Network Virtualization Overlays Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 21, 2013

Q. Wu
Huawei

February 17, 2013

# MAC address learning in NVO3 using ARP
## draft-wu-nvo3-mac-learning-arp-00

Abstract

[I.D-ietf-nvo3-framework] discusses using Dynamic data plane learning or control plane protocol to build and maintain the mapping tables and deliver encapsulated packets to their proper destination. However, there is no relevant work to discuss how those capabilities can be realized at the NVEs. This document goes into details to discuss how MAC address learning works through data plane and control plane.

Status of this Memo



Copyright Notice

Table of Contents

## 1. Introduction

[I.D-ietf-nvo3-framework] discusses using Dynamic data plane learning or control plane protocol to build and maintain the mapping tables and deliver encapsulated packets to their proper destination. However, there is no relevant work to discuss how those capability can be realized at the NVEs. This document goes into details to discuss how MAC address learning works through data plane and control plane.

## 2. Conventions used in this document

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

## 3. Overview of MAC address learning using ARP

This document addresses how to build and maintain mapping table at the NVE associated with the tenant system through data plane learning or control plane.

Figure 1 shows the example architecture for MAC learning using ARP. This example architecture assumes that:

- Tenant system A is connecting to VN by attaching to NVE X. Tenant System A knows IP address of Tenant System B but does not know MAC address of Tenant System B.

- Tenant system B is connecting to VN by attaching to NVE Y. Tenant System B knows IP address of Tenant System A but does not know MAC address of Tenant System A.

- NVE X associated with tenant system A doesn't know IP address and MAC address of tenant system B.

- NVE Y associated with tenant system B doesn't know IP address and MAC address of tenant system A.

```
                                 ,---------.
                               ,'  Backend  `.
                              (    Oracle    )
                               `.          ,'
                                 `-+------+'
                              +--+--+   +-+---+
                              |DC GW|+-+|DC GW|
                              +-+---+   +-----+
                                 |         |
                            .--..--. .--. ..
                           (    '          '.--.
                        .-.'        L3         '
                      (         Underlay       )
                       (                     '-'
    MAP Table entry  .'--'._.'.-._.'.-._)  MAP Table entry
  (TESID,MAC_X,IP_X) //               \\ (TESID,MAC_Y,IP_Y)
              +------+              +-------+
            .-|NVE X |              | NVE Y |
          (   +------+--.         ( +-------+.--.
        .-.'              '     .-.'             '
       (    DC Site X     )    (    DC Site Y     )
        (             .'-'      (             .'-'
         '--'._.'.    )          '--'._.'.    )
                 '--'              /     '--'
              /     \             /       \
           __/_      \          /_       _\__
     '--------'   '--------'   '--------'   '--------'
     : Tenant :   : Tenant :   : Tenant :   : Tenant :
     : SystemA:   : SystemC:   : SystemD:   : SystemB:
     '--------'   '--------'   '--------'   '--------'
       TESID=                                  TESID=
   (VNID,MAC_A,IP_A)                      (VNID,MAC_B,IP_B)
```

Figure 1: Figure 1 Example of MAC learning using ARP

## 4. MAC learning using ARP Resolution

MAC addresses of the Tenant systems also can be learnt by NVE through data plane and control plane. The following section outlines several examples for MAC learning using ARP resolution.

### 4.1. MAC learning using flooding without MAC hiding

The packet flow and control plane operation for MAC learning are as follows:

- o Tenant system A sends a broadcast ARP message to discover the MAC address of Tenant system B. The message contains IP_B in the ARP message payload.

- o The ARP proxy in NVE X, receiving the ARP message, will flood it on the overlay network for TESID = <VNID,IP_B,*>.

- o The ARP message will be intercepted by NVE (i.e., NVE Y) which maintain mapping table matching TESID = <VNID,IP_B,*>. NVE Y, will forward the ARP message to tenant system B. Tenant System B send ARP reply to tenant system A containing the mapping TESID=<VNID,IP_B,MAC_B>.

- o NVE X intercept ARP reply message and populates the map-table with the received entry, then send it to Tenant System A that includes MAC_B and IP_B of Tenant System B.

- o Tenant system A learns MAC_B from the ARP rely message and can now send a packet to Tenant system B by including MAC_B, and IP_B, as destination addresses.

### 4.2. MAC learning using NVE-oracle interaction

The packet flow and control plane operation for MAC learning are as follows:

- o Tenant System A sends a broadcast ARP message to discover the MAC address of Tenant system B. The message contains IP_B in the ARP message payload.

- o NVE A, receiving the ARP message, but rather than flooding it on the overlay network sends a Map-Request to the backend Oracle that maintains mapping information for entire overlay network for TESID = <VNID,IP_B,*>.

