draft-sajassi-bess-l3-optimized-irb-02.txt

A. Sajassi (Cisco), C. Wang (Cisco),
K. Ananthamurthy (Cisco),
J. Rabadan (Nokia)

IETF 119, March 2024
Brisbane

Agenda

- Rev 00 (Presented in IETF 118)
  - Overall solution
- Rev 01
  - L3-Optimized IRB Option B – ARP Packet Optimization
  - L3-Optimized IRB Flag proposal
- Rev 02
  - New Caveats
  - MAC-Only and MAC-IP Route
  - IPv6 ND Message Handling

EVPN L3-Optimized IRB (Rev 00)

- H1 & H2 are in one subnet, and H3 in a different subnet
- H1 & H2 is bridged, H1/H2 and H3 is routed
- With L3-Optimized IRB, communication between H1 & H2 is also routed

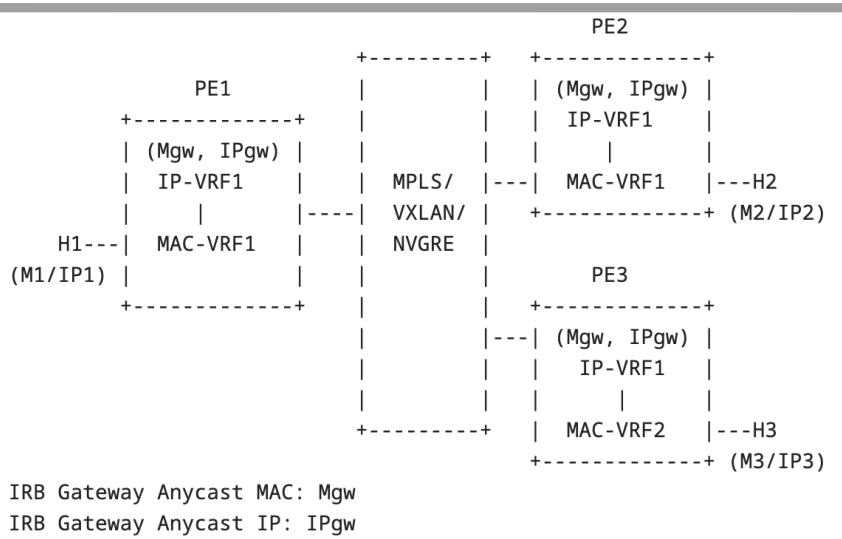


Figure 1: IRB Model with Distributed IRB Gateways

PE devices only need to maintain MAC addresses for locally-connected IP hosts, thus improving MAC scalability of customer bridges and PE devices significantly

L3-Optimized IRB Option A is proposed with Unconditional ARP response and host discovery via data packet gleaning. Interoperability with traditional-IRB is covered as well.

L3-Optimized IRB Option B (Rev 01)

- Conditional ARP response and host discovery via ARP Request Optimization
  - An L3-Optimized PE replies to an ARP Request Message from a locally connected host ONLY IF the target host is known to the PE
  - If the target host is unknown, the PE will NOT respond to the ARP Request immediately, instead, the PE re-originates the ARP Request to discover the target host

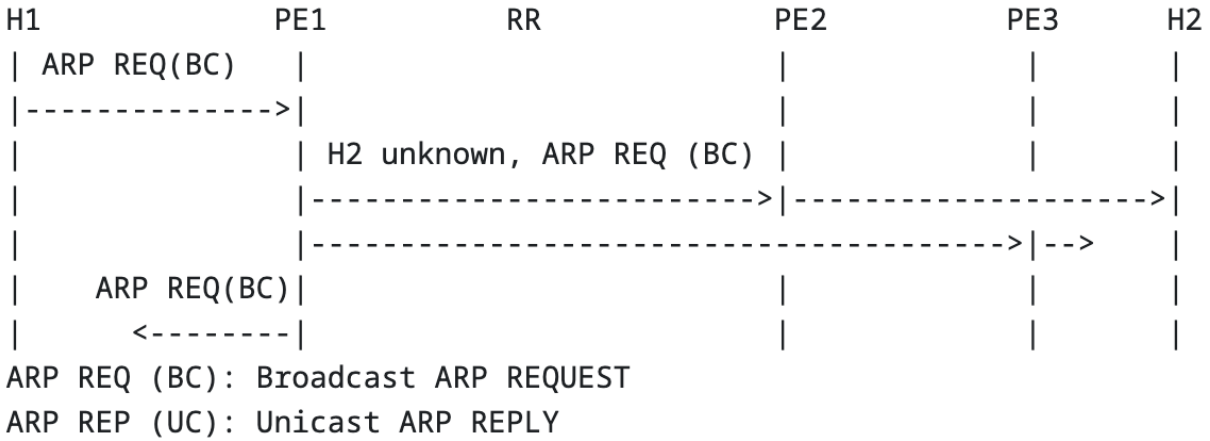


Figure 5: Host discovery via ARP Request Optimization (Option B)

