VXLAN Data Plane Learning Lab (Flood and Learn Mechanism)

Contents

Introduction

This document describes how to set up a CML with Nexus 9Kv switches using VXLAN with the Flood and Learn method.

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- Understanding of Routing and Switching
- Multicast routing concepts such as Rendezvous Point (RP) and Platform Independent Multicast (PIM)

Components Used

This document is not restricted to specific software and hardware versions.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Background Information

The document also provides guidance on deploying the lab, as well as verifying configurations and operations.

For this lab, the Cisco Modeling Lab (CML) with Nexus 9000V switches is utilized for both the Leaf and Spine.

Leaf1	Loopback0 - 1.1.1.1	Loopback1 - 10.10.10.10
Leaf2	Loopback0 - 2.2.2.2	Loopback1 - 20.20.20.20
Leaf3	Loopback0 - 3.3.3.3	Loopback1 - 30.30.30.30
Spine1	Loopback0 - 4.4.4.4	Loopback1 - 60.60.60.60 - Anycast RP
Spine2	Loopback0 - 5.5.5.5	Loopback1 - 60.60.60.60 - Anycast RP
Desktop subnet	192.168.100.0/24	

Terminlogies Used

Virtual eXtensible Local Area Network (VXLAN) Tunnel endpoint (VTEP) - Encapsulates MAC traffic into IP traffic and routes MAC traffic to other VTEPs.

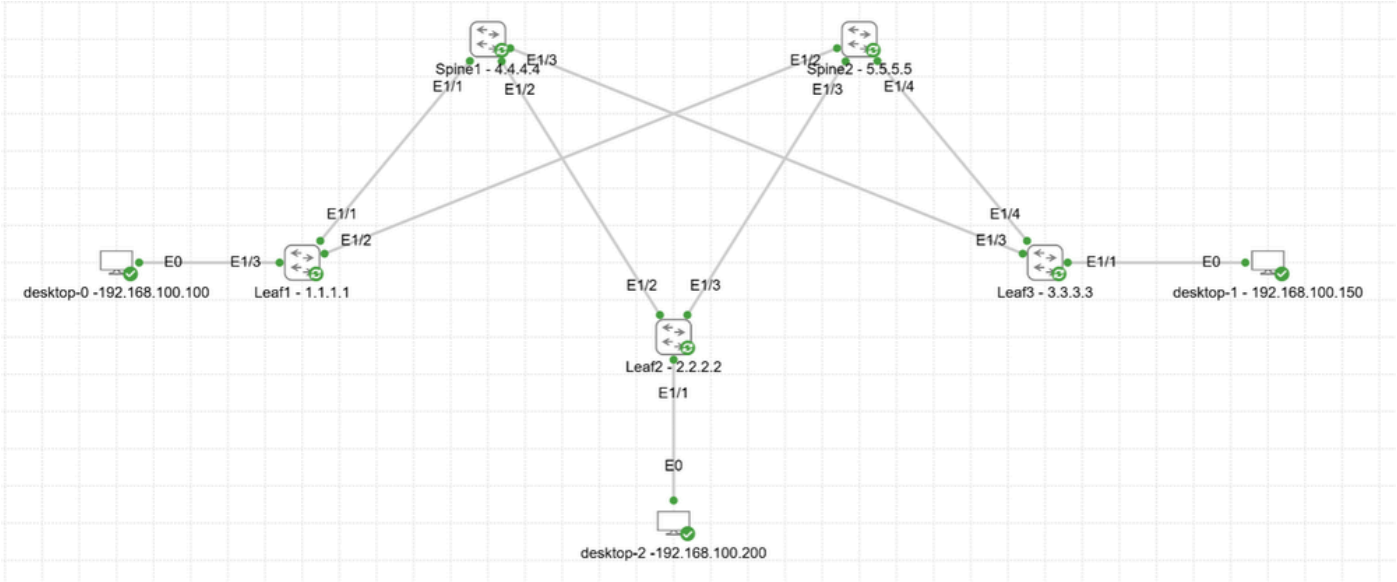
VXLAN Network Identifier (VNID) - ID within the VXLAN header which identifies the network and can be mapped to a VLAN. From a forwarding perspective, a VNID is a broadcast domain.

