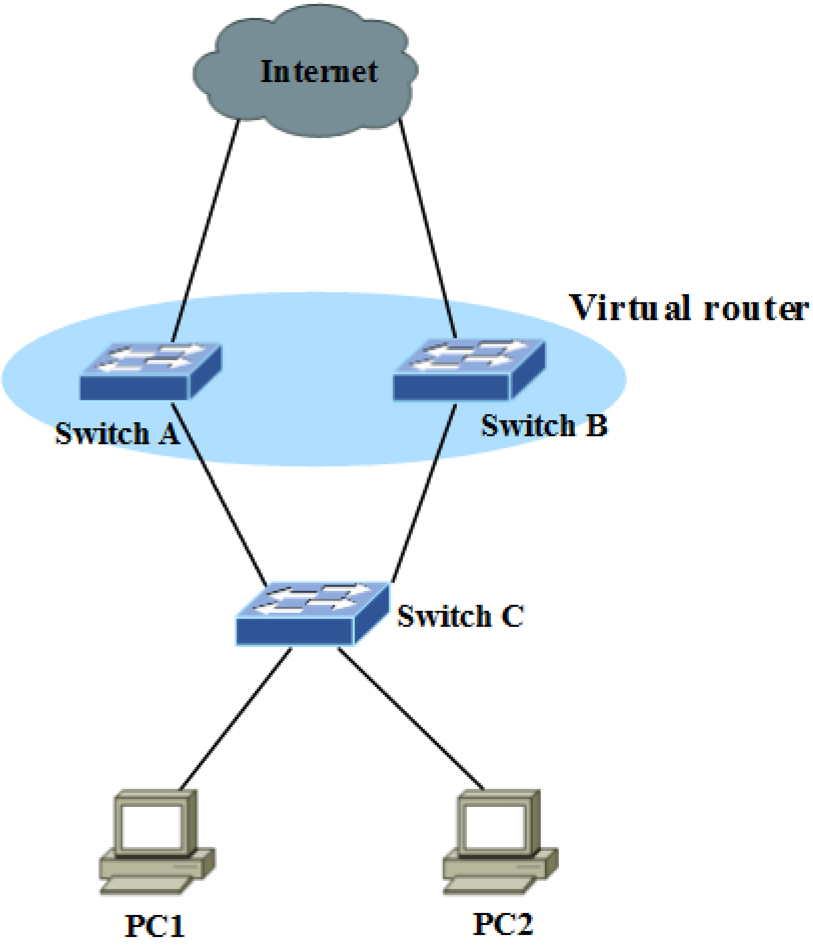
Principle of VRRP

Introduction

Virtual Router Redundancy Protocol (VRRP) is a redundancy and backup function of network devices. VRRP group combines several routing devices into one virtual routing device and uses the IP address of the virtual routing device as the default gateway to establish communication with external networks. When a gateway device in a VRRP group fails, the VRRP mechanism can elect a new gateway device as Master to transmit the data traffic to ensure reliable network communication.

Figure 1. VRRP Networking Diagram



As shown in Figure 1, Switch A and Switch B form a virtual router. This virtual router has its own IP (could be IPv4 or IPv6) address. The hosts in the LAN use the virtual router IP as their default gateway IP. The switch with the highest priority between Switch A and Switch B functions as the master and the gateway device. The other switch functions as the backup router. For IPv4, the VRRP packets will use multicast MAC 01:00:5E:00:00:12 as destination MAC. For IPv6, the VRRP packets will use multicast MAC 33:33:00:00:00:12 as destination MAC. Switch C should support layer 2 switching function.

Besides the Standard VRRP protocol mode, PICOS also supports Active-Active VRRP mode. In the Standard VRRP protocol mode, only the Virtual Master Router can forward packets whereas the Virtual Backup Routers cannot forward packets. By adding a new working mechanism based on the VRRP standard protocol mode, Active-Active VRRP mode provides load balancing between the master and backup switches in the VRRP group, both of which are active, thus avoid the situation where the backup switches are always idle in the Standard VRRP protocol mode. This greatly improves usage efficiency of network resources.

