Technical Guide

Internet Protocol (IP) Addressing and Protocols

Feature Overview and Configuration Guide

Introduction

This guide describes how to configure IPv4 addressing and the protocols used to help IP function on your network.

As well as the familiar Internet (with uppercase “I”), the term internet (with lowercase “i”) can refer to any network (usually a wide area network) that uses the Internet Protocol. This guide concentrates on this definition—a generalized network that uses IP as its network protocol.

Products and software version that apply to this guide

This guide applies to all AlliedWare Plus™ products, running version **5.4.4** or later.

However, feature support and implementation varies between products. To see whether a product supports a particular feature or command, see the following documents:

- The product’s Datasheet
- The AlliedWare Plus Datasheet
- The product’s Command Reference

These documents are available from the above links on our website at alliedtelesis.com.

Feature support may change in later software versions. For the latest information, see the above documents.

Content

Assigning an IP Address

To configure your device to perform IP routing (for example, to access the Internet) you need to configure IP. You also need to configure IP if you want to manage your device from any IP-based management process (such as SSH, Telnet, or SNMP).

Add an IP address to each of the interfaces to or from which you want to route IP, or which you want to use as a management interface on the device. You can configure an interface on your device with a static IP address, or with a dynamic IP address assigned using your device's DHCP client.

Static IP addresses

To add a static IP address to an interface, enter interface mode for the interface that you want to configure, then use the command:

```
awplus(config-if)# ip address <ip-addr/prefix-length> [secondary] [label
<label>]
```

where ***<ip-address/prefix-length>*** is the IP address followed by a slash then the prefix length. Note that you cannot specify the mask in dotted decimal notation in this command.

Note: On SBx8100 Series switches, the subnet 192.168.255.0/28 is internally reserved and cannot be configured.

For example, to give the interface vlan1 an address of 192.168.10.10, with a class C subnet mask, use the command:

```
awplus(config-if)#ip address 192.168.10.10/24
```

The secondary parameter allows you to add multiple IP addresses to an interface using this command. Each interface must have a primary IP address before you can add a secondary address. Your device treats secondary addresses the same as primary addresses in most respects, such as responding for ARP requests for the IP address. However, the only packets generated that have a secondary address as source address are routing updates. You can define up to 32 secondary addresses on a single interface.

DHCP dynamic addresses

When you use the DHCP client, it obtains the IP address and subnet mask for the interface, and other IP configuration parameters, from a DHCP server. To configure an interface to gain its IP configuration using the DHCP client, use the command:

```
awplus(config-if)# ip address dhcp [client-id <interface>]
[hostname <hostname>]
```

- If an IP interface is configured to get its IP address and subnet mask from DHCP, the interface does not take part in IP routing until the IP address and subnet mask have been set by DHCP.
- If you need to make a static entry in the DHCP server from which the device is obtaining its IP address, you need your device's MAC address, which you can display by using the command: **show interface**

IPv4 /31 bit subnetting

IPv4 /31 bit subnetting is used when peering to ISP equipment where there is a strong need to conserve IPv4 address space. IPv4 /31 bit subnetting conserves address space because there is no concept of IP subnet and IP broadcast addresses associated with a /31 subnet. A /31 bit IPv4 subnet supports only two IPv4 addresses, each of which is allocated to a host.

IPv4 /31 bit subnetting is sometimes used on point-to-point (PPP) links where IP subnet and IP broadcast addresses are unnecessary.

From software release **5.4.9-1** onwards, IPv4 /31 bit subnetting is available on all AlliedWare Plus switch and router platforms that support Layer 3 interfaces. The points to note include:

- You can configure /31 subnets on **non**-PPP/tunnel Layer 3 interfaces, such as VLAN, Ethernet, and Bridge. Prior to this release, this was only supported on AlliedWare Plus router PPP interfaces.
- The following warning message is displayed when a /31 subnet is configured on non-PPP interfaces:

  ```
  Warning: use /31 mask on non point-to-point interface cautiously.
  ```

- IP addresses in a /31 subnet can be learned on a device via DHCP configuration.
- The RIPv1 routing protocol has no concept of VLSM. RIPv1 and RIPv1-compatible updates cannot be sent on a /31 subnet because there is no subnet broadcast address in such a subnet, which is the destination address AlliedWare Plus uses for RIPv1 updates.

  RIPv2 updates can be sent on a /31 subnet because they use a multicast destination address. RIPv1 directed broadcast updates will not be sent on a /31 subnet and instead a log will be printed each time the update fails to be sent, to alert the administrator that the configuration needs to be updated. It is still possible to send RIPv1 updates on a /31 subnet by disabling the RIPv1 directed broadcast updates using the **passive-interface (RIP)** command, and then using the **neighbor (RIP)** command to add the remote host as a RIPv1 unicast neighbor.
- As documented in the RFC 3021, directed broadcasts are not possible.
- Other broadcast protocols, which use destination IP 255.255.255.255 and/or broadcast MAC (e.g. ARP/DHCP) and routing protocols to unicast or multicast addresses, operate as expected.
- General IP routing and other functionality related to reachability of hosts within a /31 subnet operate as expected.

The format of the IP command itself has not changed

```
(no)ip address <ip-addr/prefix-length> [secondary] [label <label>]
```

Host numbering starts at zero, not 1. The range for each 'subnet' is as follows:

```
x.x.x.0/31 & x.x.x.1/31 = network 1
x.x.x.2/31 & x.x.x.3/31 = network 2
x.x.x.4/31 & x.x.x.5/31 = network 3
...and so on.
```

Address Resolution Protocol (ARP)

Address Resolution Protocol (ARP) is used by your device to dynamically learn the mapping between the Layer 2 addresses and IP addresses of devices in the networks to which it is connected. Most hosts also have a MAC physical address in addition to the assigned IP address. For Ethernet, this is a 6-byte, globally unique number. ARP enables your device to learn the physical address of the host that has a given IP address.

When your device needs to forward packets to a destination whose Layer 2 address it does not know, it broadcasts an ARP request to determine the Layer 2 destination address to put on the packet. The ARP request is a broadcast packet and includes the target IP address. All stations on the LAN receive this broadcast but only one host recognizes its own IP address. It replies, thereby giving your device its physical address.

Your device creates a dynamic ARP entry in its ARP cache, to record the IP address to physical address mapping (also called a binding). It uses that ARP entry to find that host's physical address when forwarding further packets to that address.

The ARP protocol is described in RFC 826, An Ethernet Address Resolution Protocol—or—Converting Network Protocol Addresses to 48 bit Ethernet Address for Transmission on Ethernet Hardware.

Static ARP entries

If your LAN includes hosts that do not support ARP, you can add a static ARP entry to the cache. However, it is rarely necessary to add an ARP entry this way. To add a static ARP entry, use the command:

```
awplus(config)# arp <ip-addr> <mac-address> [<port-number>] [alias]
```

or, if you have VRF-lite enabled:

```
awplus(config)# arp [vrf <vrf-name>] <ip-addr> <mac-address> [<port-
number>] [alias]
```

Timing out ARP entries

Your device times out dynamic ARP entries to ensure that the cache does not fill with entries for hosts that are no longer active. If your device stops receiving traffic for a device specified in a dynamic ARP entry, it deletes the ARP entry after a configurable timeout period. Static ARP entries are not aged or automatically deleted.

Increasing the ARP timeout reduces the amount of network traffic. Decreasing the timeout makes your device more responsive to changes in network topology.

To set a timeout period, enter the interface mode, then use the command:

```
awplus(config-if)# arp-aging-timeout <0-432000>
```

Deleting ARP entries

To remove a static ARP entry, use the command:

```
awplus(config)# no arp <ip-addr>
```

or, if you have enabled VRF-lite:

```
awplus(config)# no arp [vrf <vrf-name>] <ip-addr>
```

To clear the ARP cache of dynamic entries, use the command:

```
awplus# clear arp-cache
```

This removes the dynamic ARP entries for all interfaces.

To display the entries in the ARP cache, use the command:

```
awplus# show arp
```

The ARP cache will be repopulated by the normal ARP learning mechanism. As long as the entries are relearned quickly enough, deleting dynamic ARP entries does not affect:

- routes
- OSPF neighbor status
- BGP peer status
- TCP/UDP connection status
- VRRP status

Proxy ARP

Proxy ARP (defined in RFC 1027) deals with the situation where hosts in one subnet are sending ARP requests for IP addresses that are in a different subnet. Typically, this happens when the subnet mask configured on the requesting hosts does not match the subnet mask that has actually been allocated to their subnet.

Your device intercepts these ARP broadcast packets that are requesting IP addresses that are outside the local subnet, and substitutes its own physical address for that of the remote host. This occurs only if your device has the best route to the remote host.

By responding to the ARP request, your device is effectively saying to the requesting host '***send that traffic to me, and I will ensure it gets to that requested destination***'. So, that subsequent packets from the local host, destined for the IP address outside the local subnet, are directed to your device's physical address, and it can then forward these to the remote host. The process is symmetrical.

Proxy ARP is disabled by default. To enable proxy ARP on an interface, use the commands:

```
awplus# interface <interface>
awplus(config-if)# ip proxy-arp
```

To disable Proxy ARP on an interface, use the command:

```
awplus(config-if)# no ip proxy-arp
```

To check Proxy ARP is enabled on an interface, use the **show running-config** command. If Proxy ARP has been enabled an entry shows **ip proxy-arp** below the interface it is enabled on. No **ip proxy-arp** entry below an interface in the config indicates Proxy ARP is disabled on that interface.

See the sample configuration commands and validation command with resulting output showing proxy ARP **enabled** on VLAN 2 below:

```
awplus#configure terminal
awplus(config)#interface vlan2
awplus(config-if)#ip proxy-arp
awplus(config-if)#end
awplus(config)#exit
awplus#show running-config
!
interface vlan2
 ip proxy-arp
 ip address 192.168.2.2/24
!
```

See the sample configuration commands and validation command with resulting output showing proxy ARP **disabled** on VLAN 2 below:

```
awplus#configure terminal
awplus(config)#interface vlan2
awplus(config-if)#no ip proxy-arp
awplus(config-if)#end
awplus(config)#exit
awplus#show running-config
!
interface vlan2
 ip address 192.168.2.2/24
!
```

Local Proxy ARP

Local Proxy ARP lets you stop MAC address resolution between hosts within an interface’s subnet. This ensures that traffic between hosts in an environment where hosts are isolated from each other (e.g. a Private VLAN) is directed through one forwarding point. This lets you monitor, filter, and control traffic between devices in the same subnet.

Local Proxy ARP extends proxy ARP by intercepting and responding to ARP requests between hosts within a subnet. Local proxy ARP responds to ARP requests with your device’s own MAC address details instead of those from the destination host. This stops hosts from learning the MAC address of other hosts within its subnet.

When Local Proxy ARP is operating on an interface, your device does not generate or forward any ICMP-Redirect messages on that interface.

Local Proxy ARP is disabled by default. To enable local proxy ARP on an interface, use the commands:

```
awplus# interface <interface>
awplus(config-if)# ip local-proxy-arp
```

To disable local proxy ARP on an interface, use the command:

```
awplus(config-if)# no ip local-proxy-arp
```

To check Local Proxy ARP is enabled on an interface, use the **show running-config** command. If Local Proxy ARP has been enabled an entry shows **ip local-proxy-arp** below the interface it is enabled on. If there is no **ip local-proxy-arp** entry below an interface in the config, that indicates Local Proxy ARP is disabled on it.

See the sample configuration commands and validation command with resulting output showing local proxy ARP **enabled** on VLAN 1 below:

```
awplus#configure terminal
awplus(config)#interface vlan1
awplus(config-if)#ip local-proxy-arp
awplus(config-if)#end
awplus(config)#exit
awplus#show running-config
!
interface vlan1
 ip local-proxy-arp
 ip address 192.168.1.2/24
!
```

See the sample configuration commands and validation command with resulting output showing Local Proxy ARP **disabled** on VLAN 1 below:

```
awplus#configure terminal
awplus(config)#interface vlan1
awplus(config-if)#no ip local-proxy-arp
awplus(config-if)#end
awplus(config)#exit
awplus#show running-config
!
interface vlan1
 ip address 192.168.1.2/24
!
```

ARP logging

You can enable your device to log events that happen in the ARP cache, like the adding and deleting of static and dynamic ARP entries, and you can select either default hexadecimal notation (HHHH.HHHH.HHHH) or standard IEEE format hexadecimal notation (HH-HH-HH-HH-HH-HH) for the MAC addresses displayed in the ARP log output.

If this feature is enabled, ARP log messages are stored on the device in RAM. If the device is rebooted the ARP log messages are lost. ARP logging is disabled by default.

To enable ARP logging, use the command:

```
awplus(config)# arp log [mac-address-format ieee]
```

You can specify whether the MAC address is displayed in the default hexadecimal notation HHHH.HHHH.HHHH or in the standard IEEE format HH-HH-HH-HH-HH-HH.

