INTERNET-DRAFT
June 22, 1997

S. Knight
Ascend Communications, Inc.
D. Weaver
Ascend Communications, Inc.
D. Whipple
Microsoft, Inc.
R. Hinden
Ipsilon Networks, Inc.

Virtual Router Redundancy Protocol

<draft-ietf-vrrp-spec-00.txt>

Status of this Memo

Abstract

The memo documents the Virtual Router Redundancy Protocol. This is a protocol which allows several routers to utilize the same virtual IP address. One router will be elected as a master, with X routers acting as backups in case of failure of the master router. The primary advantage to utilizing this protocol, is that host systems may be configured with a single default gateway, rather than running

an active routing protocol. Each interface on each router within a VRRP cluster, will be configured with a real IP address, and the virtual IP address for the particular cluster. Overall, this protocol adds to the options for providing fault redundancy for router networks.

Table Of Contents

1. Introduction

The reason for the development of VRRP is to create a standard protocol, with multi-vendor support to resolve the problem of router failure. Specifically, when a single router is utilized as a default gateway, and all hosts are statically configured to this default gateway, a failure is catastrophic. VRRP resolves this problem by creating virtual clusters, where each cluster is configured with a set of member routers. Each member router is either a master router for the cluster or a backup router for the cluster, but not both simultaneously. In addition, there MUST only be a single master router per cluster, at any given time. All member routers are configured to be part of a cluster, with a given virtual IP address. This virtual IP address is utilized as the default gateway on all of the host systems. Given a failure on the current master router, the next appropriate backup router will become the master router for the given cluster. When routers are configured with the equal priority the router which is master will stay master as long as it is up.

Of course this problem could be solved by running a standard routing protocol such as OSPF, RIP, or RIPv2 on the hosts. However, this is not always feasible due to either security issues, when hosts are multihomed, or in some cases implementations of these routing protocols simply do not exist.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119].

2. Scope

This memo describes the Virtual Router Redundancy Protocol.

This protocol is intended for IPv4 only. A version for IPv6 will be defined in a separate specification.

Within the scope of this specification are:

1. Packet format and header contents.
2. State Diagrams and Descriptions
3. Network Design Samples

Outside of the scope are

1. Network management
2. Host internal optimizations

3. Definitions

Cluster

Used to describe a set of routers who all have membership to the set of routers S, where S contains all routers configured with the same virtual IP address.

Master Router

Used to describe the currently active router, for a particular cluster, with a particular virtual IP address. Their can only be one master router in a particular cluster.

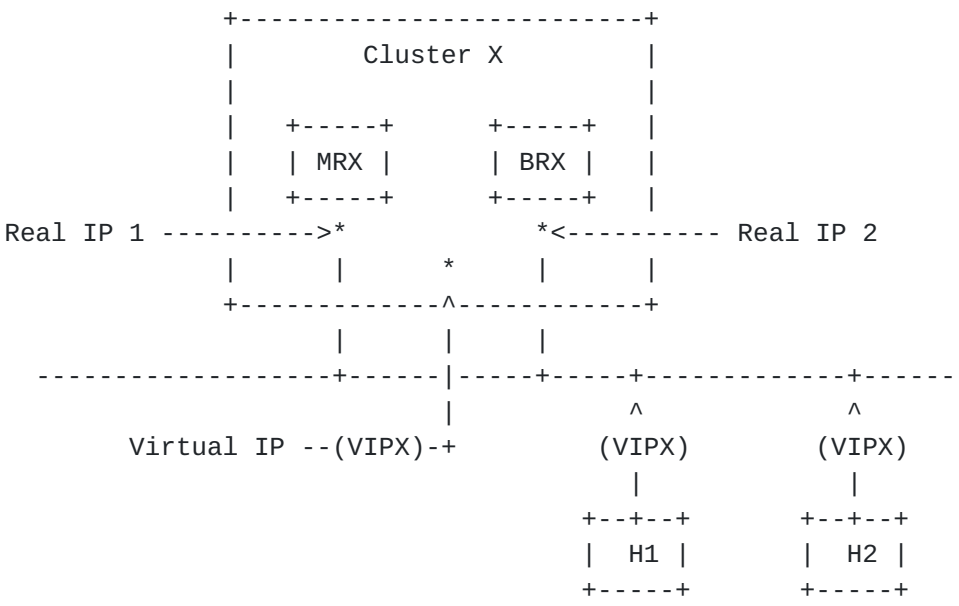
Backup Router

Used to describe a router which is configured to act as a backup for a particular cluster. There can be several backup routers in a single cluster.

