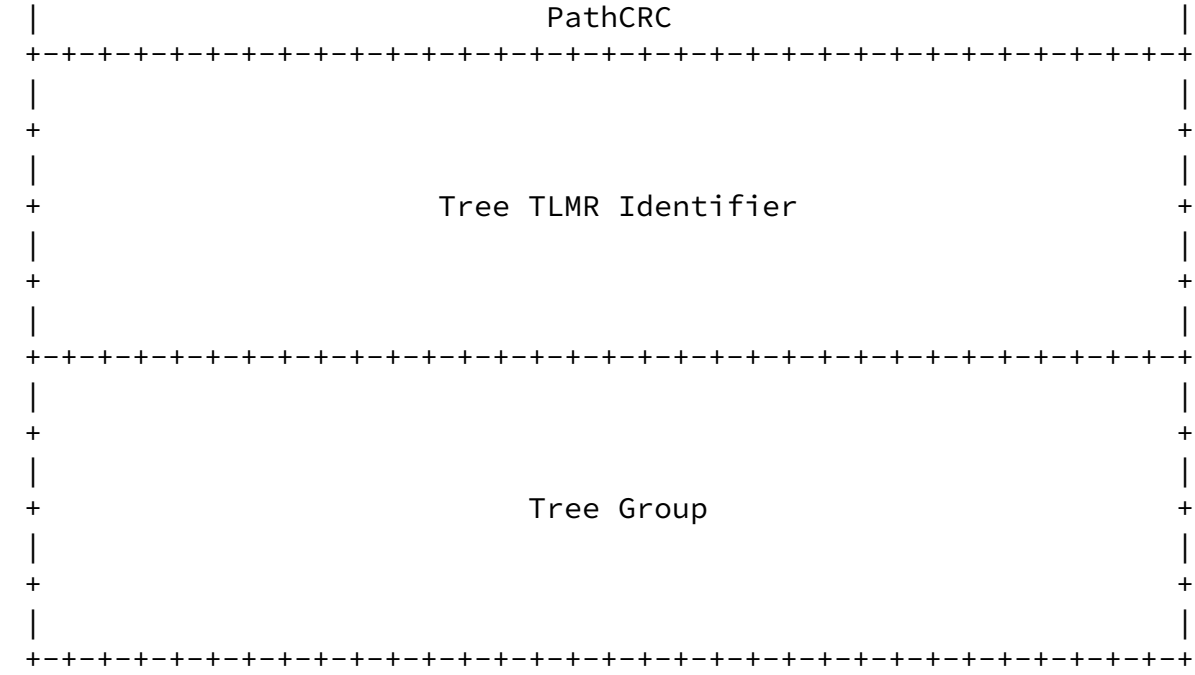
The NEMO Capable (N) bit is set in a Router Advertisement to indicate that the router sending this Router Advertisement is also functioning as a Mobile Router on this link, so that the link is a Mobile Network, possibly away from home.

6.2 New Tree Information Option Format

This draft defines a new Tree Information option, used in Router Advertisement messages. Fields set by the TLMR are propagated transparently by the MRs. Mobile Routers SHOULD add that option to the Router Advertisement messages sent over the ingress interfaces.

