the Availability Digest

VRRP – Virtual Router Redundancy Protocol

October 2008

The quest for continuous availability is not localized to the computer room. Mission-critical systems are unavailable if users do not have access to them. Therefore, the network interconnecting the users to their servers as well as the communication links between dispersed data-processing elements must be equally reliable.

This means, of course, that all network paths must be redundant so that there is a backup path to route around any component that should fail.[1] Furthermore, failover to a backup path should be very fast so that users are not inconvenienced. In large IP networks, high reliability is achieved typically by dynamic routing.[2] Using protocols such as RIP (Routing Information Protocol), network routers continually keep their neighbors informed about their connectivity so that each router can maintain a map of the network. Thus, even in the presence of network reconfigurations and failures, traffic can be routed around problem areas to reliably reach its destination.

However, dynamic routing protocols are complex and impose a significant burden on the network routing components. In addition, it may take many minutes to discover a new network topology. As a result, it is impractical to carry these techniques all the way back to clients and servers (especially users' laptops and desktops). So how at the first-hop level, where packets are trying to make their way from users to the larger network, can redundancy be achieved?

This is the role of the Virtual Router Redundancy Protocol (VRRP). VRRP provides virtual routers comprising multiple physical routers with a common IP address so that first-hop routing survives in the presence of a physical-router failure. It does so in seconds and with complete transparency to the users and servers that the virtual router is supporting. In addition, the physical routers comprising a virtual-router group can load-share the network traffic routed to the virtual router.

What Does VRRP Do?

VRRP provides transparent recovery from a router failure at the *first-hop* level. It is therefore designed, for instance, to support traffic being sent between a local LAN and an external network. To understand the benefits that VRRP brings, let us look at how we might implement router redundancy without VRRP by using redundant routers without virtualization.

Backup via Redundant Routers

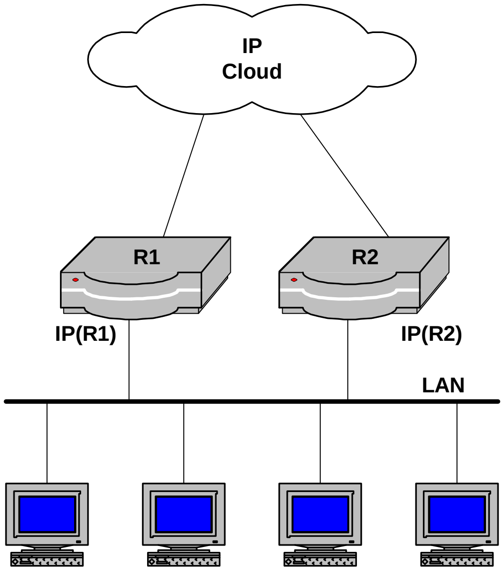
To protect against a router failure, one solution is to provide a pair of routers, one acting as the primary router and the other as a backup router. Each has its own IP address that may or may

[1] W. H. Highleyman, P. J. Holenstein, B. D. Holenstein, Chapter 5, Redundant Reliable Networks, *Breaking the Availability Barrier II: Achieving Century Uptimes with Active/Active Systems*, AuthorHouse; 2007.
[2] W. R. Stevens, Chapter 10, Dynamic Routing Protocols, *TCP/IP Illustrated: Volume 1 – The Protocols*, Addison-Wesley; 1994.

not be known by the hosts supported by the routers (we use the term *host* to include user PCs and workstations as well as the servers with which they are communicating).

For instance, let us designate the routers in a primary/backup pair as Router R1 and Router R2. Router R1 has an IP address of IP(R1), and Router R2 has an IP address of IP(R2). Normally, all hosts on the LAN are sending their traffic to IP address IP(R1); and this traffic is forwarded by Router R1. However, should Router R1 fail, the hosts will switch their outgoing traffic to Router R2. Router R2 now forwards the traffic from the LAN.