To disable ARP logging, use the command:

```
awplus(config)# no arp log [mac-address-format ieee]
```

To display the ARP log messages, use the command:

```
awplus(config)# show log | include ARP_LOG
```

See the sample ARP log output and descriptions of the fields displayed in the sample ARP log output in the **arp log** command.

Domain Name System (DNS)

The **Domain Name System** allows you to access remote systems by entering human-readable device host names rather than IP addresses. DNS works by creating a mapping between a domain name, such as “www.alliedtelesis.com”, and its IP address. These mappings are held on DNS servers. DNS translates meaningful domain names into IP addresses for networking equipment to locate and address these devices.

For information about DNS on AlliedWare Plus switches, see Domain Name System (DNS) for AlliedWare Plus Switches.

The **Dynamic Domain Name System** (DDNS) is a mechanism which allows a DDNS client to automatically update a DNS entry hosted by a DDNS Provider. When DDNS is configured on an AR-Series Firewall, DNS updates are automatically directed to the configured host name regardless of Dynamic IP address changes. This feature is available on all AR-Series Firewalls from release 5.4.7-0.1 onwards.

For information about DNS and DDNS on AlliedWare Plus AR-Series Firewalls, see Domain Name System (DNS) for AlliedWare Plus AR-Series Firewalls.

Internet Control Message Protocol (ICMP)

The Internet Control Message Protocol (ICMP) allows networking devices to send information and control messages to other devices or hosts. Your device implements all non-obsolete ICMP functions.

The following table lists the ICMP messages implemented by your device.

Table 1: ICMP messages

ICMP MESSAGE TYPE	DEVICE RESPONSE
Echo reply (0)	This is used to implement the ping command. Your device sends out an echo reply in response to an echo request.
Destination unreachable (3)	This message is sent when your device drops a packet because it did not have a route to the destination.
Redirect (5)	Your device issues this message to inform a local host that its target is located on the same LAN (no routing is required) or when it detects a host using a non-optimal route (usually because a link has failed or changed its status). For example, if your device receives a packet destined to its own MAC address, but with a destination IP address of another host in the local subnet, it returns an ICMP redirect to the originating host. ICMP redirects are disabled on interfaces on which local proxy ARP is enabled.
Echo request (8)	This is related to echo replies. If your device receives an echo request, it sends an echo reply. If you enter the ping command, your device generates echo requests.
Router Advertisements (10)	These are Router Discovery Protocol messages. If Router Discovery is enabled, your device sends these to announce the IP addresses of the sending interface.
Time to Live Exceeded (11)	If the TTL field in a packet falls to zero, your device sends this message.This occurs when there are too many hops in the path that a packet is traversing.

ICMP messages are enabled on all interfaces by default. You can control the flow of ICMP messages across different interfaces using the **access-list** commands. See the following sections in your product’s Command Reference, available on alliedtelesis.com:

- IPv4 Hardware Access Control List (ACL) Commands
- IPv4 Software Access Control List (ACL) Commands

ICMP Router Discovery Protocol (IRDP)

Router Discovery

Some AlliedWare Plus devices support the router specification sections of IRDP (RFC 1256, ICMP Router Discovery Messages). If this feature is configured, your device sends router advertisements periodically and in response to router solicitations. It does not support the Host Specification section of this RFC.

Benefits

Before an IP host can send an IP packet, the host has to know the IP address of a neighboring router that can forward the packet to its destination. ICMP Router Discovery messages let routers automatically advertise themselves to hosts. Other methods either require someone to manually keep these addresses current, or require DHCP to send router addresses.

Router Discovery process

The following table summarizes what happens when Router Discovery advertisements are enabled on an interface.

Table 2: Router Discovery advertisements

WHEN...	THEN...
Router Discovery advertising starts on an interface because: ■ your device starts up, or ■ you enable advertisements on your device or on an interface.	Your device multicasts a router advertisement and continues to multicast them periodically until router advertising is disabled.
A host starts up.	The host may send a router solicitation message.
Your device receives a router solicitation.	Your device multicasts an early router advertisement from the interface on which it received the router solicitation.
A host receives a router advertisement.	The host stores the IP address and preference level for the advertisement lifetime.
The lifetimes of all existing router advertisements on a host expire.	The host sends a router solicitation.
A host does not receive a router advertisement after sending a small number of router solicitations.	The host waits for the next unsolicited router advertisement.
A host needs a default router address.	The host uses the IP address of the router or L3 switch with the highest preference level.
Router Discovery advertising is deleted from the interface.	Your device multicasts a router advertisement with the IP address(es) that stopped advertising, and a lifetime of zero. It continues to periodically multicast router advertisements for other interfaces, if configured to.
The router receives a router advertisement from another router.	The router does nothing but silently discards the message.

Advertisement messages

A router advertisement is an ICMP (type 10) message that contains the following:

- in the destination address field of the IP header, the interface's configured advertisement address, either 224.0.0.1 or 255.255.255.255.
- in the lifetime field, the interface's configured advertisement lifetime.
- in the Router Address and Preference Level fields, the addresses and preference levels of all the logical interfaces that are set to advertise.

Your device does not send router advertisements by default.

Solicitation message

A router solicitation is an ICMP (type 10) message containing:

- **source address**: an IP address belonging to the interface from which the message is sent
- **destination address**: the configured Solicitation Address, and
- **Time-to-Live**: 1 if the Destination Address is an IP multicast address; at least 1 otherwise.

Advertisement interval

The router advertisement interval is the time between router advertisements. For the first few advertisements sent from an interface (up to 3), your device sends the router advertisements at intervals of at most 16 seconds. After these initial transmissions, it sends router advertisements at random intervals between the minimum and maximum intervals that the user configures, to reduce the probability of synchronization with the advertisements from other routers on the same link. By default, the minimum is 450 seconds (7.5 minutes), and the maximum is 600 seconds (10 minutes).

Preference level

The preference level is the preference of the advertised address as a default router address relative to other router addresses on the same subnet. By default, all routers and Layer 3 switches have the same preference level, zero. While it is entered as a decimal from 0 to 2147483647, it is encoded in router advertisements as a two’s-complement hex integer from 0x8000000 to 0x7fffffff. A higher preference level is preferred over a lower value.

Lifetime

The lifetime of a router advertisement is how long the information in the advertisement is valid. By default, the lifetime of all advertisements is 1800 seconds (30 minutes).

Address type

Your device can send its router advertisements using either a broadcast or multicast destination address. By default, your device sends router advertisements using the all-systems multicast address (224.0.0.1). However, on networks where the hosts do not support IP multicast you must use the broadcast address (255.255.255.255). To change the address type to broadcast on an interface, use the command:

```
awplus(config-if)# ip irdp broadcast
```

To change the address type back to multicast, use the **no** variant of the above command, or use the command:

```
awplus(config-if)# ip irdp multicast
```

Configuration procedure

Perform the following to configure your device to send router advertisements:

Step 1: Enter the interface to advertise.

Enter the configuration mode for the interface, using the command:

```
awplus(config)# interface <interface>
```

Step 2: Change the address type.

By default, your device sends router advertisements using a multicast destination address. If hosts on your network do not support this, change the address type to broadcast, using the command:

```
awplus(config-if)# ip irdp broadcast
```

Step 3: Configure the advertisement interval and lifetime.

By default, your device sends router advertisements every 7.5 to 10 minutes, with a lifetime of 30 minutes. These settings are likely to work well in most situations, and will not cause a large amount of extra traffic, even if there are several routers on the LAN. If you change these settings, keep the following proportions:

```
lifetime=3 x maxadvertisementinterval
minadvertisementinverval=0.75 x maxadvertisementinterval
```

You cannot set the maximum advertisement interval below the minimum interval. If you are lowering the maximum interval to a value below the current minimum interval, you must change the minimum value first. This also applies to changing the minimum interval above the current maximum interval.

To change the maximum advertisement interval, use the command:

```
awplus(config-if)# ip irdp maxadvertinterval <4-1800>
```

To change the minimum advertisement interval, use the command:

```
awplus(config-if)# ip irdp minadvertinterval <3-1800>
```

To change the lifetime for your device’s router advertisements, use the command:

```
awplus(config-if)# ip irdp lifetime <0-9000>
```

Step 4: Set preference levels.

By default, every interface has the same preference for becoming a default router. To give the interface a higher preference, increase the preference level. To give it a lower preference, decrease this value.

To set the preference level for all addresses on this interface, use the command:

```
awplus(config-if)# ip irdp preference <0-2147483647>
```

To set the preference for a specific address on the interface, use the command:

```
awplus(config-if)# ip irdp address <ip-address> preference
<0-2147483647>
```

Step 5: Enable advertising on the interface.

To enable router advertisements on an interface, enter the interface mode and use the command:

```
awplus(config-if)# ip irdp
```

Step 6: Enable advertising on your device.

To globally enable router advertisements on your device, enter global configuration mode and use the command:

```
awplus(config-if)# exit
awplus(config)# router ip irdp
```

Step 7: Check advertise settings.

To view the IRDP configuration on the interface, use the command:

```
awplus(config)# exit
awplus# show ip irdp interface [<interface-name>]
```

To view the global IRDP configuration for your device, use the command:

```
awplus# show ip irdp
```

Debugging IRDP

Information which may be useful for troubleshooting IRDP is available using the IRDP debugging function. To enable IRDP debugging, use the command:

```
awplus# debug ip irdp {event|nsm|receive|send|both|detail|all}
```

Checking IP Connections

To verify connections between networks and network devices, use the ping (Packet Internet Groper) and trace route functions on your device.

Ping

Ping tests the connectivity between two network devices to determine whether each network device can ‘see’ the other device. Echo request packets are sent to the destination addresses and responses are displayed on the console.

If you can ping the end destination, then the physical, Layer 2 and Layer 3 links are functioning, and any difficulties are in the network or higher layers.

If pinging the end destination fails, use traceroute to discover the point of failure in the route to the destination.

To ping a device, use the command:

```
awplus# ping {<hostname>|<ipaddr>}
```

Traceroute

You can use traceroute to discover the route that packets traverse between two systems running the IP protocol. Traceroute sends an initial UDP packet with the Time To Live (TTL) field in the IP header set to a starting value of 1. The TTL field is increased by one for every subsequent packet sent until the destination is reached. Each hop along the path between two systems responds with a TTL exceeded packet (ICMP type 11) and from this the path is determined.

To use traceroute, use the command:

```
awplus# traceroute {<hostname>|<ipaddr>}
```

or, if you have enabled VRF-lite:

```
awplus# traceroute [vrf <vrf-name>] {<hostname>|<ipaddr>}
```

Enter either the hostname or the IP address of the device you are trying to reach.

IP Helper (UDP Broadcast Helper)

On switches that support it, the IP Helper feature allows the switch to receive UDP broadcasts on one subnet, and forward them as broadcasts or unicasts into another subnet, so a client can use an application which uses UDP broadcast (such as Net-BIOS) when the client and server are located in different subnets. The IP Helper feature forwards UDP broadcast network traffic to specific hosts on another subnet and/or to the broadcast address of another subnet.

When the IP Helper feature is enabled on a VLAN interface, the UDP broadcast packets received on the interface are processed for forwarding out through another interface into another subnet. Depending on the nature of the ip-helper addresses configured, the UDP broadcasts will be unicast forwarded to a single host in the destination subnet, or unicast forwarded to multiple hosts in the destination subnet, or broadcast to the broadcast address of the destination subnet. Not all UDP broadcasts will be forwarded when IP Helper is configured. The set of broadcasts to be forwarded can be defined by specifying the destination UDP port(s) of the packets you wish to forward.

The command to enable the forwarding of UDP broadcasts received on a given interface is **ip helper-address** (entered in interface configuration mode). The **ip forward-protocol udp** command specifies types of broadcast packets to forward.

Multiple different destination addresses can be specified by using multiple instances of the **ip helper-address** command under the same interface. If a destination address is specified that is actually the broadcast address of one of the subnets directly connected to the switch, then the UDP packets will be forwarded as broadcasts onto that subnet.

Likewise, multiple different types of UDP packet can be specified for forwarding by specifying multiple different destination ports using the **ip forward-protocol udp** command.

Note: The types of UDP broadcast packets that the switch will forward are only those specified by the **ip forward-protocol** command(s). The IP Helper process does not forward any other UDP packet types by default.

IP Directed Broadcast

IP directed-broadcast is enabled and disabled per VLAN interface. When enabled, a directed broadcast packet is forwarded to an enabled VLAN interface if received on another subnet.

An IP directed broadcast is an IP packet whose destination address is a broadcast address for some IP subnet, but originates from a node that is not itself part of that destination subnet. When a directed broadcast packet reaches a switch that is directly connected to its destination subnet, and IP directed-broadcast is enabled on the interface via which the switch connects to that destination subnet, the packet is flooded as a broadcast on the destination subnet.