- o The Map-Request is routed by the backend Oracle to NVE Y, that will send a Map-Reply back to NVE X containing the mapping

TESID=<VNID,IP_B,MAC_B>. Alternatively, depending on the Backend Oracle configuration, the backend Oracle may send directly a Map-Reply to NVE X.

o NVE X populates the map-table with the received entry, and sends an ARP-Agent Reply to Tenant System A that includes MAC_B and IP_B.

o Tenant system A learns MAC_B from the ARP message and can now send a packet to Tenant system B by including MAC_B, and IP_B, as destination addresses.

### 4.3. MAC learning using control plane operation and MAC hiding

MAC addresses of the Tenant systems also can be learnt by NVE through control plane.

When tenant system A is attached to NVE X, the mapping table TESID=<VNID,IP_A,MAC_A> should be populated at the local NVE A. In order to enable tenant system A to communicate with any tenant system that is not under NVE X, the mapping table should be distributed to all the remote NVEs that belong to the same VN even through there is no tenant system which communicates with tenant system A behind the remote NVE. In order to achieve this, NVE X should know the list of remote NVE that belong to the same VN as NVE X and distribute the mapping table to each remote NVE. Alternatively, backend Oracle may know a list of tenant systems that is in communication with tenant system A and which remote NVE these tenant systems are attached to. So NVE X distribute the mapping table to the backend Oracle which in turn determine which remote NVE should populate mapping table and distribute mapping table to the corresponding remote NVE. The packet flow for MAC learning in data plane are as follows:

o Tenant system A sends a broadcast ARP message to discover the MAC address of Tenant system B. The message contains IP_B in the ARP message payload.

o The ARP proxy in NVE X, will terminate the ARP message, and create a ARP reply message, set the inner destination MAC address in the inner Ethernet header and sender MAC address in the payload of ARP reply message to NVE X's MAC address then send it back to tenant system A. Therefore ARP message is restricted within layer 2 network behind NVE X and will not be flooded to the entire overlay network at the outsider of NVE X.

o Tenant system A learns MAC_B from the ARP rely message and send a packet to Tenant system B by including MAC_X, and IP_B, as destination addresses.

o NVE X, will intercept the packet from tenant system A and perform a lookup operation in its map table for the destination TESID=<VNID, IP_B> and determine which tunnel the packet needs to be sent to. Then NVE X encapsulate the packet from tenant system A into tunnel header with NVE Y IP_Y as the destination address NVE X IP_X as the source address and transmit it across overlay network.

o NVE Y decapsulates the tunnel packet from NVE X and take out the packet from tenant system A and send to the tenant system B.

## 4.4. MAC learning using control plane operation without MAC hiding

MAC addresses of the Tenant systems also can be learnt by NVE through control plane.

When tenant system A is attached to NVE X, the mapping table TESID=<VNID,IP_A,MAC_A> should be populated at the local NVE A. In order to enable tenant system A to communicate with any tenant system that is not under NVE X, the mapping table should be distributed to all the remote NVEs that belong to the same VN even through there is no tenant system which communicate with tenant system A behind the remote NVE. In order to achieve this, NVE X should know the list of remote NVE that belong to the same VN as NVE X and distribute the mapping table to each remote NVE. Alternatively, backend Oracle may know a list of tenant systems that is in communication with tenant system A and which remote NVE these tenant systems are attached to. So NVE X distribute the mapping table to the backend Oracle which in turn determine which remote NVE should populate mapping table and distribute mapping table to the corresponding remote NVE. The packet flow for MAC learning in data plane are as follows:

o Tenant system A sends a broadcast ARP message to discover the MAC address of Tenant system B. The message contains IP_B in the ARP message payload.

o The ARP proxy in NVE X, will terminate the ARP message, and look up the MAC_B in the local mapping table send the ARP reply message to tenant system A that includes MAC_B and IP_B. Therefore ARP message is restricted within layer 2 network behind NVE X and will not be flooded to the entire overlay network at the outsider of NVE X.

o Tenant system A learns MAC_B from the ARP rely message and send a packet to Tenant system B by including MAC_B, and IP_B, as destination addresses.

- o NVE X, will intercept the packet from tenant system A and perform a lookup operation in its map table for the destination TESID=<VNID, IP_B> and determine which tunnel the packet needs to be sent to. Then NVE X encapsulate the packet from tenant system A into tunnel header with NVE Y IP_Y as the destination address NVE X IP_X as the source address and transmit it across overlay network.

- o NVE Y decapsulates the tunnel packet from NVE X and take out the packet from tenant system A and send to the tenant system B.

## 5. IANA Considerations

This document has no actions for IANA.

## 6. Security Considerations

TBC.

## 7. Normative References

[I.D-ietf-nvo3-framework]
Lasserre, M., "Framework for DC Network Virtualization", ID draft-ietf-nvo3-framework-00, September 2012.

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

Author's Address

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu  210012
China

Email: bill.wu@huawei.com