Conceptually, this is straightforward. However, the problem is how does a host know when a router has failed; and how does it know what the IP address is to which it should switch? It can detect a router failure when it fails to get a response to a request. However, response timeout might have to be a minute or so; and how is it known that this is a router problem and not a remote server problem?



Redundant Routers

Furthermore, once a host has decided that there is a router failure, how does it know to which IP address to switch? Two solutions to this problem are for each host to already know the two IP addresses of the primary/backup pair or for a network administrator to reconfigure the hosts to use the backup IP address.

The first solution requires that some network topology information be embedded in each host. This is not a normal capability for thin clients. It will have to be added as a plug-in to each user's browser, which will be a management nightmare in a large network.

The second solution may result in switchover times that will be painfully long. First, the network manager must realize that there is a problem. He must then analyze the situation and decide that a failover is appropriate. Finally, he must issue a reconfiguration command to the routers. This could take many minutes.

VRRP solves these problems. It automatically detects a primary router failure in just a few seconds (typically three to four seconds) and automatically routes all traffic to the backup router. This is all transparent to the clients and servers which VRRP supports.

Backup via Virtual Routers

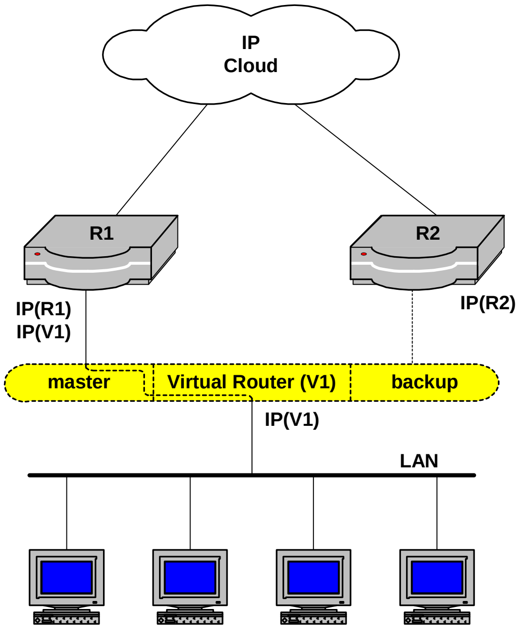
Using VRRP, the tasks of detecting a router failure and switching over to the backup router are handled within the routers themselves. To clients and servers on the LAN, the primary/backup router pair appears as a single router with a single IP address. The hosts on the LAN are completely unaware that they are not talking to a single router.

For instance, as in the previous example, primary Router R1 and backup router R2 will have physical IP addresses IP(R1) and IP(R2), respectively. However, as a virtualized pair, a virtual router V1 can be created with IP address IP(V1). When R1 is operational, all traffic routed to IP(V1) will be forwarded by Router R1. Should Router R1 fail, all traffic routed to IP(V1) will instead be forwarded by Router R2. The failover is fast and is transparent to the hosts on the LAN.

The virtual IP address IP(V1) is the real physical address of the router that is to be the primary router. In this example, IP(V1) will be equal to IP(R1). Router R1 is known as the *owner* of the virtual router.

In VRRP, the router which is currently active and carrying the communication load is known as the *master*. Ownership leads to some special characteristics in a virtual router. For one, it is the owner that always becomes the master router at startup.

For another, if the owner fails and is returned to service, master status is always returned to the owner.



