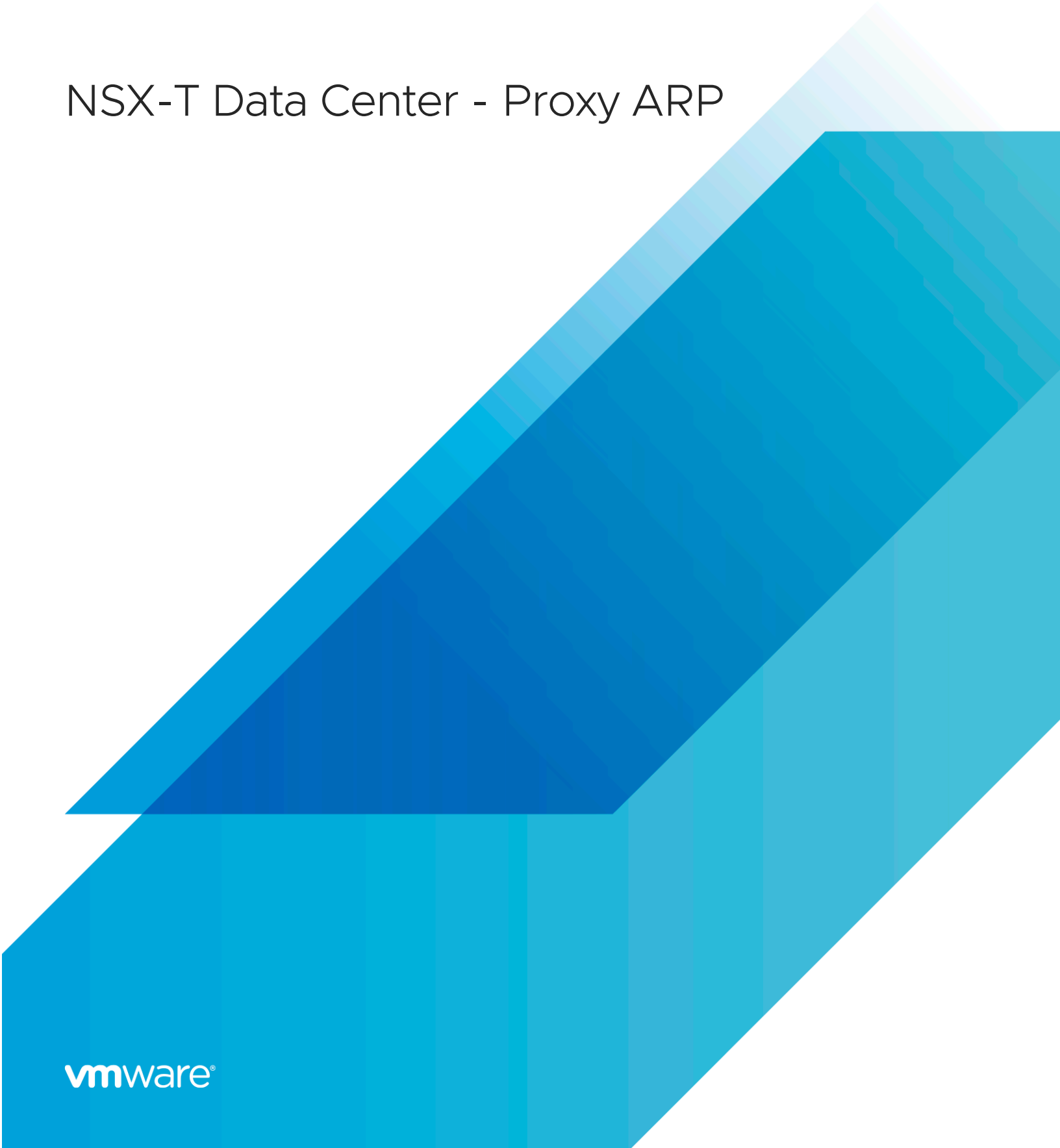
WHITE PAPER–JULY 2020

NSX-T Data Center - Proxy ARP



vmware®

Table of contents

Proxy ARP introduction

Proxy ARP is a method that consist of answering an ARP request on behalf of another host. This method is performed by a layer 3 networking device (usually a router). The purpose is to provide connectivity between 2 hosts when routing wouldn't be possible for various reasons.

To illustrate that introduction, we will consider the following network diagram:

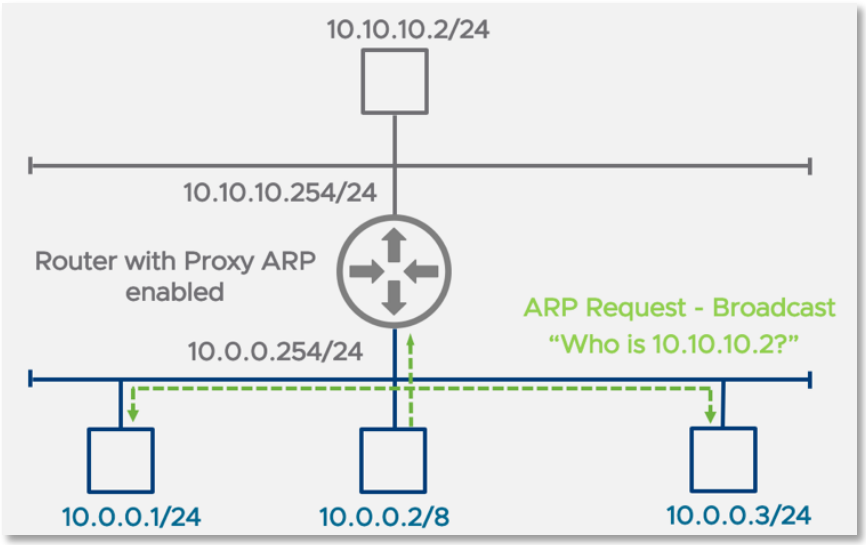


Figure 1 - Proxy ARP - ARP Request

In the following example, we will assume that the source host with the IP address 10.0.0.2 is configured with a subnet mask of 255.0.0.0. The destination host is configured with an IP address of 10.10.10.2 with a subnet mask of 255.255.255.0.

