BGP-EVPN VxLAN Lab - Part 2 - BGP-EVPN Control Plane & Route Advertisement

CSE PRACTICALS

visit http://www.csepracticals.com **25+ Courses on System Programming and Network Development**

Udemy Link : https://www.udemy.com/user/abhishek-sagar-8/

Syllabus

Introduction

This is a detailed GNS3 LAB in which **BGP-EVPN VxLAN** is Implemented on Cisco NXOSv L3 Switches. We Explain the Concept involved in BGP-EVPN VxLAN, show the various output, and packet captures, and explain the data plane operations.

In this LAB, We will Explain how EVPN Routes are shared among VTEPs using BGP EVPN Control Plane.

Software Used: GNS3 version 2.2.39 (*You can use latest available on GNS3 website* **)**

Virtualization: VMWare workstation PRO (*You can use the latest available on the VMWare website*. VMWare WS PRO is a paid software, go for *Work Station Player* instead which is free **)**

System: My system has 64 GB RAM. At least 16 GB RAM is recommended to run this LAB.

Cisco NXOSv Image: NXOSv9k-93000v-10.1.1.1 (This is L3 Switch)

A Great Resource of GNS3 Image Collection is here

The NXOSv Image that I used in this lab is here .

For this LAB you are recommended to use the same version of NXOSv Image as mine to avoid any discrepancy during lab time.

Books: We use two Books as a reference :

Building Data Centers with VxLAN BGP EVPN (Cisco Book)

Virtual Extensible LAN - VxLA A Practical Guide (*Strongly Recommend this one , I followed this book Most of the time*)

Troubleshooting PPT (*Somebody created this nice PPT on VxLAN*)

Pre-Requisites

To do this Tutorial You must know the following :

A solid understanding of ARP and PING functionality is essential.

Proficiency in L2 switching and L3 routing for IPV4 packets is crucial.

BGP basic knowledge is required, Route Reflector, Route distinguished, Route targets

Having a willingness to learn and the ability to ask questions promptly will greatly aid in your professional development.

Topology

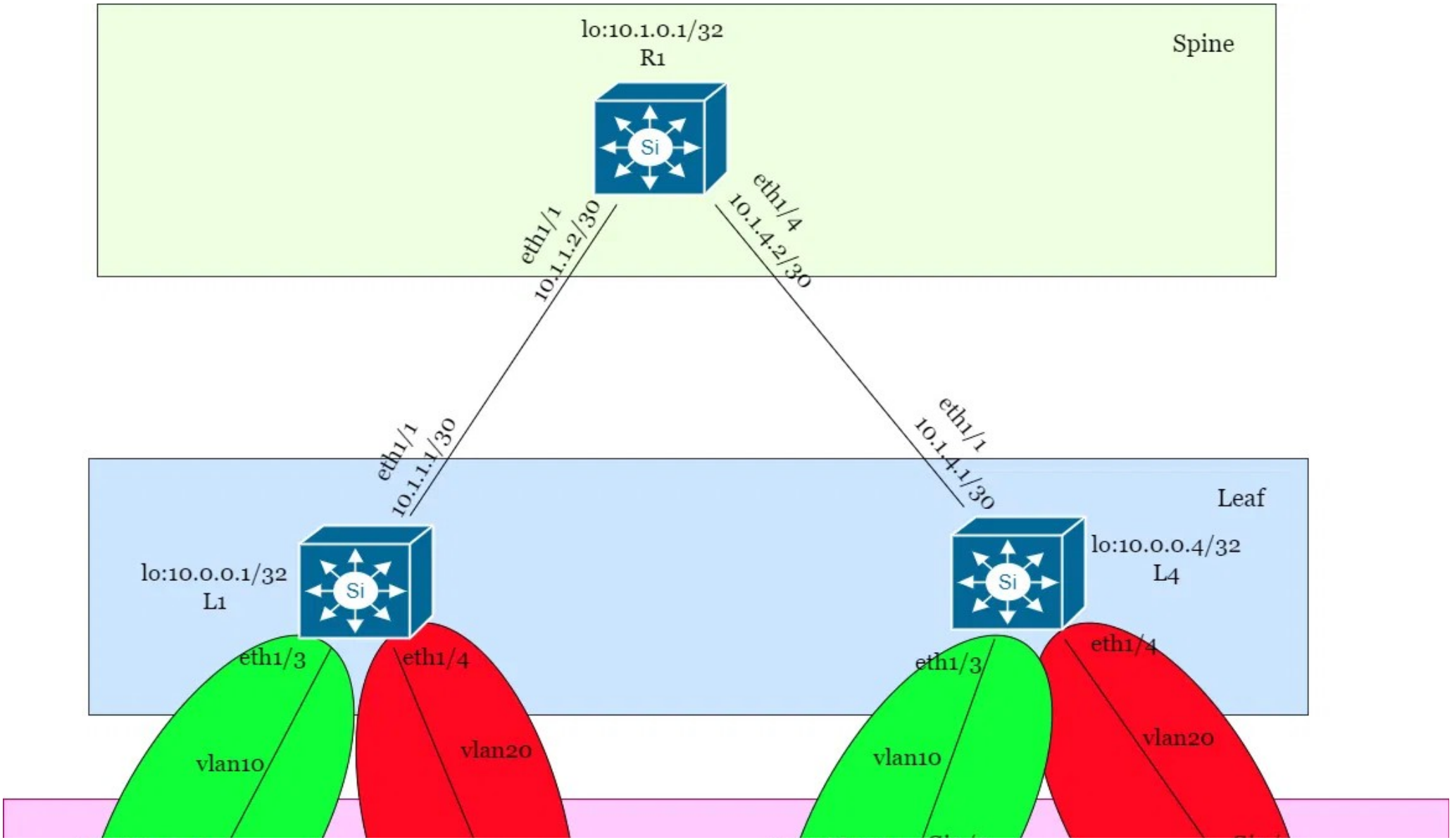
We will construct the exactly below topology.

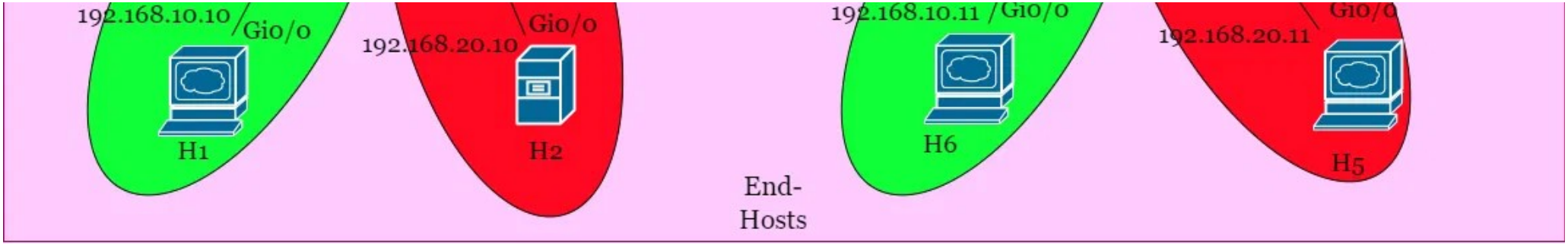
You are strongly recommended to establish the connections and name the interfaces exactly in the same way as is done in the topology so that our

discussion during LABs aligns well.

Configure the same IP Addresses as shown in the topology. Do not Deviate.

The Topology is Layered - **Spine layer, Leaf Layer** and **End host Layer** as you can see in the Diagram.





Lab Topology

At Minimum, you must deploy **R1, L1,L4, H1, H2, H5, H6** nodes as shown in the above diagram. Spine R1 and Leaf Nodes (L1 and L4) are the same NXOSv L3 Switches. So, Total 3 NXOSv nodes you need to deploy - R1, L1, and L4.

End-Hosts could be any node that supports minimum Network functionality like ping, ARP, etc. You can use VPCS hosts (come inbuilt with GNS3, no need to install anything). In my lab, I prefer to use Cisco ASAv Firewall Images. The end host Image type doesn't matter.