4. Sample Configurations

4.1 Sample Configuration 1

The following figure shows a simple VRRP network.



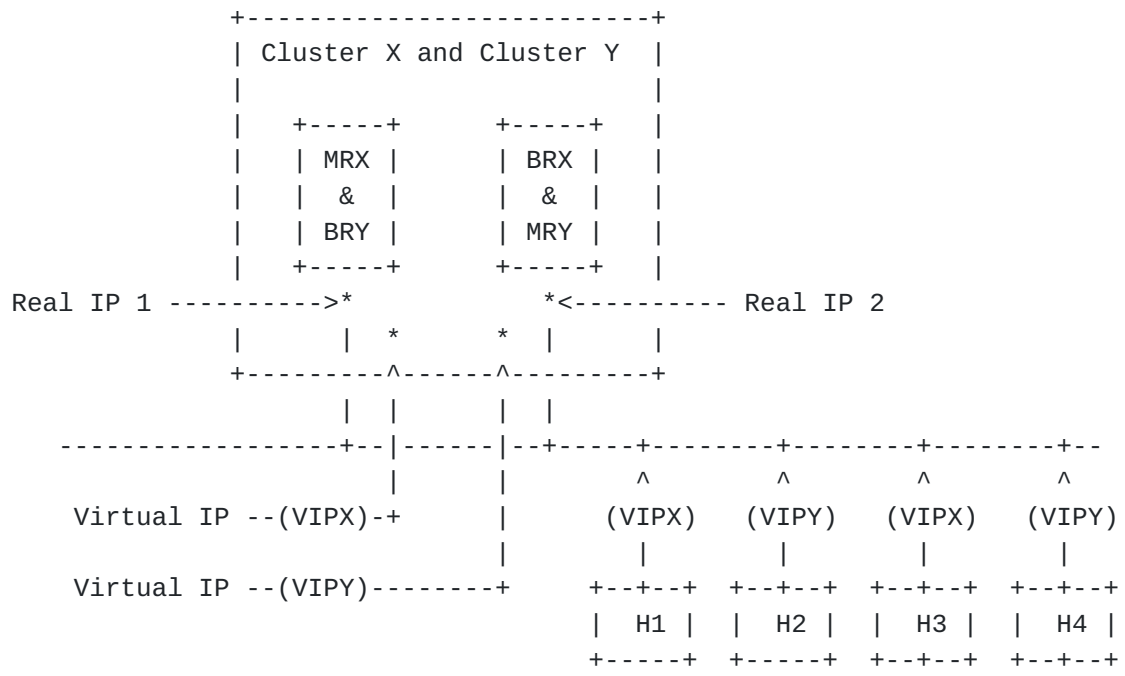
The above configuration shows the most likely utilization of the VRRP protocol. In this configuration, the hosts simply point their default routes at the virtual IP address X (VIPX), and the routers run VRRP

between themselves. The router on the left is the default master router (MRX), and the router on the right is the backup router (BRX).

```
   Legend:  ---+---+---+--  =  802 network, Ethernet or FDDI
                         H  =  Host computer
                        MR  =  Master Router
                        BR  =  Backup Router
                         *  =  IP Address
                       VIP  =  default gateway for hosts (Virtual IP)
```

4.2 Sample Configuration 2

The following figure shows a more interesting VRRP network.