Based on its subnet mask, the source host thinks that the destination host is in the same subnet and therefore will need to have knowledge of its MAC address.

The source host will then send an ARP request (broadcast) that will reach the entire broadcast domain. It is crucial to note that the ARP request is not received by the destination host since it sits on a different broadcast domain from a router point of view.

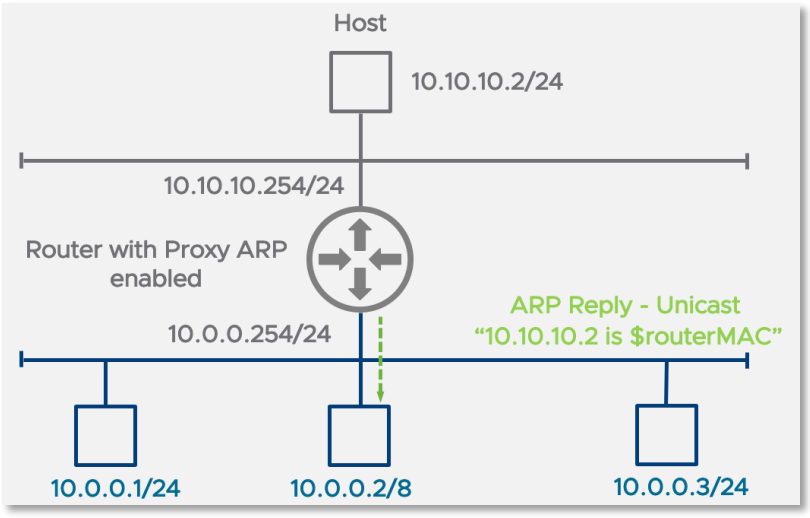


Figure 2 - Proxy ARP - ARP Reply

The Router has the "Proxy ARP" feature enabled and will reply with a unicast ARP reply to the source host on behalf of the destination host because it has a valid directly connected path to the requested IP. The MAC address specified in the ARP reply message is owned by the router physical interface.

After that sequence, the source host has knowledge of the destination host MAC address and a new entry in its ARP table has been added. The IP address of the destination host is mapped to the MAC Address of the router with Proxy ARP enabled. IP traffic can now be exchanged between the hosts.

Proxy ARP use case with NSX-T

In this case, the virtual machine connected to the NSX overlay segment is in a different subnet than the virtual machine connected to the VLAN segment.

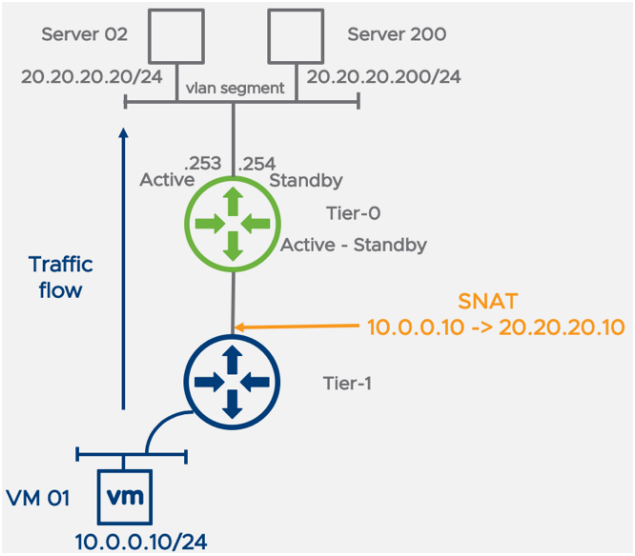


Figure 3 - Proxy ARP use case.

By enabling proxy-ARP, hosts on the overlay segments and hosts on a VLAN segment can exchange network traffic together without implementing any change in the physical networking fabric.

Proxy ARP is automatically enabled when a NAT rule or a load balancer VIP uses an IP address from the subnet of the Tier-0 gateway uplink.

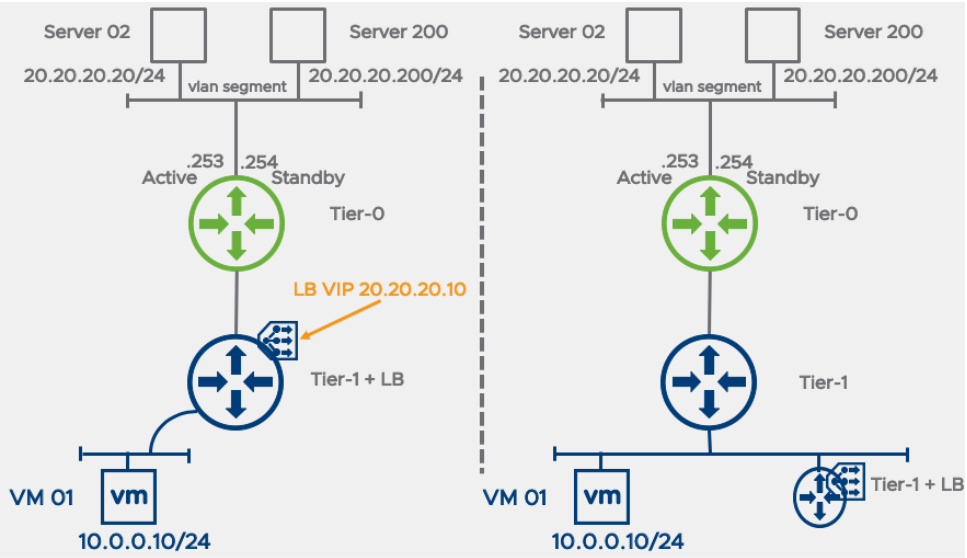


Figure 4 - Proxy ARP load balancer topologies.

Proxy ARP can be considered in environments where IP subnets are limited and where it is problematic to use new subnets easily and rapidly (either by using static routes or BGP).

Proof of concepts and VMware Enterprise PKS environments are usually using Proxy-ARP to simplify the network topology.

For production environment, VMware recommends implementing proper routing between a physical fabric and the NSX-T Tier-0 by using either static routes or Border Gateway Protocol with BFD. If proper routing is used between the Tier-0 gateway and the physical fabric, BFD with its sub-second timers will converge faster.

