The VRRP Project

(Virtual Router Redundancy Protocol)

Design Document

Cathy.Zhou@Sun.Com

Solaris Networking

Sun Microsystems, Inc.

Revision 1.0
July 7, 2009

Contents

1 Introduction

This project is to implement the VRRP protocol (Virtual Router Redundancy Protocol) version 3 for IPv4 and IPv6 (draft-ietf-vrrp-unified-spec-02), which is based on RFC 3768, VRRP version 2 for IPv4, with the goal of providing HA (high availability) on Solaris. The project provides the VRRP service, the administrative tool which configures and manages the service, and the VRRP library interfaces that can be potentially used by the third-party software.

VRRP specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers within a LAN. At one time, only one VRRP router is controlling the IPv4 or IPv6 address(es) associated with a virtual router (Master), and it is forwarding packets sent to these IP addresses. The election process provides dynamic fail over in the forwarding responsibility should the Master become unavailable. The protocol is designed to eliminate the single point of failure inherent in the static default routed environment. The advantage gained from using VRRP is a higher availability default path without requiring configuration of dynamic routing or router discovery protocols on every end-host.

2 Requirements

Key requirements for the VRRP implementation on Solaris are outlined below:

1. Conforming to the VRRP standard
2. High availability support of network services, minimize the disruption of network services during VRRP fail over
3. Extension to the standard for use on non-router appliances
4. VRRP support on VLANs and link aggregations
5. Independence of hardware (NICs)
6. Easy administrative interfaces to deploy the service

3 VRRP Configuration

3.1 Configuration Overview

A VRRP router is a router which runs the VRRP protocol. It works with other VRRP routers participating in the same virtual router, and protects a set of virtual IP addresses.

Within a LAN, each virtual router is uniquely identified by the <VRID, address_family> tuple and is associated with a set of protected virtual IP addresses. Each participating VRRP router has additional parameters such as priority, advertisement interval and accept mode. At one time, only one VRRP router (Master) will assume the responsibility of the virtual router, and forward the packets sent to the virtual IP addresses. Whenever the Master fails, the other VRRP routers participated will detect its absence and another VRRP router will be elected as the Master and take over the responsibility.

All the VRRP routers of the same virtual router share the same VRRP virtual MAC address. The virtual MAC address is calculated based on the address family and the VRID of the virtual router (in hex in internet standard bit-order):

```
IPv4: 00-00-5E-00-01-{VRID}
IPv6: 00-00-5E-00-02-{VRID}
```

Therefore, a special VRRP VNIC with the virtual MAC address must be first created in order for the VRRP router to work properly. All the IP addresses reside on this VNIC are regarded

as virtual IP addresses protected by the VRRP router. Those virtual IP addresses will be staying down at the Backup router and will be brought up when the router becomes Master, thus providing the high availability for these virtual IP addresses.

3.2 VRRP VNIC Creation

The existing `dladm create-vnic` subcommand will be extended to create the VRRP VNIC:

```
dladm create-vnic [-t] [-R <root-dir>] [-l <link>] [-m vrrp -V <VRID> -A
[inet | inet6]] [-v <vlan-id>] [-p <prop>=<value>[,...]] <vnic-link>
```

A new VNIC address type "`vrrp`" is introduced, and VRID and address family need to be specified with this new VNIC address type.

As a result, a VNIC with the well-known virtual router MAC address will be created.

3.3 `vrrpadm` Configuration

The following is a summary of the `vrrpadm` subcommands, and the detail can be found in the `vrrpadm` manpage. All the below subcommands will be persistent except `vrrpadm show-router`. For example, the VRRP router created by "`vrrpadm create-router`" will persist across reboot.

3.3.1 `vrrpadm create-router`

```
vrrpadm create-router -V <vrid> -l <link> -A [inet | inet6] [-a
<assoc_ipaddrs>][-P <primary_ipaddr>] [-p <priority>] [-i
<adv_interval>] [-o <flags>] [-f]<router_name>
```

The `vrrpadm create-router` subcommand creates a VRRP router of the specified `VRID` and address family with the given parameters. As we described in section 3.1, each VRRP router requires a special VRRP VNIC to be created, and the VNIC can be created by "`dladm create-vnic`". Alternatively, administrators can also specify the '`-f`' option to explicitly request the VNIC to be created and plumbed as part of the "`vrrpadm create-router`" operation. In this case, `vrrpadm` will first check the existence of the VNIC with the given `VRID` and address family, and create/plumb the VNIC if necessary. The name of VNICs will be generated internally in the form of `v<vrid>_<link>_v<4|6>`.