The **ip directed-broadcast** command only controls the flooding of directed broadcasts when they reach target subnets. The command affects the final transmission of the directed broadcast on its destination subnet. It does not affect the transit unicast routing of IP directed broadcasts.

If the **no ip directed-broadcast** command is configured for an interface, directed broadcasts destined for the subnet where the interface is attached will be dropped instead of broadcast.

Support for Network Load Balancing (NLB)

Introduction

NLB is one of the clustering technologies available from Microsoft®. It provides high availability for services such as HTTP and FTP, by grouping identical servers into a cluster and sharing the network load between all currently-available servers in the cluster.

Network Load Balancing clustering modes

There are two modes for Network Load Balancing with Windows 2003®: unicast and multicast. Multicast mode has a further option of IGMP Multicasting. This section discusses how these modes work.

Note that clustering only works, in multicast or unicast mode, if all packets sent to the cluster's IP address are sent to all nodes in the cluster. This means that the cluster-side switch must flood traffic to all ports that are connected to the members of the cluster.

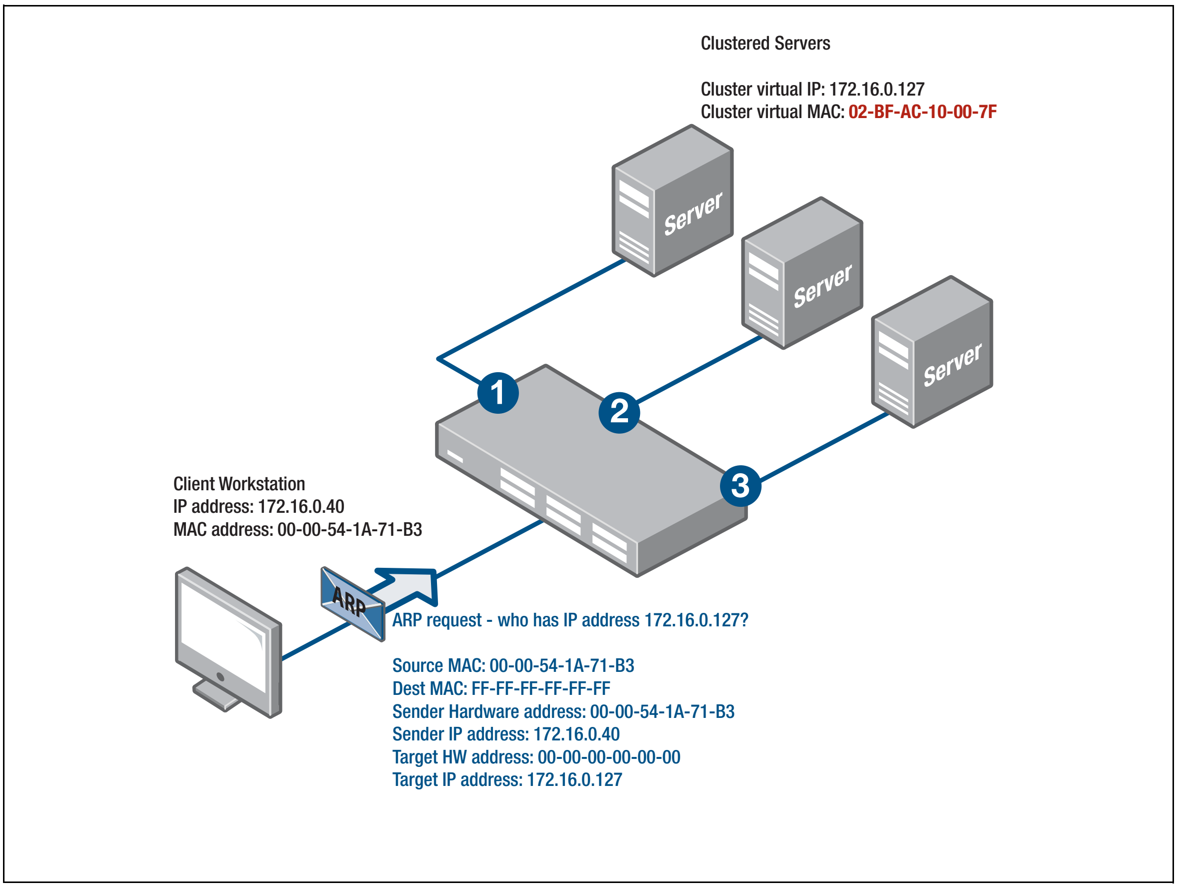
Unicast mode

In **unicast mode**, all hosts in the cluster share a single unicast 'cluster' MAC address, to go with the cluster IP address. This mode has the disadvantage that it stops cluster hosts from using their own 'burned-in' MAC addresses, so hosts cannot contact each other unless you install a second NIC card on each host and use that for intra-cluster communication.

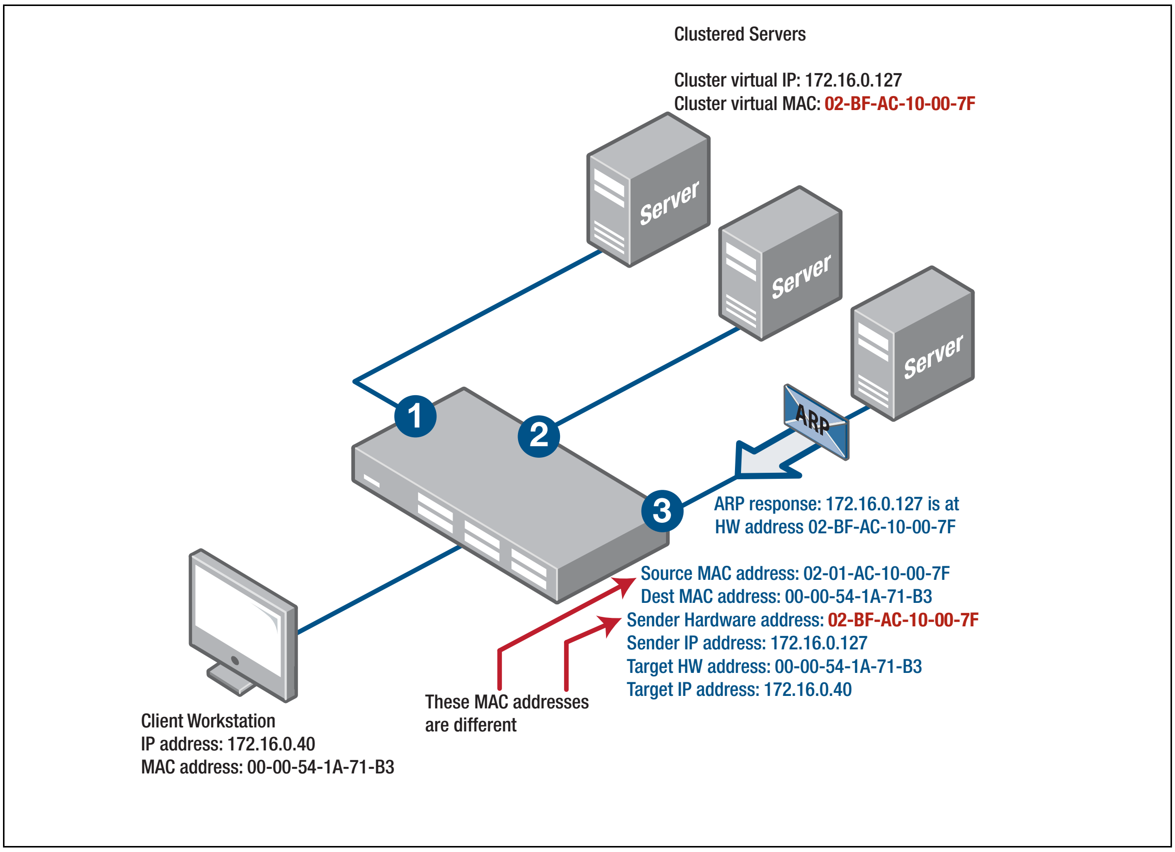
The cluster operation forces the switch to flood all packets that are destined for the cluster, by stopping the switch from learning the cluster MAC address. The switch cannot learn the cluster MAC address because it never appears in the source field of the Ethernet headers of packets sent from the clustered servers.

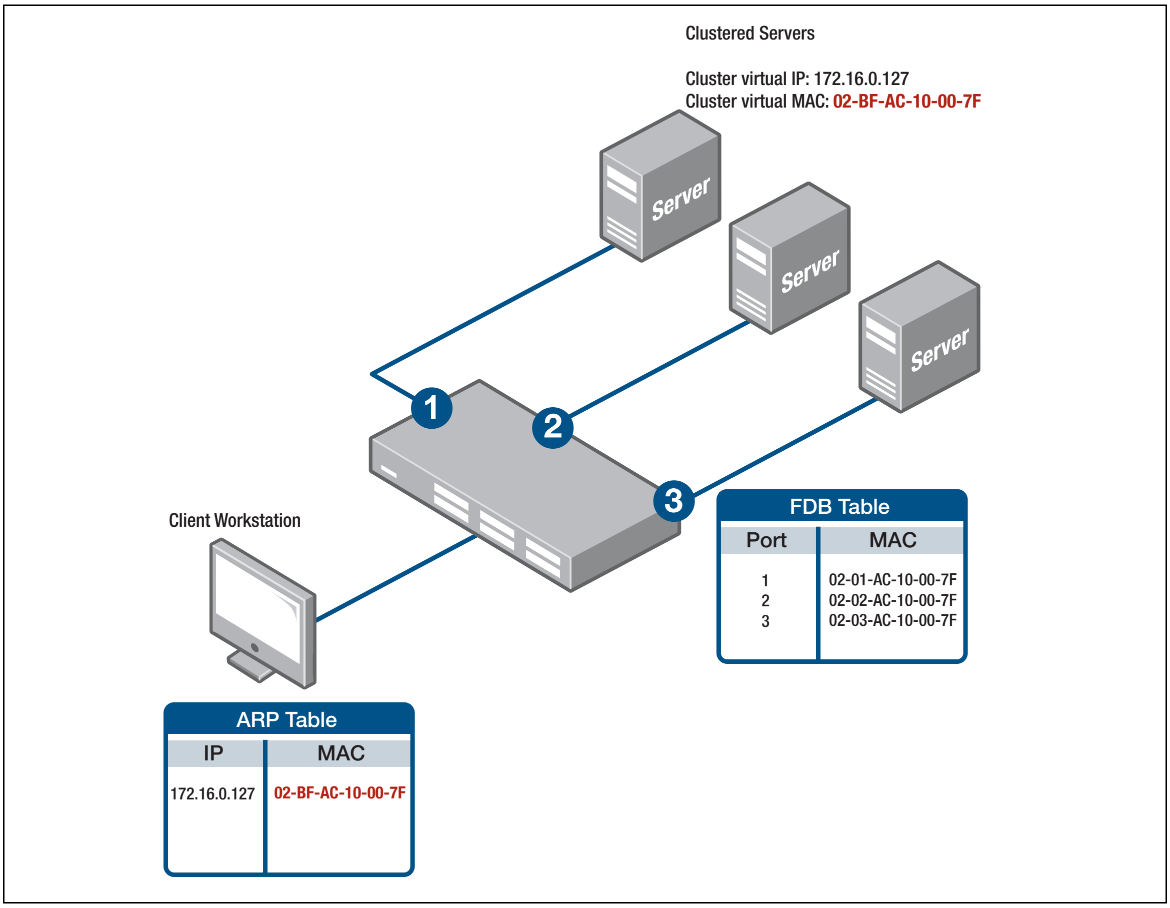
Instead, each NIC uses a special unicast MAC address in the source field of the Ethernet header. The cluster MAC address must never be found in the source field of the Ethernet header, because otherwise the switch will learn the cluster MAC address, and stop flooding packets to all ports, and only one node in the cluster will receive traffic for the cluster's IP.

The following diagrams illustrate unicast mode:


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
Server
Server
Server
1
2
3
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3
ARP
ARP request - who has IP address 172.16.0.127?