The Tree Information option has the following format:

```
 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 = 6   | TreePreference|   TreeDepth   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|H| Reserved  |   Bandwidth   |           DelayTime           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| MRPreference  |               BootTimeRandom                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```



Type

8-bit unsigned integer set to 10 by the TLMR.

Length

8-bit unsigned integer set to 6 by the TLMR. The length of the option (including the type and length fields) in units of 8 octets.

TreePreference

8-bit unsigned integer set by the TLMR to its configured preference. Range from 0 = lowest to 255 = highest.

TreeDepth

8-bit unsigned integer set to 0 by the TLMR and incremented by 1 by each MR down the tree.

Grounded (G)

1-bit flag. Set by the TLMR to indicate that it is either attached to a fixed network or at home.

Home Agent (H)

1-bit flag. Set by the TLMR to indicate that it is also functioning as a Home Agent, for re-homing purposes.

Reserved

6-bit unsigned integer, set to 0 by the TLMR.

Bandwidth

8-bit unsigned integer set by the TLMR and decremented by MRs with lower egress bandwidth. This is a power of 2 so that the available egress bandwidth in bps is between 2^Bandwidth and 2^(Bandwidth+1). 0 means 'unspecified' and can not be modified down the tree.

DelayTime

16-bit unsigned integer set by the TLMR. Tree time constant in milliseconds.

MRPreference

8-bit signed integer. Set by each MR to its configured preference. Range from 0 = lowest to 255 = highest.

BootTimeRandom

24-bit unsigned integer set by each MR to a random value that the MR generates at boot time.

PathCRC

32-bit unsigned integer CRC, updated by each MR. This is the result of a CRC-32c computation on a bit string obtained by

appending the received value and the MR CareOf Address. TLMRs use a 'previous value' of zeroes to initially set the pathCRC.

Tree TLMR Identifier

IPv6 global address, set by the TLMR. Identifier of the tree.

Tree Group

IPv6 global address, set by the TLMR. Identifier of the tree group. A MR may use the Tree Group in its tree selection algorithm.

The TLMR MUST include this option in its Router Advertisements.

A MR receiving this option from its Attachment Router MUST update the TreeDepth, MRPreference, BootTimeRandom and PathCRC fields, and MUST propagate it on its ingress interface(s), as described in Section 9.4.

The alignment requirement of the Tree Information option is 8n.

7. Binding Cache Management

7.1 Binding Updates

A MIPv6 or NEmo Binding Update provides more information than just the path in the nested cloud so they are still used as described in MIPv6 [1] for Home Registration and de-registration. The only difference when using RRH is that the Home Address Destination Option and the alternate CareOf MIP option MUST be omitted.

7.2 RRH Heartbeat

Intermediate updates (or just refreshes) between BUs are obtained as one of the results of processing the RRH by the HA.

When the MR becomes aware of a topology change in the tree (for examples it changes point of attachment, it obtains a new CoA, it receives a Tree Information Option in an RA message that indicates a change in the attachment tree) or in the absence of traffic (detected by a timeout) to the HA, it must send an RRH Heartbeat, which is a binding update with a RRH.

8. Home Agent Operation

This section inherits from chapter 10 of MIPv6 [1], which is kept unmodified except for parts 10.5 and 10.6 which are extended. This draft mostly adds the opportunity for a MN to update the Binding Cache of its Home Agent using RRH, though it does not change the fact that MNs still need to select a home agent, register and deregister to it, using the MIP Bind Update.

This draft extends [1] section 10.6 as follows:

- The entry point of the tunnel is now checked against the TLMR as opposed to the primary CoA.

- The Binding Cache can be updated based on RRH with proper AH authentication.