If "`-a`" is specified, `vrrpadm` will configure the specified IP address over the VRRP VNIC, and those IP addresses (together with other IP addresses configured over that VNIC) will be regarded as the virtual IP addresses that are protected by the VRRP router.

If "`-P`" is specified, `vrrpadm` will configure and bring up the given IP address over the `<link>` IP interface, and this IP addresses will participate the VRRP primary IP address selection process (see 5.3) and will potentially be used as the VRRP primary IP address that is used to send the VRRP advertisement.

In the case of IPv6, only link-local IP addresses can be specified by the "`-P`" option, because only link-local IP address will be selected as the primary IP address. Further, in the general case, specifying the "`-P`" and "`-a`" option will not be needed for an IPv6 router, since the link-local IP addresses will be automatically configured when the underlying interface and the VRRP VNIC are plumbed, and they will be used as primary IP address and virtual IP address of the router.

The "`-o`" option will be used to configure the preempt and accept modes of the given VRRP router. Values can be: `preempt`, `no_preempt`, `accept`, `no_accept`. By default , both modes are set to `true`.

The <router_name> will be used as the unique identification of this VRRP router and used in other vrrpadm subcommands.

3.3.2 vrrpadm modify-router

```
vrrpadm modify-router [-p <priority>] [-i <adv_interval>] [-o <flags>]
      <router_name>
```

The vrrpadm modify-router subcommand will be used to change the configuration of a given VRRP router.

3.3.3 vrrpadm delete-router

```
vrrpadm delete-router <router_name>
```

The vrrpadm delete-router subcommand will be used to delete a given VRRP router.

3.3.4 vrrpadm enable-router

```
vrrpadm enable-router <router_name>
```

A VRRP router does not start to take effect until it is enabled. The underlying data-link the VRRP router is created over (specified with the -l option when the router is created by vrrpadm create-router) and router's VRRP VNIC must exist when the router is enabled, or the enable operation will fail. Further, the VNIC and the underlying data-link will be held open to prevent them being deleted or renamed unless the VRRP router is disabled (see 5.6).

3.3.5 vrrpadm disable-router

```
vrrpadm disable-router <router_name>
```

Sometimes it is useful to temporarily disable a VRRP router, so that the administrator can make some changes of the configuration and then reenable the router.

3.3.6 vrrpadm show-router

```
vrrpadm show-router [-P | -x] [-p] [-o field[,...]] [<router_name>]
```

The vrrpadm show-router subcommand will be used to show the configuration and status of a given VRRP router. See more details in vrrpadm(1M). Below shows some examples of the vrrpadm show-router output:

```
$ vrrpadm show-router vrrp1
NAME    VRID LINK    AF   PRIO ADV_INTV MODE  STATE VNIC
vrrp1   1    bge1    IPv4 100  1000     e-pa- BACK  vnic1
$ vrrpadm show-router -x vrrp1
NAME    STATE PRV_STAT STAT_LAST VNIC    PRIMARY_IP          VIRTUAL_IPS
vrrp1   BACK  MAST        1m17s  vnic1   10.0.0.100          10.0.0.1
$ vrrpadm show-router -P vrrp1
NAME    PEER                P_PRIO P_INTV   P_ADV_LAST M_DOWN_INTV
vrrp1   10.0.0.123          120    1000        0.313s  3609
```