In the above configuration, half of the hosts point their default gateway at cluster X's virtual IP address (VIPX), and half the hosts point their default gateway at cluster Y's virtual IP address (VIPY). This has the effect of load balancing the outgoing traffic, while also providing full redundancy.


```
   Legend:  ---+---+---+--  =  802 network, Ethernet or FDDI
                         H  =  Host computer
                        MR  =  Master Router
                        BR  =  Backup Router
                         *  =  IP Address
                       VIP  =  default gateway for hosts (Virtual IP)
```

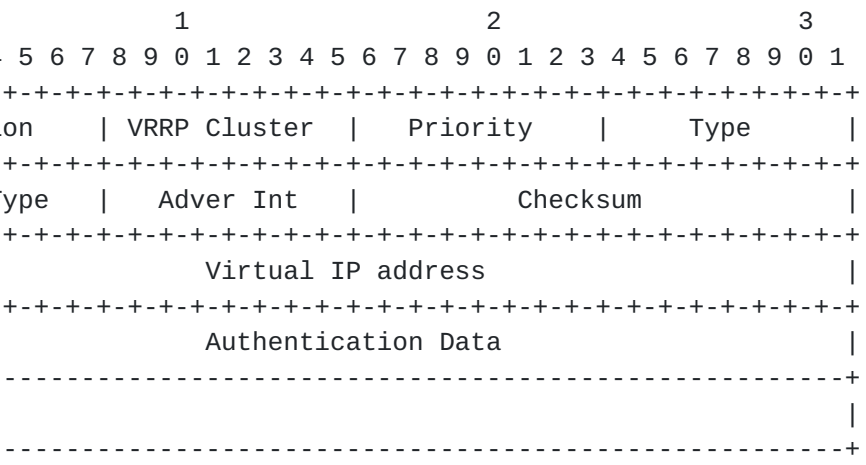
5. Protocol

The purpose of the VRRP packet is to communicate to all other VRRP routers both the priority and the state of the master's associated interface.

VRRP packets are sent encapsulated in IP packets. They are sent to an IPv4 multicast address assigned for VRRP.

5.1 VRRP Packet Format

This section defines the format of the VRRP packet and the relevant fields in the IP header.



5.2 IP Field Descriptions

5.2.1 Source Address

The real IP address of the interface the packet is being sent from.

5.2.2 Destination Address

The VRRP IP multicast address assigned by the IANA. It is defined to be:

224.0.0.(TBD IANA assignment)

This is a link local scope multicast address. Routers should not forward a datagram with this destination address regardless of its TTL.

5.2.3 TTL

The TTL should be set to 255. A VRRP router receiving a packet with the TTL not equal to 255 MUST discard the packet.

5.2.4 Protocol

The VRRP IP protocol number assigned by the IANA. It is defined to be (TBD).

5.3 VRRP Field Descriptions

5.3.1 Version

The version field specifies the VRRP protocol version of this packet. This document defines version 1.

5.3.2 VRRP Cluster

The VRRP Cluster field specifies the cluster this packet applies to. Note: The interface may participate in more than one VRRP cluster simultaneously, perhaps serving as master in one cluster, while simultaneously serving as backup in other clusters.

5.3.3 Priority

The priority field specifies the currently configured VRRP priority value for this interface and cluster. Higher values equal higher priority. This field is an 8 bit unsigned field, giving 1 as the minimum priority, and 255 as the maximum priority. The default priority is 100 (decimal).

Priority value of zero (0) has a special meaning. It means that the current master had decided to stop running VRRP. This is used to cause other backup routers to quickly become master with out having to timeout the current master.

In the event that two or more routers within a cluster have equal priority, and that priority is the highest priority in the cluster, initially the router with the higher real interface IP address (interpreted as a 32 bit unsigned integer) will become master. Any new router joining the cluster with the same priority will not become master even if it has a higher IP address unless the current master goes down.

5.3.4 Type

The type field specifies the type of this VRRP packet. The only packet type defined in this version of the protocol is:

1 ADVERTISEMENT

All other values are currently unknown, and if a packet is received with a value not listed, it should be discarded.

5.3.5 Authentication Type

The authentication type field identifies the authentication method being utilized. The current supported authentications are listed below:

0 - No authentication
1 - Simple text authentication
2 - IP Security Option Authentication

For simple text authentication any VRRP packet with an authentication string that does not match its configured authentication string should be discarded.

The authentication type field is an 8 bit number and must be one of the above listed values.