In case of failover with proxy ARP, the convergence relies on gratuitous ARP (broadcast) to update all hosts on the VLAN segment with the new MAC Address to use. If the Tier-0 gateway has proxy ARP enabled for 100 IP addresses, the newly active Tier-0 SR needs to send 100 Gratuitous ARP packets.

With such a mechanism, we can conclude that convergence time depends on the number IP addresses enabled for proxy ARP.

Proxy ARP Support and Summary

Proxy ARP on the NSX-T edge node is a feature supported since *NSX-T 2.4*.

The following table summarizes the design option and support for the Proxy ARP feature:

NSX-T PROXY ARP SUPPORT	
MODE	SUPPORT
Active / Standby	Supported[1]
Active / Active	Not Supported

Proxy ARP implementation in NSX-T Data Center – Active Standby architecture

We will be using the following diagram to illustrate this topology. In our demonstration, we will configure the SNAT rule on the Tier-1 SR Gateway.

Virtual Machine 01 is connected to the overlay segment and has an IP address of 10.0.0.10/24.

Server 02 and Server 200 are connected to the VLAN segment with an IP address of 20.20.20.20/24 and 20.20.20.200/24 respectively.

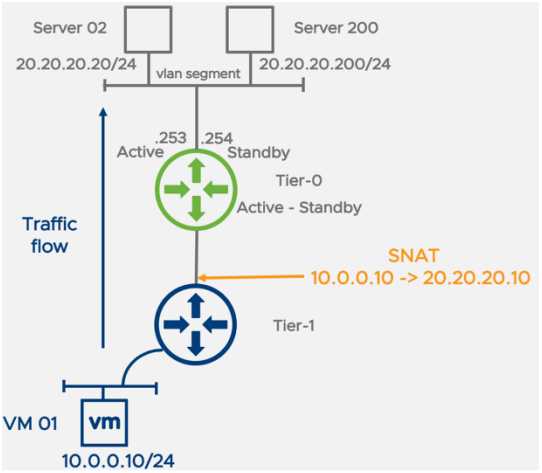


Figure 5 - Proxy ARP - Multi Tier Architecture.

In our example, the traffic is initiated by the virtual machine on the overlay segment. In order to enable the Proxy-ARP feature on the Tier-0 gateway uplinks, a Source NAT (SNAT) rule must be implemented on the Tier-1 (or Tier-0) gateway as described in the diagram above.

If the traffic is initiated by the server on the VLAN segment, a Destination NAT (DNAT) rule must be implemented on the Tier-1 gateway.

[1] In case of 1:1 SNAT, ICMP echo sent by the host on the VLAN segment will be answered by the Tier-0 SR gateway.

The overlay segment is connected to the Tier-1 gateway. The route advertisement statements must include the "NAT IPs" to redistribute these routes from the Tier-1 gateway to the Tier-0 gateway.

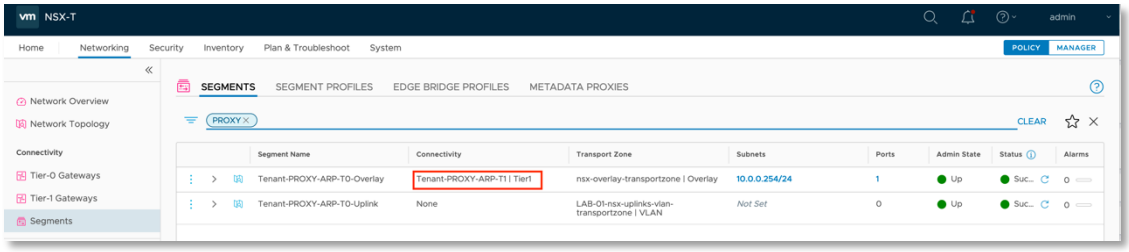


Figure 6 - Active Standby - Segments

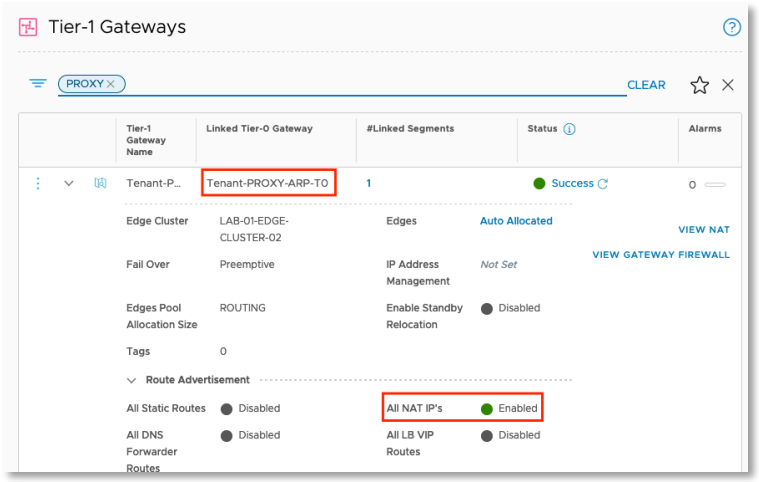


Figure 7 - Active/Standby - Tier-1 Configuration.