End Hosts H1 and H2 are sitting behind Leaf node L1

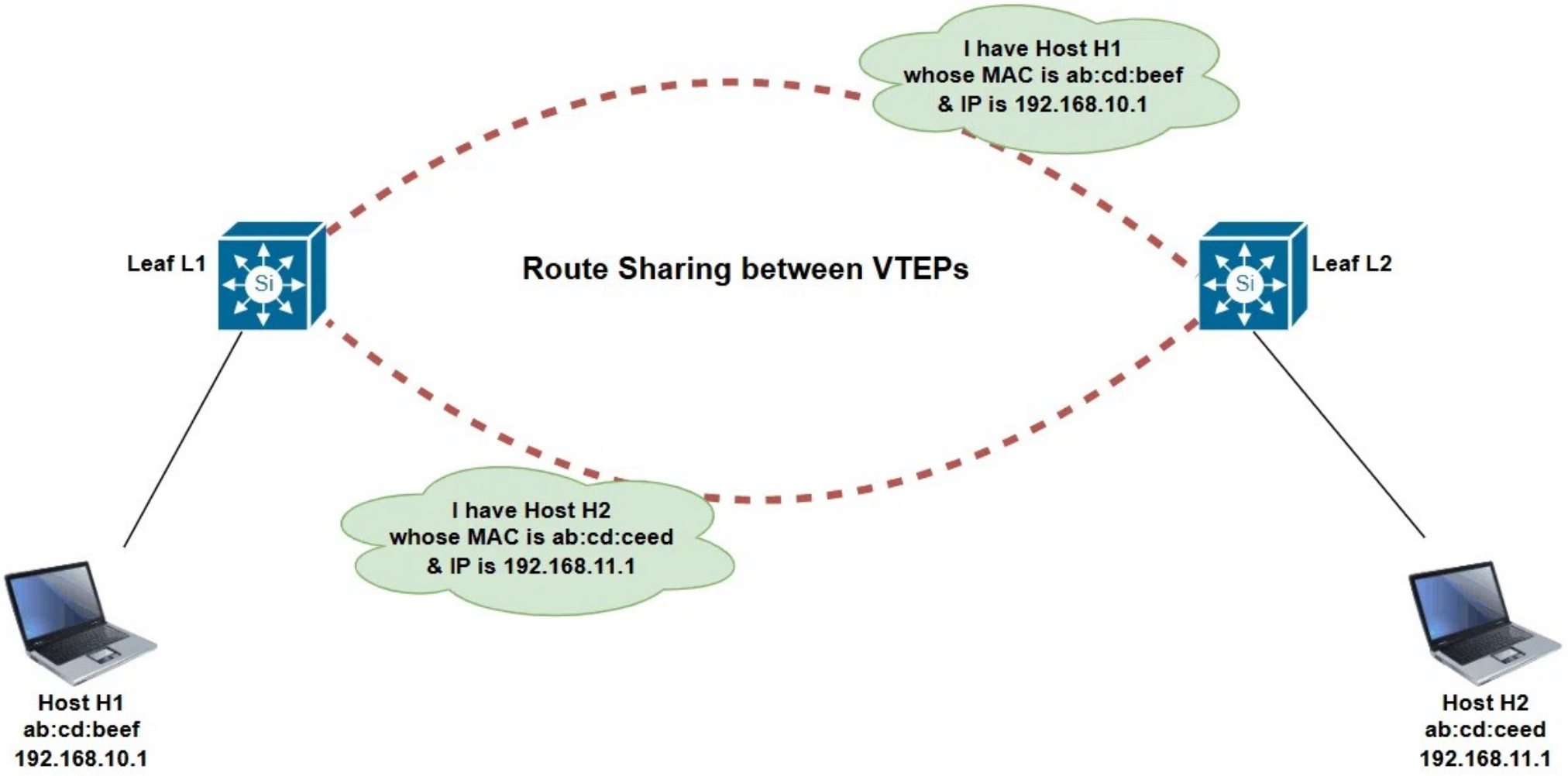
End Hosts H5 and H6 are sitting behind Leaf node L4

End Goal of the LAB

To put it in a few words, our Goal is to :

Understand how Leaf Node L1 knows about the existence of remote hosts H2 and how Leaf Node L2 knows about the existence of end hosts H1 as shown in below diagram. This is accomplished through the advertisement of routes. Unless Leafs collect the required data about their directly connected hosts (for Ex, Leaf L1 collecting route information (MAC/IP) about its directly connected Host H1 and advertise them through some means to other leafs in VxLAN fabric (in this case, Leaf L2), Leafs would not know about the existence of other hosts present in the network.

If Leaf L1 could know - not only the address of hosts H2 but also how to reach host H2 (Routes), then only Leaf L1 could facilitate communication between Hosts H1 and H2. The same is true for Leaf L2. In this Blog, we will learn the mechanism using which End-Hosts's Identities are shared by Leaf VTEPs among themselves so that, all Leafs know about all end-hosts in the network and how to reach them.



Route Sharing/Distrbution

Identities of the Hosts

Identities are the set of parameters that uniquely identify the host in the network. Ideally, they are Mac addresses at Layer 2 and IP Addresses at Layer 3. In VxLAN Network, there can be multiple hosts that may have the same Mac Addresses (if they are Virtual Machines) or the same IP

Addresses (IP Address Reuse). Therefore, Mac address/IP address alone is not sufficient to represent the identities of the host. In VxLAN network, the identity of the host is defined by Two parameters which are (Mac Address Or IP Address) and VTEP Address.

For Example, Layer 2 Address of Host H1 would be : Mac(H1) + 10.0.0.1 and Layer 3 address of Host H1 would be : 192.168.10.10./32 and 10.0.0.1/32. So, we have two types of addresses that uniquely identify the host in the VxLAN network - Layer 2 address and Layer 3 address as described.

Creating VRFs and MAC VRFs on Cisco NXOSv Switches

In this section we will discuss the configuration required to create VRFs and MAC VRFs on Cisco NXOSv L3 switches in the context of BGP-EVPN VxLAN. We Create VRFs and MAC VRFs only on Leaf Switches and not on Spine.

Creating VRF

Leaf L1/L4	Comments
`conf`	Configure Anycast gateway
`fabric forwarding anycast-gateway-mac 2.2.2`	Enable the creation of SVI interfaces
`feature interface-vlan`	Create VRF Cust-A
`vrf context Cust-A`	Assign VNI to this VRF, this is VxLAN based routing requirement. This Also Creates MAC-IP VRF EVPN RIB.
`vni 99999`	Assign Route Distinguisher to this VRF (This is BGP advertisement requirement)
`rd auto`	This is the requirement for appropriate redistribution of VRF routes amongst Leafs.
`address-family ipv4 unicast`	This vlan is created as a way to implement VxLAN Support. This is Cisco-specific. Any pkt which lands in the vlan999 subnet will be subjected to L3 routing as against the traditional L2 routing.
`route-target import 9999:9999`	Configuring this vlan to enforce L3 routing to traffic landing in this vlan.
`route-target export 9999:9999`	This vlan is in VRF Cust-A.
`end`	Create SVI 10
`conf`	Assign this SVI to VRF
`vlan 999`	

```
name L3-VNI
vn-segment 99999
end
conf
interface vlan 999
ip forward
no shutdown
vrf member Cust-A
end
conf
interface vlan 10
no shut
vrf member Cust-A
ip address 192.168.10.1/24
fabric forwarding mode
anycast-gateway
end
conf
interface vlan 20
no shut
vrf member Cust-A
ip address 192.168.20.1/24
fabric forwarding mode
anycast-gateway
end
conf
interface nve 1
member vni 99999 associate-vrf
```

Assign Anycast GW Address 2.2.2 to SVI 10 interface.

Tell VXLAN process to use VNI 99999 as L3VNI

Use BGP as a transport protocol for all VNI members of this NVE interface

host-reachability protocol bgp end	

This config is exactly same on Leaf L1 and L4.

Creating MAC VRFs

Next, we will create MAC VRFs (also called EVPN instances) in the context of EVPN. MAC VRFs are L2 EVPN RIBs created which store MAC Routes of hosts. MAC VRF is created per L2 VNI.

Leaf L1	Comments	Leaf L4
config evpn vni 5010 l2 rd auto route-target both 65501:10 vni 5020 l2 rd auto route-target both 65501:20 end	Create MAC VRF with ID 5010 Assign RD to MAC VRF Assign Route target values to MAC VRF	config evpn vni 5010 l2 rd auto route-target both 65501:10 vni 5020 l2 rd auto route-target both 65501:20 end

Thats it. We have successfully configured MAC VRFs with VNI value, RD, and Import Export Route Target values. For Example, MAC VRF 5010 has RD ‘auto’ and import and export route-target value as 65501:10 (BGP Autonomous system 65501 , vlan 10)

BGP Configuration

We need to configure internal-BGP (ibgp) on Leafs and Spine. Spine should configure Leafs as Route-Reflector. On Leafs, Spine should be configured as IBGP neighbor. On Spine, Leafs must be configured as Neighbors. The whole purpose for configuring I-BGP across spines and Leafs is to transport and redistribute routes of Hosts on all Leafs. I-BGP is a carrier or Transporter of Routes.

Spine R1	Leaf L1	Leaf L4
config feature bgp feature nv overlay nv overlay evpn router bgp 65501 #address-family ipv4 unicast address-family l2vpn evpn neighbor 10.0.0.1 remote-as 65501 update-source loopback 0 address-family l2vpn evpn send-community send-community extended route-reflector-client neighbor 10.0.0.4 remote-as 65501 update-source loopback 0 address-family l2vpn evpn send-community send-community extended route-reflector-client end	config feature bgp feature nv overlay nv overlay evpn router bgp 65501 #address-family ipv4 unicast address-family l2vpn evpn neighbor 10.1.0.1 remote-as 65501 update-source loopback 0 address-family l2vpn evpn send-community send-community extended end	config feature bgp feature nv overlay nv overlay evpn router bgp 65501 #address-family ipv4 unicast address-family l2vpn evpn neighbor 10.1.0.1 remote-as 65501 update-source loopback 0 address-family l2vpn evpn send-community send-community extended end

Thats it, We are finished with all the required configuration. Now, We will learn how route advertisement happens across Leafs.

Verification

To verify that BGP peering between the spine and leafs is indeed established successfully, use the below command, preferably on spine as it will show bgp peering status with each leaf.

```
spine1# show bgp l2vpn evpn neighbors | include "BGP state"
BGP state = Established, up for 00:55:01
BGP state = Established, up for 00:16:07
spine1#
```

The BGP-EVPN Control plane is set up fully. Now we will see how BGP contributes to Route Advertisement.