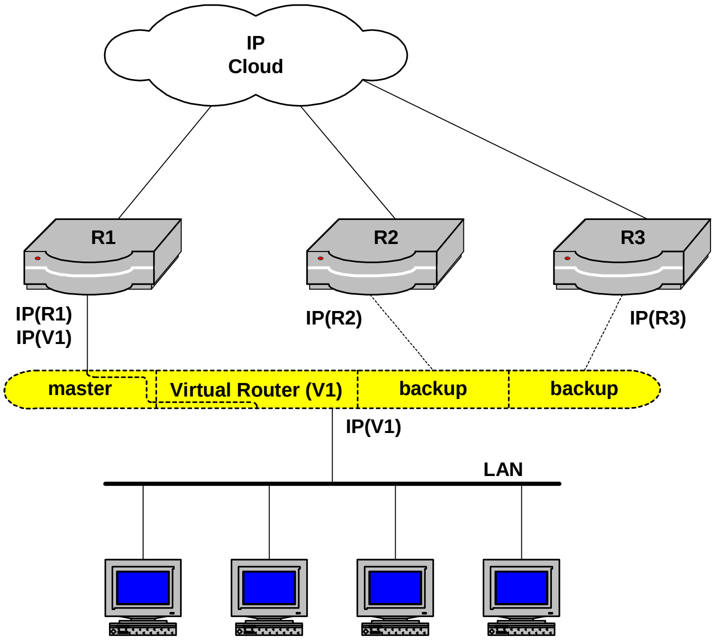
VRRP Redundant Routers

Multiple Virtual Backups

In a similar way, VRRP can support multiple backups. Let us add another backup router to our configuration – Router R3 with physical IP address IP(R3). In normal operation, traffic routed to the virtual address IP(V1) will be forwarded by master Router R1.

Should it fail, one of the backup routers, backup Router R2 or backup Router R3, will take over the master role. VRRP determines which backup router to promote to master based on a priority assignment made to all routers, as described later. Should the new master router fail, the remaining backup router will assume the role of master.

This VRRP configuration can be extended to any number of backup routers.[3]



VRRP Router with Multiple Backups

Recovery from multiple failures is a little more involved than in the case of the dual-router configuration described above and depends upon ownership. The owner, if operational, is always the master. Let us call the current master the *incumbent* and the newly repaired router the *contestant*. If the incumbent is the owner, then any new contestant becomes a backup. If the contestant is the owner, then it becomes master, preempting the role of master from the current incumbent.

However, if neither the incumbent nor the contestant is the owner (i.e., the owner is still down), the new master is determined by priority and preemption mode. If the virtual router is configured with preemption not allowed, the current incumbent remains the

[3] Up to 254 backup routers, or many more than will probably ever be used.

master. However, if the configuration allows preemption, the router with the highest priority becomes the master.

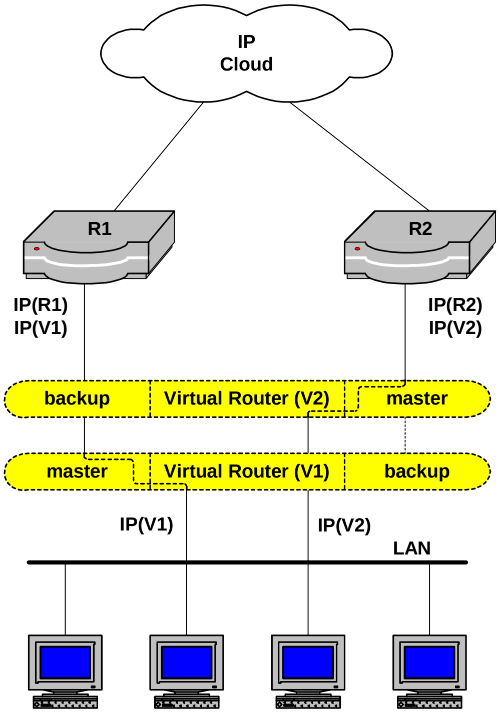
Router Load Sharing

The physical routers comprising a virtual router can share the communication load addressed to the virtual router. To do this, multiple virtual networks are configured with different virtual IP addresses but with each using some or all of the physical routers in the virtual router group.

For instance, consider a two-router virtualized pair comprising Router R1 and Router R2. One virtual router V1 is configured with IP address IP(V1) and uses Router R1 as its master and Router R2 as its backup. Another virtual router V2 is configured with IP address IP(V2). This virtual router uses Router R2 as its master and Router R1 as its backup.

The hosts on the LAN are split between these virtual routers. Some hosts are configured to use virtual router V1 and send their traffic to IP address IP(V1). The other hosts are configured to use virtual router V2 and send their traffic to IP address IP(V2).