3.4 VRRP Configuration Example

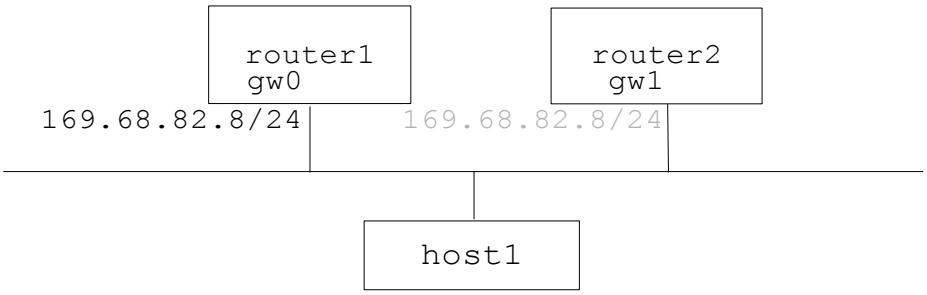

Figure 1 VRRP setup example

The above figure shows a typical VRRP setup. In the example, `169.68.82.8` is configured as the default gateway for `host1`, and this IP address is the virtual IP address protected by the virtual router composed by two VRRP routers: `router1` and `router2`. At one time, only one of the two routers behaves as the Master and assumes the responsibilities of the virtual router and forwards packets come from `host1`.

Assuming that the `VRID` of the virtual router is `12`, the following shows the configuration steps to configure the above VRRP setup on `router1` and `router2` - `router1` is the owner of the virtual IP address (`169.68.82.8`) and its priority is the default value (`255`); `router2` is the backup whose priority is `100`:

```
Router1:
# dladm create-vnic -m vrrp -V 12 -A inet -l gw0 vnic1
# vrrpadm create-router -V 12 -A inet -l gw0 vrrp1
# vrrpadm enable-router vrrp1
# ifconfig vnic1 plumb 169.68.82.8/24
# ifconfig gw0 plumb 169.68.82.100/24 up
# vrrpadm show-router -x vrrp1
NAME    STATE PRV_STAT STAT_LAST VNIC    PRIMARY_IP       VIRTUAL_IPS
vrrp1   MAST  BACK        1m17s  vnic1   169.68.82.100    169.68.82.8
Router2:
# dladm create-vnic -m vrrp -V 12 -A inet -l gw1 vnic1
# vrrpadm create-router -V 12 -A inet -l gw1 -p 100 vrrp1
# vrrpadm enable-router vrrp1
# ifconfig vnic1 plumb 169.68.82.8/24
# ifconfig gw0 plumb 169.68.82.101/24 up
# vrrpadm show-router -x vrrp1
NAME    STATE PRV_STAT STAT_LAST VNIC    PRIMARY_IP       VIRTUAL_IPS
vrrp1   BACK  INIT       2m32s   vnic1   169.68.82.101    169.68.82.8
```

Using `router1`'s configuration as the example: note that we must configure at least one IP address over `gw0`, and this IP address is selected as the primary IP address which is used to send the VRRP advertisement packets in this example:

```
# vrrpadm show-router -x vrrp1
NAME    STATE PRV_STAT STAT_LAST VNIC    PRIMARY_IP       VIRTUAL_IPS
vrrp1   MAST  BACK        1m17s  vnic1   169.68.82.100    169.68.82.8
```

In some cases, the VRRP configuration is pretty simple and the administrator would rather to have one single administrative tool to configure everything related to a VRRP router. Below shows an alternative way to set up the same VRRP router using a single command:

```
# vrrpadm create-router -V 12 -A inet -l gw0 -f -a 169.68.82.8/24 -P
169.68.82.100/24 vrrp1
# vrrpadm enable-router vrrp1
# vrrpadm show-router -x vrrp1
NAME  STATE PRV_STAT STAT_LAST VNIC          PRIMARY_IP     VIRTUAL_IPS
vrrp1 MAST  BACK     2m26s     v12_gw0_v4   169.68.82.100  169.68.82.8
```

Note that results of the above two approach are the same. In the second example, the `vrrpadm create-router` subcommand does "`create-vnic`" and "`ifconfig`" internally.

While the second configuration approach looks simpler, `vrrpadm` does not have the same flexibility as the existing data-link and IP configuration utilities. For example, the `vrrpadm` cannot specify the bandwidth limitation over the VNIC, nor specify a very virtual IP address to be "`preferred`". In those cases, the first configuration approach can be used to make some complicated configuration.

4 Architecture Overview

4.1 VRRP router

As specified in the draft, VRRP specifies an election protocol that dynamically assigns the responsibility of a virtual router to one of the VRRP routers on a LAN. The VRRP protocol

runs on each VRRP router, and manages the router's state. A host can have multiple VRRP routers configured, where each VRRP router belongs to a different virtual router.