Route Advertisement

Route Advertisement is a process using which VTEPs/Leafs advertise the information about directly connected hosts to other VTEPs. The responsibility to transport the routes from VTEP to other VTEP is done via Spine using BGP as a transport protocol. In the diagram below, BGP process transport route about host H1 to from VTEP 10.0.0.1 to Spine first. Spine, as a BGP route reflector, will advertise these routes to all other VTEPs, in this case, VTEP 10.0.0.4. There is no BGP peering between VTEPs. In subsequent sections, we will try to find answers to the following questions :

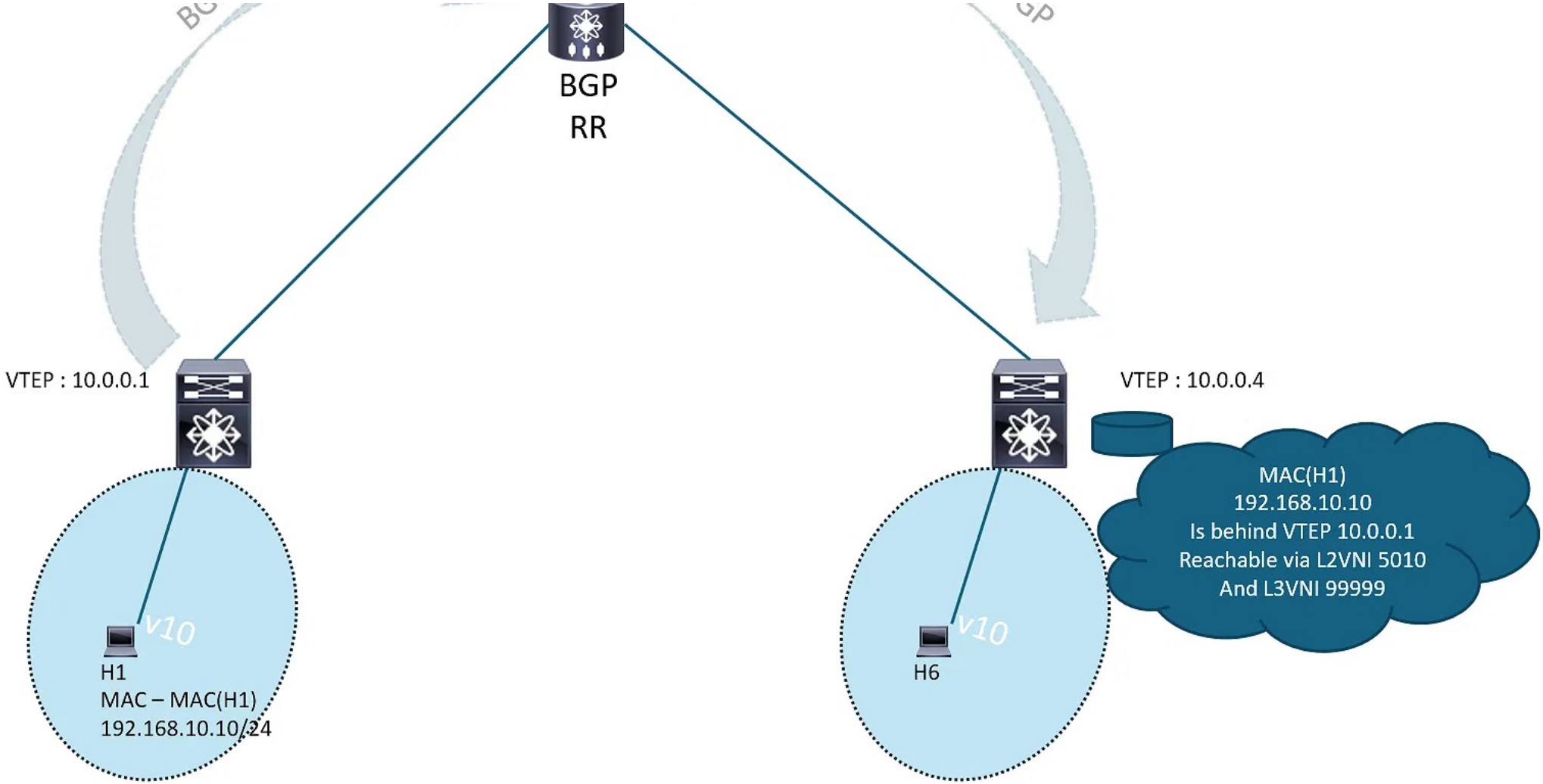
How VTEPs know about their directly connected hosts?

How routes are pushed to BGP process?

How does BGP advertise these routes?

How routes are stored on the recipient VTEP ?

How do Recvd routes on VTEP contribute to VxLAN-based routing ?



In the above example, VTEP 10.0.0.1 has advertised the credentials of Host H1 (mac address, ip address, L2VNI, L3VNI) to the BGP process. The process then transports them to all other VTEPs or which it is configured as Route reflectors without any change. These Credentials are then processed and stored on VTEP 10.0.0.4 when received through BGP.

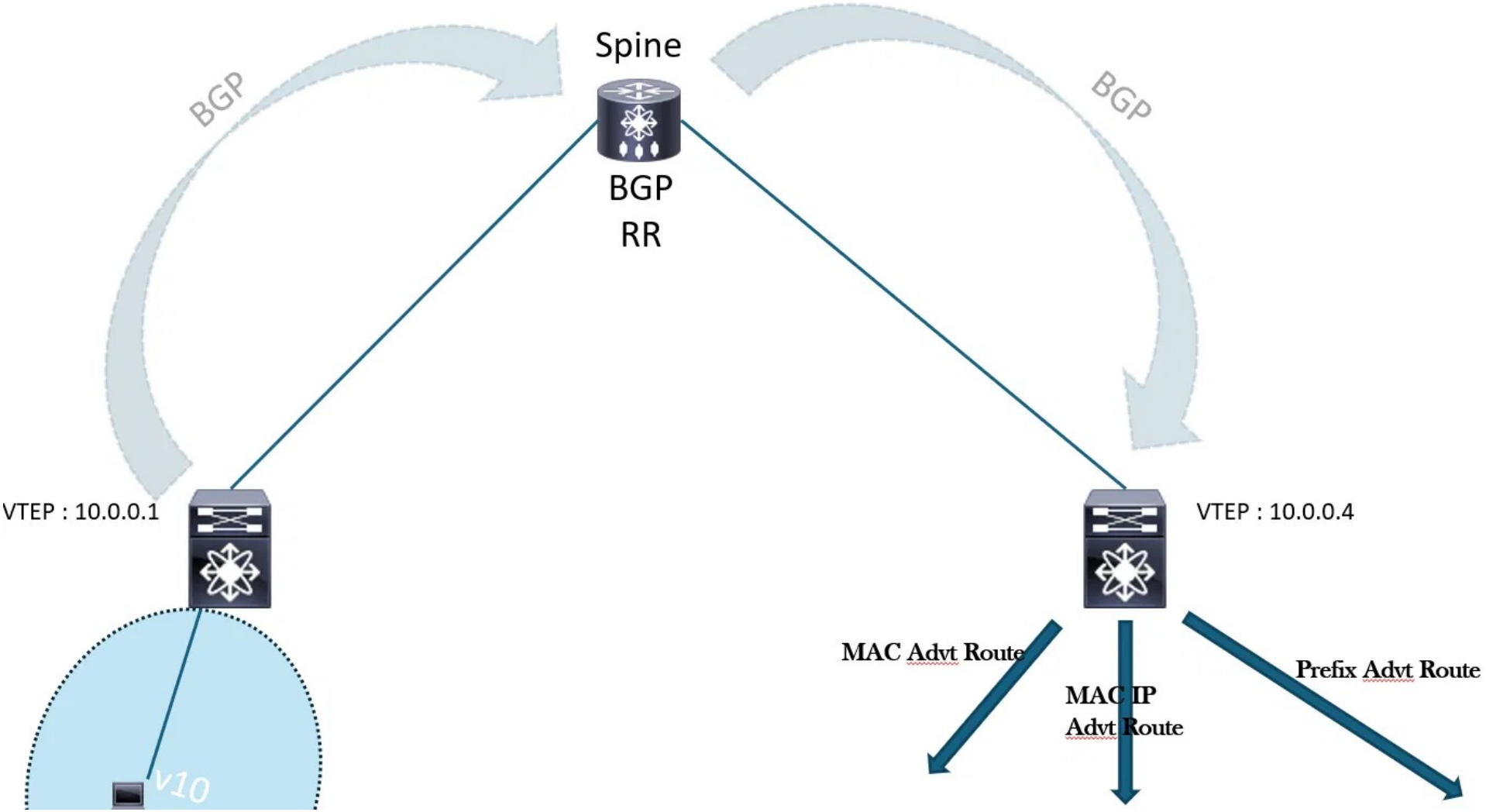
Our Route advertisement discussion is divided into three parts as below. So, Leaf advertises three types of routes per directly connected host

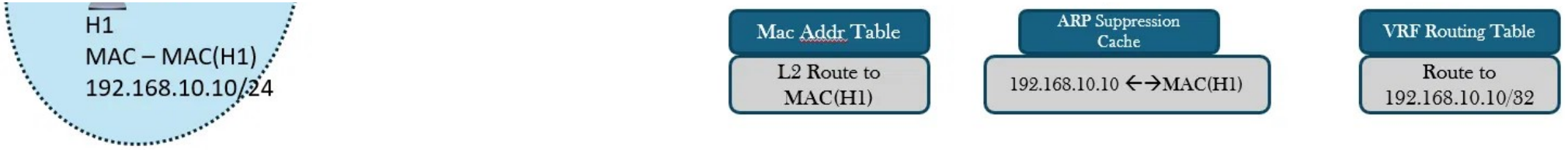
using BGP.