5.3.5.1 IP Security Option Authentication

When authentication is performed by using the IP Authentication Header as specified in [AUTH], the Authentication type should be set to "2". If packet is received with the Authentication type set to "2" indicating IP security option authentication and no authentication header is present in the packet, the packet should be discarded.

5.3.6 Advertisement Interval (Adver Int)

This field is the time interval for Master to Send ADVERTISEMENTS. Default is 1 second. This field is used for troubleshooting misconfigured routers.

5.3.7 Checksum

The checksum field is used to detect data corruption in the VRRP message.

The checksum is the 16-bit one's complement of the one's complement

sum of the entire VRRP message starting with the version field. For computing the checksum, the checksum field is set to zero.

5.3.8 Virtual IP address

The virtual IP address field specifies the Virtual IP (VIP) address associated with the particular cluster. This field is used for troubleshooting misconfigured routers.

The VIP should be an IP address assigned from the subnet that the interface is attached.

5.3.9 Authentication Data

The authentication string is currently utilized for simple text authentication, similar to the simple text authentication found in OSPF. It is up to 8 characters of plain text. If the configured authentication string is shorter than 8 bytes, the remaining space MUST be zero-filled. Any VRRP packet with an authentication string that does not match its configured authentication string should be discarded. The authentication string is unique on a per cluster basis.

6. Protocol State Machine

6.1 Parameters

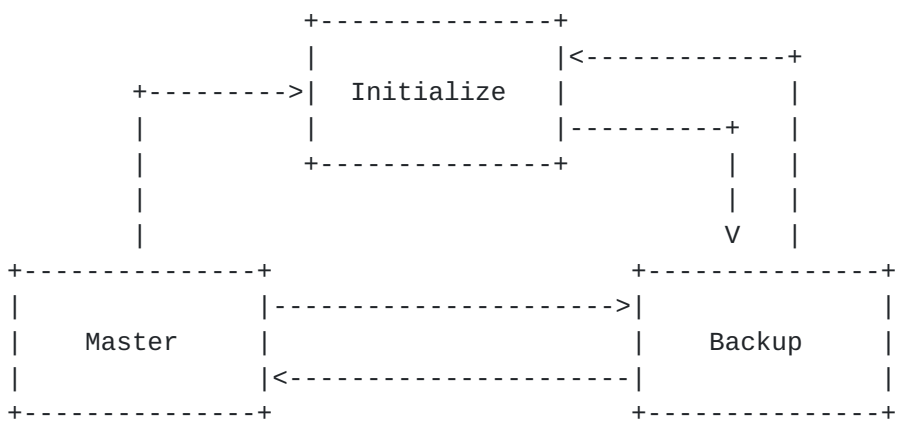
Cluster_ID	Cluster identifier. Configured item.
Priority	Priority value for this cluster. Configured item. Default is 100 (decimal).
Virtual_IP	Virtual IP Address for this cluster. Configured item.
Advertisement_Interval	Time interval for Master to Send ADVERTISEMENTS. Default is 1 second.
Skew_Time	Calculated time to skew Master_Down_Interval. Defined to be: ((256 - Priority) / 256)

Master_Down_Interval	Time interval for Backup to declare Master down. Defined to be: (3 * Advertisement_Interval) + Skew_time seconds.

6.2 Timers

Master_Down_Timer	Timer which fires when Master has not been heard for Master_Down_Interval.
Adver_Timer	Timer which fires when time to send next ADVERTISEMENT based on Advertisement_Interval.

6.3 State Transition Diagram



6.4 State Descriptions

In the below state descriptions, the state names will be identified as follows {state-name}, and the packets will be identified by utilizing all upper case characters.

6.4.1 Initialize

{Initialize} is the initial state an interface takes when VRRP is enabled or disabled. The basic function of the state is to wait for a startup event. When that is received it:

- Set the Master_Down_Timer to Master_Down_Interval

- Set state to {Backup} state.

6.4.2 Backup

The main purpose of {Backup} state is for an interface to wait for the current master to stop sending ADVERTISEMENT packets.

While in this state, an interface should do the following:

- Should not respond to ARP request for the interface VIP router address

- Should discard packets with destination link layer MAC address equal to virtual router MAC.