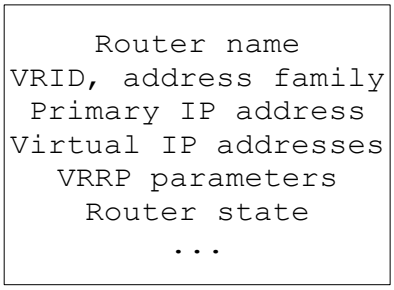

Figure 2 A VRRP Router Instance

A VRRP router consists of:

- The router name (a system-wide unique identifier)
- VRID, address family (identify a virtual router within a LAN)
- Primary IP used as the source IP address of the VRRP advertisement
- Virtual IPs to be protected
- VRRP parameters: priority, advertise interval, preempt mode, accept mode
- VRRP State information and statistics

4.2 VRRP Architecture Components

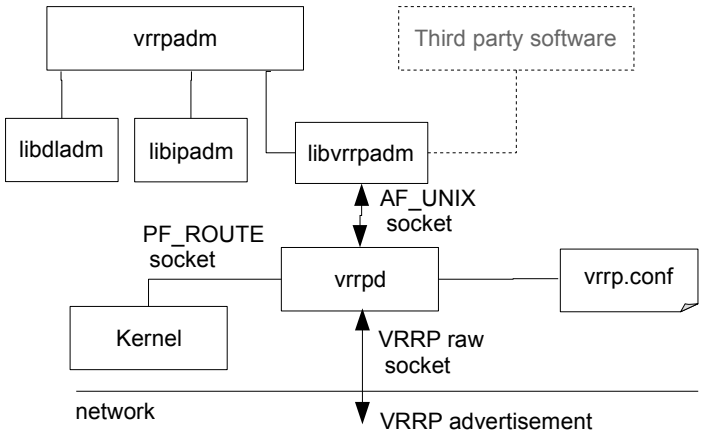

Figure 3 VRRP Architecture

The VRRP implementation includes the following major components:

- An administrative tool - `vrrpadm` – which administrators the VRRP service
- A daemon - `vrrpd` – which is controlled by the VRRP SMF service
- A private library `libvrrpadm` that can be potentially used by the third party programs in the future
- A private configuration file `vrrp.conf` for persistent configuration

- The libdladm library is used by vrrpadm to create VRRP VNICs if "-f" is specified in the create-router command.
- The libipadm will be implemented by another project and it is currently under development. It will be used by vrrpadm to plumb the interface and configure the IP addresses.

The vrrpd daemon is started by the svc:/network/vrrp:default service, which depends on the svc:/network/physical and svc:/system/filesystem/minimal services. The former service is required to configure the data-links and the IP addresses needed by the VRRP routers configuration, and the latter service is required since the vrrpd, vrrpadm and libvrrpadm binaries reside under the /usr directory, and the AF_UNIX socket file resides under the /var/run directory

Each system running VRRP service will have the network/vrrp service enabled, which starts the vrrpd daemon. The daemon maintains all the VRRP routers on the system and manages their states. It also processes the administrative requests, receives and sends the VRRP advertisements and implements the VRRP state machine.

The vrrpd daemon also maintains the list of virtual IP addresses protected by each VRRP router, and select its primary IP address which is used as the source IP address of the VRRP advertisements. In order to do that, vrrpd tracks all the IP addresses configured on the system and their up/down state using a PF_ROUTER socket (see 5.3).

The vrrpadm tool is used by the administrator to configure VRRP. It calls into the libvrrpadm library which communicates with vrrpd through a AF_UNIX socket (see 5.1).

The libvrrpadm library remain as private for now, but can be potentially make public so that third-party software can use it to interact with VRRP service.

All VRRP configuration is persisted in the a private configuration file – vrrp.conf.

5 Implementation Details

5.1 Communication Between vrrpadm and vrrpd

The vrrpadm tool is used to administrate the VRRP service. vrrpadm calls into the libvrrpadm library, which communicates with vrrpd using a AF_UNIX socket. The vrrpd daemon will process the administrative requests received from the AF_UNIX socket and send the result back to libvrrpadm using the same socket.