L3-Optimized IRB Option B (Rev 01)

- Once the target host is learned via EVPN RT-2 route, it then becomes known to the PE and the PE can respond the the original or subsequent ARP Request

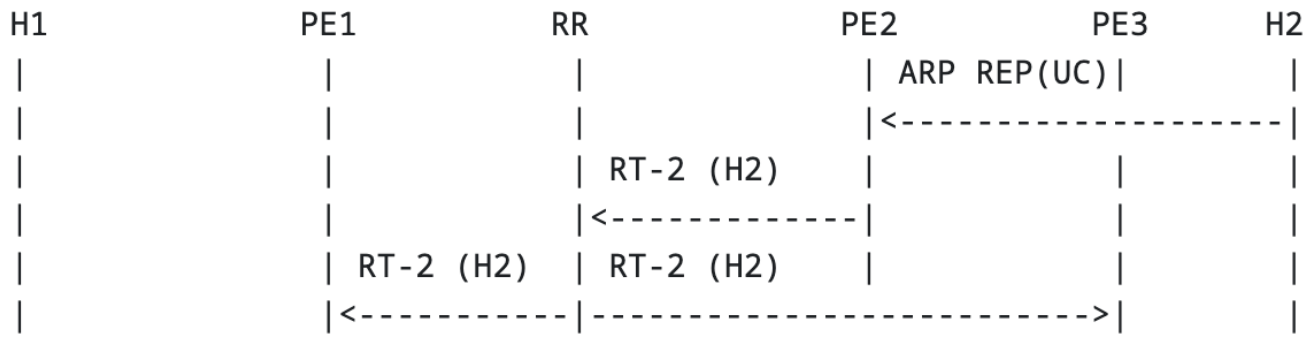


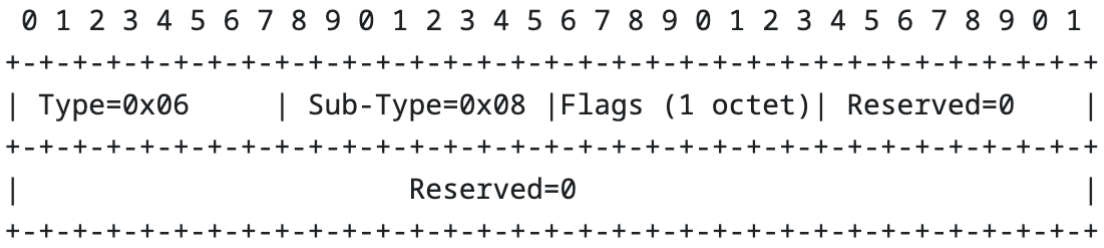
Figure 6: ARP Response from an IP host

L3-Optimized IRB Option B (Rev 01)

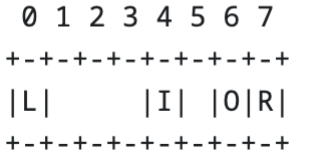
- Advantages of Option B
  - Solves the first data packet loss issue, which might be critical to some applications
  - Won’t create unnecessary ARP cache entries on CE side or attract traffic from the CE if the target host is unknown
  - Choosing Option A or Option B won’t introduce any interoperability problem, ultimately the target host is discovered by the ARP Request from the PE, either being originated (A) or re-originated (B)
  - Option B is the recommended approach despite that it’s relatively more complicated to implement

L3-Optimized IRB Flag proposal (Rev 01)

- This "L3-Optimized IRB flag" can be carried in an extended flag field in "EVPN ARP/ND Extended Community" (RFC 9047):



Flags field:



R: Router flag (corresponds to Bit 23 of the EC)
O: Override flag (corresponds to Bit 22 of the EC)
I: Immutable ARP/ND Binding flag (corresponds to Bit 20 of the Extended Community)

Proposed New flag (TBD)
L: L3-Optimized IRB flag (corresponds to Bit 16 of the EC)

Figure 2: EVPN ARP/ND Extended Community

New Caveats (Rev 02)

- On top of the existing Caveats to Consider, like TTL, Source MAC Rewrite, Subnet Broadcast, IPv6 LL and DAD, MAC Duplication Detection, Static ARP/ND, two new caveats are added:
  - IPv4 Address Conflict Detection (ARP Probe in RFC5227)
  - Neighbor Unreachability Detection (RFC4861) received from a traditional IRB PE

The above messages are not allowed to reach target host to avoid polluting CE MAC Table.