As further explained in Section 7.1, this specification modifies MIP so that the HA can rely on the RH type 4 (RRH) to update its Bind Cache Entry (BCE), when the Mobile Node moves. The conceptual content of the BCE is extended to contain a sequence counter, and the sequence of hops within the --potentially nested-- Mobile Network to a given Mobile Node. The sequence counter is initially set to 0.

When the HA receives a packet destined to itself, it checks for the presence of a Routing Header of type 3 or 4. Both contain as least the entry for the home address of the MN in slot 0; this replaces the MIP Home Address Option and allows the HA to determine the actual source of the packet, to access the corresponding security association.

As explained in Section 11.2, the HA MUST verify the authenticity of the packet using IPSEC AH and drop packets that were not issued by the proper Mobile Node. An RRH is considered only if the packet is authenticated and if its sequence number is higher than the one saved in the BCE.

Also, an RRH is considered only if an initial Bind Update exchange has been successfully completed between the Mobile Node and its Home Agent for Home Registration. If the RRH is valid, then the Bind Cache

Entry is revalidated for a lifetime as configured from the initial Bind Update.

The BCE abstract data is updated as follows:

> The first hop for the return path is the last hop on the path of the incoming packet, that is between the HA and the Top Level Mobile Router (TLMR) of the Mobile Network. The HA saves the IP address of the TLMR from the source field in the IP header.
>
> The rest of the path to the MN is found in the RRH.
>
> The sequence counter semantics is changed as described in Section 4.2

This draft extends [1] section 10.5 as follows:

> A Home Agent advertises the prefixes of its registered Mobile Routers, during the registration period, on the local Interior Gateway Protocol (IGP).
>
> The Routing Header type 2 is extended to be multi-hop.

The Home Agent is extended to support routes to prefixes that are owned by Mobile Routers. This can be configured statically, or can be exchanged using a routing protocol as in [3], which is out of the scope of this document. As a consequence of this process, the Home Agent which is selected by a Mobile Router advertises reachability of the MR prefixes for the duration of the registration over the local IGP.

When a HA gets a packet for which the destination is a node behind a Mobile Router, it places the packet in the tunnel to the associated MR. This ends up with a packet which destination address in the IP Header is the TLMR, and with a Routing Header of type 2 for the rest

of the way to the Mobile Router, which may be multi-hop.

To build the RH type 2 from the RRH, the HA sets the type to 2, and clears the bits 32-63 (byte 4 to 7).

9. Mobile Router Operation

This section inherits from chapter 11 of [1], which is extended to support Mobile Networks and Mobile Routers as a specific case of Mobile Node.

This draft extends section 11.2.1 of MIPv6 [1] as follows:

- When not at home, an MR uses a reverse tunnel with its HA for all the traffic that is sourced in its mobile network(s); traffic originated further down a nested network is not tunneled twice but for exception cases.

- The full path to and within the Mobile Network is piggy-backed with the traffic on a per-packet basis to cope with rapid movement. This makes the packet construction different from MIPv6.

The MR when not at home sets up a bi-directional tunnel with its HA. The reverse direction MR -> HA is needed to assure transparent topological correctness to LFNs, as in [3]. But, as opposed to that solution, nested tunnels are generally avoided.

9.1 Processing of ICMP "RRH too small"

The New ICMP message "RRH too Small" is presented in Section 5. This message is addressed to the MR which performs the tunnel

encapsulation and generates the RRH.

Hence, a MR that receives the ICMP "RRH too small" MUST NOT propagate it to the originating LFN or inner tunnel source, but MUST process it for itself.

If the Current Size in the ICMP messages matches the actual current number of slots in RRH, and if the ICMP passes some safety checks as described in Section 5, then the MR MAY adapt the number of slots to the Proposed Size.

9.2 Processing of ICMP error