Network Virtual Interface (NVE) - Logical interface where the encapsulation and de-encapsulation occur.

Broadcast, Uknown unicast and multicast (BUM)

Configure

Network Diagram



Network Connectivity Diagram

Configurations

Step 1.

- Enable feature Open Shortest Path First (OSPF).
- Add loopbacks to all device.
- Enable OSPF on Ethernet interfaces and loopbacks.

```
leaf1# show  running-config interface  ethernet 1/1

!Command: show running-config interface Ethernet1/1
!Running configuration last done at: Tue Dec 24 13:12:55 2024
!Time: Wed Dec 25 05:24:23 2024

version 9.3(8) Bios:version

interface Ethernet1/1
  no switchport
  ip address 70.0.0.2/30
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
  no shutdown

leaf1# show  running-config interface  loopback 0

!Command: show running-config interface loopback0
!Running configuration last done at: Tue Dec 24 13:12:55 2024
!Time: Wed Dec 25 05:24:24 2024

version 9.3(8) Bios:version

interface loopback0
  ip address 1.1.1.1/32
  ip router ospf 1 area 0.0.0.0
```

Enabling OSPF on Leaf Interface

```
spine1# show  running-config interface ethernet 1/1

!Command: show running-config interface Ethernet1/1
!Running configuration last done at: Tue Dec 24 13:16:16 2024
!Time: Wed Dec 25 05:25:46 2024

version 9.3(8) Bios:version

interface Ethernet1/1
  no switchport
  ip address 70.0.0.1/30
  ip router ospf 5 area 0.0.0.0
  ip pim sparse-mode
  no shutdown

spine1# show  running-config interface loopback 0

!Command: show running-config interface loopback0
!Running configuration last done at: Tue Dec 24 13:16:16 2024
!Time: Wed Dec 25 05:25:51 2024

version 9.3(8) Bios:version

interface loopback0
  ip address 4.4.4.4/32
  ip router ospf 5 area 0.0.0.0
```

Enabling OSPF on Spine Interface

OSPF neighborship is established between Leaf and Spine switches.

```
spine1# show  ip ospf neighbors
 OSPF Process ID 5 VRF default
 Total number of neighbors: 3
 Neighbor ID     Pri State            Up Time  Address         Interface
 1.1.1.1           1 FULL/DR          16:22:51 70.0.0.2        Eth1/1
 2.2.2.2           1 FULL/DR          16:22:52 50.0.0.2        Eth1/2
 3.3.3.3           1 FULL/DR          16:22:52 30.0.0.2        Eth1/3
```

OSPF neighbor established with Leaf Switches

```
leaf1# show  ip ospf neighbors
 OSPF Process ID 1 VRF default
 Total number of neighbors: 2
 Neighbor ID     Pri State            Up Time  Address         Interface
 4.4.4.4           1 FULL/BDR         16:15:40 70.0.0.1        Eth1/1
 5.5.5.5           1 FULL/BDR         16:15:10 80.0.0.1        Eth1/2
```

OSPF neighbor established with Spine Switches

```
leaf1# ping 3.3.3.3 source-interface loopback 0
PING 3.3.3.3 (3.3.3.3): 56 data bytes
64 bytes from 3.3.3.3: icmp_seq=0 ttl=253 time=6.616 ms
64 bytes from 3.3.3.3: icmp_seq=1 ttl=253 time=6.695 ms
64 bytes from 3.3.3.3: icmp_seq=2 ttl=253 time=6.018 ms
64 bytes from 3.3.3.3: icmp_seq=3 ttl=253 time=6.52 ms
64 bytes from 3.3.3.3: icmp_seq=4 ttl=253 time=7.768 ms
```

Ping reachability from Leaf1 to Leaf3

Step 2.