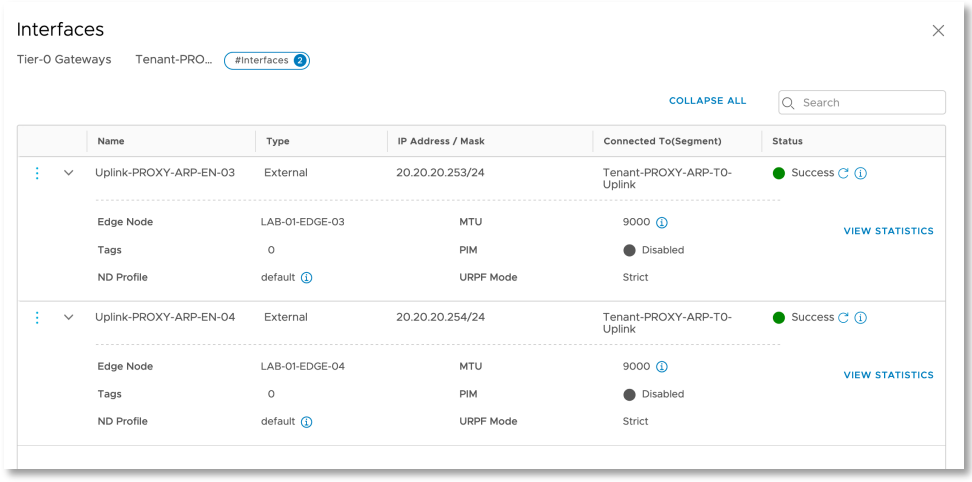


Figure 8 - Active/Standby - Tier-0 Uplink Interfaces

Assuming that the Source NAT rule is not yet configured, the following output demonstrates that Proxy ARP is not enabled on the uplink. The Tier-0 SR will not answer to the ARP requests on the VLAN segment that are needed to reach 20.20.20.10.

```
lab-01-edge-03> get logical-router
Logical Router
UUID                                   VRF    LR-ID  Name                        Type                        Ports
736a80e3-23f6-5a2d-81d6-bbefb2786666   0      0                                  TUNNEL                      4
52510271-dbbb-4a84-a8a3-e5f28146c926   1      22     SR-Tenant-PROXY-ARP-T0      SERVICE_ROUTER_TIER0        5
15b05770-78a9-4a36-b018-2bd8ae81c067   3      24     DR-Tenant-PROXY-ARP-T1      DISTRIBUTED_ROUTER_TIER1    4
8bc5b902-d446-4aac-a8cb-43e75bc77631   4      34     SR-Tenant-PROXY-ARP-T1      SERVICE_ROUTER_TIER1        5
c5ff5633-c4f3-459c-bf31-f6b013243e15   5      21     DR-Tenant-PROXY-ARP-T0      DISTRIBUTED_ROUTER_TIER0    4
```

Figure 9 - Finding the VRF ID on the Tier-0 Service Router

```
lab-01-edge-03> vrf 1
lab-01-edge-03(tier0_sr)> get interfaces
Logical Router
UUID                                   VRF    LR-ID  Name                        Type
52510271-dbbb-4a84-a8a3-e5f28146c926   1      22     SR-Tenant-PROXY-ARP-T0      SERVICE_ROUTER_TIER0
Interfaces (IPv6 DAD Status A-Assigned, D-Duplicate, T-Tentative)
    Interface     : 311f0f31-9571-5ba7-9da3-a36cc3884988
    Ifuid         : 266
    Mode          : cpu
    Port-type     : cpu

    Interface     : 1115faa8-5789-557d-9f2a-aa3f286bd6db
    Ifuid         : 267
    Mode          : blackhole
    Port-type     : blackhole

    Interface     : 6bc194a8-6cf7-43a4-80d8-4fb858c1777b
    Ifuid         : 286
    Mode          : loopback
    Port-type     : loopback
    IP/Mask       : 127.0.0.1/8;::1/128(NA)

    Interface     : 6bef3b7d-ef70-416b-9d74-31146ec1f837
    Ifuid         : 281
    Name          : Uplink-PROXY-ARP-EN-03
    Fwd-mode      : IPV4_AND_IPV6
    Internal name : uplink-281
    Mode          : lif
    Port-type     : uplink
    IP/Mask       : 20.20.20.253/24
    MAC           : 00:50:56:83:8e:90
    VLAN          : None
    Access-VLAN   : None
    LS port       : 3750d92c-d40a-4af5-936c-f579b8e8c42a
    Urpf-mode     : STRICT_MODE
    DAD-mode      : LOOSE
    RA-mode       : SLAAC_DNS_THROUGH_RA(M=0, O=0)
    Admin         : up
    Op_state      : up
    MTU           : 1500
```

Figure 10 - Finding the Interface UUID for the Tier-0 uplink

```
lab-01-edge-03> get logical-router interface 6bef3b7d-ef70-416b-9d74-31146ec1f837
interface    : 6bef3b7d-ef70-416b-9d74-31146ec1f837
ifuid        : 310
VRF          : 52510271-dbbb-4a84-a8a3-e5f28146c926
name         : Uplink-PROXY-ARP-EN-03
mode         : lif
IP/Mask      : 20.20.20.253/24
Fwd-mode     : IPV4_AND_IPV6
MAC          : 00:50:56:83:8e:90
VLAN         : 2000
LS port      : 3750d92c-d40a-4af5-936c-f579b8e8c42a
urpf-mode    : STRICT_MODE
admin        : up
op_state     : up
MTU          : 1500

lab-01-edge-03> get logical-router interface 6bef3b7d-ef70-416b-9d74-31146ec1f837 arp-proxy
interface    : 6bef3b7d-ef70-416b-9d74-31146ec1f837
ifuid        : 310
VRF          : 52510271-dbbb-4a84-a8a3-e5f28146c926
name         : Uplink-PROXY-ARP-EN-03
mode         : lif
MAC          : 00:50:56:83:8e:90
admin        : up
op_state     : up
arp_proxy    :
```

Figure 11 - Active Tier-0 Interfaces CLI output before NAT rule configuration

The next step is to configure the source NAT rule that matches the specific host 10.0.0.10. According to the rule, the Tier-01 SR will translate the source IP address of every packet sent by VM 01 with a new

value of “20.20.20.10”.

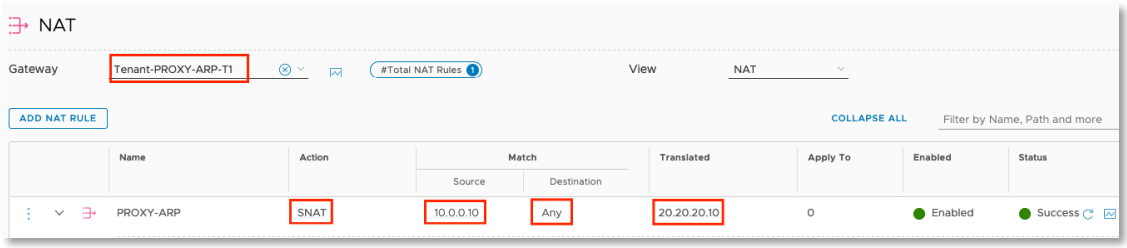


Figure 12 - Source NAT Rule configured on the Tier-1 gateway.