- Should discard packets addressed to the interface VIP address.

- If Master_Down_Timer fires, Send ADVERTISEMENT, set Adver_Timer to Advertisement_Interval, and set state to {Master} state

- If ADVERTISEMENT received,

    If Priority of the received ADVERTISEMENT is Zero, then set Mater_Down_Timer to Skew_Time.

    If Priority of the received ADVERTISEMENT is greater than this interfaces Priority, then reset Master_Down_Timer.

    If Priority of the received ADVERTISEMENT is equal to this interfaces Priority, then reset Master_Down_Timer.

    If Priority of the received ADVERTISEMENT is lower than this interfaces Priority, then discard ADVERTISEMENT.

6.4.3 Master

In {Master} state an interface is functioning as the actual physical router for the virtual router IP and MAC address.

While in this state, an interface should do the following:

- Accept and forward traffic for the virtual router MAC address.

- Respond to ARP requests for the VIP address with the virtual router MAC address.

- Respond to packets addressed to the VIP address.

- If Adver_Timer fires, send a ADVERTISEMENT and reset Adver_Timer.

- If ADVERTISEMENT received,

  If Priority of the received ADVERTISEMENT is higher than this interfaces Priority, then cancel Adver_Timer, Set Master_Down_Timer, and set state to {Backup}.

  If Priority of the received ADVERTISEMENT is equal to this interfaces Priority, then:

    If IP Address of sender of ADVERTISEMENT is higher than this interfaces IP Address, then cancel Adver_Timer, Set Master_Down_Timer, and set state to {Backup}.

    If IP Address of sender of ADVERTISEMENT is lower than this interfaces IP Address, discard ADVERTISEMENT.

  If Priority of the received ADVERTISEMENT is lower than this interfaces Priority, discard ADVERTISEMENT.

6.5 State Table

Current State-> / Event	{Initialize}	{Backup}	{Master}
Startup	Set Master_ Down_Timer State = Backup		
Shutdown	Ignore Event	Cancel Master_ Down_Timer State = Initialize	Cancel Adver_ Timer Send ADVER w/ Priority=0 State = Init.
Master_Down_ Timer fires		Send ADVERTISEMENT Set Adver_ Timer State = Master	

Adver_Timer fires			Send ADVER. Reset Adver_ Timer
Receive VRRP ADVERTISEMENT with Priority equal Zero		Set Master_ Down_Timer= Skew_Timer	Send ADVER. Reset Adver_ Timer
Receive VRRP ADVERTISEMENT with Higher Priority		Reset Master_Down_ Timer	Cancel Adver_ Timer Set Master__ Down_Timer State = Backup
Receive VRRP ADVERTISEMENT with Equal Priority and Higher IP Address		Reset Master_Down_ Timer	Cancel Adver_ Timer Set Master__ Down_Timer State = Backup
Receive VRRP ADVERTISEMENT with Equal Priority and Lower IP Address		Reset Master_Down Timer	Discard Packet
Receive VRRP ADVERTISEMENT with Lower Priority		Discard Packet	Discard Packet
Receive ARP Request for VIP address		Discard Packet	Send ARP Reply w/ VMAC
Receive IP packet w/ Destination = VIP			Process as Normal IP Packet sent to Router
Receive IP packet w/ Dest. MAC = VMAC			Process and Forward as Normal IP Packet

Unknown VRRP packet		Discard Packet	Discard Packet

7. Sending and Receiving VRRP Packets

7.1 Receiving VRRP Packets

The following rules must be performed when a VRRP packet is received:

- Verify TTL = 255.
- Verify that received packet length is greater or equal to VRRP header length.
- Verify checksum in packet
- Verify version
- Verify Source address does not equal interface IP address
- Verify Cluster identifier valid on received interface
- Perform indicated authentication
- Verify VIP in packet is same as configured VIP for this cluster
- Verify Adver Interval in packet is same as configured VIP for this cluster

If one of these checks fails, the receiver should discard the packet, log the event and indicate via network management that an error occurred.

7.2 Transmitting Packets

The following operations must be performed prior to transmitting a VRRP packet.