Add an additional loopback interface that will be used for VXLAN on the leaf switches. Also, verify the reachability from all the leaf switches into the fabric.

```
leaf2# show  running-config interface loopback 1

!Command: show running-config interface loopback1
!Running configuration last done at: Wed Dec 25 05:41:46 2024
!Time: Wed Dec 25 05:42:33 2024

version 9.3(8) Bios:version

interface loopback1
  description Vxlan loopback
  ip address 20.20.20.20/32
  ip router ospf 2 area 0.0.0.0
```

Loopback for VXLAN

```
leaf2# ping 10.10.10.10 source 20.20.20.20
PING 10.10.10.10 (10.10.10.10) from 20.20.20.20: 56 data bytes
64 bytes from 10.10.10.10: icmp_seq=0 ttl=253 time=7.187 ms
64 bytes from 10.10.10.10: icmp_seq=1 ttl=253 time=6.248 ms
64 bytes from 10.10.10.10: icmp_seq=2 ttl=253 time=5.472 ms
64 bytes from 10.10.10.10: icmp_seq=3 ttl=253 time=4.741 ms
64 bytes from 10.10.10.10: icmp_seq=4 ttl=253 time=4.887 ms

--- 10.10.10.10 ping statistics ---
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min/avg/max = 4.741/5.707/7.187 ms
```

Overlay reachability from Leaf2 to Leaf1

Step 3.

Configure PIM Any-Source Multicast (ASM) and Anycast RP on the spines:

- Enable the PIM feature.

- Enable PIM on all the underlay links.
- Create new loopback on spines to be used for Anycast RP.
- Advertise this loopback into OSPF.
- Configure Anycast RP (Nexus feature) on spines.
- Configure the RP on all devices.

ANYCAST RP:

Anycast RP is a mechanism for providing rapid RP failover and RP load-sharing. Anycast RP involves using the same IP address (rp-address) on two or more routers that will function as the RP. This IP address must be advertised in the Interior Gateway Protocol (IGP) so that other routers can choose the best path to the rp-address. In the event of a failure, the convergence time will be the same as the IGP.

Having multiple RPs with the same IP address ensures that sources and receivers will always be routed to the nearest RP based on the unicast routing table. PIM Join messages from receivers can be sent to one RP, while PIM-designated routers register their local sources to another RP.
It is important to synchronize information between the different RPs because some senders and receivers can join router 1 as the RP, while others can join router 2 as the RP. If the routers do not have complete information about all the sources, multicast communication can be disrupted. In order to address this problem, a mechanism is required to synchronize information about sources among all the routers acting as an RP.
There are two protocols that can serve this purpose: Multicast Source Discovery Protocol (MSDP) and PIM.

```
spine1# show  running-config interface loopback 1

!Command: show running-config interface loopback1
!Running configuration last done at: Tue Dec 24 13:16:16 2024
!Time: Wed Dec 25 05:50:44 2024

version 9.3(8) Bios:version

interface loopback1
  description Anycast RP loopback
  ip address 60.60.60.60/32
  ip router ospf 5 area 0.0.0.0
  ip pim sparse-mode

spine1# show  running-config | section  rp
ip pim rp-address 60.60.60.60 group-list 224.0.0.0/4
ip pim anycast-rp 60.60.60.60 4.4.4.4
ip pim anycast-rp 60.60.60.60 5.5.5.5
```

60.60.60.60 is the Anycast RP IP and 4.4.4.4/5.5.5.5 are the Loopback IP of Spine 1 and Spine 2

```
leaf2# show  running-config interface  ethernet 1/2

!Command: show running-config interface Ethernet1/2
!Running configuration last done at: Wed Dec 25 05:41:46 2024
!Time: Wed Dec 25 05:51:18 2024

version 9.3(8) Bios:version

interface Ethernet1/2
  no switchport
  ip address 50.0.0.2/30
  ip router ospf 2 area 0.0.0.0
  ip pim sparse-mode
  no shutdown

leaf2# show  running-config | section rp-address
ip pim rp-address 60.60.60.60 group-list 224.0.0.0/4
```