Advertisement of MAC address of Host - BGP EVPN Route Type 2 (used for intra-vlan routing Or Extending the VLANs)

Advertisement of MAC-IP Address of Host - BGP EVPN Route Type 2 (used in ARP suppression)

Prefix Advertisement - BGP EVPN Route Type 5 (used for inter-vlan routing)





As you can see in the above Diagram, the Route advertised by VTEP 10.0.0.1 for its direct Host H1 is settled into 3 different tables of Remote VTEPs (VTEP 10.0.0.4) -

The Mac Address Table containing L2 route to Host H1

The ARP suppression Cache containing IP to MAC mapping for Host H1

VRF L3 routing table containing L3 route to 192.168.10.10/32 , Layer 3 IP Address of Host H1

Mac Advertisement

To start with how the MAC address of the Host H1 is advertised and eventually settles in the MAC address table of remote VTEPS, let's inspect the tables involved in advertising VTEP L1.

The Tables involved are :

Mac address Table

MAC VRF table (L2 EVPN RIB)

BGP Local RIBs

The Data is pushed from the previous table to the next table in sequence as listed above :

```
MAC Table --------> MAC VRF Table --------> BGP Local RIBs
```

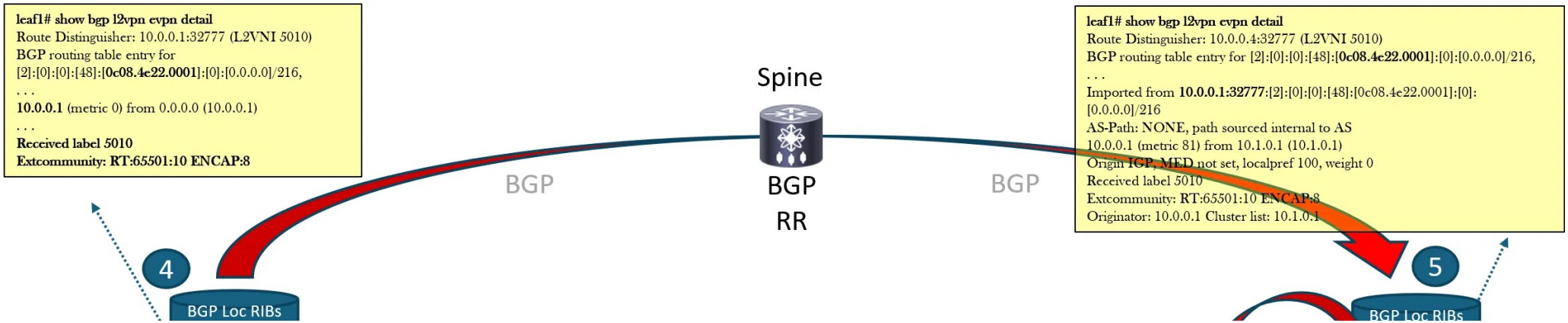
MAC VRF table is created per VNI. So, we have two MAC VRF tables - one for VNI5010 and one for VNI5020. Since VLAN 10 is associated with VNI 5010, all MAC addresses present in VLAN 10 are pushed into MAC VRF table for VNI5010. This table contains **MAC-only** routes as

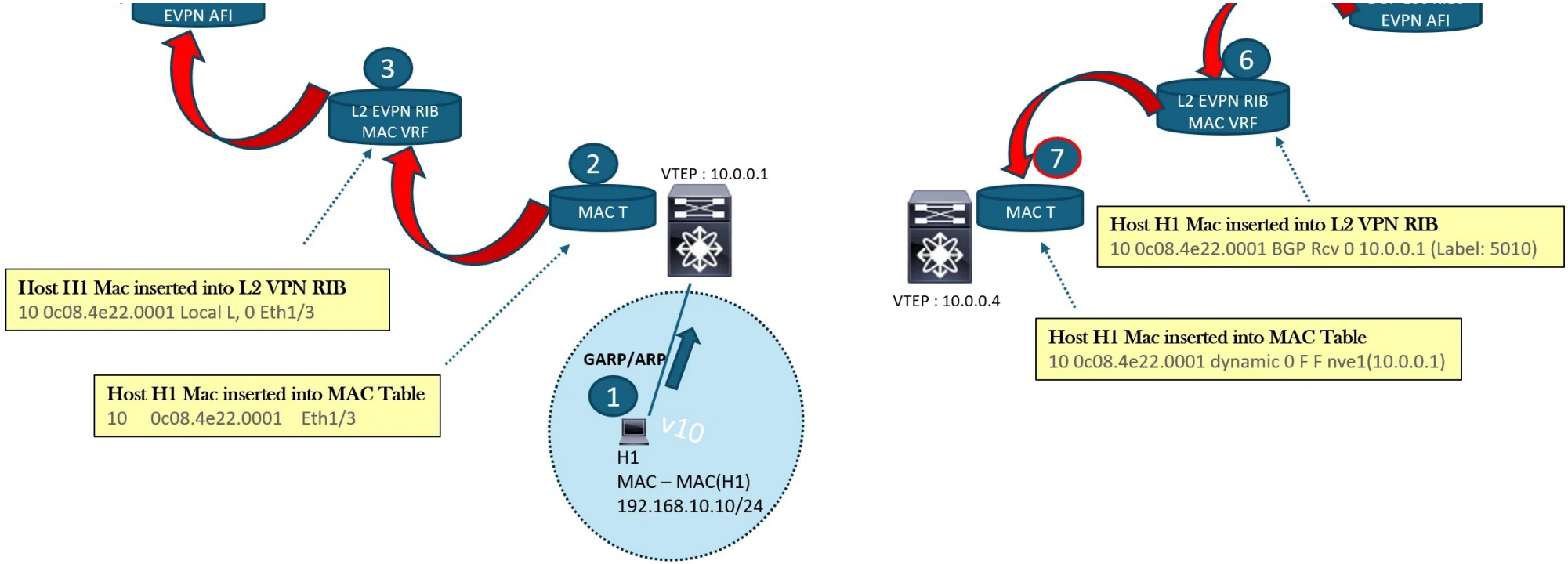
exported from mac-address table. Each MAC VRF table is also configured with RD and RT. BGP process is listening on these MAC VRF tables to export MAC routes as **EVPN TYPE 2 Routes** along with RDs and RTs configured on these tables.

```
vlan 10 << MAC VRF table is created with this configuration
vn-segment 5010
vlan 20
vn-segment 5020
```

```
evpn
vni 5010 l2 << Setting the VNI Attributes : RD, RT, ARP Supppression, Ingress Replication etc.
rd auto
route-target both 65501:10
vni 5020 l2
rd auto
route-target both 65501:20
```

The Diagram below Captures the flow of MAC Routes on Sending VTEP and Receiving VTEP.





MAC Route Export-Import Flow Diagram

Initial Table States

The MAC Address table is empty, except that it contains a Router-MAC entry for each VLAN created on VTEP. The Router MAC address is the same for all VLANs.

```
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
---------+-----------------+--------+---------+------+----+------------------
G - 0002.0002.0002 static - F F sup-eth1(R)
G - 0ca5.0000.1b08 static - F F sup-eth1(R)
G 10 0ca5.0000.1b08 static - F F sup-eth1(R)
G 20 0ca5.0000.1b08 static - F F sup-eth1(R)
G 999 0ca5.0000.1b08 static - F F sup-eth1(R)
```

The L2 EVPN RIB is empty for all VLANs

```
leaf1# show l2route evpn mac all
Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
(Dup):Duplicate (Spl):Split (Rcv):Recv (AD):Auto-Delete (D):Del Pending
(S):Stale (C):Clear, (Ps):Peer Sync (O):Re-Originated (Nho):NH-Override
(Pf):Permanently-Frozen, (Orp): Orphan
Topology Mac Address Prod Flags Seq No Next-Hops
----------- -------------- ------ ------------ ---------- ---------------------
------------------
```

BGP L2EVPN address family RIBs are also empty.

```
leaf1# show bgp l2vpn evpn
leaf1#
```

Route Export Procedure

Let's start step by step :

Host H1, has two ways to notify VTEP 10.0.0.1 about its presence :

Generate GARP when booting up

Initiate communication (ARP msg will be generated)

Either of the two methods will send VTEP the ARP message from Host H1 **STEP1**. VTEP analyzes only the Src MAC Address embedded in the ARP message. In this case, Src mac address is **MAC(H1) = 0c08.4e22.0001**

VTEP processes the ARP message in a way we described in the previous article, we will confine our discussion here regarding how this MAC address is learned and advertised by VTEP

As soon as VTEP receives ARP message from Host H1, it inserts an entry into its mac address table **STEP2**

```
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 0c08.4e22.0001 dynamic 0 F F Eth1/3 <<< New Entry which says a hosts with mac 0c08.4e22.0001 is present in vlan 10 and is reachable via local intf Eth1/3.

G - 0002.0002.0002 static - F F sup-eth1(R)
```

```
G - 0ca5.0000.1b08 static - F F sup-eth1(R)

G 10 0ca5.0000.1b08 static - F F sup-eth1(R)

G 20 0ca5.0000.1b08 static - F F sup-eth1(R)

G 999 0ca5.0000.1b08 static - F F sup-eth1(R)
```

This new entry is pushed from MAC Address table to MAC VRF table corresponding to vlan 10 **STEP3**

```
leaf1# show l2route evpn mac evi 10 detail

Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link

(Dup):Duplicate (Spl):Split (Rcv):Recv (AD):Auto-Delete (D):Del Pending

(S):Stale (C):Clear, (Ps):Peer Sync (O):Re-Originated (Nho):NH-Override

(Pf):Permanently-Frozen, (Orp): Orphan

Topology Mac Address Prod Flags Seq No Next-Hops

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

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

10 0c08.4e22.0001 Local L, 0 Eth1/3 <<< New Entry in MAC VRF table.

Route Resolution Type: Regular

Forwarding State: Resolved

Sent To: BGP
```

Finally, BGP reads the routes from MAC VRF table and exports the MAC Routes along with their attributes (VNI, RD, RT) **STEP4.** BGP maintains its EVPN Address Family RIBS keyed by VNIID.

```
leaf1# show bgp l2vpn evpn detail
```

```
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 10.0.0.1:32777 (L2VNI 5010) << Route Distinguisher Value which is 10.0.0.1 : 32767 + vlan-id
BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[0]:[0.0.0.0]/216, version 25 <<< Actual Route which is mac address.
Paths: (1 available, best #1)
Flags: (0x000102) (high32 00000000) on xmit-list, is not in l2rib/evpn
Advertised path-id 1
Path type: local, path is valid, is best path, no labeled nexthop
AS-Path: NONE, path locally originated
10.0.0.1 (metric 0) from 0.0.0.0 (10.0.0.1) <<< Originator of the route
Origin IGP, MED not set, localpref 100, weight 32768
Received label 5010 <<< L2VNI this route is associated with
Extcommunity: RT:65501:10 ENCAP:8 <<< Route Target Value as per the configuration, ENCAP 8 is VxLAN encapsulation
Path-id 1 advertised to peers:
10.1.0.1
```

Route Import Procedure

Checking the BGP EVPN Address Family RIBS on VTEP 10.0.0.4.