```
ICMP back {

   if RRH is present {
      compute RH type 2 based on RRH
      get packet source from IP header
      send ICMP error to source including RH type 2.
   }
   else {
      get packet source from IP header
      send ICMP error to source with no RH.
   }
}
```

When the MR receives an ICMP error message, it checks whether it is the final destination of the packet by looking at the included

packet. If the included packet has an RRH, then the MR will use the RRH to forward the ICMP to the original source of the packet.

9.3 Processing of RHH for Outbound Packets

```
if no RRH in outer header              /* First Mobile Router specific */
   or RRH present but saturated {      /* Need a nested encapsulation */

   if RRH is saturated {
      do ICMP back (RRH too small)
   }

   /* put packet in sliding reverse tunnel */
   insert new IP header plus RRH
   set source address to the MR Home Address
   set destination address to the MR Home Agent Address
   add an RRH with all slots zeroed out
   compute IPsec AH on the resulting packet
}

/* All MRs including first */
if packet size <= MTU {
   select first free slot in RRH bottom up
   set it to source address from IP header
   overwrite source address in IP header with MR CareOf
   transmit packet
} else {
   do ICMP back (Packet too Big)
}
```


If the packet already contains an RRH in the outer header, and has a spare slot, the MR adds the source address from the packet IP header to the RRH and overwrites the source address in the IP header with its CoA. As a result, the packets are always topologically correct.

Else, if the RRH is present but is saturated, and therefore the source IP can not be added, the MR sends a ICMP 'RRH too small' to the tunnel endpoint which originated the outer packet, using the RRH info to route it back. The ICMP message is a warning, and the packet is not discarded. Rather, the MR does a nested encapsulation of the packet in its own reverse tunnel home with an additional RRH.

Else, if the packet does not have an RRH, the MR puts it in its reverse tunnel, sourced at the CoA, with an RRH indicating in slot 0 the Home Address of the MR, and with proper IPsec AH as described further in Section 11.1.

9.4 Processing of Tree Information Option

The Tree Information option in Router Advertisement messages allows the Mobile Router to select a tree and learn about its capabilities. The treeDepth can be used to compute the optimum number of slots in the RRH.

The RRH contains an entry for the home address in slot 0, and one for every CareOf on the way but that of the last Mobile Router (TLMR). As the TLMR sets the treeDepth to 0 and each MR increments it on the way down the tree, the optimum number of slots is normally (treeDepth+1), where treeDepth is the depth advertised by the MR over its Mobile Networks.

9.5 Processing of the extended Routing Header Type 2