In the version before PICOS 2.11.10, PICA8 switch supports only VRRPv2. From PICOS 2.11.10, PICA8 switch supports both VRRPv2 and VRRPv3. The differences between VRRPv2 and VRRPv3 are as follows:

- Apply to different networks. VRRPv3 supports IPv4 and IPv6 address families while VRRPv2 only supports IPv4 addresses.
- Packet format is different between VRRPv2 and VRRPv3, for details please refer to VRRP Packet Format.
- VRRPv3 supports accept mode while VRRPv2 does NOT support this mode. Accept mode controls whether a virtual router in master state will accept packets addressed to the virtual IPvX address of the VRRP group if it is not the IP address owner. However, in VRRPv2, the master switch always accepts packets addressed to the virtual IPvX address. For details about accept mode, please refer to Accept Mode.

NOTE:

- VRRPv2 and VRRPv3 interoperation is not supported, VRRP version configured on all devices in a VRRPv3 group must be the same.
- Standard VRRP mode supports one Master and several Backup switches in a VRRP group, while Active-Active VRRP mode supports only one Master and one Backup switch in a VRRP group.
- The same VRID must be configured on all devices for the same VRRP group.
- The IP address of the virtual router can be either an unassigned IP address in the network segment where the VRRP group resides or the IP address of an interface on a router in the VRRP group. A router whose interface IP address is same as the virtual IP address is called an "IP address owner". When the router is an IP address owner, its priority is always 255.
- Only one IP address owner can be configured in the same VRRP group.

The VRRP principle included in this document describes only the working mechanism of Active-Active VRRP mode for IPvX (In this document, the term "IPvX" (where X is 4 or 6) is introduced to mean either "IPv4" or "IPv6"). For details about standard VRRPv2 protocol mode, please refer to the RFC3768. For details about Standard VRRPv3 protocol mode, please refer to RFC5798.

Active-Active VRRP Mode

The basic principle in Active-Active VRRP is that one virtual IPvX address corresponds to two virtual MAC addresses, and each switch in the VRRP group corresponds to one virtual MAC address to support load balancing between the master and backup switches. This helps in avoiding the backup switch from being in idle state as is the case in Standard VRRP mode.

VRRP Working Mechanism

Active-Active VRRP mode working process is as follows:

1. Devices in a VRRP group elect the Master and Backup based on their priorities and VLAN interface IPvX addresses by exchanging VRRP advertisement packets.

2. The master and backup devices periodically send VRRP Advertisement packets to each other to advertise its configuration (such as priority) and running status. This can also notify the downstream devices to refresh the MAC entries.

3. All the ARP requests / Neighbor Solicitation (NS) packets from the downstream devices or hosts are responded to by the master. ARP reply / Neighbor Advertisement (NA) packets carry one of the virtual MAC addresses of the VRRP group. For details about virtual MAC address allocation, please see Virtual MAC Address Allocation.

4. To notify the downstream devices of the virtual MAC, the master and backup devices periodically send virtual MAC update messages. For details, please refer to Virtual MAC Updating.

5. For IPv6, master sends Router Advertisements for the link-local addresses of virtual router IP address on the local area network to announce its availability for routing.

6. The VRRP device uses its real MAC address as source MAC for traffic forwarding.

VRRP Packet Format

VRRP packets are sent encapsulated in IPvX packets. For IPv4, in Layer 3 IP header of VRRP packets, source IP is the real IP (same segment with Virtual IP) of the VLAN interface, destination IP is IPv4 multicast address 224.0.0.18 that is assigned to VRRP. In Ethernet header, source MAC is virtual MAC, and destination MAC is multicast MAC 01:00:5E:00:00:12. For IPv6, destination IP is IPv6 multicast address FF02::12 and destination MAC is multicast MAC 33:33:00:00:00:12.