Source MAC: 00-00-54-1A-71-B3
Dest MAC: FF-FF-FF-FF-FF-FF
Sender Hardware address: 00-00-54-1A-71-B3
Sender IP address: 172.16.0.40
Target HW address: 00-00-00-00-00-00
Target IP address: 172.16.0.127


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
Server
Server
Server
ARP
1
2
3
ARP response: 172.16.0.127 is at
HW address 02-BF-AC-10-00-7F
Source MAC address: 02-01-AC-10-00-7F
Dest MAC address: 00-00-54-1A-71-B3
Sender Hardware address: 02-BF-AC-10-00-7F
Sender IP address: 172.16.0.127
Target HW address: 00-00-54-1A-71-B3
Target IP address: 172.16.0.40
These MAC addresses
are different
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

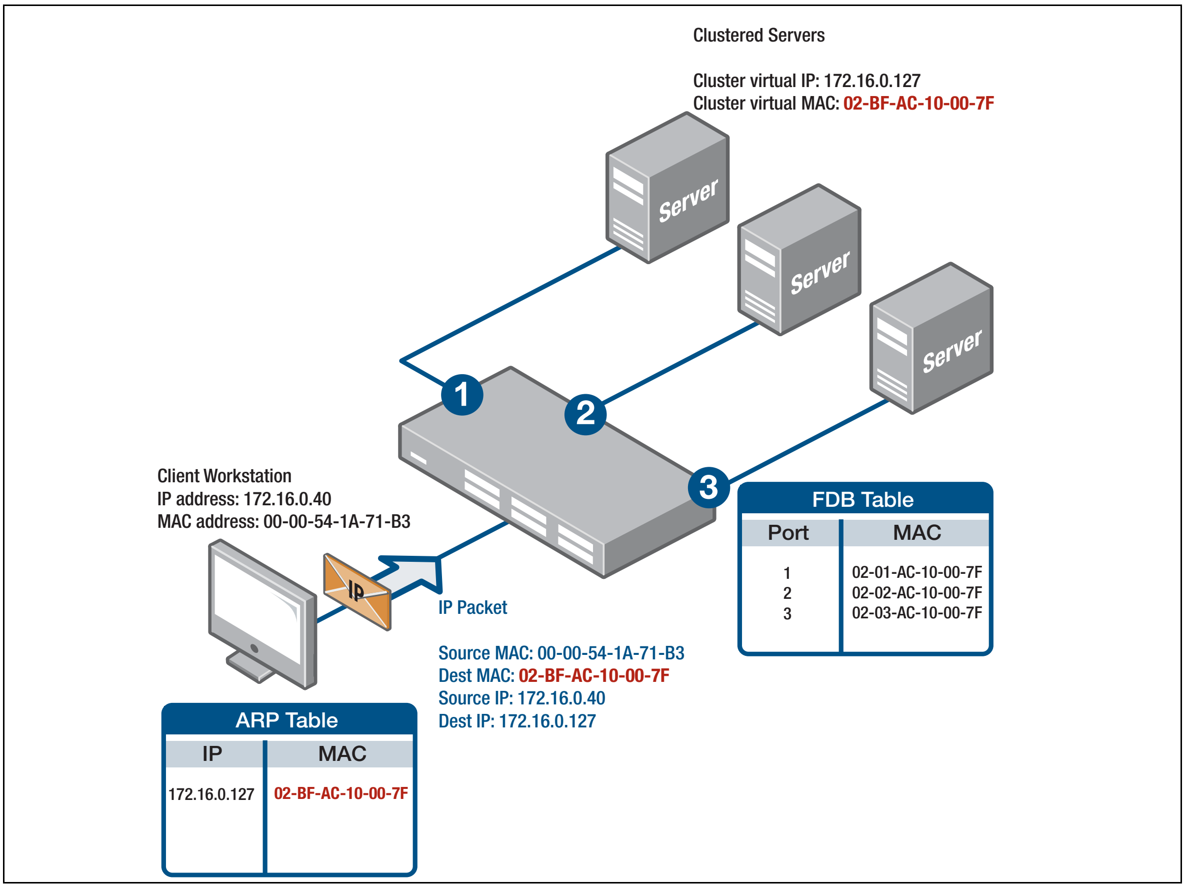

Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
Server
Server
Server
1
2
3
Client Workstation

FDB Table

Port	MAC
1	02-01-AC-10-00-7F
2	02-02-AC-10-00-7F
3	02-03-AC-10-00-7F

ARP Table

IP	MAC
172.16.0.127	02-BF-AC-10-00-7F


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
Server
Server
Server
1
2
3
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3
IP
IP Packet

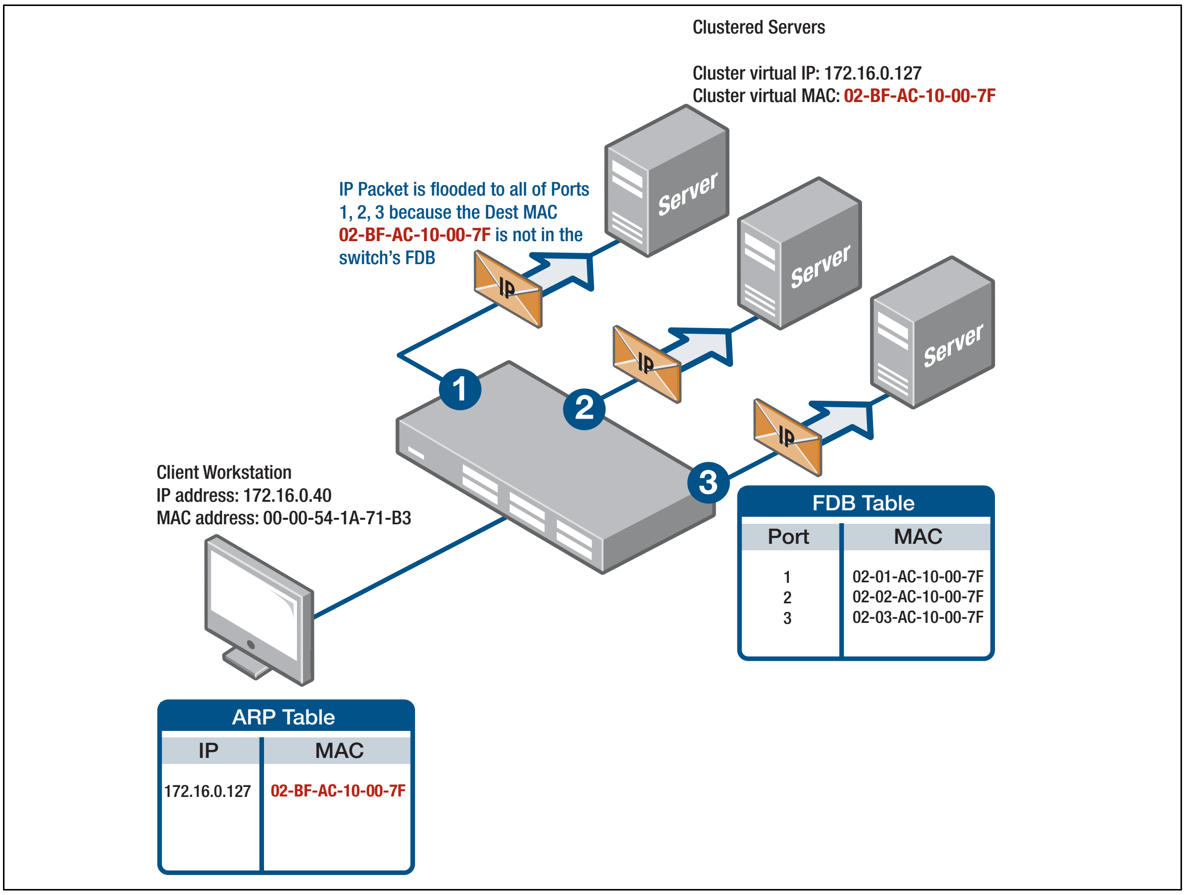
Source MAC: 00-00-54-1A-71-B3
Dest MAC: 02-BF-AC-10-00-7F
Source IP: 172.16.0.40
Dest IP: 172.16.0.127

FDB Table

Port	MAC
1	02-01-AC-10-00-7F
2	02-02-AC-10-00-7F
3	02-03-AC-10-00-7F

ARP Table

IP	MAC
172.16.0.127	02-BF-AC-10-00-7F


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
Server
Server
Server
IP
IP
IP
1
2
3
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

IP Packet is flooded to all of Ports 1, 2, 3 because the Dest MAC 02-BF-AC-10-00-7F is not in the switch's FDB