```
if Segments Left = 0 {

   /* new check: packet must be looped back internally */
   if packet doesn't come from a loopback interface {
       discard the packet
       return
   }

   proceed to process the next header in the packet, whose type is
   identified by the Next Header field in the Routing header
}
else if Hdr Ext Len is odd {
      send an ICMP Parameter Problem, Code 0, message to the Source
      Address, pointing to the Hdr Ext Len field, and discard the
      packet
```



```
}
else {
   compute n, the number of addresses in the Routing header, by
   dividing Hdr Ext Len by 2
```

```
      if Segments Left is greater than n {
         send an ICMP Parameter Problem, Code 0, message to the Source
         Address, pointing to the Segments Left field, and discard the
         packet
      }
      else {
         decrement Segments Left by 1;

         compute i, the index of the next address to be visited in
         the address vector, by subtracting Segments Left from n

         if Address [i] or the IPv6 Destination Address is multicast {
            discard the packet
         }
         else {
            /* new security check */
            if Address [i] doesn't belong to one of the Mobile Network prefixes
                discard the packet
                return
            }

            /* new check: keep MIPv6 behavior: prevent packets from being
             * forwarded outside the node.
             */
            if Segments Left equals 0 and Address[i] isn't the node's own
            home address {
                discard the packet
                return
            }
            swap the IPv6 Destination Address and Address[i]
            if the IPv6 Hop Limit is less than or equal to 1 {
               send an ICMP Time Exceeded -- Hop Limit Exceeded in
               Transit message to the Source Address and discard the
               packet
            }
            else {
               decrement the Hop Limit by 1
               resubmit the packet to the IPv6 module for transmission
               to the new destination;
            }
         }
      }
   }
```

9.6 Decapsulation

A MR when decapsulating a packet from its HA must perform the following checks

1. Destination address

   The destination address of the inner packet must belong to one of the Mobile Network prefixes.

10. Mobile Host Operation

When it is at Home, a Mobile Host issues packets with source set to its home address and with destination set to its CN, in a plain IPv6 format.

When a MH is not at home but is attached to a foreign link in the Fixed Infrastructure, it SHOULD use MIPv6 as opposed to this draft to manage its mobility.

When a MH is visiting a foreign Mobile Network, it forwards its outbound packets over the reverse tunnel (including RRH) to its HA. One can view that operation as a first MR process applied on a plain IPv6 packet issued by a LFN.

As a result, the encapsulating header include:

with source set to the MH COA and destination set to the MH HA

with slot 0 set to the MH Home Address

The inner packet is the plain IPv6 packet from the MH Home Address to the CN.

11. Security Considerations

This section is not complete; further work is needed to analyse and solve the security problems of record and source route.

Compared to MIPv6, the main security problem seems to be the fact that the RRH can be modified in transit by an attacker on the path. It has to be noted that such an attacker (for example any MR in the Mobile Network) can perform more effective attacks than modifying the RRH.

11.1 IPsec Processing

The IPsec [7] AH [8] and ESP [9] can be used in tunnel mode to provide different security services to the tunnel between a MR and its HA. ESP tunnel mode SHOULD be used to provide confidentiality and authentication to the inner packet. AH tunnel mode MUST be used to provide authentication of the outer IP header fields, especially the Routing Headers.

11.1.1 Routing Header type 2

Due to the possible usage of Doors [5] to enable IPv4 traversal, the Routing Header type 2 cannot be treated as type 0 for the purpose of IPsec processing (i.e. it cannot be included in its intierity in the Integrity Check Value (ICV) computation, because NAT/PAT may mangle one of the MR care-of-addresses along the HA-MR path.

The sender (the HA) will put the slot 0 entry (the MR Home Address) of the RH as destination of the outer packet, will zero out completely the Routing Header and will perform the ICV computation.

The receiver (the MR) will put the slot 0 entry as destination of the outer packet, will zero out the Routing Header and will perform the ICV verification.

11.1.2 Routing Header type 4

The Routing Header type 4 is "partially mutable", and as such can be included in the Authentication Data calculation. Given the way type 4 is processed, the sender cannot order the field so that it appears as it will at the receiver; this means the receiver will have to shuffle the fields.

The sender (the MR) will zero out all the slots and the Segment Used field of the RRH, and will put as source address of the outer packet its Home Address, and then will perform the ICV computation.

The receiver (the HA) will put the entry in slot 0 (the MR Home Address) in the source address and will zero out all the slots and the Segment Used field of the RRH, and then will perform the ICV verification.

11.2 New Threats

The RH type 4 is used to construct a MIPv6 RH type 2 with additional semantics, as described in Section 4.1. Since RH type 2 becomes a multi hop option like RH type 0, care must be applied to avoid the spoofing attack that can be performed with the IPv4 source route option. This is why IPv6 [10] takes special care in responding to packets carrying Routing Headers.

AH authenticates the MR Home Address identity and the RRH sequence number. The RRH sequence number is to be used to check the freshness of the RRH; anti-replay protection can be obtained if the receiver enables the anti-replay service of AH [8].

In particular, if IPSec is being used, the content is protected and can not be read or modified, so there is no point in redirecting the traffic just to screen it.

Say a MR in a nested structure modifies the RRH in order to bomb a target outside of the tree. If that MR forwards the packet with itself as source address, the MR above it will make sure that the response packets come back to the attacker first, since that source is prepended to the RRH. If it forges the source address, then the ingress filtering at the MR above it should detect the irregularity and drop the packet. Same if the attacker is actually TLMR. The conclusion is that ingress filtering is recommended at MR and AR.