VRRPv2 and VRRPv3 packet formats are shown below.

Figure 2. VRRPv2 Packet

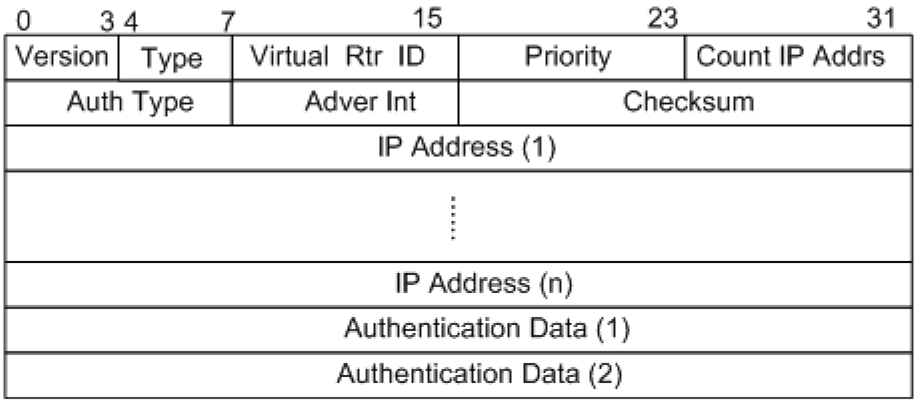
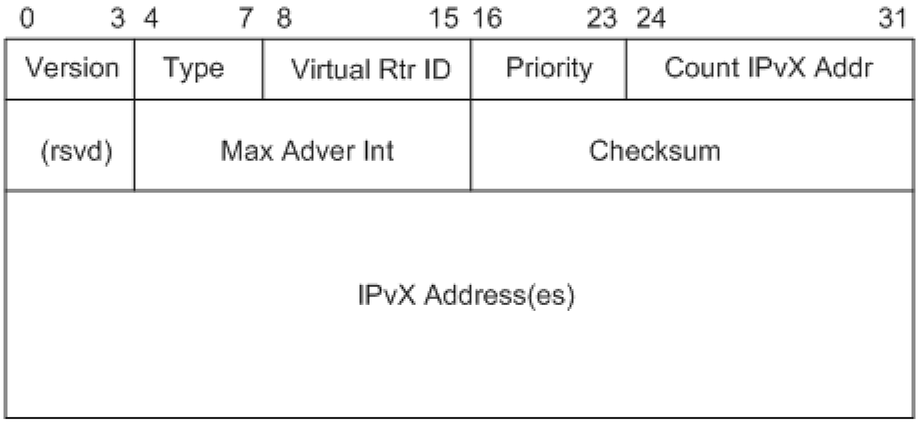


Figure 3. VRRPv3 Packet



VRRP Field	Descriptions
Version	The version field specifies the VRRP protocol mode. • VRRP_VERSION = 2, Standard VRRPv2 mode. • VRRP_VERSION = 3, Standard VRRPv3 mode. • VRRP_LOAD_BALANCE_VERSION = 4, Active-Active VRRPv2 mode. • VRRP_LOAD_BALANCE_VERSION = 5, Active-Active VRRPv3 mode.
Type	The type field specifies the type of VRRP packet. • VRRP_TYPE_ADVERTISEMENT = 1, advertisement packet sent by the master. • VRRP_BACKUP_ADVERTISEMENT = 2, advertisement packet sent by the backup device. • VRRP_UPDATE_ADVERTISEMENT = 3, MAC address update packet, sent periodically both by the master and backup devices. • VRRP_SYNC_ARP_ADVERTISEMENT / VRRP_SYNC_ NEIGHBOR_ADVERTISEMENT = 4, ARP synchronization /IPv6 Neighbor synchronization packet.
Virtual Rtr ID	Virtual router ID, range from 1 to 254.
Priority	Device priority in the VRRP group, range from 1 to 254. The default value is 100.
Count IP Addrs / Count IPvX Addr	The number of virtual IPvX addresses in the VRRP group.