This Route is shown as it is without any change on Receiving side in BGP Local Ribs. This is what was sent.

```
leaf4# show bgp l2vpn evpn detail
```

```
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 10.0.0.1:32777
BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[0]:[0.0.0.0]/216,
version 32
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not i
n HW
Advertised path-id 1
Path type: internal, path is valid, is best path, no labeled nexthop
Imported to 1 destination(s)
Imported paths list: L2-5010
AS-Path: NONE, path sourced internal to AS
10.0.0.1 (metric 81) from 10.1.0.1 (10.1.0.1)
Origin IGP, MED not set, localpref 100, weight 0
Received label 5010
Extcommunity: RT:65501:10 ENCAP:8
Originator: 10.0.0.1 Cluster list: 10.1.0.1
Path-id 1 not advertised to any peer
```

BGP then import the route receieved above into its local RIBs as per the route's VNI values. The above route has L2VNI 5010, hence below is the MAC Route in BGP Local EVPN AFI RIB for VNI 5010. In this process, BGP updates the RD value of the route to that of the local RD value.

STEP5

```
Route Distinguisher: 10.0.0.4:32777 (L2VNI 5010) <<< RD value updated to that of local VTEP
BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[0]:[0.0.0.0]/216, version 33
Paths: (1 available, best #1)
Flags: (0x000212) (high32 00000000) on xmit-list, is in l2rib/evpn, is not in HW
Advertised path-id 1
Path type: internal, path is valid, is best path, no labeled nexthop, in rib
Imported from 10.0.0.1:32777:[2]:[0]:[0]:[48]:[0c08.4e22.0001]:[0]:[0.0.0.0]/216
AS-Path: NONE, path sourced internal to AS
10.0.0.1 (metric 81) from 10.1.0.1 (10.1.0.1)
Origin IGP, MED not set, localpref 100, weight 0
Received label 5010
Extcommunity: RT:65501:10 ENCAP:8
Originator: 10.0.0.1 Cluster list: 10.1.0.1
Path-id 1 not advertised to any peer
```

The next Step is to insert the BGP Route in L2 EVPN RIBs / MAC VRF table based on the Route's RT value **STEP6**

This Route's Route target value is 10, so matching MAC VRF table on local VTEP with same RT value is MAC VRF table for VNI 5010 (as per the config). Hence route will be imported in MAC VRF table for VNI 5010.

```
leaf4# show l2route evpn mac all detail
Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
```

```
(Dup):Duplicate (Spl):Split (Rcv):Recv (AD):Auto-Delete (D):Del Pending
(S):Stale (C):Clear, (Ps):Peer Sync (O):Re-Originated (Nho):NH-Override
(Pf):Permanently-Frozen, (Orp): Orphan
Topology Mac Address Prod Flags Seq No Next-Hops
---------- -------------- ------ ------------ ---------- --------------------
------------------
10 0c08.4e22.0001 BGP Rcv 0 10.0.0.1 (Label: 5010)
Route Resolution Type: Regular
Forwarding State: Resolved (PeerID: 1)
Sent To: L2FM
Encap: 1
```

Finally, the route is installed in MAC Address table of VTEP 10.0.0.4 **STEP7**

```
leaf4# 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
---------+-----------------+--------+---------+------+----+------------------
```

```
C 10 0c08.4e22.0001 dynamic 0 F F nve1(10.0.0.1) << new entry *
G - 0002.0002.0002 static - F F sup-eth1(R)
G - 0c48.0000.1b08 static - F F sup-eth1(R)
G 10 0c48.0000.1b08 static - F F sup-eth1(R)
G 20 0c48.0000.1b08 static - F F sup-eth1(R)
G 999 0c48.0000.1b08 static - F F sup-eth1(R)
leaf4#
```

*Since 5010 is the VNI value being used in encapsulation by NVE interface nve1

```
interface nve 1
. . .
member vni 5010 mcast-group 239.1.1.1
member vni 5020 mcast-group 239.2.2.2
```

This completes the Learning of MAC Routes by Destination VTEPs, starting from Advertising it by Src VTEP, Receiving it by Remote VTEP, and installing it in the data plane of Remote VTEPs. The Host is identified by its MAC Address only, this type of routes are referred to as **MAC-Only routes.**

Now that, VTEP 10.0.0.4 knows the L2 Addressing of Host H1. If the local Host H6 sitting behind VTEP 10.0.0.4 also initiates ARP/GARP, then its Layer 2 addressing would have been learned by VTEP 10.0.0.1.

This route enables routing of the intra-vlan traffic over VxLAN tunnel (as Explained in Part-1)

MAC-IP Advertisement

MAC-IP Route Advertisement aims at advertising not only the MAC addresses as well as the IP Address of the Host in the same BGP Msg Type 2 by the VTEP. MAC-IP routes contain all information which is already there in MAC-only routes, plus in addition they contain Layer 3 information of the route, such as IP-Address, L3VNI, Route Targets for L3VNI etc. Thus MAC-IP routes is super-set of Mac-only routes. The main aim of advertising these routes is to populate the **ARP Suppression cache** of remote VTEPs. ARP Suppression Cache is used by VTEP is optimize the BUM Traffic flooding. For example, if Host H1 sitting behind VTEP 1 generated ARP-B request for 192.168.10.11 , i.e. for Host 6, then if ARP suppression feature is enabled on VTEP 1 (10.0.0.1), VTEP1 itself respond to such ARP broadcast request provided it has ARP resolution entry in its ARP suppression cache which is learned by BGP-EVPN MAC-IP route advertisement.

To start with how MAC + IPV4 addresses of the Host H1 are advertised and eventually settle in the ARP suppression table of remote VTEPS, let's inspect the tables involved in advertising of VTEP L1.

The Tables involved are :

ARP Table

ARP Suppression Cache

IP VRF table (L2 EVPN RIB for MAC-IP Routes with L2 VNI)