In normal operation, Router V1 forwards all traffic addressed to IP(V1); and Router V2 forwards all traffic addressed to IP(V2). However, should a router fail, its traffic is handled by its backup router.



VRRP Load Sharing

An additional advantage of router load-sharing is that there is no idle backup router that may have suffered an undetected problem and is therefore not operational. The failure of a backup router may not be discovered until a failover is required, leading to a downed system. Since all routers in a load-sharing virtualized router are being used, it is known that they are all operational. This is similar to active/active systems, in which all processing nodes are being actively used and are known to be available if they have to assume the load of a failed processing node.

This configuration can be extended to many routers and many virtual networks. Of course, a backup router must be configured with a capacity that will allow it to handle its own communication load plus that of the master that it is backing up. Alternatively, there must be some strategy to shed noncritical load in the event of a router failure.

Other Configurations

A VRRP virtual router may be configured in other ways. For instance, one backup router might protect many primary routers (N:1); or many backup routers might cooperate in protecting many primary routers (N:M).

In addition, a single VRRP router can be configured to protect many IP addresses.

How Does VRRP Work?

VRRP is a protocol that enables the creation of redundancy between two or more routers. VRRP controls its master/backup configuration via the multicast of IP *advertising* messages from the current master to the backup router(s) in the virtual router. Except for receiving and analyzing the advertisements, a backup router is passive unless it must take over the master role.

The IP Container

The IP header used to address the VRRP virtual router by a client uses the standard IP header format[4] with one restriction. The TTL (Time to Live) field in the IP header is used to limit the number of hops that a packet can take through the network. It is decremented with each hop and is discarded should it reach zero. For instance, a TTL value of sixteen will limit a packet's life to sixteen hops through the network. This prevents a packet from circulating endlessly in a network with inadvertent loops.

A packet sent by a client to a VRRP virtual router must have a TTL value of 255 (its maximum value). Since VRRP virtual routers are only used as first-hop routers in a network, the receipt of any IP message with a TTL value less than 255 will be discarded because it has taken other hops before arriving at the virtual router. In addition to compensating for network misconfigurations, this has the added advantage of preventing some types of hacker attacks since the attack packets, if not originated locally on the router's LAN, will have gone through one or more hops before reaching the virtual router and will be discarded.

Using a TTL value of 255 has the disadvantage that a VRRP packet inadvertently sent to the network by a faulty router may circulate for a long time in the network. However, the developers of VRRP felt that it was more important to protect against hackers than it was to protect against faulty router configurations.

VRRP Advertisements

If the Protocol field in the IP header is equal to 112, the contents of the IP packet is a VRRP multicast advertisement message. Its fields contain the following information relevant to this brief description:[5]

- Priority is the rank in the election process for the VRRP routers and determines which router will be the master router. Priority values range from 1 to 254. Priority values of 0 and 255 have special meaning. The owning router always has priority 255 and is master if it is operational. A master router (owner or not) will change its priority to 0 if it is relinquishing master status to one of its backup routers.

- Authentication may be none, may require only a simple clear-text password, or may require strong authentication.

- Advertisement Interval is the time in seconds between advertisements multicast by the master router. Its default value is one second.

- IP Addresses are the one or more addresses protected by the virtual router.

Furthermore, the source and destination fields in the IP header have specified meanings:

- IP Source Address is the primary IP address of the router sending the packet.

- IP Destination Address is always 224.0.0.18, the multicast address assigned to VRRP packets.

[4] W. R. Stevens, Chapter 3, IP: Internet Protocol, *TCP/IP Illustrated: Volume 1 – The Protocols*, Addison-Wesley; 1994.
[5] A. Srikanth, A. A. Onart, Chapter 3, VRRP Messages, *VRRP: Increasing Reliability and Failover with the Virtual Router Redundancy Protocol*, Pearson Education, Inc.; 2003.

The Advertisement Process

Failover

Once every advertising interval, the master router multicasts an advertisement containing its priority. All other routers in the virtual router group (the backup routers) receive and analyze this advertisement. They take no other action unless a failover is called for.