Auth Type	Used only for VRRPv2 packets. The authentication type field identifies the authentication method being utilized. There are three types: • 0: Non Authentication. • 1: Simple Text Password. • 2: IP Authentication Header, which indicates the MD5 authentication mode. By default, there's no authentication.
rsvd	Reserved field for VRRPv3 packets.
Adver Int / Max Adver Int	VRRP advertisement interval, in second for IPv4 network and centisecond for IPv6 network. The default value is 1 second for both IPv4 and IPv6.
Checksum	The checksum field is used to detect data corruption in the VRRP packet.
IP Address / IPvX Addresses	IPvX address list that is associated with the virtual router. The number of addresses included is specified in the "Count IPvX Addr" field.
Authentication Data	Authentication key.

VRRP Packet Types

There are four types of VRRP packets in Active-Active VRRP mode: master advertisement, backup advertisement, virtual MAC update packet, and ARP synchronization / IPv6 neighbor synchronization packet.

1) Master advertisement where type = 1

The master device periodically sends VRRP advertisement packets to the backup device to advertise its configuration information (such as priority) and working status. The backup device determines whether the master is working properly by receiving VRRP advertisement packets.

When the master device fails to send advertisement packets due to some network fault, the backup device can not immediately get the working status of the master. It will wait until the Master_Down_Interval timer expires before it considers that the master device has failed, then switches its own state to Master. The value of the Master_Down_Interval timer is 3 × Advertisement_Interval + Skew_time, Skew_Time = (256 - Priority)/256, in seconds.

2) Backup advertisement message where type = 2

Backup device periodically sends VRRP advertisements to advertise its configuration information (such as priority) and working status in the VRRP group. The master device determines whether the backup device is working properly by receiving Backup advertisement packets. The interval for sending advertisement messages is the same as that of the Master advertisement message.

3) Virtual MAC update message where type = 3

To notify the downstream devices of the virtual MAC, the master and backup devices periodically send virtual MAC update messages. The virtual MAC address in Active-Active VRRP mode is used in the Ethernet header of virtual MAC update packet as the source MAC address. The connected network devices of the VRRP group refresh the MAC entries in time to perform packet forwarding. The default interval is 120s.

4) ARP synchronization message or IPv6 neighbor synchronization where type = 4

The VRRP device learns ARP entries / IPv6 Neighbor entries from the receiving traffic.

In order to ensure that both devices in the VRRP group can forward downstream data rapidly, ARP **/** IPv6 Neighbor synchronization packets are sent between the Master and Backup device when the ARP **/** IPv6 Neighbor table on the VRRP group device is updated.

For IPv4, the IPv4 addresses in the ARP table that are in the same network segment with virtual IPv4 address are synchronized to the peer VRRP device.

For IPv6, two types of IPv6 address in the Neighbor table are synchronized to the peer VRRP device,

- The global IPv6 addresses in the IPv6 Neighbor table that are in the same network segment with virtual IPv6 address.
- The link-local addresses that are in the same VLAN with the VRRP group.

Upon receiving the ARP synchronization **/** IPv6 Neighbor synchronization packet, the peer device initiates ARP **/** IPv6 Neighbor learning process. The ARP **/** IPv6 Neighbor entries on both VRRP devices will eventually reach a consistent state and both devices in the VRRP group can forward downstream data rapidly.

VRRP States

VRRP protocol defines three state machines: Initialize, Master and Backup.

Initialize

- VRRP is unavailable. The device in Initialize state cannot process VRRP advertisement packets.