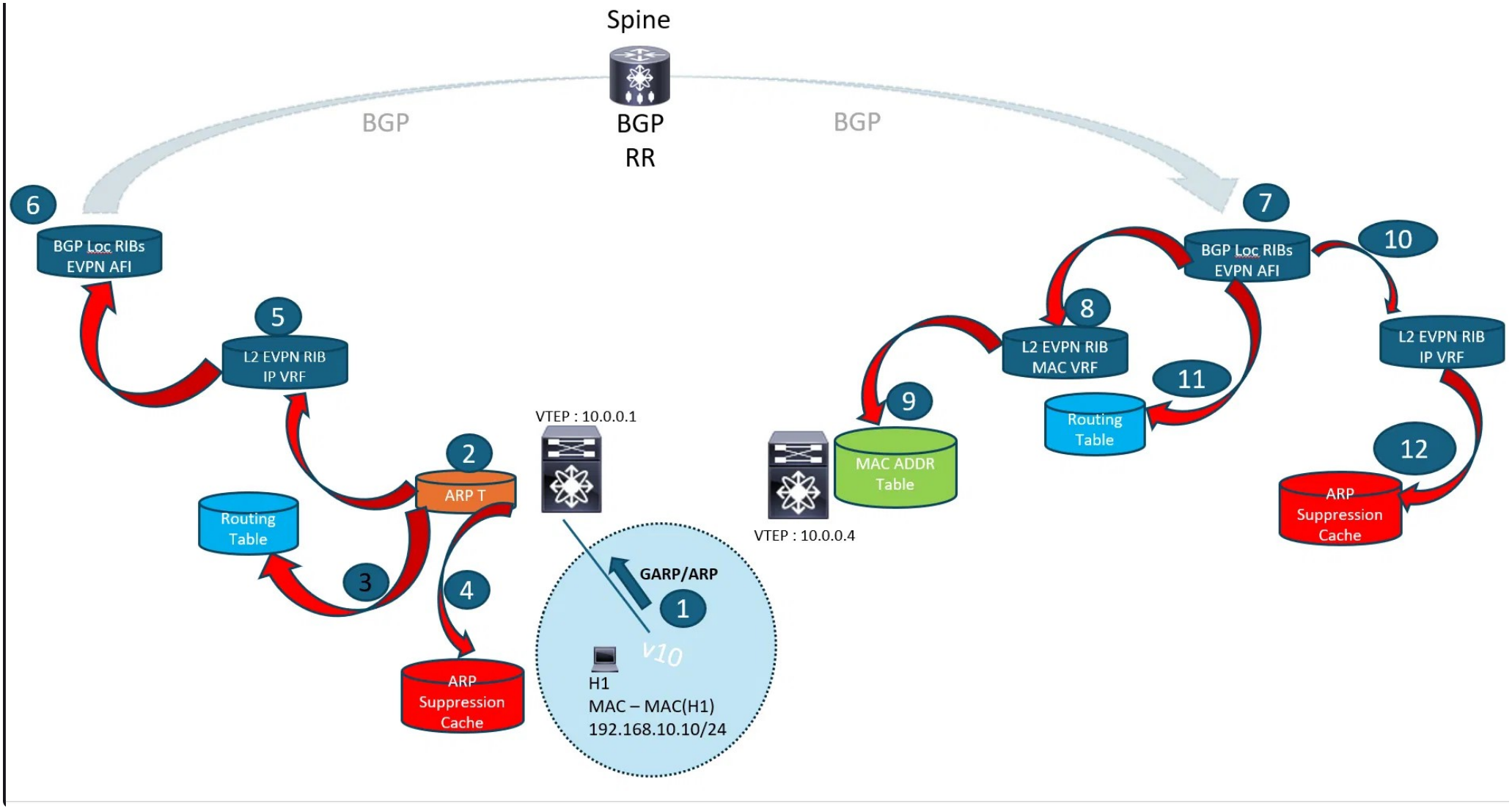
BGP Local RIBs

Routing Table

The Data is pushed from the previous table to the next table in sequence as listed below on Sending VTEP. As soon as local ARP table is updated, Route Update is pushed from ARP table to IP VRF table, Local Routing table and Local ARP suppression Cache.

```
ARP Table
|
|--------> IP VRF Table --------> BGP Local RIBs
|--------> Routing Table
|--------> ARP Suppresion Cache
```

IP VRF tables are created per L3 VRF. The internal structure of both tables is almost the same (MAC VRF and IP VRF). In this case, one IP VRF table is created for VRF Cust-A. The IP VRF table contains MAC-IP route information for all the hosts present in VLANs associated with the VRF along with their L2VNI Value. In this case, Host H1 and Host H2 are in vlan 10 and vlan 20 respectively, but both vlans are housed under VRF **Cust-A** (see config excerpts below). Therefore, IP VRF table is expected to contain MAC-IP routes for both the Hosts. On Sending VTEP, The **MAC-IP** routes are exported from the VRF ARP table into IP VRF tables. Each IP VRF table is also configured with RD and RT. BGP process is listening on these IP VRF tables to export MAC-IP routes as **EVPN TYPE 2 Routes** along with RDs and RTs configured on these tables.



MAC-IP Route Import-Export Flow Diagram

Initial Table States

Let us check the state of the tables before start the advertisement of MAC-IP routes on Source VTEP 10.0.0.1

VRF ARP table is empty.

```
leaf1# show ip arp vrf Cust-A

IP ARP Table for context Cust-A

Total number of entries: 0

Address Age MAC Address Interface Flags
```

MAP-IP Table is also empty

```
leaf1# show l2route evpn mac-ip all

leaf1#
```

BGB L2 EVPN Address family RIBs for VNIID 99999 and 5010 are also empty.

```
leaf1# show bgp l2vpn evp vni-id 99999

leaf1# show bgp l2vpn evp vni-id 5010

leaf1#
```

VRF Routing Table on VTEP 10.0.0.1.

```
leaf1# show ip route vrf Cust-A

IP Route Table for VRF "Cust-A"

192.168.10.0/24, ubest/mbest: 1/0, attached

*via 192.168.10.1, Vlan10, [0/0], 01:19:58, direct

192.168.10.1/32, ubest/mbest: 1/0, attached

*via 192.168.10.1, Vlan10, [0/0], 01:19:58, local
```

```
192.168.20.0/24, ubest/mbest: 1/0, attached
*via 192.168.20.1, Vlan20, [0/0], 01:19:58, direct
192.168.20.1/32, ubest/mbest: 1/0, attached
*via 192.168.20.1, Vlan20, [0/0], 01:19:58, local
leaf1#
```

ARP Suppression Cache on VTEP 10.0.0.1

```
leaf1# show ip arp suppression-cache vlan 10
Flags: + - Adjacencies synced via CFSoE
L - Local Adjacency
R - Remote Adjacency
L2 - Learnt over L2 interface
PS - Added via L2RIB, Peer Sync
RO - Dervied from L2RIB Peer Sync Entry
Ip Address Age Mac Address Vlan Physical-ifindex Flags Remote Vtep Addrs
```

Route Export Procedure

Let’s start step by step :

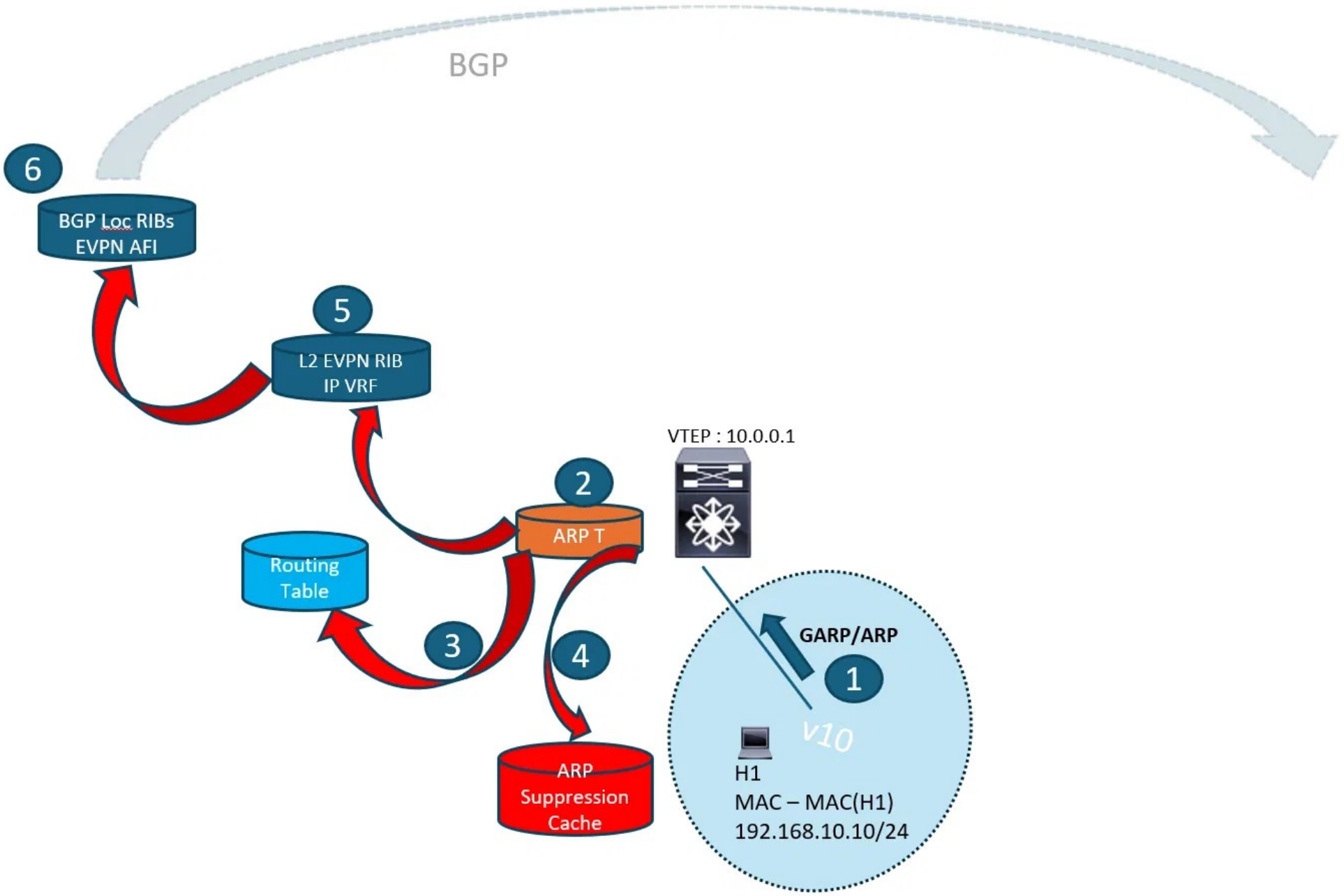
Host H1, has two ways to notify VTEP 10.0.0.1 about its presence :

Generate GARP when booting up

Initiate communication (ARP msg will be generated)

Either of the two methods will send VTEP the ARP message from Host H1 **STEP1**. VTEP analyzes only the Src MAC Address embedded in the ARP message and Src IP Address Embedded in ARP msg. In this case, Src mac address is **MAC(H1) = 0c08.4e22.0001 and IP(H1) = 192.168.10.10**

VTEP processes the ARP message in a way we described in the previous article, we will confine our discussion here regarding how these MAC & IP addresses are learned and advertised by VTEP


BGP
6
BGP Loc RIBs
EVPN AFI
5
L2 EVPN RIB
IP VRF
VTEP : 10.0.0.1
2
ARP T
Routing
Table
3
4
ARP
Suppression
Cache
GARP/ARP
1
v10
H1
MAC – MAC(H1)
192.168.10.10/24

