Troubleshoot IGMP for NLB Deployments on Catalyst 9000 Switches

Contents

Introduction

This document describes how the IGMP feature on Catalyst 9000 series switches behaves in a Microsoft Network Load Balancer (NLB) deployment.

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- Microsoft NLB modes of operation
- IGMP Multicast

Components Used

The information in this document is based on these software and hardware versions:

- Catalyst 9200
- Catalyst 9300
- Catalyst 9400
- Catalyst 9500
- Catalyst 9600

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

NLB is a cluster technology available in all Windows 2000 Server and Windows 2003 Server family systems. It provides a single virtual IP address for all clients as the destination IP address for the entire cluster.

NLB can be used in order to distribute client requests across a set of servers. In order to ensure clients experience acceptable performance levels, NLB provides the ability to add additional servers to scale out stateless applications (such as IIS-based web servers) as client load increases. In addition, it reduces downtime that is caused by server malfunction.

You can configure NLB to work in one of these three modes:

- Unicast mode
- Multicast mode
- Internet Group Management Protocol (IGMP) mode

> **Tip**: The unicast mode and multicast mode deployments have the same configuration and verification described in the document Catalyst Switches for Microsoft Network Load Balancing Configuration Example

---

This document is focused on the Internet Group Management Protocol (IGMP) mode.

Best Practices

Catalyst 9000 Series switches snoop the layer 3 headers of IGMP packets in order to populate the Snooping table. Due to how NLB must be configured on the switch using a static multicast MAC, the IGMP Snooping table is not populated and flooding in the destination VLAN does occur. In other words, IGMP Snooping in Catalyst 9000 does not automatically contain the multicast flood when the NLB Server is in IGMP mode (forwarding in Catalyst 9000 is based on multicast IP and not on multicast MAC address).

> **Note**: On Catalyst 9000 flooding occurs in all three modes of NLB. Flooding does not occur in the user VLAN, given that the destination of the packets has to be their default gateway. Only after header rewrite to the destination VLAN, the flood occurs.

Therefore, consider these best practices for successful deployments:

- Use a dedicated VLAN to constrain the flooding only to the NLB cluster.
- Utilize static MAC entries to limit the ports in which the flood occurs within the NLB VLAN.

IGMP Mode

In this mode, the virtual MAC of the NLB cluster falls within the Internet Assigned Numbers Authority (IANA) range, and it starts with 0100.5exx.xxxx. The IGMP Snooping feature configured on the switch does not program in the MAC address table the virtual multicast MAC address of the cluster. Since this dynamic programming is absent, the multicast traffic received by the switch from the NLB cluster is flooded to all the ports members of the same VLAN. Cisco bug ID CSCvw18989.

For the topologies where the NLB servers are in different VLAN than the users, since the virtual IP address of the cluster uses a multicast MAC address, it is unreachable outside of the local subnet. In order to address this, you must configure a static ARP entry on each device with a layer 3 interface in the cluster VLAN.

IGMP Snooping feature in the Catalyst 9000 series switches do not use the multicast MAC address for forwarding. They use the multicast IP address, this is why it is unable to automatically program multicast MAC address in the MAC table as other legacy platforms do (such as Catalyst 6000 series). All new platforms use the multicast IP address forwarding method to avoid the overlapping addresses issue found on legacy switches.

> **Note**: An ethernet multicast MAC address has some overlap. The same MAC address is assigned to 32 different multicast groups. If one user on an ethernet segment subscribes to multicast group

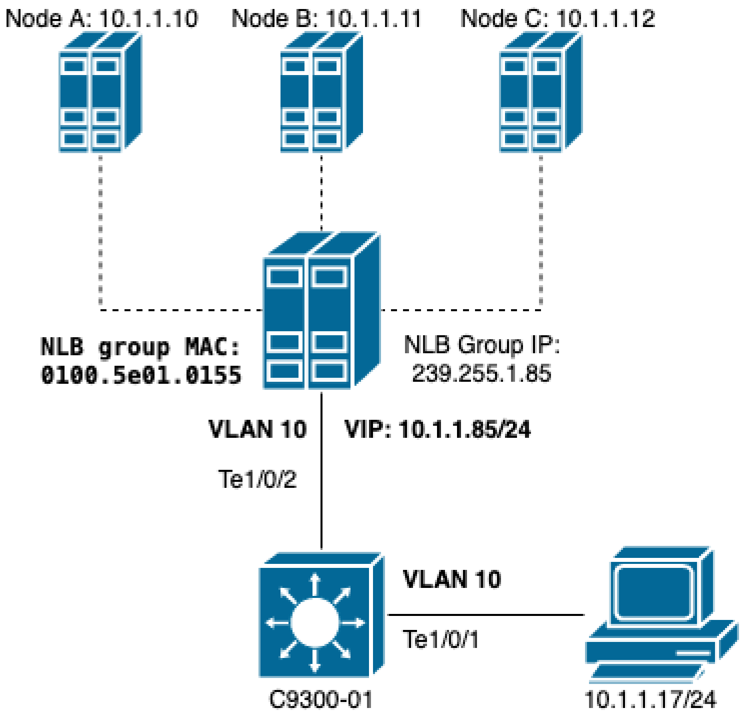
---

225.1.1.1, and another user subscribes to 230.1.1.1, both users receives both multicast streams (MAC address is the same 01-00-5e-01-01-01). In engineering multicast networks on LAN segments, this overlap needs to be specifically watched for and engineered to avoid the problem.