FDB Table

Port	MAC
1	02-01-AC-10-00-7F
2	02-02-AC-10-00-7F
3	02-03-AC-10-00-7F

ARP Table

IP	MAC
172.16.0.127	02-BF-AC-10-00-7F

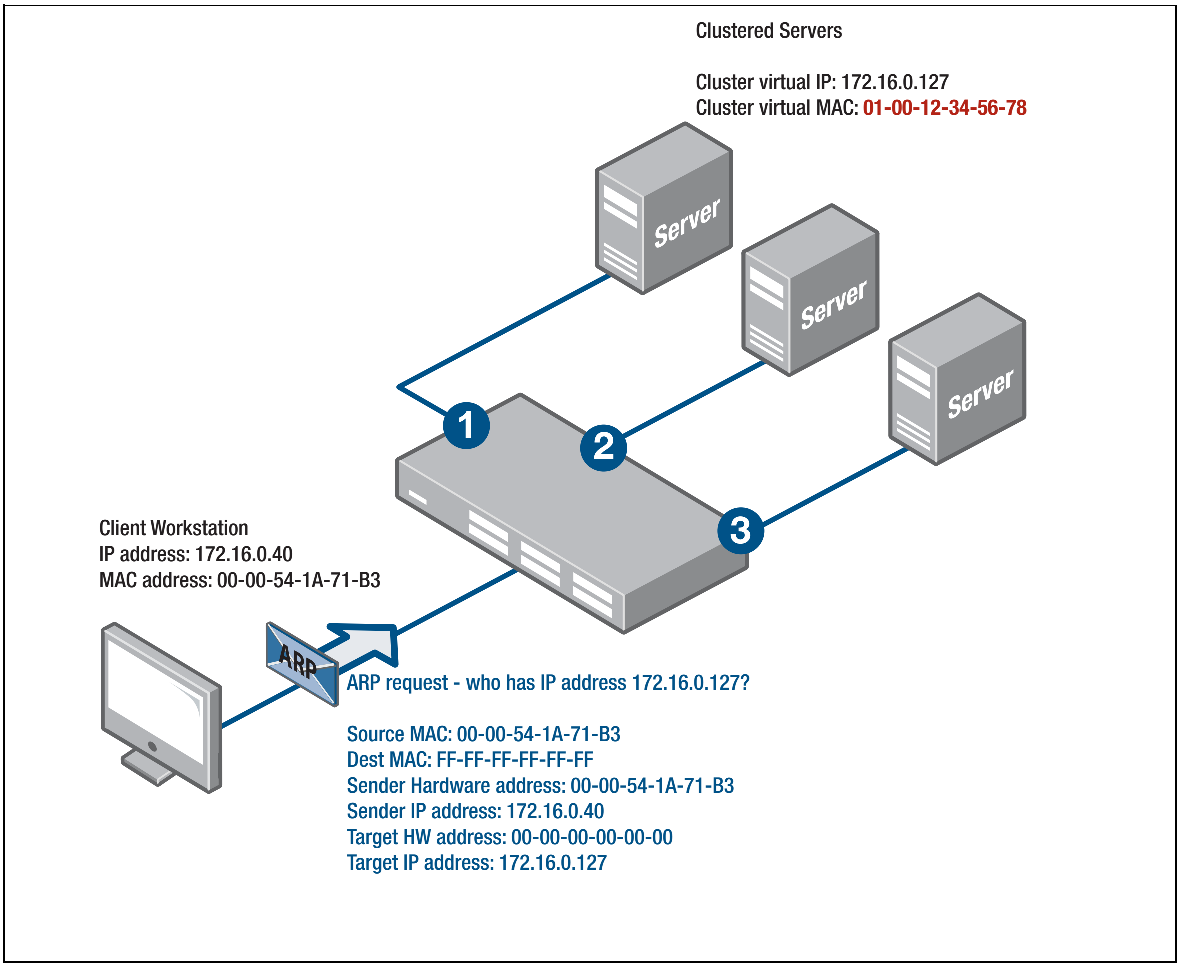
Multicast mode

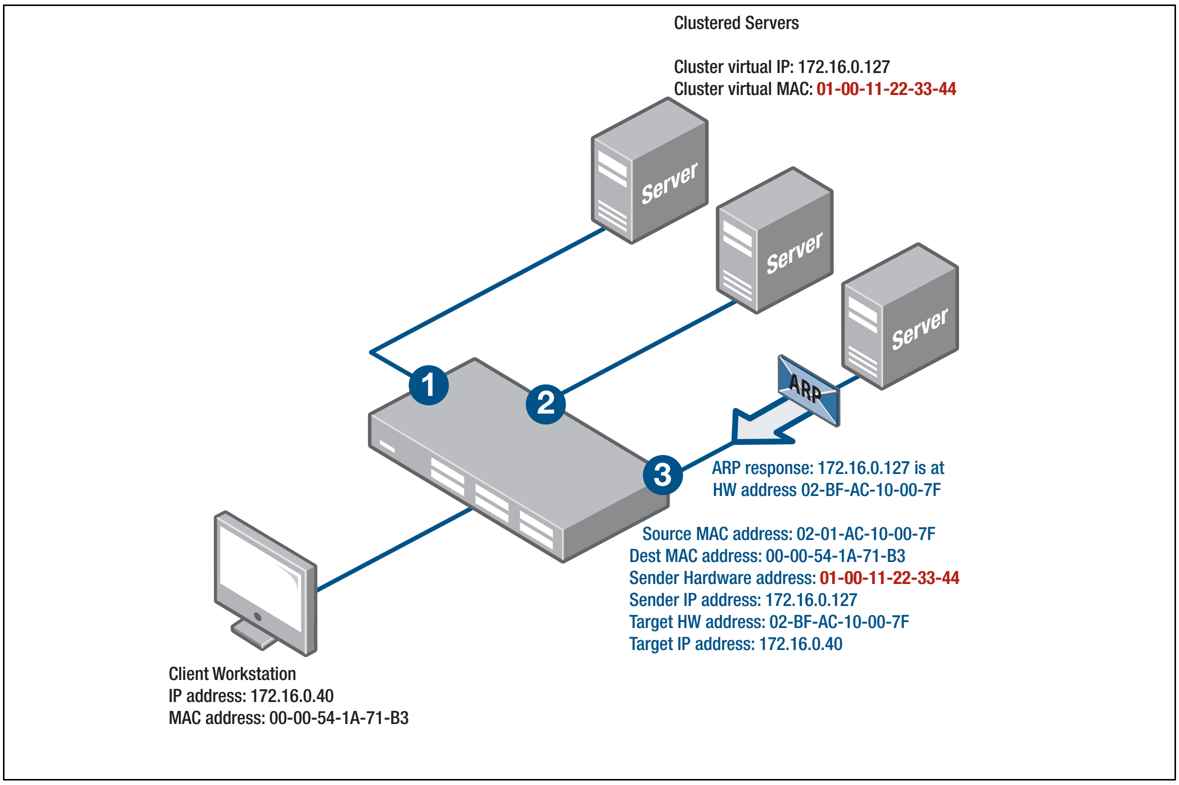
In **multicast mode**, hosts in the cluster use their real ‘burned-in’ MAC address in the source field of the Ethernet header. However, they answer ARP requests for the cluster IP address with a multicast MAC in the ARP packet's payload, while the Ethernet header on the ARP reply still has the real MAC address. This allows cluster hosts to contact each other.

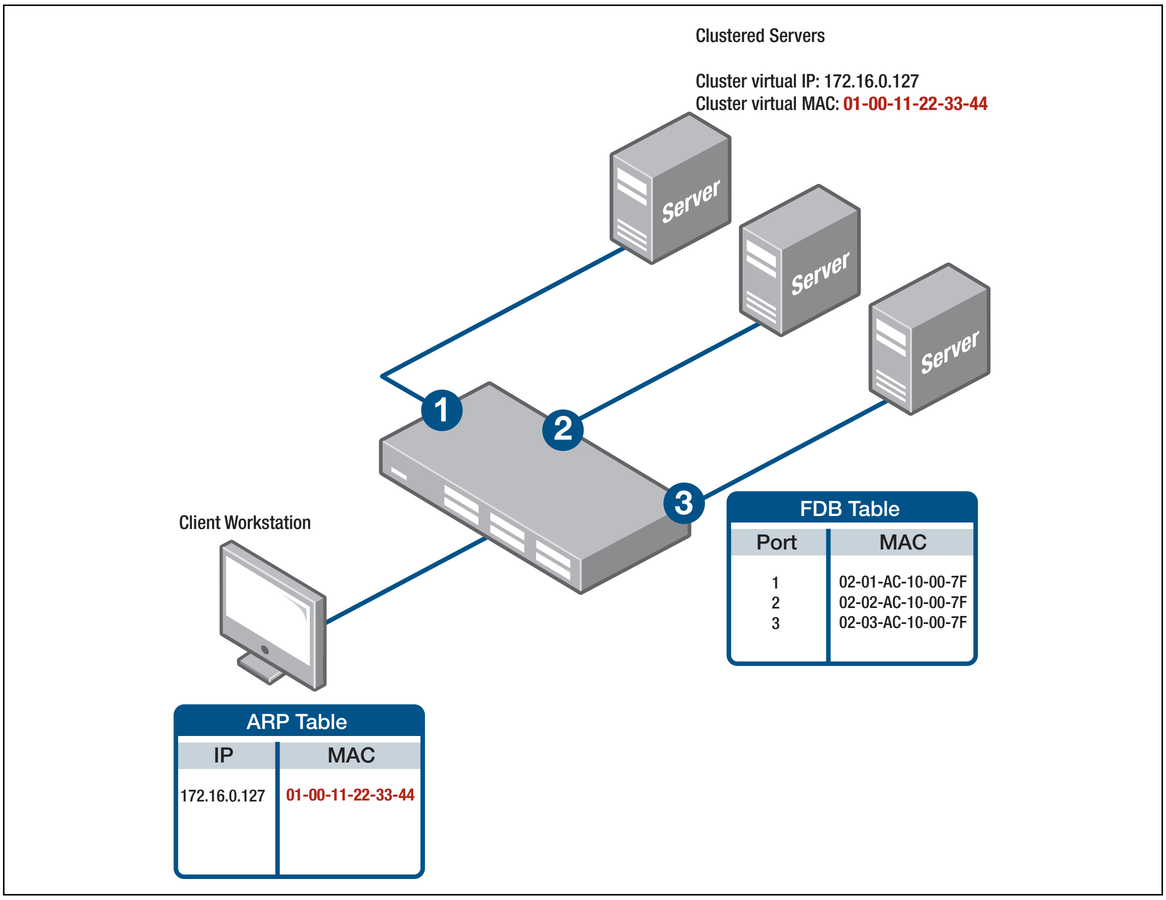
In both multicast and unicast mode the mechanism to cause flooding is as such:

- ARP requests for the cluster IP are answered with the virtual cluster MAC address in the ARP packet's payload.
- The node must send all packets, including ARP requests and replies, with a different source MAC in the Ethernet header.
- This prevents the switch from entering the virtual cluster MAC into the forwarding database (FDB).

The following diagrams illustrate multicast mode:


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-12-34-56-78
Server
Server
Server
1
2
3
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3
ARP
ARP request - who has IP address 172.16.0.127?
Source MAC: 00-00-54-1A-71-B3
Dest MAC: FF-FF-FF-FF-FF-FF
Sender Hardware address: 00-00-54-1A-71-B3
Sender IP address: 172.16.0.40
Target HW address: 00-00-00-00-00-00
Target IP address: 172.16.0.127


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-11-22-33-44
Server
Server
Server
ARP
1
2
3
ARP response: 172.16.0.127 is at
HW address 02-BF-AC-10-00-7F
Source MAC address: 02-01-AC-10-00-7F
Dest MAC address: 00-00-54-1A-71-B3
Sender Hardware address: 01-00-11-22-33-44
Sender IP address: 172.16.0.127
Target HW address: 02-BF-AC-10-00-7F
Target IP address: 172.16.0.40
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

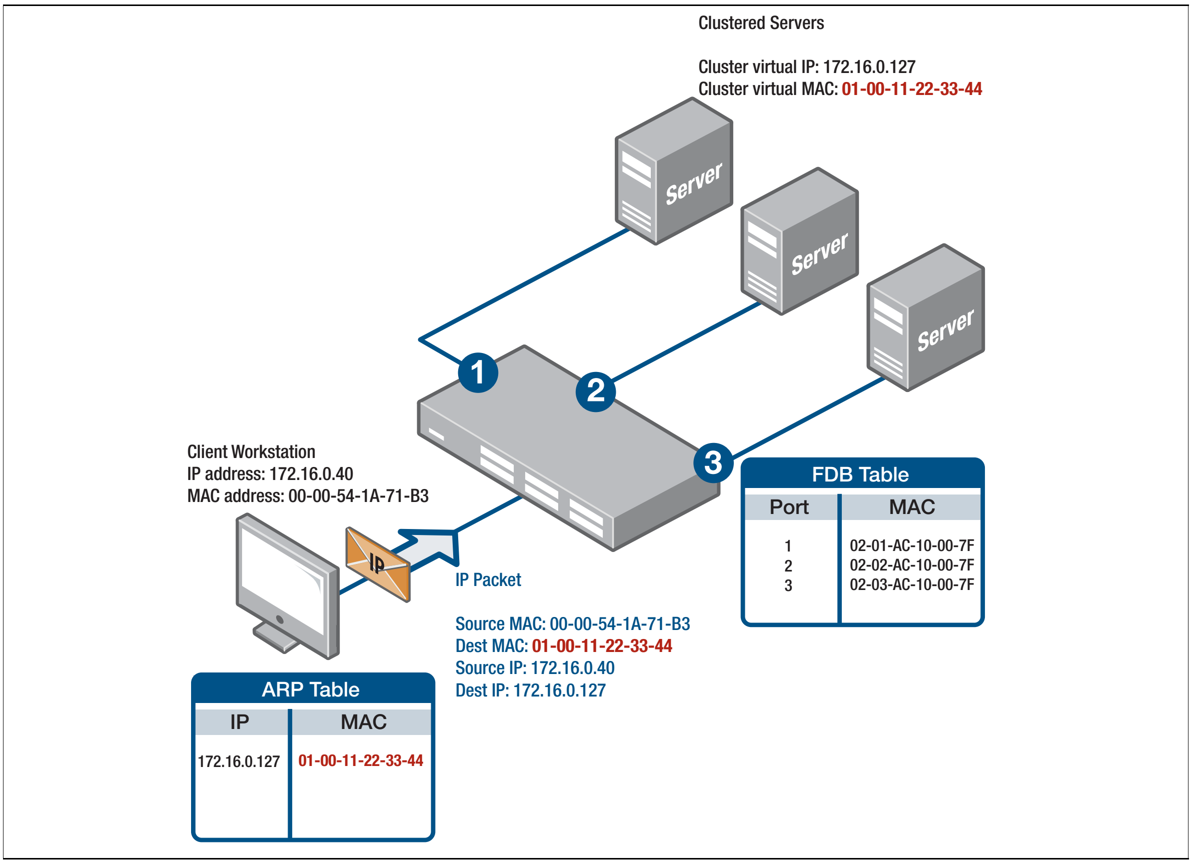

Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-11-22-33-44
Server
Server
Server
1
2
3
Client Workstation

FDB Table

Port	MAC
1	02-01-AC-10-00-7F
2	02-02-AC-10-00-7F
3	02-03-AC-10-00-7F

ARP Table

IP	MAC
172.16.0.127	01-00-11-22-33-44

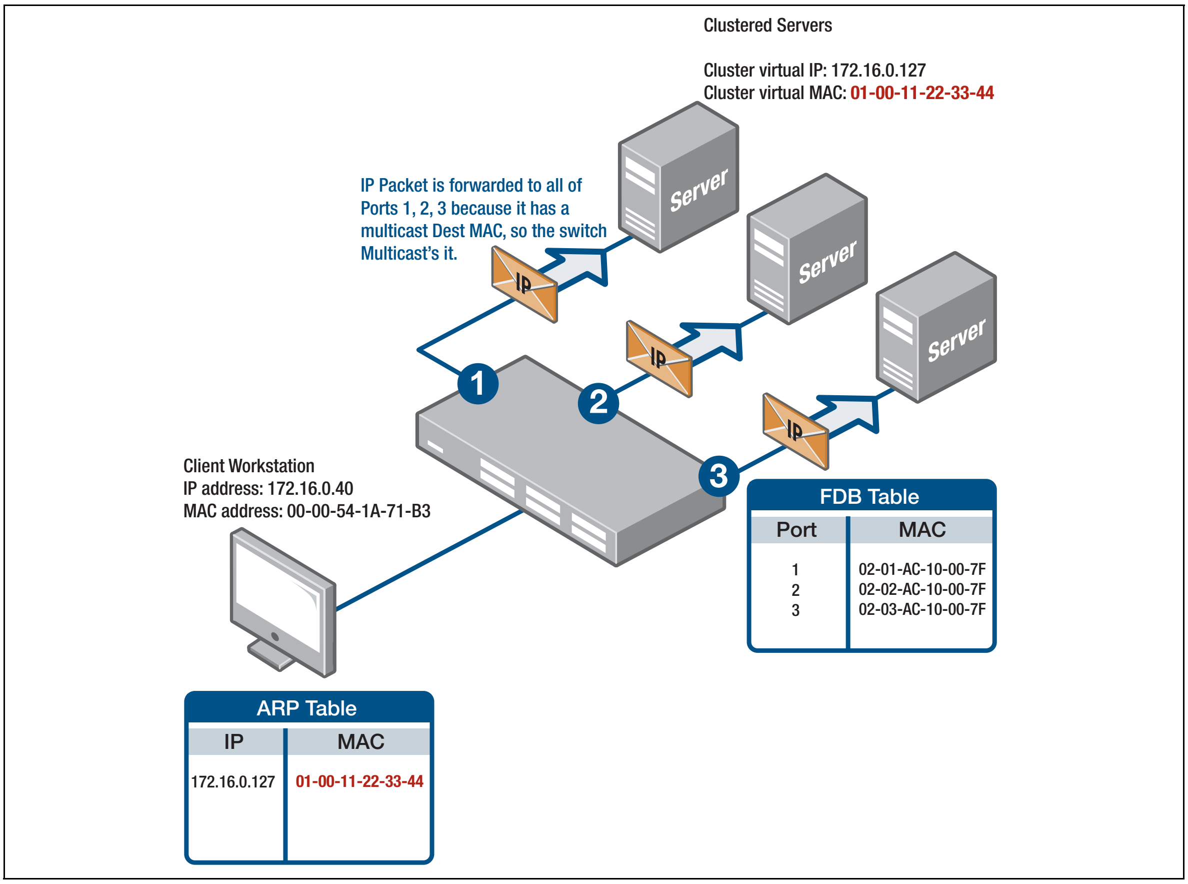

Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-11-22-33-44
Server
Server
Server
1
2
3
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3
IP
IP Packet
Source MAC: 00-00-54-1A-71-B3
Dest MAC: 01-00-11-22-33-44
Source IP: 172.16.0.40
Dest IP: 172.16.0.127