If no advertisement is received from the master within three advertisement intervals (typically three seconds), the master is deemed to have failed. Each backup router will prepare itself to broadcast an advertisement telling the other routers that it is taking over the role of master. However, each backup router will delay its advertisement by an interval inversely proportional to its priority (with the lowest priority router, priority 1, delaying for an advertisement interval). Thus, the highest priority router will respond first and will become master. The other routers will note that a higher priority router than itself has taken over the master role and will remain backup routers.

Another failover trigger occurs when the current master sends an advertisement with a priority of zero. This signals the backup routers that a failover is to occur immediately. As with a missing advertisement, the backup router with the highest priority will respond with its advertisement first and will become the new master.

Recovery

When a router is returned to service, it returns initially as a backup router and monitors advertisements from the master router. When it receives the next advertisement, it determines whether it should become the new master.

If the restored router is the owner (priority 255), it preempts master status by multicasting its own advertisement. The current master (the incumbent) will receive the advertisement and will relinquish master status to the owner.

As described earlier, if the current master is the owner, the newly restored router will take no action and will remain a backup router.

If neither the incumbent nor the new contestant is the owner, the resulting action depends upon whether preemption has or has not been allowed in the router configuration. If it has not been allowed, no change in status takes place. However, if preemption has been allowed, the contestant will take over master status from the incumbent if it has a higher priority.

Network Management

A VRRP virtual router is configured and maintained via SNMP (Simple Network Management Protocol). It uses a MIB (Management Information Base) that has been standardized by an IETF RFC (the Internet Engineering Task Force sets standards for the Internet via Request for Comment standards).

Standardization

The VRRP standards being set by the IETF are still in draft form. The current RFC for VRRP is Virtual Router Redundancy Protocol (VRRP) RFC 3768.

However, implementations of the draft standard are available in products today.

VRRP Product Availability

VRRP is available from several router providers, including Cisco, Nortel, and Force 10. A Linux version is offered by Red Hat as part of its Cluster Suite.

VRRP is similar to HSRP (Hot Standby Router Protocol), a proprietary Cisco protocol. There may be some patent infringement issues to be resolved.

The Definitive VRRP Reference

The definitive book on VRRP is *VRRP: Increasing Reliability and Failover with the Virtual Router Redundancy Protocol*, authored by Ayikudy Srikanth and Adnan Adam Onart.[6] In addition to a clear description of the inner workings of VRRP with UML diagrams, SDL diagrams, and logical symbolism, the book delves into the details of the VRRP packet structure and the MIB structure.

It discusses the implementation of VRRP over network technologies other than IP. They include Ethernet, FDDI, Token Ring, and ATM. It also describes the considerations to be used with VRRP in environments using firewalls and VPN tunnels.

The book deals extensively with the management of VRRP virtual routers with SNMP and demonstrates this with examples of the Command Line Interface (CLI) and GUI interfaces provided by HP Openview and Nortel Networks.

It describes in detail Cisco's equivalent proprietary protocol HSRP and compares its features to those of VRRP. It looks at the future of VRRP, including its support for IPv6 and extensions for higher availability. These extensions include the ability of a router to shut down if it loses its WAN connection so that failover to a redundant WAN link serviced by a backup router can be made. Other topics include firewall synchronization, server load balancing, mobile IP, and preservation of state during failover.

Appendices include a detailed (100 page) review of TCP/IP, the MIB structure, and open source VRRP. Detailed descriptions of the VRRP protocol are provided via SDL flowcharts, logical pseudocode, and UML diagrams.

The authors are uniquely qualified to write such a book. Ayikudy Srikanth is a member of the IETF VRRP working group. He was the architect of the Nortel Bay series VRRP router and was the manager of the team that implemented it. He holds two VRRP-related patents. Adnan Adam Onart managed the implementation of the VRRP and early high-availability strategy for Nortel. Prior to that, he had been Director of Internetworking Software at Fujitsu-Nexion.

[6] Pearson Education, Inc.; 2003.