Configure

Source and Destination in the same VLAN

Network Diagram



This section describes how to configure NLB when the cluster and the users are in the same VLAN.

1. Verify the NLB VLAN is created. It is suggested to have a dedicated VLAN for NLB traffic due to the flood.

<#root>

C9300-01#

show vlan id 10

```
VLAN Name                             Status    Ports
---- -------------------------------- --------- -------------------------------
10   NLB                              active    Te1/0/1, Te1/0/2, Te1/0/3

VLAN Type  SAID       MTU   Parent RingNo BridgeNo Stp  BrdgMode Trans1 Trans2
---- ----- ---------- ----- ------ ------ -------- ---- -------- ------ ------
10   enet  100010     1500  -      -      -        -    -        0      0

Remote SPAN VLAN
----------------
Disabled
```

```
Primary Secondary Type              Ports
------- --------- ----------------- ------------------------------------------
```

2. Configure a static MAC address entry for the ports that must get this NLB traffic. This command must include all trunk ports or access ports in the path towards the NLB Cluster in the NLB VLAN. In the diagram, there is only one path towards the NLB via Tengig1/0/2.

<#root>

C9300-01(config)#

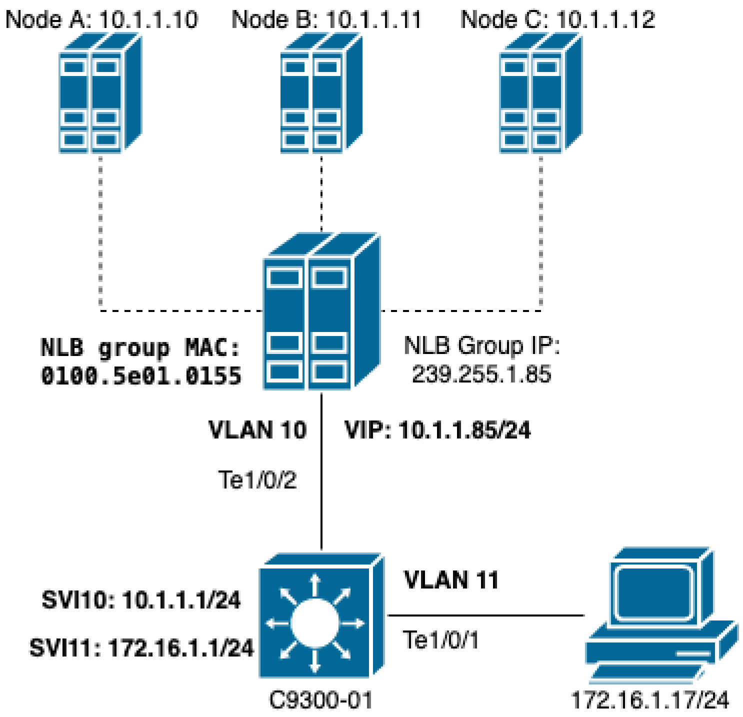
```
mac address-table static 0100.5e01.0155 vlan 10 interface TenGigabitEthernet 1/0/2
```

```
C9300-01#

show run | in mac

mac address-table static 0100.5e01.0155 vlan 10 interface TenGigabitEthernet1/0/2
```

> **Note**: You can have as many mapped ports in the static MAC address entry as you need. This map of ports reduces the expected flood within the VLAN of the NLB. In the example, the static MAC entry can avoid the traffic towards the NLB Cluster to be flooded out of Te1/0/3.

Source and Destination in different VLAN

Network Diagram



This section describes how to configure NLB when the cluster and the users are in different VLANs.

1. Configure the NLB VLAN and an IP address to be the default gateway of the NLB cluster.

<#root>

```
C9300-01#

show vlan id 10

VLAN Name                             Status    Ports
---- -------------------------------- --------- -------------------------------
10   NLB                              active    Te1/0/2, Te1/0/3

VLAN Type  SAID       MTU   Parent RingNo BridgeNo Stp  BrdgMode Trans1 Trans2
---- ----- ---------- ----- ------ ------ -------- ---- -------- ------ ------
10   enet  100010     1500  -      -      -        -    -        0      0

Remote SPAN VLAN
----------------
Disabled

Primary Secondary Type              Ports
------- --------- ----------------- ------------------------------------------

C9300-01#

show run interface vlan 10

Building configuration...

Current configuration : 59 bytes
!
interface Vlan10
 ip address 10.1.1.1 255.255.255.0
end
```

2. Configure a static ARP entry for the virtual IP address of the NLB cluster servers. The static ARP must be configured on all of the layer 3 devices that have a Switch Virtual Interface (SVI) in the cluster VLAN. The purpose of the static ARP is to allow the switch to have the rewrite information necessary to send routed packets towards the NLB VLAN.