The following output demonstrates that Proxy ARP has been automatically enabled on the uplink interface of the active Tier-0 SR facing the 20.20.20.0/24 subnet.

```
lab-01-edge-03> get logical-router interface 6bef3b7d-ef70-416b-9d74-31146ec1f837
interface    : 6bef3b7d-ef70-416b-9d74-31146ec1f837
ifuid        : 310
VRF          : 52510271-dbbb-4a84-a8a3-e5f28146c926
name         : Uplink-PROXY-ARP-EN-03
mode         : lif
IP/Mask      : 20.20.20.253/24
fwd-mode     : IPV4_AND_IPV6
MAC          : 00:50:56:83:8e:90
VLAN         : 2000
LS port      : 3750d92c-d40a-4af5-936c-f579b8e8c42a
urpf-mode    : STRICT_MODE
admin        : up
op_state     : up
MTU          : 1500

lab-01-edge-03> get logical-router interface 6bef3b7d-ef70-416b-9d74-31146ec1f837 arp-proxy
interface    : 6bef3b7d-ef70-416b-9d74-31146ec1f837
ifuid        : 310
VRF          : 52510271-dbbb-4a84-a8a3-e5f28146c926
name         : Uplink-PROXY-ARP-EN-03
mode         : lif
MAC          : 00:50:56:83:8e:90
admin        : up
op_state     : up
arp_proxy    : 20.20.20.10
```

Figure 13 - Active Tier-0 Interfaces CLI output after NAT rule configuration.

The standby Tier-0 SR has also the “ARP_Proxy” field defined but it will not answer to any ARP request for that IP address while running in standby mode.

```
lab-01-edge-04> get logical-router interface a4280164-4633-4e1e-8f43-7a0e04044699
interface    : a4280164-4633-4e1e-8f43-7a0e04044699
ifuid        : 502
VRF          : 11f72fa5-3d06-49bd-8f9d-0b09d81849d6
name         : Uplink-PROXY-ARP-EN-04
mode         : lif
IP/Mask      : 20.20.20.254/24
fwd-mode     : IPV4_AND_IPV6
MAC          : 00:50:56:83:50:42
VLAN         : 2000
LS port      : 93d9f3b1-d8db-49b4-9df1-c126a0e1b9e0
urpf-mode    : STRICT_MODE
admin        : up
op_state     : up
MTU          : 1500

lab-01-edge-04> get logical-router interface a4280164-4633-4e1e-8f43-7a0e04044699 arp-proxy
interface    : a4280164-4633-4e1e-8f43-7a0e04044699
ifuid        : 502
VRF          : 11f72fa5-3d06-49bd-8f9d-0b09d81849d6
name         : Uplink-PROXY-ARP-EN-04
mode         : lif
MAC          : 00:50:56:83:50:42
admin        : up
op_state     : up
arp_proxy    : 20.20.20.10
```

Figure 14 - Standby Tier-0 Interfaces CLI output after NAT rule configuration.

Server 02 on the VLAN segment has the following ARP table:

```
C:\Users\admin>arp -a

Interface: 20.20.20.20 --- 0x3
  Internet Address      Physical Address      Type
  20.20.20.253          00-50-56-83-8e-90     dynamic
  20.20.20.254          00-50-56-83-50-42     dynamic
  20.20.20.255          ff-ff-ff-ff-ff-ff     static
  224.0.0.22            01-00-5e-00-00-16     static
  224.0.0.251           01-00-5e-00-00-fb     static
  224.0.0.252           01-00-5e-00-00-fc     static
  239.255.255.250       01-00-5e-7f-ff-fa     static
```

Figure 15 - Server 02 - ARP entries before any traffic is initiated

The Active Tier-0 SR interface IP address of 20.20.20.253 corresponds to the mac address "00:50:56:83:8E:90".

The Standby Tier-0 SR interface IP address of 20.20.20.254 corresponds to the mac address "00:50:56:83:50:42".

The previous output demonstrates that the Server 02 has already the knowledge of both Tier-0 Uplink interfaces. The virtual machine 01 on the overlay segment initiate some network traffic to the destination of Server 02 on the VLAN segment. At first, the traffic is sent to the Tier-01 DR on the transport node and then sent to the Tier-01 SR on an edge node. The IP traffic will match the SNAT rule previously configured on the Tier-01 SR. All the IP packets sourced by the virtual machine 01 are sent to the Tier-0 DR/SR with a source IP address of 20.20.20.10.

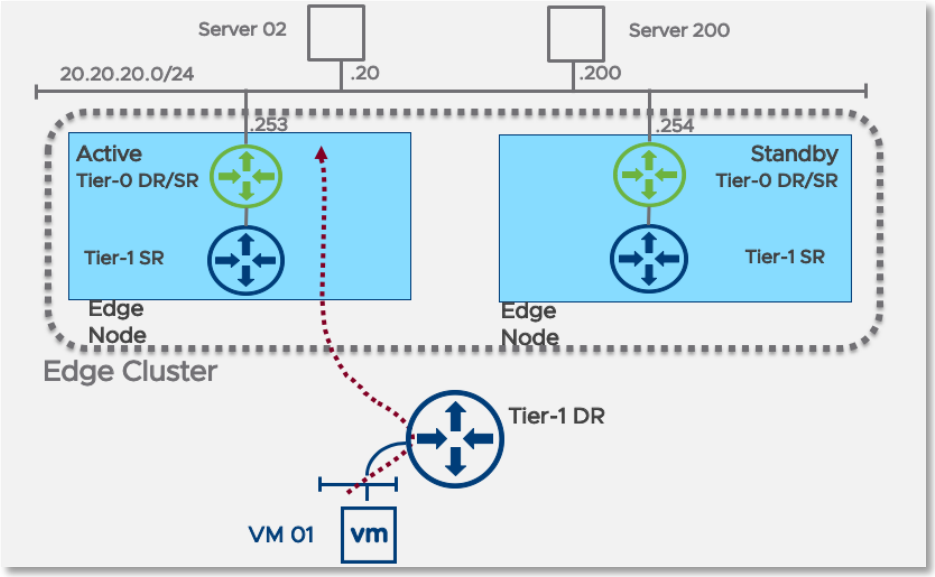


Figure 16 - Traffic initiated by virtual mach 01 and traversing the NSX fabric.