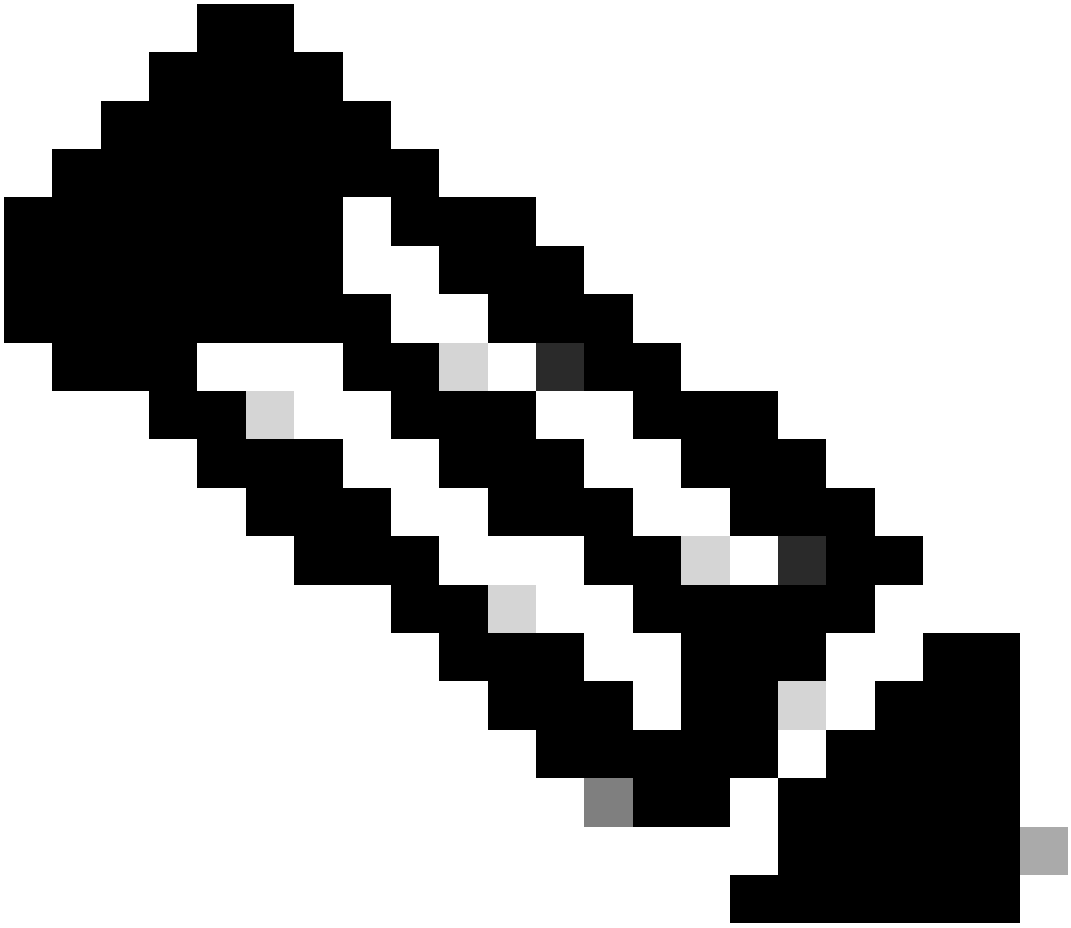
Anycast RP configuration on Leaf Switches

```
spine1# show  ip pim neighbor
PIM Neighbor Status for VRF "default"
Neighbor        Interface            Uptime    Expires   DR       Bidir-  BFD
 ECMP Redirect
                                                         Priority Capable State
    Capable
70.0.0.2        Ethernet1/1          16:44:28  00:01:39  1        yes     n/a
  no
50.0.0.2        Ethernet1/2          16:44:28  00:01:44  1        yes     n/a
  no
30.0.0.2        Ethernet1/3          16:44:28  00:01:25  1        yes     n/a
  no
spine1# show  ip pim  rp
PIM RP Status Information for VRF "default"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

Anycast-RP 60.60.60.60 members:
  4.4.4.4*  5.5.5.5

RP: 60.60.60.60*, (0),
 uptime: 16:57:06   priority: 255,
 RP-source: (local),
 group ranges:
 224.0.0.0/4
```

PIM neighborship formed



Note: Do not forget to also put PIM on the loopback to be used for VXLAN on leaf switches.