<#root>

```
C9300-01(config)#

arp 10.1.1.85 0100.5e01.0155 arpa
```

3. Verify the user VLAN created at the access layer and its default gateway. It is important that you configure the default gateway on both parties. (NLB cluster and users).

<#root>

```
C9300-01#

show vlan id 11
```

```
VLAN Name                             Status    Ports
---- -------------------------------- --------- -------------------------------
11   Users2                           active    Te1/0/1, Te1/0/4

VLAN Type  SAID       MTU   Parent RingNo BridgeNo Stp  BrdgMode Trans1 Trans2
---- ----- ---------- ----- ------ ------ -------- ---- -------- ------ ------
11   enet  100011     1500  -      -      -        -    -        0      0

Remote SPAN VLAN
----------------
Disabled

Primary Secondary Type              Ports
------- --------- ----------------- ------------------------------------------

C9300-01#
```

```
show run interface vlan 11
```

```
Building configuration...

Current configuration : 59 bytes
!
interface Vlan11
 ip address 172.16.1.1 255.255.255.0
end
```

> **Note**: Any packet that is routed after MAC header rewrite whose destination MAC is not learnt in egress SVI, the packet is then flooded in the corresponding VLAN. In order to mitigate the flood, you need to create a gateway and a separate VLAN just for the NLB servers. If you do not want to configure a dedicated VLAN for the NLB traffic, then you can configure a static MAC address entry for the ports that must receive the NLB traffic, that is, **mac address-table static** *0100.5exx.xxxx* **vlan** *#* **interface** *interface_name*

Troubleshoot

1. Verify if the static MAC address are configured to all the destination ports that need to forward the traffic to the NLB.

<#root>

C9300-01#

```
show mac address multicast
```

```
Vlan Mac Address Type Ports
---- ----------- ---- -----
10 0100.5e01.0155 USER Te1/0/2
```

2. For deployments where the NLB cluster is in different subnet than the clients, verify if there are static ARP entries that map the Virtual IP of the NLB server with its multicast MAC address.

<#root>

```
C9300-01#

show run | in arp

arp 10.1.1.85 0100.5e01.0155 ARPA

C9300-01#

show ip arp

Protocol Address Age (min) Hardware Addr Type Interface
Internet 10.1.1.1 - c4c6.0309.cf46 ARPA Vlan10
Internet 10.1.1.85 - 0100.5e01.0155 ARPA
Internet 172.16.1.1 - c4c6.0309.cf54 ARPA Vlan11
```

3. Do a ping to the NLB Server IP with a size that is not been frequently used. Clear the controllers of the port and check with multiple iterations of the command which size is not been used that much.

<#root>

```
C9300-01#

show controllers ethernet-controller Te1/0/2 | in 1024

0 1024 to 1518 byte frames 0 1024 to 1518 byte frames

C9300-01#

clear controllers ethernet-controller Te1/0/2


HOST#

ping 10.1.1.85 time 0 rep 1000 size 1024

Type escape sequence to abort.
Sending 1000, 1024-byte ICMP Echos to 10.1.1.85, timeout is 0 seconds:
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
......................................................................
....................
Success rate is 0 percent (0/1000)
```

4. Check if the pings are been forwarded out of the destination port properly and if the same ping test is been flooded to other ports. Verify this with the same controllers counters command on the different interfaces.

<#root>

```
C9300-01#

show controllers ethernet-controller Te1/0/1 | in 1024

 <-- Ingress Host port
0 1024 to 1518 byte frames 1000 1024 to 1518 byte frames

C9300-01#

show controllers ethernet-controller Te1/0/2 | in 1024

 <-- Egress port to NLB
1000 1024 to 1518 byte frames 0 1024 to 1518 byte frames
```

5. Take packet captures on the ingress port with EPC and on the egress port with SPAN and check if the switch is forwarding the data or not.

<#root>

```
C9300-01#

monitor capture tac buffer size 10 match any interface Te1/0/1 in

C9300-01#

monitor capture tac start

C9300-01#

monitor capture tac stop

C9300-01#

monitor capture tac export location flash:DataTraffic.pcap
```

> **Tip**: Embedded Packet Capture (EPC) functionality is reliable when packets are forwarded in layer 2 ingress or egress direction. However, if the traffic is routed by the switch and then forwarded to the egress port, EPC is not be reliable. To capture packets in egress after layer 3 routing occurs, use Switch Port Analyzer (SPAN) feature.

<#root>

```
C9300-01(config)#

monitor session 1 source interface Te1/0/2 tx

C9300-01(config)#

monitor session 1 destination interface Te1/0/3 encapsulation replicate


C9300-01#

show monitor session all

Session 1
---------
Type : Local Session
Source Ports :
```

```
TX Only : Te1/0/2
Destination Ports : Te1/0/3
Encapsulation : Replicate
Ingress : Disabled
```

Related Information

- **Catalyst Switches for Microsoft Network Load Balancing Configuration Example**
- **Cisco Technical Support & Downloads**