The active Tier-0 DR sends the traffic to the active Tier-0 SR. When the traffic reaches the Tier-0 SR, the MAC address for Server 02 is unknown from a Tier-0 point of view.

The following network trace will demonstrate how the ARP request/replies and the ICMP packets are exchanged between the Tier-0 and Server 02.

Packet #`1 – Broadcast: it shows the ARP request sent by the active Tier-0 SR needed in order to map the MAC address of Server 02.

Packet #2 – Unicast: shows the ARP reply sent by Server 02 to the active Tier-0 SR with its own MAC Address embedded.

Packet #3 – Unicast: ICMP Echo request type 8 code 0 – Sent by the Active Tier-0 SR.

Packet #4 – Broadcast: it shows the ARP request sent by Server 02 needed in order to map the MAC address of 20.20.20.10.[2]

Packet #5 – Unicast: shows the ARP reply sent by the active Tier-0 SR showing that the Proxy ARP feature is enabled and effective.

Packet #6 – Unicast: ICMP Echo reply type 0 code 0 – Sent by Server 02.

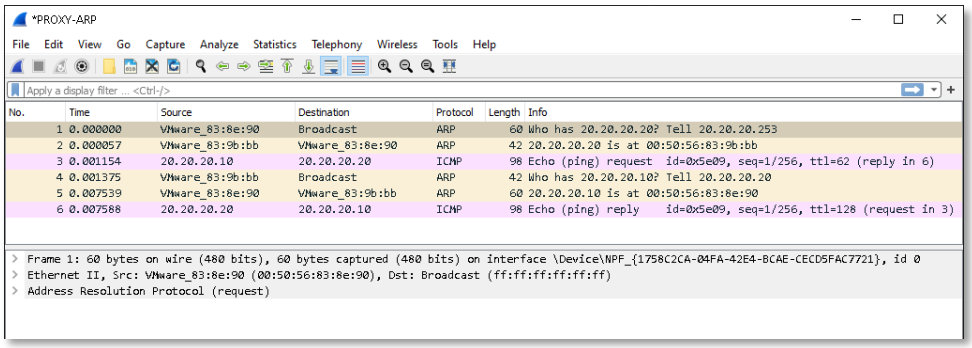


Figure 17 – ICMP trace on host 02 – Traffic initiated by host 01

The following ARP table output will demonstrate that the MAC Address mapped to the respective IP Addresses of 20.20.20.10 and 20.20.20.253 are identical.

```
C:\Users\admin>arp -a

Interface: 20.20.20.20 --- 0x3
  Internet Address      Physical Address      Type
  20.20.20.10           00-50-56-83-8e-90     dynamic
  20.20.20.253          00-50-56-83-8e-90     dynamic
  20.20.20.255          ff-ff-ff-ff-ff-ff     static
  224.0.0.22            01-00-5e-00-00-16     static
  224.0.0.251           01-00-5e-00-00-fb     static
  224.0.0.252           01-00-5e-00-00-fc     static
  239.255.255.250       01-00-5e-7f-ff-fa     static
```

Figure 18 – ARP table Host 02

[2] ARP caching behavior has been changed in Windows Vista. The TCP/IP stack implementations in Windows Vista (and later versions) comply with RFC4861 (Neighbor Discovery protocol for IP version 6 [Ipv6]) for both the IPv4 and the IPv6 Neighbor Discovery process. : Source *https://support.microsoft.com/en-us/help/949589/description-of-address-resolution-protocol-arp-caching-behavior-in-win.*

As a reminder, it is crucial to note that in our case, the traffic is initiated by the virtual machine 01 which is connected to the overlay segment on the Tier-01. If the initial traffic was initiated by a server on the VLAN segment, a Destination NAT rule would have been required on the Tier-1/Tier-0 since the initial traffic would not match the SNAT rule that has been configured previously.

If the edge nodes hosting the Tier-0 gateways experience an outage, a failover will be triggered, the newly active Tier-0 gateway will send a gratuitous ARP to announce the new MAC address to be used by the hosts on the VLAN segment in order to reach virtual machine 01. It is critical to fathom that the newly active Tier-0 will send a Gratuitous ARP for each IP address that are configured for Proxy ARP.

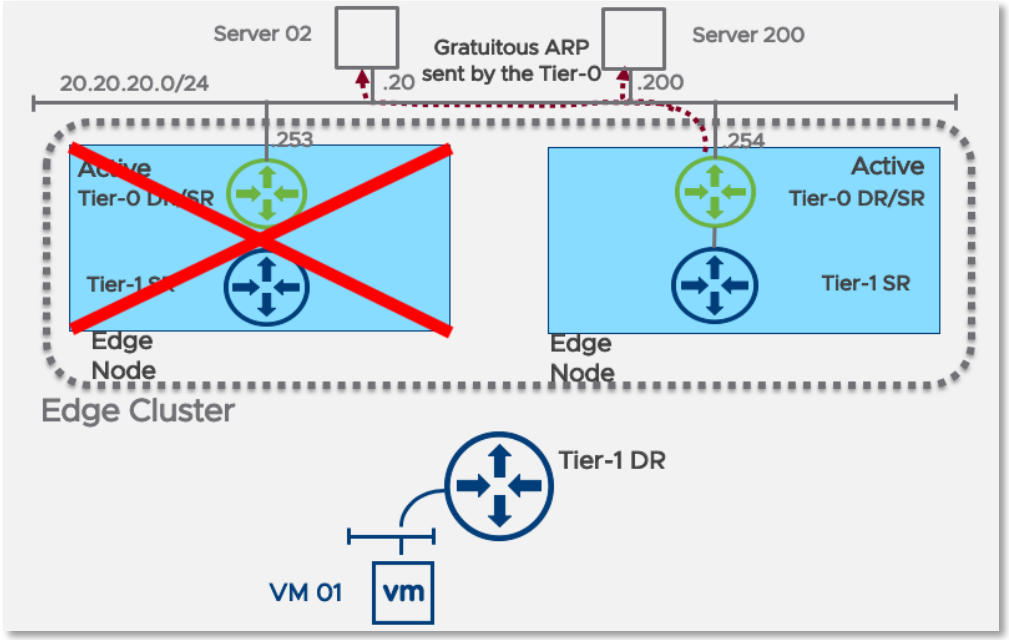


Figure 19 - Edge node failover - Gratuitous ARP sent by the Tier-0.