Step 4.

- Enable VXLAN feature.
- Enable feature in order to map VLANs to Virtual Network Identifiers (VNIs).
- Create NVE.
- Configure access port towards desktop.

```
feature vn-segment-vlan-based
feature nv overlay
```

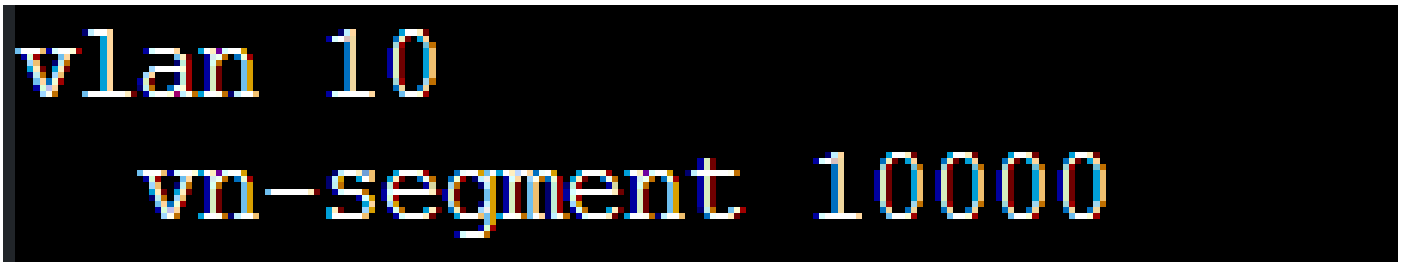
Enabling VXLAN Features

```
!Command: show running-config interface nve1
!Running configuration last done at: Wed Dec 25 06:08:01 2024
!Time: Wed Dec 25 06:08:04 2024

version 9.3(8) Bios:version

interface nve1
  no shutdown
  source-interface loopback1
  member vni 10000
    mcast-group 239.0.0.1
```

Creating NVE Interface



VLAN to VN Segment Mapping

```
leaf1# show  interface nve 1
nve1 is up
admin state is up,  Hardware: NVE
  MTU 9216 bytes
  Encapsulation VXLAN
  Auto-mdix is turned off
  RX
    ucast: 39 pkts, 3346 bytes - mcast: 0 pkts, 0 bytes
  TX
    ucast: 18 pkts, 2216 bytes - mcast: 0 pkts, 0 bytes
```

Status of NVE Interface

Initiate a ping from Desktop0 to Desktop1 and Desktop2 in order to verify their reachability.

When an Address Resolution Protocol (ARP) request is initiated from Desktop0 to Desktop1, the ARP packet is sent to Leaf1. Leaf1 will then forward the packet towards the Spine device using multicast address 239.0.0.1, which is used for VNI1000. The Spine device will multicast the packets to all the leaf devices that are part of the same VNI 10000.

Ping from Desktop0 (192.168.100.100) to Desktop1 (192.168.100.150) and Desktop2 (192.168.100.200).

```
inserthostname-here:~$ ping 192.168.100.150
PING 192.168.100.150 (192.168.100.150): 56 data bytes
64 bytes from 192.168.100.150: seq=0 ttl=42 time=8.477 ms
64 bytes from 192.168.100.150: seq=1 ttl=42 time=12.791 ms
64 bytes from 192.168.100.150: seq=2 ttl=42 time=8.352 ms
^C
--- 192.168.100.150 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max = 8.352/9.873/12.791 ms
inserthostname-here:~$
inserthostname-here:~$
inserthostname-here:~$ ping 192.168.100.200
PING 192.168.100.200 (192.168.100.200): 56 data bytes
64 bytes from 192.168.100.200: seq=0 ttl=42 time=15.432 ms
64 bytes from 192.168.100.200: seq=1 ttl=42 time=9.228 ms
64 bytes from 192.168.100.200: seq=2 ttl=42 time=7.133 ms
^C
--- 192.168.100.200 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max = 7.133/10.597/15.432 ms
inserthostname-here:~$ arp -a
? (192.168.100.150) at 52:54:00:05:84:a2 [ether]  on eth0
? (192.168.100.1) at 00:01:00:01:00:01 [ether]  on eth0
? (192.168.100.200) at 52:54:00:10:70:ae [ether]  on eth0
```