MAC-Only and MAC-IP Route (Rev 02)

- EVPN Allows advertising RT-2s with MAC address only and with MAC and IP address. For L3-Optimized IRB mode:
  - PE SHOULD NOT advertise MAC-Only routes
  - PE MUST advertise MAC-IP routes with L3-Optimized IRB Flag
  - If a PE chooses to advertise both MAC-Only and MAC-IP routes, then both MAC-Only and MAC-IP routes MUST carry L3-Optimized IRB Flag

IPv6 ND Message Handling (Rev 02)

- As described in section 2.2, IPv6 (or dual stack) ND behaviors are generally the same as ARP behaviors, with some specifics:
  - NS messages use multicast address, while ARP Requests are broadcast messages
  - NA generated on PE1 in response to NS from H1 to resolve a remote H2, always set to 1 the flags R (Router), O (Override) and S (Solicited)
  - Egress PE MUST set R (Router) and O (Override) flags to zero or one as per RFC9047 when it generates a RT-2 with L3-Optimized IRB set
  - Unsolicited NA from hosts will be handled the same way as Gratuitous ARP, and the PE will modify the R (Router) and O (Override) flags as per RFC9047 when it generates a RT-2 with L3-Optimized IRB set

Next Steps

- Solicit more input on the mailing list
- Asking for WG call

THANK YOU!

L3-Optimized IRB Option A (Rev 00)

- Unconditional ARP Reply and host discovery via data packet gleaning
  - An L3-Optimized PE replies to an ARP Request Message from a locally connected host unconditionally

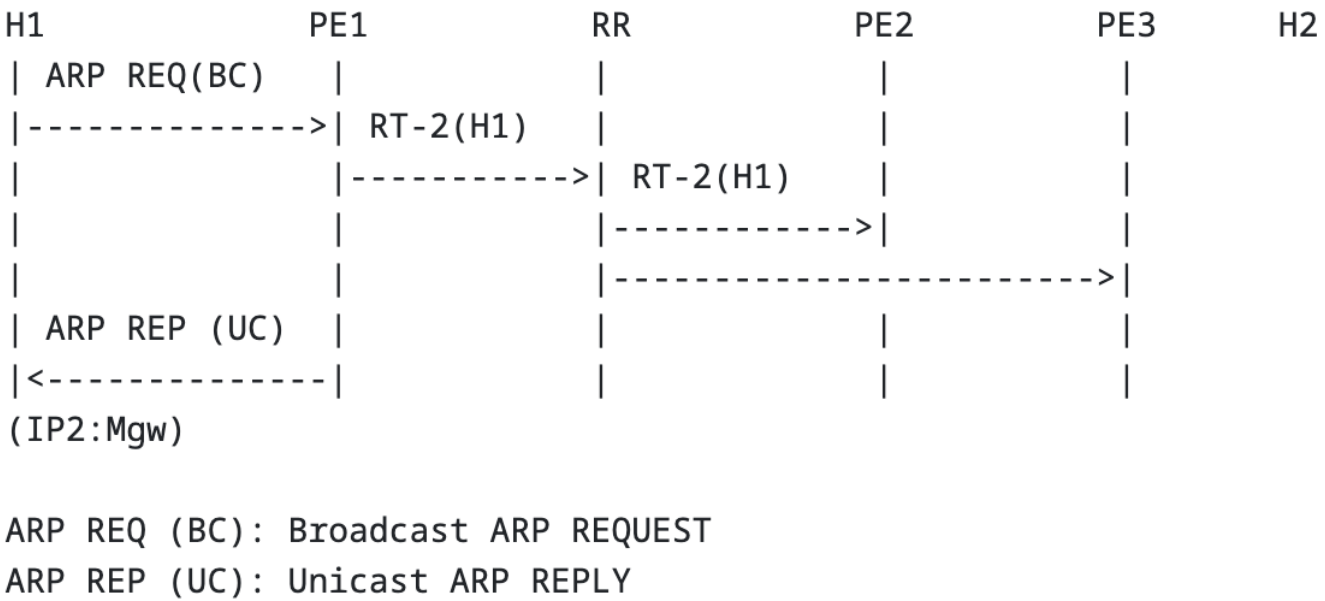


Figure 2: ARP Request from an IP Host

L3-Optimized IRB Option A (Rev 00)