The following capture shows a single gratuitous ARP sent by the newly active Tier-0 for the IP address 20.20.20.10.

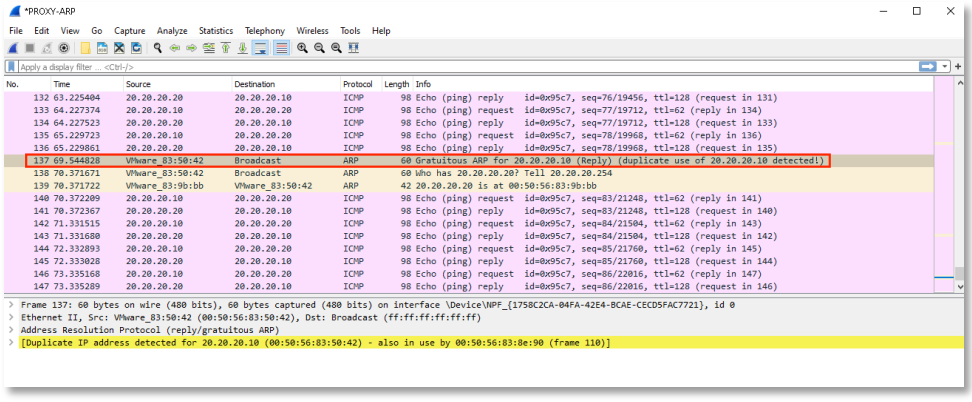


Figure 20 - Edge node failover - GARP sent by the newly active Tier-0 Gateway.

After announcing the new MAC address to use, Server 02 updates (and all other hosts on the VLAN segment with an entry for 20.20.20.10) its ARP table entry for 20.20.20.10 and traffic will be sent to the newly active Tier-0 SR.

```
C:\Users\admin>arp -a

Interface: 20.20.20.20 --- 0x3
  Internet Address      Physical Address      Type
  20.20.20.10           00-50-56-83-50-42     dynamic
  20.20.20.253          00-50-56-83-8e-90     dynamic
  20.20.20.254          00-50-56-83-50-42     dynamic
  20.20.20.255          ff-ff-ff-ff-ff-ff     static
  224.0.0.22            01-00-5e-00-00-16     static
  224.0.0.251           01-00-5e-00-00-fb     static
  224.0.0.252           01-00-5e-00-00-fc     static
  239.255.255.250       01-00-5e-7f-ff-fa     static
```

Figure 21 - ARP Table Server 02 - After Tier-0 failover.

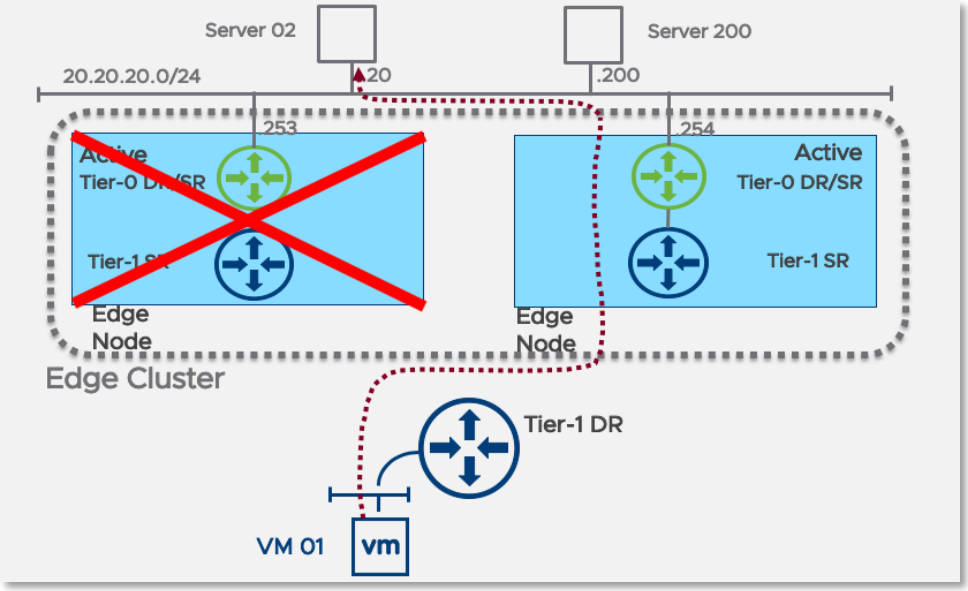


Figure 22 - Edge node failover - Traffic initiated by Host 01.

Proxy ARP implementation in NSX-T Data Center – Active Active architecture

In an Active/Active mode, IP traffic uses all the equal cost and available paths to reach the physical fabric through the Tier-0 gateway. In this mode, stateful NAT is not supported because of asymmetrical path might cause issues. It is possible to configure reflexive NAT. The stateless NAT rule will rewrite the source IP address field of the IP packets sent by VM 01.