- When the VRRP process starts it goes into initialize state. When the device is in master or backup states and it detects a fault, it enters the Initialize state.
- After receiving an interface Up message, the VRRP-enabled device first switches to the Backup state and then switches to the Master state after the Master_Down_Interval timer expires.

Master

The VRRP device in Master state performs the following operations:

- Sends VRRP advertisement packets periodically.
- Receives VRRP advertisement packets from the backup and determines whether the backup is working properly.
- Uses the virtual MAC address to respond to ARP request destined for the virtual IPv4 address.
- Uses the virtual MAC address to respond to ND Neighbor Solicitation messages destined for the virtual IPv6 address.
- Forwards IPvX packets destined for the virtual MAC address.
- Becomes the backup if the device receives a VRRP advertisement packet with a higher priority than its VRRP priority.
- Becomes the backup if the device receives a VRRP advertisement packet with the same priority as its VRRP priority and the local VLAN interface address is smaller than the connected interface address on the peer VRRP device.
- For VRRPv3, accepts packets addressed to the IPvX address(es) associated with the virtual router if it is the IPvX address owner or if Accept Mode is True. For details about Accept Mode, please see Accept Mode.
- Sends Router Advertisement for the link-local addresses in virtual IP address list, as the source IP address of the RA packet, on the local area network to announce its virtual IP address as available for routing. In the prefix information field of the RA packet, it carries the global addresses in virtual IP address list for applying IPv6 stateless address auto-configuration protocol (refer to RFC2462 IPv6 Stateless Address Autoconfiguration).

Backup

The VRRP device in Backup state performs the following operations:

- Sends VRRP advertisement packets periodically.
- Receives VRRP advertisement packets from the master and determines whether the master is working properly.
- Does NOT respond to ARP request / ND Neighbor Solicitation messages destined for the virtual IPv4 / IPv6 address.
- Forwards IPvX packets destined for the virtual MAC address in Active-Active mode for load balancing.
- When it receives a packet of lower priority, it immediately switches to the Master state by default. If non-preemptive is configured, the device remains in the backup state.
- Master_Down_Interval timer: If the backup receives no advertisement packet after the timer expires, the backup takes the role of the master. The calculation formula is as follows:

  Master_Down_Interval = 3*Advertisement_Interval + Skew_time (offset time)

  Skew_Time = (256 - Priority) / 256

- For VRRPv3, does NOT accept packets addressed to the IPvX address(es) associated with the virtual router.
- Does NOT send Router Advertisement messages for the virtual router.

Virtual MAC Address Allocation

In Active-Active VRRP mode, the master switch no longer sends gratuitous ARP request or unsolicited ND Neighbor Advertisement packets as Standard VRRP protocol mode does. Instead, the master switch receives the ARP request / ND Neighbor Solicitation message sent by the connected hosts or devices and responds with ARP reply / ND Neighbor Solicitation message using the virtual MAC address instead of the real MAC address of the interface.

The virtual MAC address is allocated according to the hash value calculated from the source MAC address in the ARP request packet / ND Neighbor Solicitation message. This results in half of the hosts end up learning the virtual MAC address of Master and the other half of the hosts learn the virtual MAC address of the Backup. In this way, traffic from the hosts is shared between two VRRP devices to achieve load balancing.

The virtual router generates the virtual MAC address based on the virtual router ID, the format is 00: 00: 5E: 00: 0X: VRID. For IPv4, X is 1 in master device and 2 in backup device; For IPv6, X is 2 in master device and 1 in backup device.

For example, assuming that VRID 5 enables Active-Active VRRP mode for IPv4 and VRID 6 enables Active-Active VRRP mode for IPv6, then the virtual MAC addresses of VRID 5 will be 00-00-5E-00-01-05 and 00-00-5E-00-02-05, and that of VRID 6 will be 00-00-5E-00-02-06 and 00-00-5E-00-01-06.