FDB Table

Port	MAC
1	02-01-AC-10-00-7F
2	02-02-AC-10-00-7F
3	02-03-AC-10-00-7F

ARP Table

IP	MAC
172.16.0.127	01-00-11-22-33-44


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-11-22-33-44
IP Packet is forwarded to all of Ports 1, 2, 3 because it has a multicast Dest MAC, so the switch Multicast's it.
Server
Server
Server
IP
IP
IP
1
2
3
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

FDB Table

Port	MAC
1	02-01-AC-10-00-7F
2	02-02-AC-10-00-7F
3	02-03-AC-10-00-7F

ARP Table

IP	MAC
172.16.0.127	01-00-11-22-33-44

Multicast mode with the IGMP option selected

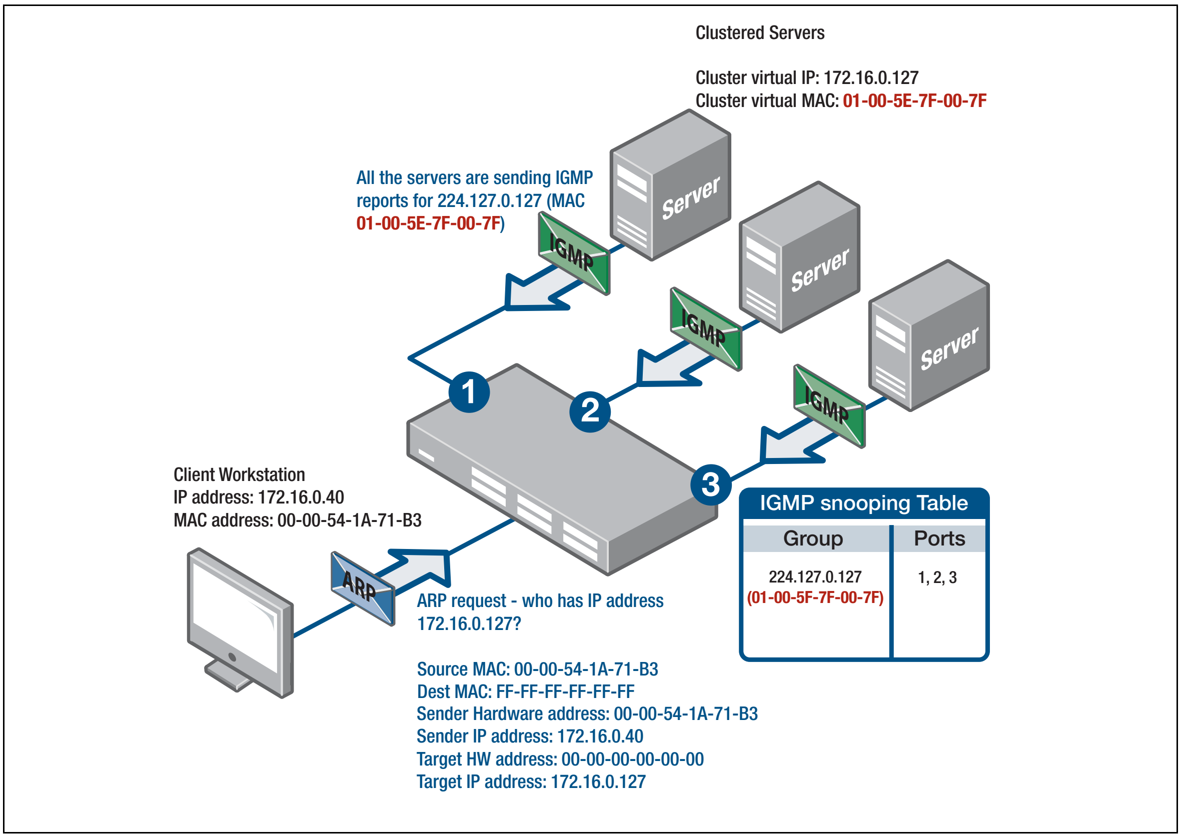
This mode uses IGMP (Internet Group Management Protocol) to prevent the switch from flooding all ports; instead traffic only goes to NLB ports.

This mode also causes the cluster to use its real ‘burned in’ MAC address in the source field of the Ethernet header. The cluster will answer ARP requests for the cluster IP address with a multicast MAC in the ARP packet's payload, while the Ethernet header on the ARP reply still has the real MAC address. Note though, that the MAC address is slightly different to the previously discussed multicast mode—the MAC address starts with the bytes 01:00:5e, which still identifies it as an IP-multicast MAC address:

The Windows client will happily add this IP / MAC combination to its internal ARP table (on a Windows or Linux PC, use the command **arp-a** to see all ARP entries). All IP traffic destined for 172.16.0.127 is then sent with an Ethernet destination of 01:00:5e:7f:00:7f. The cluster node responds with its burned in MAC in the Ethernet source field.

Microsoft's IGMP multicasting mode cleverly turns the concept of IGMP clients and hosts on its head. As far as an Allied Telesis switch is concerned, the clustered servers are seen as IGMP clients and the IGMP multicast data (stream) is sent by the workstations trying to access the cluster. To ensure that the switches do correctly forward the multicast data, the servers send IGMP reports for a group that corresponds to the MAC address they are putting in their ARP responses. Reports are sent frequently by the NLB servers so there is no concern about IGMP entries timing out on the switch.

The following diagrams illustrate multicast mode with the IGMP option selected:



Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-5E-7F-00-7F

All the servers are sending IGMP reports for 224.127.0.127 (MAC 01-00-5E-7F-00-7F)

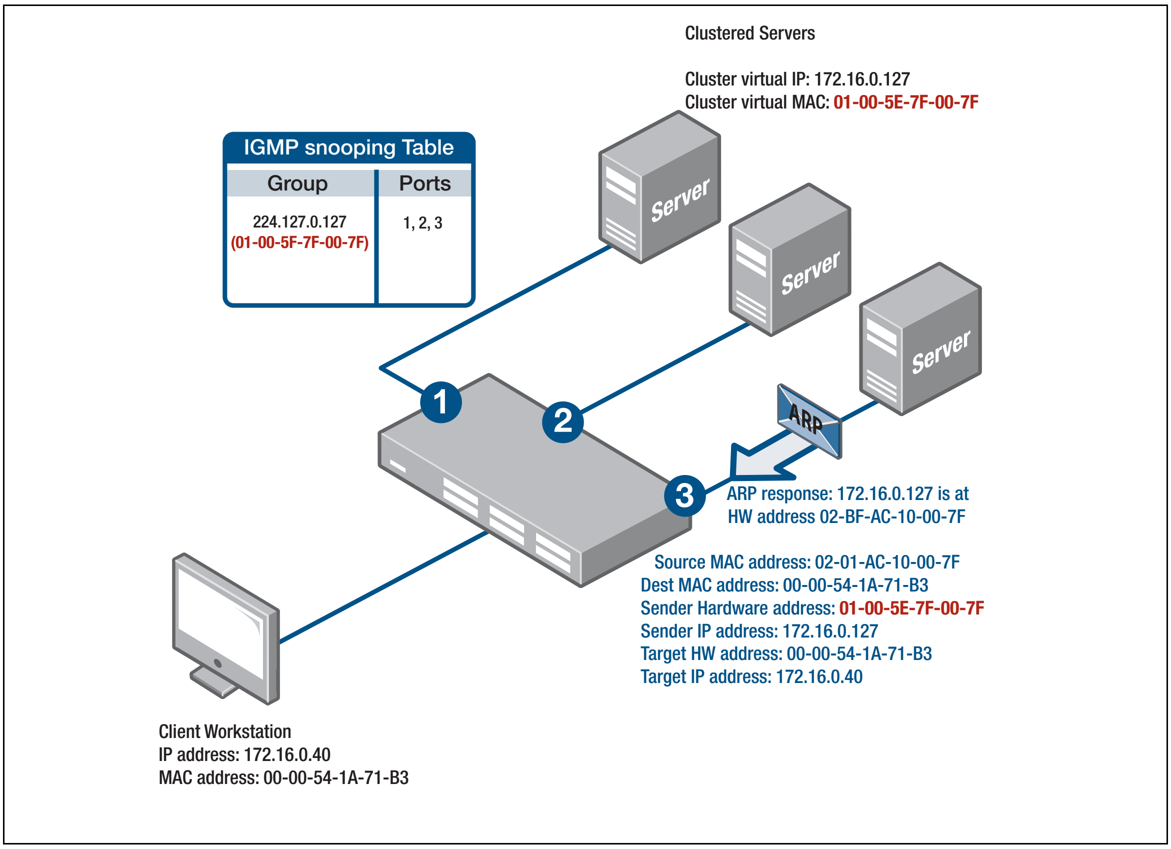
Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

IGMP snooping Table

Group	Ports
224.127.0.127 (01-00-5F-7F-00-7F)	1, 2, 3

ARP request - who has IP address 172.16.0.127?

Source MAC: 00-00-54-1A-71-B3
Dest MAC: FF-FF-FF-FF-FF-FF
Sender Hardware address: 00-00-54-1A-71-B3
Sender IP address: 172.16.0.40
Target HW address: 00-00-00-00-00-00
Target IP address: 172.16.0.127



Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-5E-7F-00-7F

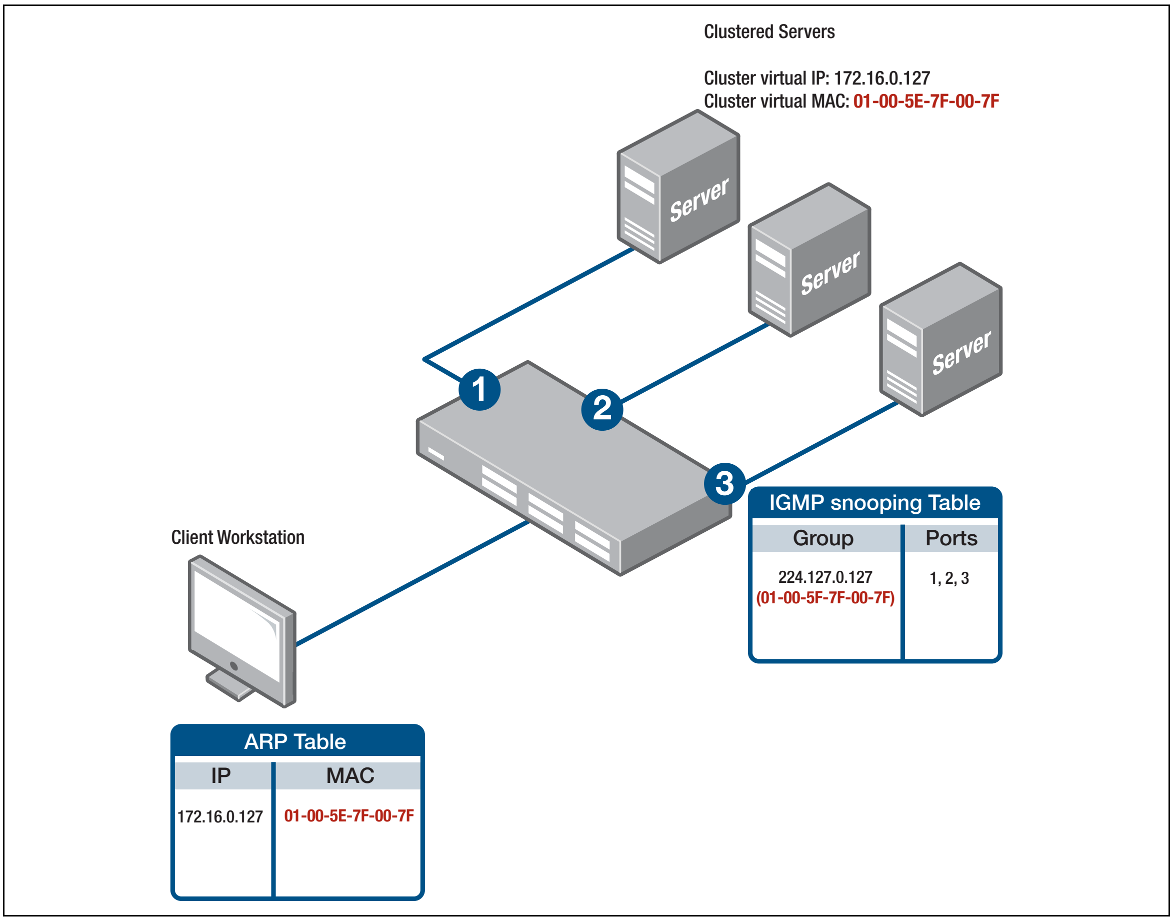
IGMP snooping Table

Group	Ports
224.127.0.127 (01-00-5F-7F-00-7F)	1, 2, 3

ARP response: 172.16.0.127 is at
HW address 02-BF-AC-10-00-7F

Source MAC address: 02-01-AC-10-00-7F
Dest MAC address: 00-00-54-1A-71-B3
Sender Hardware address: 01-00-5E-7F-00-7F
Sender IP address: 172.16.0.127
Target HW address: 00-00-54-1A-71-B3
Target IP address: 172.16.0.40

IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3



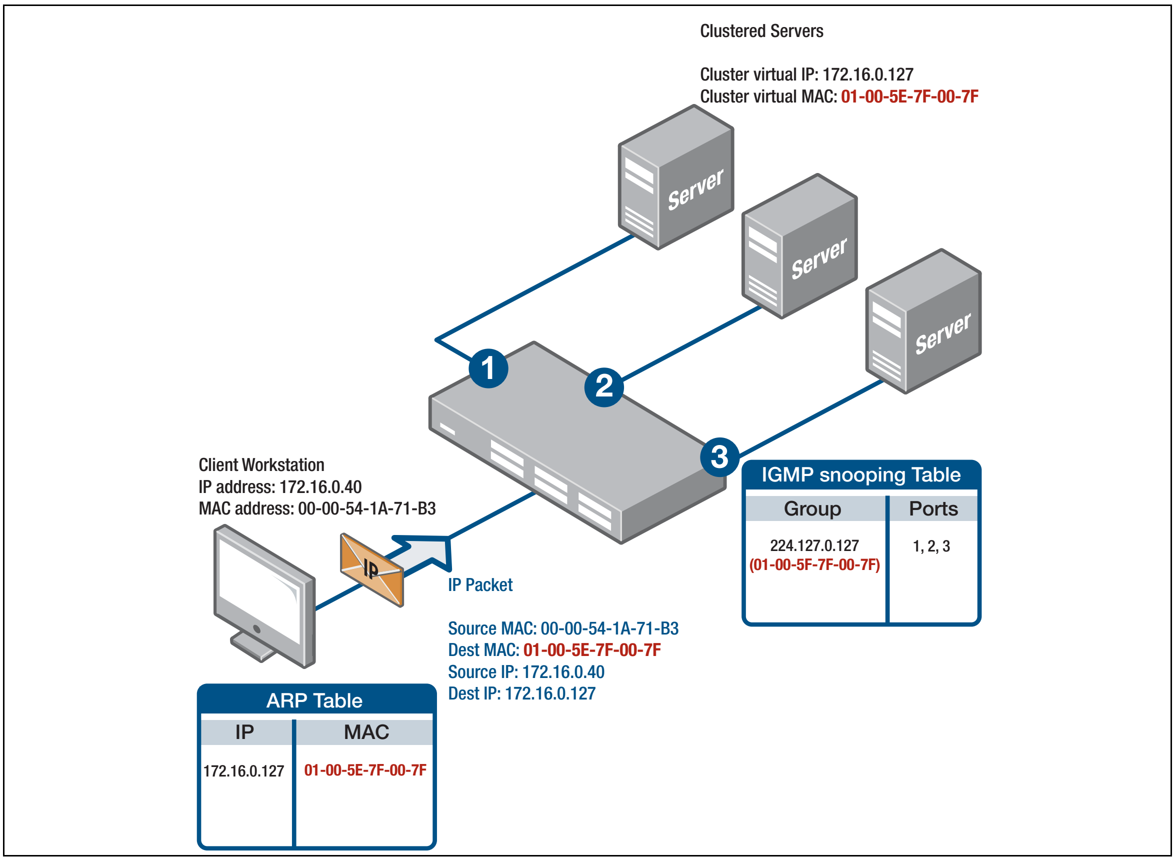
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-5E-7F-00-7F

IGMP snooping Table

Group	Ports
224.127.0.127 (01-00-5F-7F-00-7F)	1, 2, 3

ARP Table

IP	MAC
172.16.0.127	01-00-5E-7F-00-7F



Clustered Servers

Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-5E-7F-00-7F

Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

IP Packet

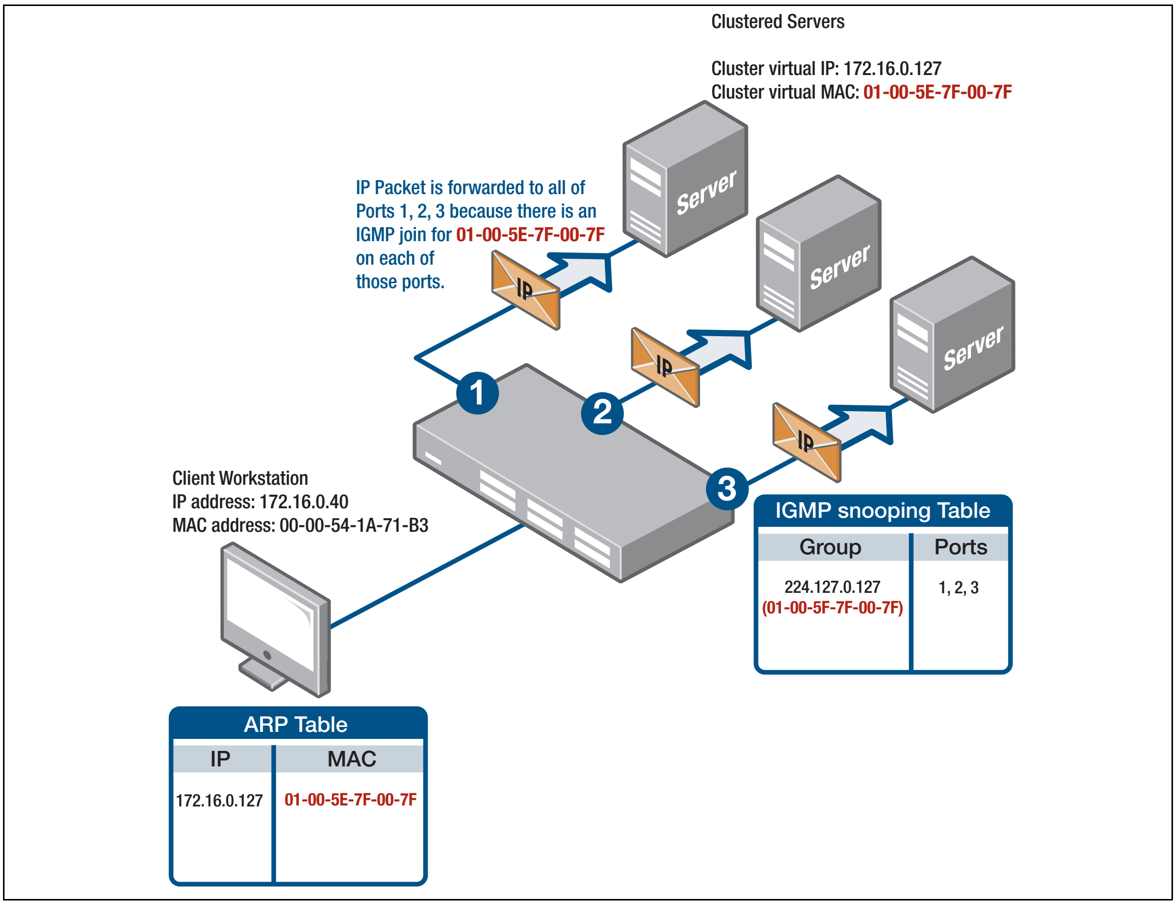
Source MAC: 00-00-54-1A-71-B3
Dest MAC: 01-00-5E-7F-00-7F
Source IP: 172.16.0.40
Dest IP: 172.16.0.127

IGMP snooping Table

Group	Ports
224.127.0.127 (01-00-5F-7F-00-7F)	1, 2, 3

ARP Table

IP	MAC
172.16.0.127	01-00-5E-7F-00-7F



Clustered Servers

Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 01-00-5E-7F-00-7F

IP Packet is forwarded to all of Ports 1, 2, 3 because there is an IGMP join for 01-00-5E-7F-00-7F on each of those ports.

Client Workstation
IP address: 172.16.0.40
MAC address: 00-00-54-1A-71-B3

IGMP snooping Table

Group	Ports
224.127.0.127 (01-00-5F-7F-00-7F)	1, 2, 3

ARP Table

IP	MAC
172.16.0.127	01-00-5E-7F-00-7F

Support for NLB in x-Series Switches

The operation of the NLB modes is optimized for Layer 2 networks. That is, the modes all work smoothly if the client work-stations and clustered servers are all in the same IP subnet, and any switches in between them are just operating at Layer 2.

In the purely Layer 2 environment, it is the client workstation that learns the ARP entry for the cluster's virtual IP/virtual MAC, and then sends cluster-directed packets to the virtual MAC. The Layer 2 switches in between then behave as described in the illustrations above.

For a Layer 2 network, the requirements on the switches are simple. In fact, NLB has been designed to make use of standard behaviors of Layer 2 switches. Therefore, in this environment, the switches don't even need to be aware that NLB is being used in the network.

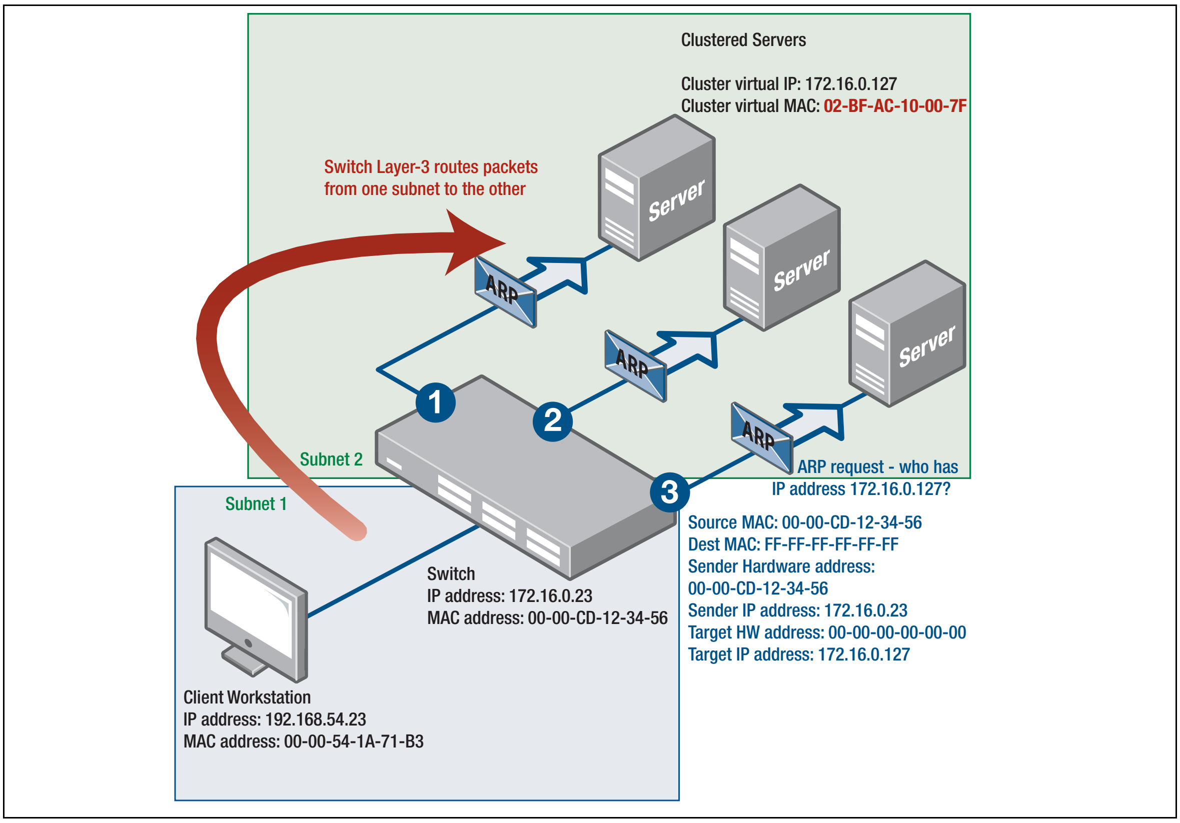
Allied Telesis x-series switches, like almost all switches in the market, provide good support for NLB when operating at Layer 2.