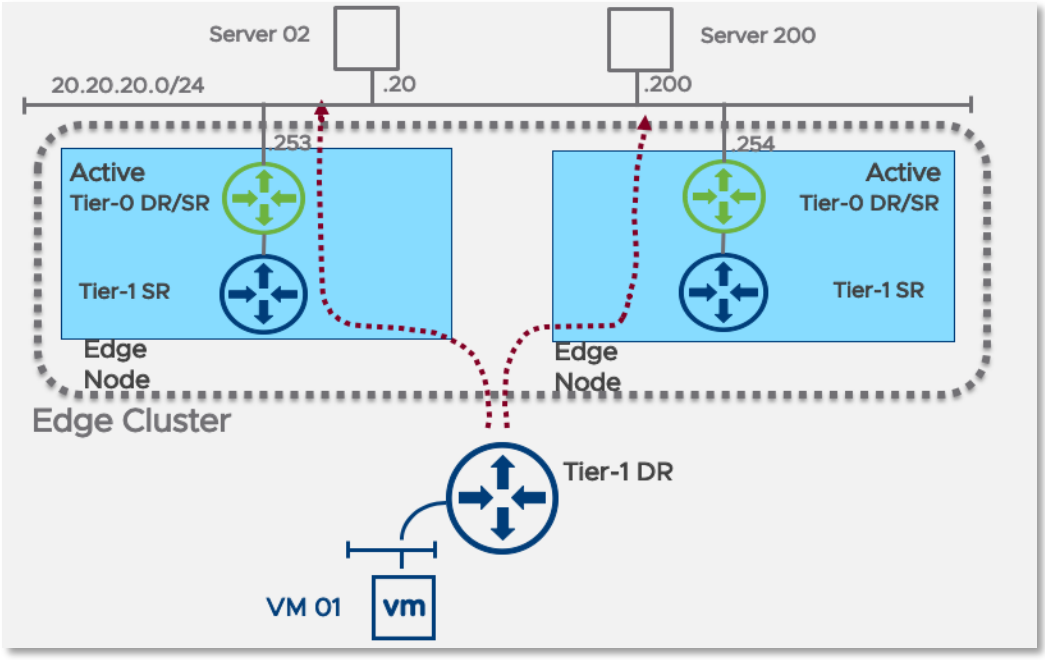


Figure 23 - Active/Active Tier-0 - Traffic initiated by VM 01

In contrast with our previous example, none of the Tier-0 SR will respond to the ARP Requests sent by the servers on the VLAN segments for the IP 20.20.20.10.

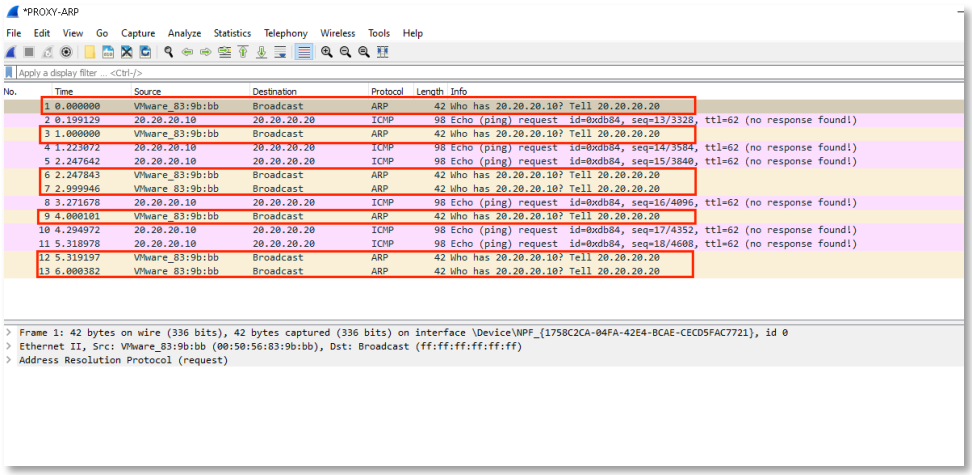


Figure 24 - Active/Active Topology – Tier-0 does not answer to ARP requests sent for 20.20.20.10

```
lab-01-edge-03> get logical-router interface 6bef3b7d-ef70-416b-9d74-31146ec1f837
interface    : 6bef3b7d-ef70-416b-9d74-31146ec1f837
ifuid        : 281
VRF          : 52510271-dbbb-4a84-a8a3-e5f28146c926
name         : Uplink-PROXY-ARP-EN-03
mode         : lif
IP/Mask      : 20.20.20.253/24
Fwd-mode     : IPV4_AND_IPV6
MAC          : 00:50:56:83:8e:90
VLAN         : 2000
LS port      : 3750d92c-d40a-4af5-936c-f579b8e8c42a
urpf-mode    : STRICT_MODE
admin        : up
op_state     : up
MTU          : 1500

lab-01-edge-03> get logical-router interface 6bef3b7d-ef70-416b-9d74-31146ec1f837 arp-proxy
interface    : 6bef3b7d-ef70-416b-9d74-31146ec1f837
ifuid        : 281
VRF          : 52510271-dbbb-4a84-a8a3-e5f28146c926
name         : Uplink-PROXY-ARP-EN-03
mode         : lif
MAC          : 00:50:56:83:8e:90
admin        : up
op_state     : up
arp_proxy    :
```

Figure 25 - Active Tier-0 Interfaces CLI - Edge Node 03 - output after NAT rule configuration.

```
lab-01-edge-04> get logical-router interface eb12d624-2c32-4354-b8f3-39aea8d1455d
interface    : eb12d624-2c32-4354-b8f3-39aea8d1455d
ifuid        : 612
VRF          : b2acc729-b19f-4607-84db-ffe394883ac2
name         : Uplink-PROXY-ARP-EN-04
mode         : lif
IP/Mask      : 20.20.20.254/24
Fwd-mode     : IPV4_AND_IPV6
MAC          : 00:50:56:83:50:42
VLAN         : 2000
LS port      : 67040350-10fc-4f4e-b9d5-350b02abd057
urpf-mode    : STRICT_MODE
admin        : up
op_state     : up
MTU          : 1500

lab-01-edge-04> get logical-router interface eb12d624-2c32-4354-b8f3-39aea8d1455d  arp-proxy
interface    : eb12d624-2c32-4354-b8f3-39aea8d1455d
ifuid        : 612
VRF          : b2acc729-b19f-4607-84db-ffe394883ac2
name         : Uplink-PROXY-ARP-EN-04
mode         : lif
MAC          : 00:50:56:83:50:42
admin        : up
op_state     : up
arp_proxy    :
```

Figure 26 - Standby Tier-0 Interfaces CLI - Edge Node 03 - output after NAT rule configuration

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 vmware.com.