Role Elect of Master and Backup

VRRP determines the device role in the VRRP group based on device priority and VLAN interface by exchanging VRRP advertisement packets. The device with a higher priority is more likely to become the master. If two devices have the same priority, the device with a larger VLAN interface IP address becomes the master.

The VRRP-enabled device in a VRRP group initially works in Initialize state. After receiving an interface Up message, if the priority of the device is 255, it will become the master directly; if the priority of the device is less than 255, the VRRP-enabled device first switches to the Backup state and then switches to the Master state when the Master_Down_Interval timer expires.

The device that first switches to the Master state obtains the priorities of other devices in the group by exchanging VRRP advertisement packet and then elect the master router.

- If the master priority in VRRP advertisement packets is higher than or equal to the priority of the device, the backup remains in Backup state.
- If the backup device has a higher priority than the master, the working mode of the backup (preemptive or non-preemptive) determines whether the master is re-elected.
    - Preemptive mode: If the priority of the backup router is higher than the priority of the current master router, the backup router automatically becomes the master router.
    - Non-preemptive mode: As long as the master router is working properly, the backup router with a higher priority cannot become the master router.
- The IP address owner's running priority is always 255; the IP address owner always works in preemptive mode, regardless of whether the preemption function is enabled. If the VRRP device is the IP address owner, it will switch to the master state immediately after receiving the interface Up message.

Virtual MAC Updating

To notify the downstream devices of the virtual MAC, the master and backup devices periodically send virtual MAC update messages. The virtual MAC address in Active-Active VRRP mode is used in the Ethernet header of virtual MAC update packet as the source MAC address. The connected network devices of the VRRP group refresh the MAC entries in time to perform packet forwarding. The default interval is 120s.

ARP / IPv6 Neighbor synchronization

The VRRP devices learn ARP entries / IPv6 Neighbor entries when receiving downstream traffic.

In order to ensure that both devices in the VRRP group can forward downstream data rapidly, ARP **/** IPv6 Neighbor synchronization packets are sent between the Master and Backup device when the ARP **/** IPv6 Neighbor table on the VRRP group device is updated.

For IPv4, the IPv4 addresses in the ARP table that are in the same network segment with virtual IPv4 address are synchronized to the peer VRRP device.

For IPv6, two IP addresses in the Neighbor table are synchronized to the peer VRRP device,

- The global IPv6 addresses in the IPv6 Neighbor table that are in the same network segment with virtual IPv6 address.
- The link-local addresses that are in the same VLAN with the VRRP group.

Upon receiving the ARP synchronization **/** IPv6 Neighbor synchronization packet, the peer device initiates ARP **/** IPv6 Neighbor learning process. The ARP **/** IPv6 Neighbor entries on both VRRP devices will eventually reach a consistent state and both devices in the VRRP group can forward downstream data rapidly.

Accept Mode

VRRPv3 supports Accept Mode which controls whether a virtual router in Master state will accept packets addressed to the virtual IPvX address of a VRRP group if it is not the IP address owner (the IP address owner is the router that has the interface whose actual IP address is used as the virtual router’s IP address).

By default, the Accept Mode is disabled, if the master is not the IP address owner, it only accepts the ARP requests/ARP replies or NS/NA messages addressed to the virtual IP, any other messages whose destination IP is the virtual IP are not accepted. But when accept mode is enabled, it can accept all packets whose destination IP is a virtual IP.

Deployments that rely on, for example, pinging the address owner's IPvX address may choose to configure Accept Mode to True.

NOTE:

- Accept Mode is only supported in VRRPv3 while VRRPv2 does NOT support. In VRRPv2, the master switch always accepts packets addressed to the virtual IPvX address.
- When Accept Mode is disabled, PICOS can still accept and process IPv6 Neighbor Solicitations / Neighbor Advertisements packets and ARP Request / ARP Reply packets.
- If the master is the IP address owner, it accepts all the packets addressed to the IPvX address(es) associated with the virtual router even though Accept Mode is disabled.