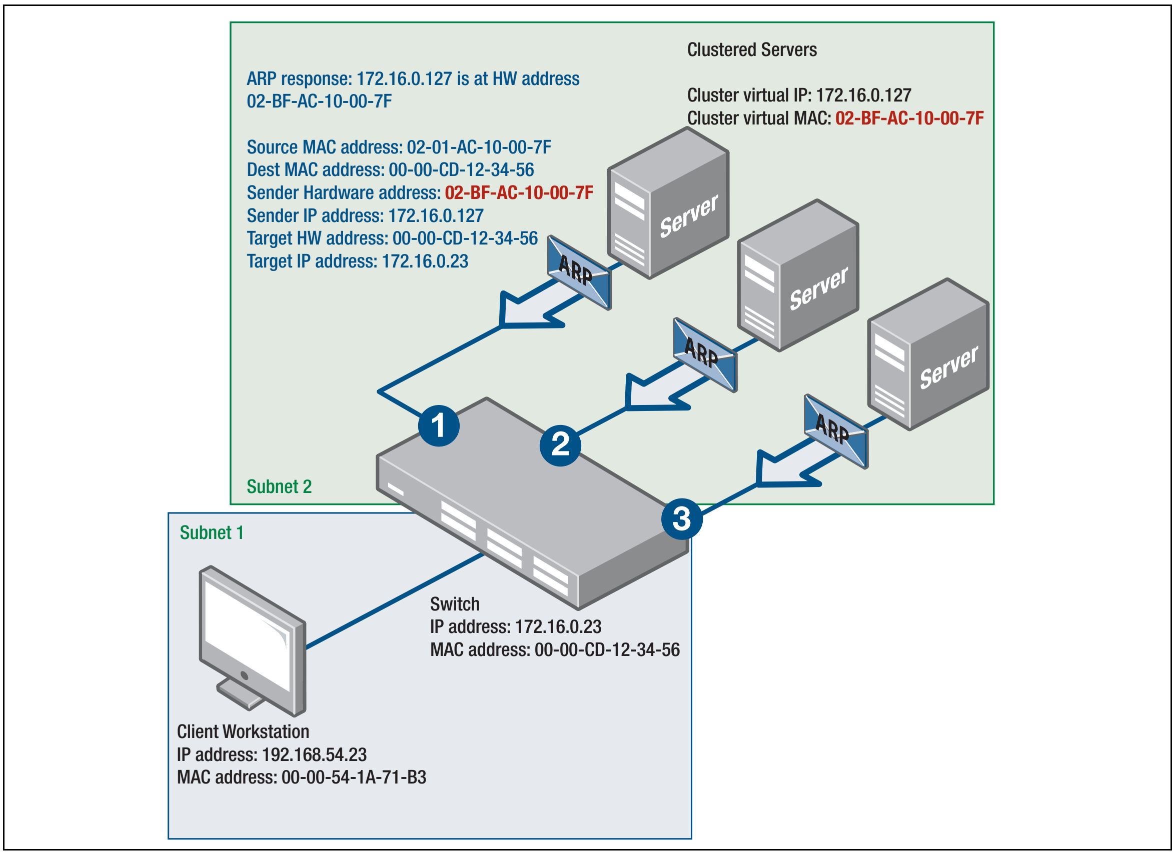
When the switches are operating at Layer 3, the situation is rather different. If the client workstations are in one subnet, and the clustered servers are in another subnet, and a switch in between is performing the Layer 3 routing between those subnets, then the switch needs to be explicitly NLB-aware.

The key point is that the switch needs to learn the ARP entries for the Cluster's virtual IP and Virtual MAC, and apply those ARP entries to multiple egress ports. It is not standard for a switch to apply the same ARP entry to multiple ports, and, as will be discussed below, each of the NLB modes required other non-standard behavior on the part of the Layer 3 switch.

Unicast Mode

When a switch operating at Layer 3 sits in front of a set of clustered servers operating in Unicast Mode, the ARP conversation is as below:


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
Switch Layer-3 routes packets from one subnet to the other
Server
Server
Server
ARP
ARP
ARP
1
2
3
Subnet 2
Subnet 1
ARP request - who has IP address 172.16.0.127?
Source MAC: 00-00-CD-12-34-56
Dest MAC: FF-FF-FF-FF-FF-FF
Sender Hardware address: 00-00-CD-12-34-56
Sender IP address: 172.16.0.23
Target HW address: 00-00-00-00-00-00
Target IP address: 172.16.0.127
Switch
IP address: 172.16.0.23
MAC address: 00-00-CD-12-34-56
Client Workstation
IP address: 192.168.54.23
MAC address: 00-00-54-1A-71-B3


Clustered Servers
Cluster virtual IP: 172.16.0.127
Cluster virtual MAC: 02-BF-AC-10-00-7F
ARP response: 172.16.0.127 is at HW address 02-BF-AC-10-00-7F
Source MAC address: 02-01-AC-10-00-7F
Dest MAC address: 00-00-CD-12-34-56
Sender Hardware address: 02-BF-AC-10-00-7F
Sender IP address: 172.16.0.127
Target HW address: 00-00-CD-12-34-56
Target IP address: 172.16.0.23
Server
Server
Server
ARP
ARP
ARP
1
2
3
Subnet 2
Subnet 1
Switch
IP address: 172.16.0.23
MAC address: 00-00-CD-12-34-56
Client Workstation
IP address: 192.168.54.23
MAC address: 00-00-54-1A-71-B3

Disparate ARP

The ARP responses in which the Sender MAC address does not match the Source MAC in the Ethernet header is referred to as a **Disparate ARP**.

In Allied Telesis x-series switches, the default behavior upon receiving a Disparate ARP response is:

- An ARP entry is created, pointing to the IP/MAC address in the content of the ARP packet, with the egress port being the port on which the ARP packet was received.
- An FDB entry is created for the MAC address in the content of the ARP packet, associated with the port on which the ARP was received. This is despite the fact that the MAC address in question was not the source MAC in the Ethernet header of the ARP packet. This FDB entry is not created by normal MAC learning, but is specifically created when the ARP entry is created, so that the hardware forwarding process (which uses the FDB as the way to find the egress port) works correctly.

The problem, though, is that the ARP entry (and FDB entry) is only ever associated with one egress port. So, rather than associating itself with multiple egress ports, it will jump from port to port as the replies from multiple servers arrive in quick succession. This fails to provide the behavior that the cluster requires—whereby packets destined to the cluster are sent to all cluster members at once.

So, AlliedWare Plus provides a specific, NLB-friendly, mode of dealing with Disparate ARP responses.

An ARP entry is created, pointing to the IP/MAC address in the content of the ARP packet, with no specific egress port - the 'egress port' is set to ‘flood', which means that packets matching this ARP entry are flooded to all ports in the egress VLAN.

They appear in the output of the commands **show arp** and **show ip flooding-nexthops** as:

```
ATX8106M2#show arp
IP Address       LL Address       Interface     Port         Type
192.168.200.34   001a.a004.25ef   vlan200       port1.0.1    dynamic
172.21.0.3       02bf.ac15.0005   vlan40        port1.0.2    dynamic
172.21.0.4       02bf.ac15.0005   vlan40        port1.0.2    dynamic
```

```
ATX8106M2#show ip flooding-nexthops
IP Address       MAC Address      Interface            Flooding Mode    Type
192.168.200.34   001a.a004.25ef   vlan200              vlan             dynamic
172.21.0.3       02bf.ac15.0005   vlan40               vlan             dynamic
172.21.0.4       02bf.ac15.0005   vlan40               vlan             dynamic
```

- No FDB entry is created for the MAC address in the content of the ARP packet. This ensures that packets destined to this MAC address are flooded to all ports of the VLAN.

This mode is configured by the command **arp-mac-disparity unicast** on the VLAN that faces the clustered servers.

For example:

```
con t
 int vlan10
  arp-mac-disparity unicast
```

Note: For the SBx8100 platform: The default behavior upon receiving a Disparate ARP response depends on whether **arp-mac-disparity multicast** or **arp-mac-disparity multicast-igmp** has been configured on an interface.

- If this has been configured, then the default action is to flood the packets.
- If this has not been configured, then the default action is to drop the packets.
- The command **platform multicast-address-mismatch-action [drop|bridge]** will override the default behaviour regardless of the arp-mac-disparity configuration.

Multicast Mode

In multicast mode, not only is the ARP response from the clustered servers disparate (i.e. the source MAC on the Ethernet header differs from the sender MAC within the ARP packet content) it also has the unusual property that the sender IP address in the packet is unicast, but the sender MAC address is multicast.

By default, x-series switches simply drop such ARP responses, deeming a multicast sender MAC to be invalid.

However, the command **arp-mac-disparity multicast**, entered on the VLAN that faces the clustered servers, puts the switch into a mode whereby it will accept such ARP responses, and upon receiving them will do the same as it does when receiving unicast disparate ARPs when **arp-mac-disparity unicast** has been configured. That is, an ARP entry is created for the IP/MAC in the content of the ARP packet, but with the egress port set to ‘flood’, and no FDB entry is created for the MAC address.

As a result, packets destined to the IP address in question are flooded to the destination VLAN.

This behavior makes no distinction between clusters operating in Multicast mode and those operating in multicast mode with the IGMP option. The fact that a cluster is sending IGMP reports to the switch does not make any difference, it still floods the packets to all ports in the egress VLAN.

In both cases, when the ARP entries are learnt, they appear in the output of the commands **show arp** and **show ip flooding-nexthops** as:

```
ATX8106M2#show arp
IP Address      LL Address      Interface    Port        Type
10.100.0.50     03bf.0a64.0032  vlan200      port1.0.1   dynamic
10.100.0.56     0100.5e7f.0013  vlan200      port1.0.1   dynamic
10.100.0.75     03bf.0a64.0044  vlan200      port1.0.1   dynamic
10.100.0.76     03bf.0a64.0032  vlan200      port1.0.1   dynamic
10.100.0.60     0100.5e7f.0013  vlan200      port1.0.1   dynamic
```

```
ATX8106M2#show ip flooding-nexthops
IP Address      MAC Address     Interface        Flooding Mode    Type
10.100.0.50     03bf.0a64.0032  vlan200          vlan             dynamic
10.100.0.56     0100.5e7f.0013  vlan200          vlan             dynamic
10.100.0.75     03bf.0a64.0044  vlan200          vlan             dynamic
10.100.0.76     03bf.0a64.0032  vlan200          vlan             dynamic
10.100.0.60     0100.5e7f.0013  vlan200          vlan             dynamic
```

Note: If Layer 2 load balancing is enabled, the commands **arp-mac-disparity multicast** and **arp-mac-disparity unicast** cannot be configured on the following platforms: SBx8100, x530, x220, and FS980. This is due to a hashing incompatibility.

However, you can use these commands if you first run the command **no platform load-balancing src-dst-mac**.

Static multicast ARP entries

It is also possible to create static ARP entries with multicast destination MAC, and multiple egress ports, with the **arp** command:

```
arp UNICAST_IP MULTICAST_MAC PORT_LIST
```

```
For example:
arp 10.10.1.100 010e.11ff.2222 port1.0.1,port1.0.3,port1.0.8
```

In this case, an ARP entry is created for the IP/MAC in question, associated with all the egress ports specified in the command. At Layer 2 a special hardware forwarding entry is created, that associates the MAC address with the set of egress ports. So, packets destined to the IP/MAC in question are forwarded just to the egress ports specified in the command, and are not flooded to the whole egress VLAN.

Allied Telesis NETWORK SMARTER

North America Headquarters | 19800 North Creek Parkway | Suite 100 | Bothell | WA 98011 | USA | T: +1 800 424 4284 | F: +1 425 481 3895
Asia-Pacific Headquarters | 11 Tai Seng Link | Singapore | 534182 | T: +65 6383 3832 | F: +65 6383 3830
EMEA & CSA Operations | Incheonweg 7 | 1437 EK Rozenburg | The Netherlands | T: +31 20 7950020 | F: +31 20 7950021

alliedtelesis.com