Route Import Procedure

As soon as VTEP receives ARP message from Host H1, it inserts an entry into its VRF Specific ARP table **STEP2**

```
leaf1# show ip arp vrf Cust-A

. . .

IP ARP Table for context Cust-A

Total number of entries: 1

Address Age MAC Address Interface Flags

192.168.10.10 0.693378 0c08.4e22.0001 Vlan10 <<< New Entry learned
```

The HMM (Host Mobility Manager) component (Cisco Specific) installs the local route to host H1 (192.168.10.10/32) in VRF routing table for Host H1 in Local VTEP 10.0.0.1 **STEP 3**

```
leaf1# show ip route vrf Cust-A

IP Route Table for VRF "Cust-A"

192.168.10.0/24, ubest/mbest: 1/0, attached

*via 192.168.10.1, Vlan10, [0/0], 01:23:56, direct

192.168.10.1/32, ubest/mbest: 1/0, attached

*via 192.168.10.1, Vlan10, [0/0], 01:23:56, local

192.168.10.10/32, ubest/mbest: 1/0, attached

*via 192.168.10.10, Vlan10, [190/0], 00:02:19, hmm <<< Host route of H1 installed by HMM

192.168.20.0/24, ubest/mbest: 1/0, attached
```

```
*via 192.168.20.1, Vlan20, [0/0], 01:23:56, direct
192.168.20.1/32, ubest/mbest: 1/0, attached
*via 192.168.20.1, Vlan20, [0/0], 01:23:56, local
leaf1#
```

The ARP Suppression Cache is also updated with MAC ←→ IP Mapping which is `0c08.4e22.0001 <--> 102.168.10.10` **STEP 4**

<show ip arp suppression-cache detail, Feature note enabled on my VTEP unfortunately>

This new MAC-IP route entry is pushed from VRF ARP table to IP VRF EVPN table. This is done by Cisco Specific Process called *Host-Mobilty Manager (HMM)*. **STEP5**

```
leaf1# show l2route evpn mac-ip all detail
Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
(Dup):Duplicate (Spl):Split (Rcv):Recv(D):Del Pending (S):Stale (C):Clear
(Ps):Peer Sync (Ro):Re-Originated (Orp):Orphan
Topology Mac Address Host IP Prod Flags
Seq No Next-Hops
---------- -------------- ------------------------------------ ------ ------
---- ---------- ------------------------------------
10 0c08.4e22.0001 192.168.10.10 HMM L, 0 Local <<< New Entry
L3-Info: 99999
Sent To: BGP
```

Finally, BGP reads the routes from the IP-VRF table and exports the MAC-IP Routes along with their attributes (VNI, RD, RT) **STEP 6.** BGP

maintains its EVPN Address Family RIBS keyed by VNIID.

The MAC-IP route is advertised as a separate **BGP EVPN Route Type 2** advertisement, besides the MAC-only route. If you notice the output below, the route is exactly same as its counter part Mac-only route except that it contains additional information: L3 VNI (99999), IP Address, RouteTarget Value of VRF (99999:99999). This message also advertises an additional mac address called **Router Mac.**

```
leaf1# show bgp l2vpn evpn detail

<snipped MAC-only Route>

BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]:[192.168.10.10]/272, version 12 <<< IP and MAC Address

Paths: (1 available, best #1)

Flags: (0x000102) (high32 00000000) on xmit-list, is not in l2rib/evpn

Advertised path-id 1

Path type: local, path is valid, is best path, no labeled nexthop

AS-Path: NONE, path locally originated

10.0.0.1 (metric 0) from 0.0.0.0 (10.0.0.1) <<< Advertising VTEP

Origin IGP, MED not set, localpref 100, weight 32768

Received label 5010 99999 <<< The MAC IP Route advt both, L2VNI ( for MAC Address) and L3VNI ( for IP Address )

Extcommunity: RT:65501:10 RT:65501:99999 ENCAP:8 Router MAC:0ca5.0000.1b08 <<< Two Route Target Values, implies that this route will be imported in two L2 EVPN tables - corresponding to RT value 10 which is MAC VRF for VNI5010 (RT = 65501:10) and corresponding to RT value 99999 which is IP VRF for VNI99999 ( RT = 99999:99999) on Other VTEPs.

Path-id 1 advertised to peers:

10.1.0.1
```

> **Router MAC:** `0ca5.0000.1b08` **- Used for Inner MAC Header source address for routed packets. This is needed because VXLAN is MAC in IP/UDP encapsulation tunneling mechanism and data payload over L3 border does not carry source host MAC address information. This is where the RMAC is used.**

Route Import Procedure

BGP on Receiving VTEP receives the MAC-IP Route update message and imports the routes into its BGP's local EVPN address family RIBs as per the route's VNI values. Since, MAC-IP route is associated with two VNI values (5010 and 99999) , therefore, BGP imports this route into its two EVPN Local RIBs for VNI 5010 and VNI 99999. **STEP 7**

Checking the route for host H1 192.168.10.10 on VTEP L4 in BGP Local EVPN RIB for L3 VNI 99999

```
leaf4# show bgp l2vpn evpn 192.168.10.10 vrf Cust-A

Route Distinguisher: 10.0.0.4:3 (L3VNI 99999)

BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]:[192.168.10.10]/272, version 7

Paths: (1 available, best #1)

Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

Advertised path-id 1

Path type: internal, path is valid, is best path, no labeled nexthop

Imported from 10.0.0.1:32777:[2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]

:[192.168.10.10]/272

AS-Path: NONE, path sourced internal to AS

10.0.0.1 (metric 81) from 10.1.0.1 (10.1.0.1)

Origin IGP, MED not set, localpref 100, weight 0
```

```
Received label 5010 99999
Extcommunity: RT:65501:10 RT:65501:99999 ENCAP:8 Router MAC:0ca5.0000.1b08
Originator: 10.0.0.1 Cluster list: 10.1.0.1
Path-id 1 not advertised to any peer
```

Route can be queried using L3VNI as a key also, the same output is displayed as above.

```
leaf4# show bgp l2vpn evpn vni-id 99999 detail
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 10.0.0.4:3 (L3VNI 99999)
BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]:[192.168.10.10]/272, version 7
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW
Advertised path-id 1
Path type: internal, path is valid, is best path, no labeled nexthop
Imported from 10.0.0.1:32777:[2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]
:[192.168.10.10]/272
AS-Path: NONE, path sourced internal to AS
10.0.0.1 (metric 81) from 10.1.0.1 (10.1.0.1)
Origin IGP, MED not set, localpref 100, weight 0
Received label 5010 99999
```

```
Extcommunity: RT:65501:10 RT:65501:99999 ENCAP:8 Router MAC:0ca5.0000.1b08

Originator: 10.0.0.1 Cluster list: 10.1.0.1

Path-id 1 not advertised to any peer
```

Since, it is a MAC-IP route that has VNI 5010 associated with it, the same route is installed in BGP local EVPN RIB for VNI 5010 also.

```
leaf4# show bgp l2vpn evpn vni-id 5010 detail

BGP routing table information for VRF default, address family L2VPN EVPN

Route Distinguisher: 10.0.0.4:32777 (L2VNI 5010)

BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[0]:[0.0.0.0]/216, version 8 <<< MAC-only route learned from MAC-only route BGP update

Paths: (1 available, best #1)

Flags: (0x000212) (high32 00000000) on xmit-list, is in l2rib/evpn, is not in HW

Advertised path-id 1

Path type: internal, path is valid, is best path, no labeled nexthop, in rib

Imported from 10.0.0.1:32777:[2]:[0]:[0]:[48]:[0c08.4e22.0001]:[0]:

[0.0.0.0]/216

AS-Path: NONE, path sourced internal to AS

10.0.0.1 (metric 81) from 10.1.0.1 (10.1.0.1)

Origin IGP, MED not set, localpref 100, weight 0

Received label 5010

Extcommunity: RT:65501:10 ENCAP:8
```

```
Originator: 10.0.0.1 Cluster list: 10.1.0.1

Path-id 1 not advertised to any peer

BGP routing table entry for [2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]:[192.168.10.10]/272, version 6 <<< MAC-IP route installed in EVPN local BGP RIB for VNI 5010

Paths: (1 available, best #1)

Flags: (0x000212) (high32 00000000) on xmit-list, is in l2rib/evpn, is not in HW

Advertised path-id 1

Path type: internal, path is valid, is best path, no labeled nexthop, in rib

Imported from 10.0.0.1:32777:[2]:[0]:[0]:[48]:[0c08.4e22.0001]:[32]:[192.168.10.10]/272

AS-Path: NONE, path sourced internal to AS

10.0.0.1 (metric 81) from 10.1.0.1 (10.1.0.1)

Origin IGP, MED not set, localpref 100, weight 0

Received label 5010 99999

Extcommunity: RT:65501:10 RT:65501:99999 ENCAP:8 Router MAC:0ca5.0000.1b08

Originator: 10.0.0.1 Cluster list: 10.1.0.1

Path-id 1 not advertised to any peer
```