- First data packet gets routed if the target host is known, otherwise the PE originates a new ARP Request to discover the target host
- Due to the data gleaning process, the data packet may get lost in control plane

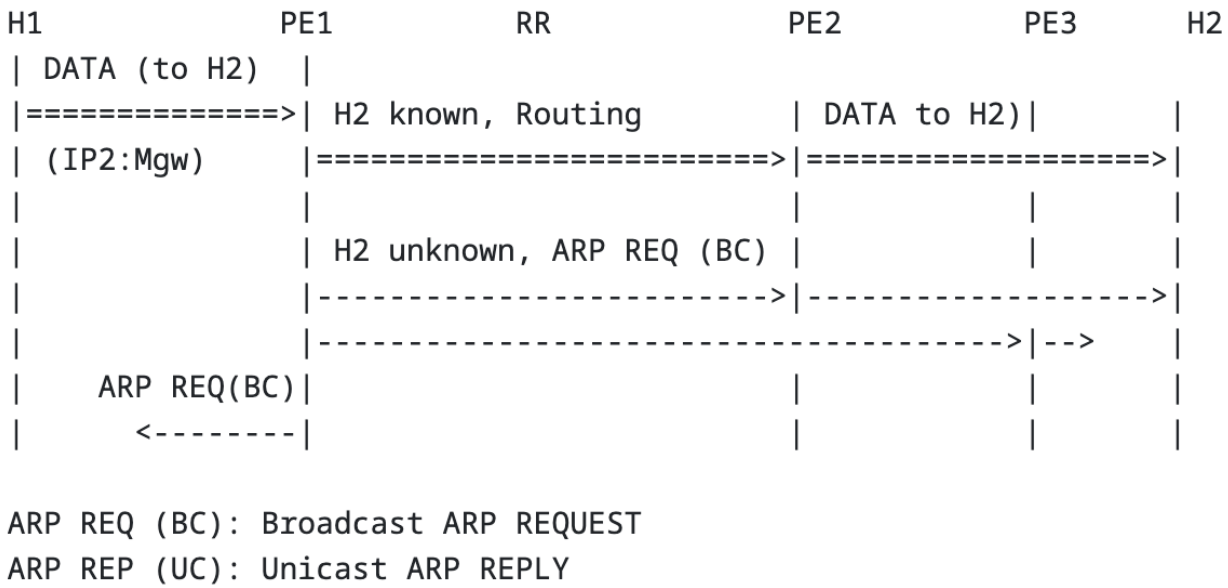


Figure 3: First data packet from an IP host

L3-Optimized IRB Option A (Rev 00)

- Once the target host is learned via EVPN RT-2 route, the target route is installed, and further data packets are getting routed

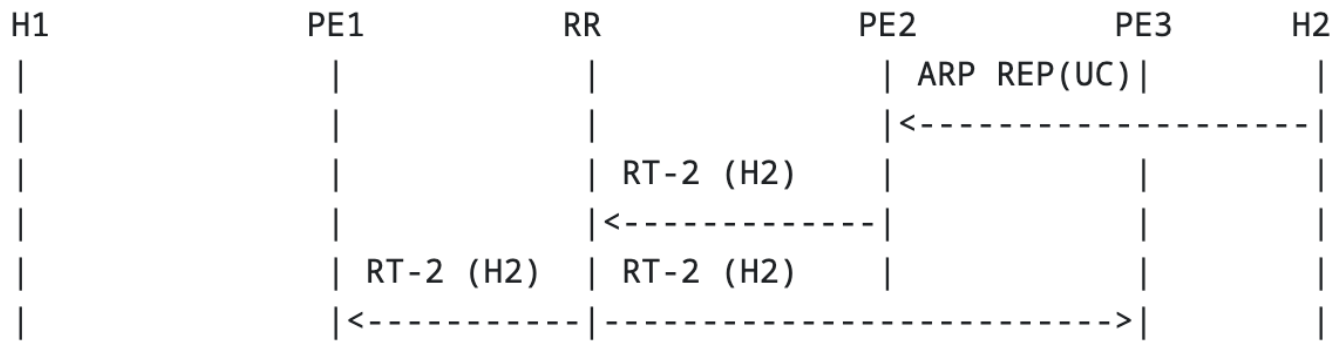


Figure 6: ARP Response from an IP host