Say that an attacker in the infrastructure and on the path of the MRHA tunnel modifies the RRH in order to redirect the response packets and bomb a target. Considering the position of the attacker - a compromised access or core router - there's a lot more it could do to send perturbations to the traffic, like changing source and destinations of packets on the fly or eventually polute the routing protocols.

Say a MR in a nested structure modifies the RH 2 in order to attack a target outside of the tree. The RH type 2 forwarding rules make sure that the packet can only go down a tree. So unless the attacker is TLMR, the packet will not be forwarded. In any case, the attacker will be bombed first.

Say that an attacker on the path of the MRHA tunnel modifies the RRH in order to black out the MR. The result could actually be accomplished by changing any bit in the packet since the IPSec signature would fail, or scrambling the radio waves in the case of wireless.

Selecting the tree to attach to is a security critical operation

outside of the scope of this draft. Note that the MR should not select a path based on trust but rather on measured service. If a better bandwidth is obtained via an untrusted access using IPSec, isn't it better than a good willing low bandwidth trusted access?

12. Acknowledgements

The authors wish to thank David Auerbach, Fred Baker, Dana Blair, Steve Deering, Dave Forster, Thomas Fossati, Francois Le Faucheur, Kent Leung, Massimo Lucchina, Vincent Ribiere, Dan Shell and Patrick Wetterwald -last but not least :)-.

References

[1] Johnson, D., Perkins, C. and J. Arkko, "Mobility Support in IPv6", draft-ietf-mobileip-ipv6-24 (work in progress), July 2003.

[2] Ernst, T. and H. Lach, "Network Mobility Support Terminology", draft-ietf-nemo-terminology-00 (work in progress), May 2003.

[3] kniveton, t., "Mobile Router Support with Mobile IP", draft-kniveton-mobrtr-02 (work in progress), July 2002.

[4] Deering, S. and B. Zill, "Redundant Address Deletion when Encapsulating IPv6 in IPv6", draft-deering-ipv6-encap-addr-deletion-00 (work in progress), November 2001.

[5] Thubert, P., Molteni, M. and P. Wetterwald, "IPv4 traversal for MIPv6 based Mobile Routers", draft-thubert-nemo-ipv4-traversal-01 (work in progress), May 2003.

[6] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

[7] Kent, S. and R. Atkinson, "Security Architecture for the Internet Protocol", RFC 2401, November 1998.

[8] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.

[9] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, November 1998.

[10] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[11] Narten, T., Nordmark, E. and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, December 1998.

[12] Conta, A. and S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 2463, December 1998.

[13] Reynolds, J., "Assigned Numbers: RFC 1700 is Replaced by an On-line Database", RFC 3232, January 2002.

Authors' Addresses

Pascal Thubert
Cisco Systems Technology Center
Village d'Entreprises Green Side
400, Avenue Roumanille
Biot - Sophia Antipolis 06410
FRANCE

EMail: pthubert@cisco.com

Marco Molteni
Cisco Systems Technology Center

Village d'Entreprises Green Side
400, Avenue Roumanille
Biot - Sophia Antipolis 06410
FRANCE

EMail: mmolteni@cisco.com

Appendix A. Optimizations

A.1 Path Optimization with RRH

The body of the draft presents RRH as a header that circulates in the reverse tunnel exclusively. The RRH format by itself has no such limitation. This section illustrates a potential optimization for

end-to-end traffic between a Mobile Network Node and its Correspondent Node.

The MNN determines that it is part of a Mobile Network by screening the Tree Information option in the RA messages from its Attachment Router. In particular, the MNN knows the TreeDepth as advertised by the AR. An initial test phase could be derived from MIPv6 to decide whether optimization with a given CN is possible.

When an MNN performs end-to-end optimization with a CN, the MNN inserts an empty RRH inside its packets, as opposed to tunneling them home, which is the default behavior of a Mobile Host as described in Section 10.

The number of slots in the RRH is initially the AR treeDepth plus 1, but all slots are clear as opposed to the MR process as described in Section 9. The source address in the header is the MNN address, and the destination is the CN.

The AR of the MNN is by definition an MR. Since an RRH is already present in the packet, the MR does not put the packets from the MNN on its reverse tunnel, but acts as an intermediate MR; it adds the source address of the packet (the MNN's address) in the RRH (in slot 0) and stamps its careOf instead in the IP header source address field. Recursively, all the MRs on a nested network trace in path in the RRH and take over the source IP.

The support required on the CN side extends MIPv6 in a way similar to the extension that this draft proposes for the HA side. The CN is required to parse the RRH when it is valid, refresh its BCE accordingly, and include an RH type 2 with the full path to its packets to the MNN.

Note that there is no Bind Update between the MNN and the CN. The RRH must be secured based on tokens exchanged in the test phase. For the sake of security, it may be necessary to add fields to the RRH or to add a separate option in the Mobility Header.

A.2 Packet Size Optimization

RRH allows to update the Correspondent BCE on a per packet basis, which is the highest resolution that we can achieve. While this may cope with highly mobile and nested configurations, it can also be an overkill in some situations.

The RRH comes at a cost: it requires processing in all intermediate Mobile Routers and in the Correspondent Node. Also, a RRH increases the packet size by more than the size of an IP address per hop in the Mobile Network.

This is why an additional Routing Header is proposed (type 3). The semantics of type 3 are very close to type 4 but:

- Type 3 has only one slot, for the Home Address of the source.
- When it can not add the source to the RH type 3 of an outbound packet, an intermediate MR:
  - MR MUST NOT send ICMP (RRH too small)
  - MUST NOT put the packet in a reverse tunnel

  Rather, it simply overwrites the source and forwards the packet up the tree as if the RRH had been properly updated.
- Since the path information is not available, the correspondent MUST NOT update its BCE based on the RH type 3. The CN (or HA) identifies the source from the entry in slot 0 and may reconstruct the initial packet using the CareOf in slot 1 as source for AH purposes.