Ping from Desktop0 to Desktop1 and Desktop2

LEAF1 is forming NVE peer with LEAF3.

<#root>

```
leaf1# show  nve  peers

Interface Peer-IP                                 State LearnType Uptime   Route

r-Mac

--------- --------------------------------------  ----- --------- -------- -----

------------

nve1

30.30.30.30

                          Up    DP        00:10:23 n/a



leaf1# show  nve  vni 10000

Codes: CP - Control Plane        DP - Data Plane

       UC - Unconfigured         SA - Suppress ARP
```

```
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -----
nve1      10000    239.0.0.1         Up    DP   L2 [10]
```

LEAF3 is forming NVE peer with LEAF1.

```
<#root>
leaf3# show  nve  peers
Interface Peer-IP                                 State LearnType Uptime   Route
r-Mac
--------- --------------------------------------  ----- --------- -------- -----
------------
nve1
10.10.10.10
                            Up    DP        00:10:56 n/a

leaf3# show  nve vni 10000
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -----
nve1      10000    239.0.0.1         Up    DP   L2 [10]
```

```
leaf1# show  mac address-table
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link,
        (T) - True, (F) - False, C - ControlPlane MAC, ~ - vsan
   VLAN     MAC Address      Type      age     Secure NTFY Ports
---------+-----------------+--------+---------+------+----+------------------
*   10
5254.0004.8b92
   dynamic  0         F      F    Eth1/3                                  ------- MAC Address of De
*   10
5254.0005.84a2
   dynamic  0         F      F    nve1(30.30.30.30)                    ------- MAC Address of Desktop1
G    -     5206.ab8a.1b08   static   -         F      F    sup-eth1(R)


leaf3# show  mac address-table
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link,
        (T) - True, (F) - False, C - ControlPlane MAC, ~ - vsan
   VLAN     MAC Address      Type      age     Secure NTFY Ports
---------+-----------------+--------+---------+------+----+------------------
*   10
5254.0004.8b92
   dynamic  0         F      F    nve1(10.10.10.10)                 ------- MAC Address of Desktop0 con
*   10
5254.0005.84a2
   dynamic  0         F      F    Eth1/1                                  ------- MAC Address of Desk
G    -     5206.0619.1b08   static   -         F      F    sup-eth1(R)
```

Here is the Wireshark snapshot when an ARP packet is initiated from Leaf1 to a multicast.

No.	Time	Source	Destination	Protocol	Length	Info
7	5.105615	52:54:00:05:84:a2	52:54:00:04:8b:92	ARP	110	192.168.100.200 is at 52:54:00:05:84:a2
8	7.019252	52:54:00:04:8b:92	ff:ff:ff:ff:ff:ff	ARP	110	Who has 192.168.100.200? Tell 192.168.100.100

```
▶ Frame 8: 110 bytes on wire (880 bits), 110 bytes captured (880 bits)
▶ Ethernet II, Src: 52:06:ab:8a:1b:08 (52:06:ab:8a:1b:08), Dst: IPv4mcast_01 (01:00:5e:00:00:01)
▶ Internet Protocol Version 4, Src: 10.10.10.10, Dst: 239.0.0.1
▶ User Datagram Protocol, Src Port: 50384, Dst Port: 4789
▶ Virtual eXtensible Local Area Network
▶ Ethernet II, Src: RealtekU_04:8b:92 (52:54:00:04:8b:92), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
▶ Address Resolution Protocol (request)
```

Wireshark Capture showing ARP request packet going to multicast group