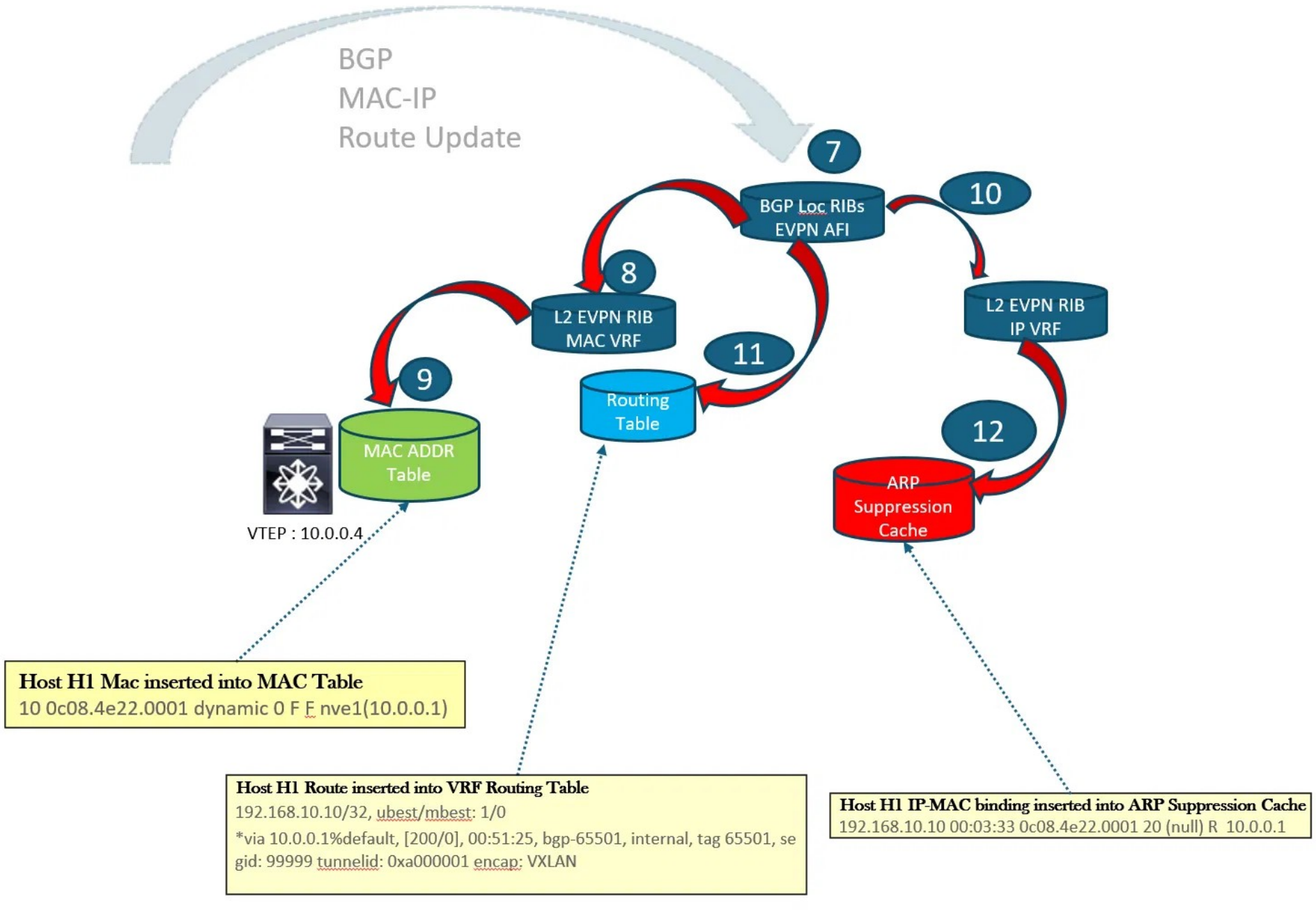
MAC-IP route has two route target values specified as Extended Community in a BGP msg. In the above output, one RT value is 65501:10. The L2 EVPN Rib is looked up using for this Route target value, which in this case is , EVPN RIB corresponding to VNI 5010. MAC-related information is pushed into this RIB (MAC VRF) **STEP 8**

```
leaf4# show l2route evpn mac all detail
```

```
Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
(Dup):Duplicate (Spl):Split (Rcv):Recv (AD):Auto-Delete (D):Del Pending
(S):Stale (C):Clear, (Ps):Peer Sync (O):Re-Originated (Nho):NH-Override
(Pf):Permanently-Frozen, (Orp): Orphan
Topology Mac Address Prod Flags Seq No Next-Hops
---------- -------------- ------ ------------ ---------- ------------------------------------
10 0c08.4e22.0001 BGP SplRcv 0 10.0.0.1 (Label: 5010) <<< This entry is creatd by Mac-only routes as well as by MAC-IP route
Route Resolution Type: Regular
Forwarding State: Resolved (PeerID: 1)
Sent To: L2FM
Encap: 1
999 0ca5.0000.1b08 VXLAN Rmac 0 10.0.0.1 <<< Router MAC is installed corresponding to Vlan mapped to L3VNI.
Route Resolution Type: Regular
Forwarding State: Resolved (PeerID: 1)
Sent To: L2FM
leaf4#
```

The other RT value is 99999:99999: The L2 EVPN RIB for this RT value is RIB specific to VNI 99999. The IP-related information is pushed into this RIB (IP-VRF table) from MAC-IP route by BGP. Therefore, the Import procedure involves the update of two Tables by MAC-IP BGP route - the MAC VRF table and IP VRF table. However, the update of MAC VRF table is redundant here as it is already updated by MAC-only route advertisement. Thus BGP redistributes routes into L2 EVPN local RIBs of Destination VTEP based on route's Target Values. **STEP 10**

```
leaf4# show l2route evpn mac-ip evi 10 detail

Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link

(Dup):Duplicate (Spl):Split (Rcv):Recv(D):Del Pending (S):Stale (C):Clear

(Ps):Peer Sync (Ro):Re-Originated (Orp):Orphan

Topology Mac Address Host IP Prod Flags

Seq No Next-Hops

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

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

10 0c08.4e22.0001 192.168.10.10 BGP --

0 10.0.0.1 (Label: 5010) <<< L2VNI of the Route

encap-type:1 << VxLAN Encap
```


BGP
MAC-IP
Route Update
7
BGP Loc RIBs
EVPN AFI
10
8
L2 EVPN RIB
MAC VRF
L2 EVPN RIB
IP VRF
11
9
Routing
Table
12
MAC ADDR
Table
ARP
Suppression
Cache
VTEP : 10.0.0.4
Host H1 Mac inserted into MAC Table
10 0c08.4e22.0001 dynamic 0 F F nve1(10.0.0.1)
Host H1 Route inserted into VRF Routing Table
192.168.10.10/32, ubest/mbest: 1/0
*via 10.0.0.1%default, [200/0], 00:51:25, bgp-65501, internal, tag 65501, se
gid: 99999 tunnelid: 0xa000001 encap: VXLAN
Host H1 IP-MAC binding inserted into ARP Suppression Cache
192.168.10.10 00:03:33 0c08.4e22.0001 20 (null) R 10.0.0.1

MAC VRF and IP VRF table updates by MAC-IP route

Installing the Route from MAC VRF table into MAC Address table on VTEP L4 **STEP 9**

```
leaf1# show mac address-table

VLAN MAC Address Type age Secure NTFY Ports

---------+-----------------+--------+---------+------+----+-----------------

* 10 0c08.4e22.0001 dynamic 0 F F Eth1/3 <<< New Entry which says a hosts with mac 0c08.4e22.0001 is present in vlan 10 and is reachable via local intf Eth1/3.

G - 0002.0002.0002 static - F F sup-eth1(R)

G - 0ca5.0000.1b08 static - F F sup-eth1(R)

G 10 0ca5.0000.1b08 static - F F sup-eth1(R)

G 20 0ca5.0000.1b08 static - F F sup-eth1(R)

G 999 0ca5.0000.1b08 static - F F sup-eth1(R)
```

Next is we need to check ARP Suppression Cache on VTEP L4. **STEP 12**

```
show ip arp suppression-cache detail
```

<No CLI on my image version ! >

At this point, the VxLAN network is able to work as a transparent Layer 2 switch for hosts participating in L2VNI 5010 and switch frames between the hosts connected to it

And Final Step is, MAC-IP Route is used by BGP to install the VRF Routing table.pm VTEP L4. This entry will be used for inter-VLAN routing. In this case, the host route (prefix /32) is installed. Therefore, this can lead to an over-whelmed routing table if there are too many hosts. **STEP 11**

```
leaf4# show ip route vrf Cust-A
```

```
IP Route Table for VRF "Cust-A"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>
192.168.10.0/24, ubest/mbest: 1/0, attached
*via 192.168.10.1, Vlan10, [0/0], 01:04:14, direct
192.168.10.1/32, ubest/mbest: 1/0, attached
*via 192.168.10.1, Vlan10, [0/0], 01:04:14, local
192.168.10.10/32, ubest/mbest: 1/0 <<< Route to Host H1 is installed in VRF routing table of VTEP L4
*via 10.0.0.1%default, [200/0], 00:51:25, bgp-65501, internal, tag 65501, se <<< Nexthop is 10.0.0.1
gid: 99999 tunnelid: 0xa000001 encap: VXLAN <<< Encapsulate L3VNI99999 using VxLAN encapsulation using tunnel 0xa000001 which is nve1
192.168.20.0/24, ubest/mbest: 1/0, attached
*via 192.168.20.1, Vlan20, [0/0], 01:04:14, direct
192.168.20.1/32, ubest/mbest: 1/0, attached
*via 192.168.20.1, Vlan20, [0/0], 01:04:14, local
```

Conclusion

In this tutorial, we learnt how local host on a VTEP report their presence to Leafs using ARP/GARP, and leaf in-turn picks up their Identities (

MAC/IP/VNI etc) and populate its MAC VRF tables which serves as the adverising point to advertise this information into iBGP domain. iBGP works as a transport protocol in control plane which distributes the Routes amont VTEPs. The VTEPs after receiving routes from iBGP reprogram their data plane to facilitate VxLAN based routing amongst hosts.

visit : http://www.csepracticals.com for more courses and offers.

Abhishek Sagar