```
   /* MR processing on outbound packet with RH type 3 support */
   {

      if no RH type 3 or 4 in outer header    /* First Mobile Router specific *
         or RH type 4 present but saturated { /* Need a nested encapsulation */

         if RRH is saturated {
            do ICMP back (RRH too small)
         }

         /* put packet in sliding reverse tunnel */
         insert new IP header plus RRH
         set source address to the MR Home Address
         set destination address to the MR Home Agent Address
         add an RRH with all slots zeroed out
         compute IPsec AH on the resulting packet
      }

      /* All MRs including first */
      if packet size > MTU {
         do ICMP back (Packet too Big)
      } else if RRH {
         select first free slot in RRH bottom up
         set it to source address from IP header
         overwrite source address in IP header with MR CareOf
         transmit packet
      } else if RH type 3 {
         if slot 0 is still free {
            /* this is end-to-end optimization */
            set it to source address from IP header
         }
         overwrite source address in IP header with MR CareOf
         transmit packet
      }
   }
```

A.2.1 Routing Header Type 3 (Home Address option replacement)

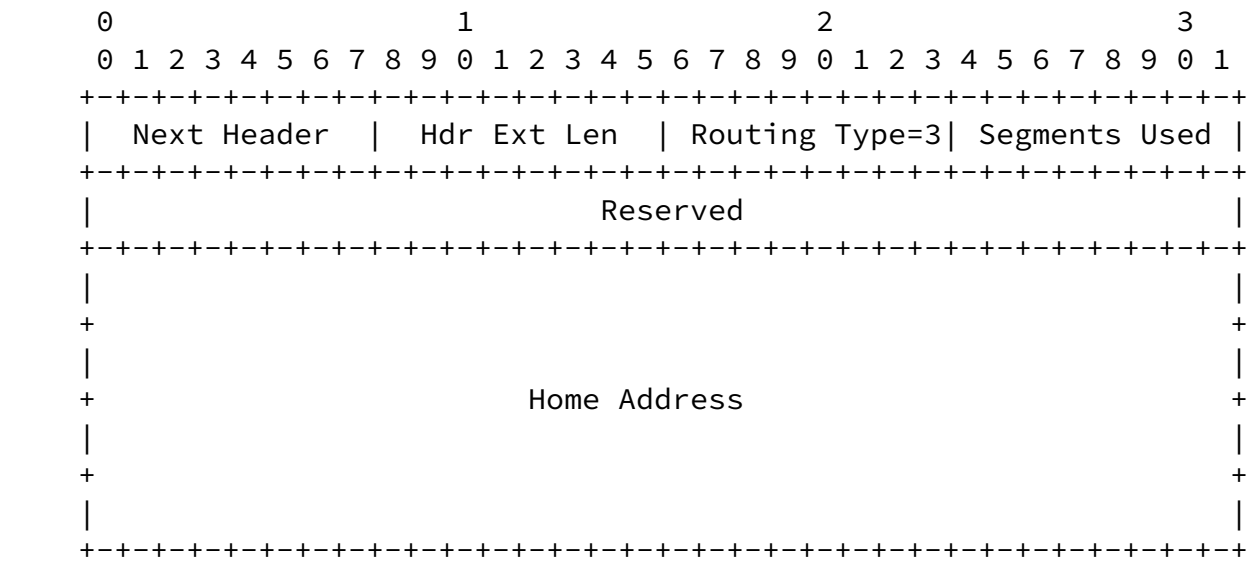
This is an RH-based alternative to the Home Address destination option. Its usage is described in Appendix A.2.

The decision to send RH type 3 or type 4 is up to the source of the

RRH. Several algorithms may apply, one out of N being the simplest.

IPsec HA processing is done as described in Section 11.1 for Type 4.

The Type 3 Routing Header has the following format:



Next Header

8-bit selector. Identifies the type of header immediately following the Routing header. Uses the same values as the IPv4 Protocol field [13].

Hdr Ext Len

8-bit unsigned integer. Length of the Routing header in 8-octet units, not including the first 8 octets. For the Type 3 Routing header, Hdr Ext Len is always 2.

Routing Type

8-bit unsigned integer. Set to 3.

Segment Used

8-bit unsigned integer. Number of slots used. Either 0 or 1. When the field is zero, then there is no MR on the path and it is valid for a CN that does not support RRH to ignore this header.

Reserved

32-bit reserved field. Initialized to zero for transmission; ignored on reception.

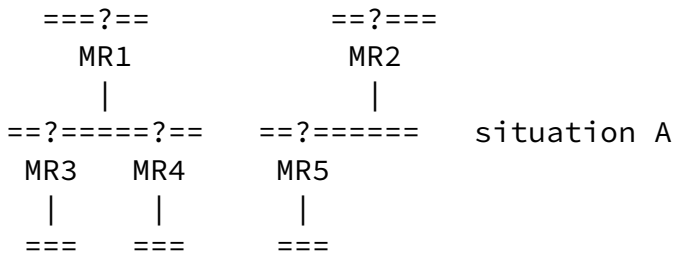
Home Address

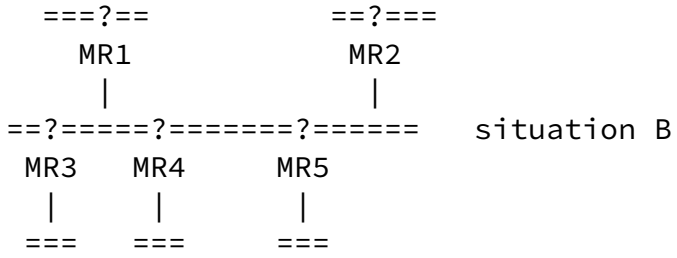
128-bit home address of the source of the packet.

Appendix B. Multi Homing

B.1 Multi-Homed Mobile Network

Consider difference between situation A and B in this diagram:





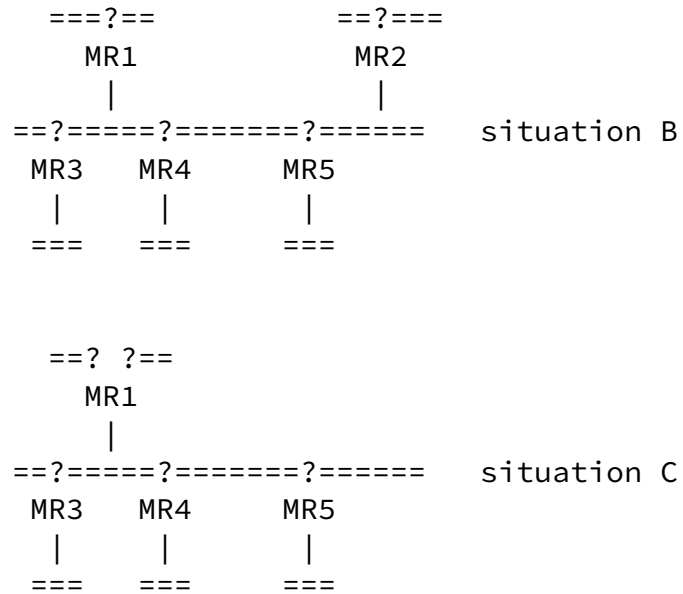
Going from A to B, MR5 may now choose between MR1 and MR2 for its Attachment (default) Router. In terms of Tree Information, MR5, as well as MR3 and MR4, now sees the MR1's tree and MR2's tree. Once MR5 selects its AR, MR2, say, MR5 belongs to the associated tree and whether MR1 can be reached or not makes no difference.