- Fill in packet fields with appropriate interface and cluster information
- Compute Checksum
- Set source MAC to Virtual MAC Address
- Send to VRRP IP Multicast Group

Note: VRRP packets are transmitted with the Virtual MAC address as the source MAC to ensure that learning bridges correctly determine the LAN segment the virtual MAC is attached to.

7.3 Virtual MAC Address

The default virtual MAC address associated with the virtual IP address is a IEEE 802 MAC Address of the following format:

00-00-5E-XX-XX-{cluster id} (in hex in internet standard bit-order)

The first three octets are the IANA's OUI. The next two octets (to be assigned by the IANA) indicate the address address block assigned to the VRRP protocol. {cluster id} in the last octet is the VRRP cluster identifier. This mapping allows for up to 255 VRRP clusters per interface.

Implementations may also allow Virtual MAC addresses to be configured for each cluster.

8. Host Operation

8.1 Host ARP Requests

When a client sends a ARP request for the virtual IP address, the appropriate master router should respond to the ARP request with the above virtual MAC address for the appropriate cluster. This allows the client to always use the same MAC address regardless of the current master router. The request should be handled as a standard ARP reply.

9. Operational Issues

9.1 ICMP Redirects

VRRP operation relies on the client host only using the Virtual IP address and corresponding Virutal MAC. It is important that client hosts do not learn the real IP address of VRRP routers on LAN segment. Consequentially routers on the same LAN segment MUST NOT send ICMP Redirects with the real IP address of any VRRP routers.

9.2 Proxy ARP

If Proxy ARP is being used on routers running VRRP, the VRRP routers must advertise the Virtual MAC address in the Proxy ARP message. Doing otherwise would cause them to learn the real IP address of the VRRP routers.

9.3 Network Management

It is important that network management tools (e.g., SNMP, Telnet, etc.) always use the real IP addresses of VRRP routers. This is necessary to insure that network management is aware of the real status of the VRRP routers (e.g., detect that a router has failed so that it can be repaired).

10. Operation over Token Ring

TBD

11. References

[AUTH] Atkinson, R., "IP Authentication Header", RFC 1826, Naval Research Laboratory, August 1995.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC2119, BCP14, March 1997.

12. Security Considerations

The protocol design supports no authentication, simple text authentication, and integrity/authentication/integrity using the IP Security options.

13. Author's Addresses

Steven Knight
Ascend Communications
High Performance Network Division
10250 Valley View Road, Suite 113
Eden Prairie, MN USA 55344
Phone: +1 612 943-8990
EMail: Steven.Knight@ascend.com

Douglas Weaver
Ascend Communications
High Performance Network Division
10250 Valley View Road, Suite 113
Eden Prairie, MN USA 55344
Phone: +1 612 943-8990
EMail: Doug.Weaver@ascend.com

David Whipple
Microsoft Corporation
One Microsoft Way
Redmond, WA USA 98052-6399
Phone: +1 206 703-3876
EMail: dwhipple@microsoft.com

Robert Hinden
Ipsilon Networks, Inc.
232 Java Drive
Sunnyvale, CA 94089
Phone: +1 408 990-2004
EMail: hinden@ipsilon.com

14. Acknowledgments

The authors would like to thank Glen Zorn, and Michael Lane, Clark Bremer, Hal Peterson, Danny Mitzel, and Peter Hunt for their comments and suggestions.

15. Changes from Previous Drafts

Changes from <draft-hinden-vrrp-00.txt>

- Changed default behavior to stay with current master when priorities are equal. This behavior can be changed by configuring explicit priorities.

- Changed Master state behavior to not send Advertisements when receiving Advertisement with lower priorty. Change reduces worst case election message overhead to "n", where "n" is number of configured equal priority VRRP routers.

- Added Skew_Time parameter and changed receiving advertisement with zero priority behavior to cause resulting advertisement sent to be skewed by priority.

- Changed sending behavior to send VRRP packets with VMAC as source MAC and added text describing why this is important for bridged environments.

- Changed definition of VMAC to be in IANA assigned unicast MAC block.

- Added Advertisement Interval to VRRP header.

- Added text regarding ICMP Redirects, Proxy ARP, and network management issues.

- Various small text clarifications.