As long as each MR has a single default router for all its outbound traffic, 2 different logical trees can be mapped over the physical configurations in both situations, and once the trees are established, both cases are equivalent for the processing of RRH.

Note that MR5 MUST use a CareOf based on a prefix owned by its AR as source of the reverse tunnel, even if other prefixes are present on the Mobile Network, to ensure that a RH type 2 can be securely routed back.

B.2 Multihomed Mobile Router

Consider the difference between situation B and C in this diagram:



In situation C, MR2's egress interface and its properties are migrated to MR1. MR1 has now 2 different Home Addresses, 2 Home Agents, and 2 active interfaces.

If MR1 uses both CareOf addresses at a given point of time, and if they belong to different prefixes to be used via different attachment routers, then MR1 actually belongs to 2 trees. It must perform some routing logic to decide whether to forward packets on either egress interface. Also, it MUST advertise both tree information sets in its RA messages.

The difference between situations C and B is that when an attached router (MR5, say) selects a tree and forwards egress packets via MR1, it can not be sure that MR1 will actually forward the packets over that tree. If MR5 has selected a given tree for a specific reason, then a new source route header is needed to enforce that path on MR1.

The other way around, MR5 may leave the decision up to MR1. If MR1 uses the same attachment router for a given flow or at least a given destination, then the destination receives consistent RRHs. Otherwise, the BCE cache will flap, but as both paths are valid, the traffic still makes it through.

Appendix C. Changes from Previous Version of the Draft

From -03 to -04

TI option: renamed the F (fixed) flag bit to G (grounded).

Binding Update: Made clear that the BU flow conforms MIPv6 and Nemo but that RRH replaces both Home address Option and Alternate CareOf option.

From -02 to -03

Reworded the security part to remove an ambiguity that let the reader think that RRH is unsafe.

From -01 to -02

Made optional the usage of ICMP warning "RRH too small" (Section 5).

Changed the IPsec processing for Routing Header type 2 (Section 11.1).

From -00 to -01

Added new Tree Information Option fields:

A 8 bits Bandwidth indication that provides an idea of the egress bandwidth.

A CRC-32 that changes with the egress path out of the tree.

a 32 bits unsigned integer, built by each MR out of a high order configured preference and 24 bits random constant. This can help as a tie break in Attachment Router selection.

Reduced the 'negative' part of the lollipop space to 0..255

Fixed acknowledgements (sorry Patrick :)

Changed the type of Tree Information Option from 7 to 10.

Intellectual Property Statement

IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in BCP-11. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementors or users of this specification can be obtained from the IETF Secretariat.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this standard. Please address the information to the IETF Executive Director.

Full Copyright Statement

MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.