5.2 Sending and Receiving VRRP Advertisements

VRRP router in the Master state sends VRRP advertisements periodically as required by the VRRP protocol. VRRP router in the Backup state receives advertisements and refreshes the Master_Down_Timer. Once that timer expires, the Backup router claims the old Master is down and will becomes the Master.

To receive the VRRP advertisement packets, when a VRRP router is enabled, a IPPROTO_VRRP SOCK_RAW socket will be created on the data-link where the VRRP router is created. The same SOCK_RAW socket will be shared by all the VRRP router created on the same data-link with the same address family. Once all the VRRP routers with the same address family are deleted on a data-link, this SOCK_RAW socket will be closed. This RX socket will join the VRRP multicast group to receive the VRRP advertisement packets. Further, since the IPv6 SOCK_RAW socket does not pass up the IP header of the received packet, the IPV6_RECVPKTINFO and IPV6_RECVHOPLIMIT socket option will be set in order to get the destination IP address and the hop limit information of a specific IPv6 advertisement packet, and verify they are valid.

On the send side, to make sure the source MAC address of the VRRP advertisement packets to be the well-known VRRP virtual MAC address, a `IPPROTO_VRRP SOCK_RAW` socket will be created for each VRRP router, and the VRRP VNIC will be set as the output interface using the `IP_MULTICAST_IF` (or `IPV6_MULTICAST_IF` for IPv6) socket option. The `IP_HDRINCL` socket option will also be set on IPv4 socket and the IP header will be composed by `vrrpd` itself when the VRRP advertisements are sent. In the case of IPv6, since there is no equivalence of `IP_HDRINCL`, the `IPV6_PKTINFO` and `IPV6_HOPLIMIT` ancillary_data will be used to specify the source IP address and IP hop limit of the advertisement packets.

5.3 Virtual IP addresses Tracking and Primary IP selection

As described in section 3, all the IP addresses configured on a VRRP VNIC (with the specific address family) will be regarded as the virtual IP addresses configured on the corresponding VRRP router, regardless the services and tools used to configure the IP addresses.

To do that, `vrrpd` must track all the IP addresses configured on the system, and by determining which VNIC is associated with which VRRP router, it maintains the list of the virtual IP addresses for each specific VRRP router.

The `PF_ROUTE` socket is usually used by the user-land applications to track the set of the IP addresses and their states. Unfortunately, the `PF_ROUTE` socket in today's Solaris does not report the changes of IP addresses unless the addresses are brought up. To satisfies the needs for `vrrpd`, `PF_ROUTE` socket will be extended and two new `PF_ROUTE` event opcodes will be introduced to report the changes of the IP addresses configuration, regardless the IP address's up/down state:

- `RTM_CHGADDR`

  The `RTM_CHGADDR` event will be generated when a new IP address is newly configured (added or updated to).

- `RTM_FREEADDR`

  The `RTM_FREEADDR` event will be generated when a IP address is removed from the configuration.

The routing socket message format of both events will be the same as the format of the `RTM_NEWADDR/RTM_DELADDR` messages.

Note that both events will not report the unspecified (all-zero) IP addresses.

Further, `vrrpd` will also track all the "`UP`" IP addresses configured over the data-link that the VRRP router is configured over, and select one as the primary IP address used as the source IP address of the VRRP advertisements. In the case of IPv6, the IPv6 link-local IP address will be selected as the primary IP address.

5.4 Master/Backup and Virtual IP Addresses Management

When a VRRP router is in the Backup state, it:

- Must not respond to ARP requests for the virtual IPv4 address(s) associated with the Ipv4 virtual router;
- Must not respond to ND Neighbor Solicitation messages for the virtual IPv6 address(es) associated with the Ipv6 virtual router;
- Must not send ND Router Advertisement messages for the virtual router;
- Must discard packets with a destination link layer MAC address equal to the virtual router MAC address;
- Must not accept packets addressed to the IPvX address(es) associated with the virtual router.

When a VRRP router is in the Master state, it functions as the forwarding router for the IPvX address(es) associated with the virtual router:

- Must respond to ARP requests for the IPv4 address(es) associated with the Ipv4 virtual router.
- Must be a member of the Solicited-Node multicast address for the IPv6 address(es) associated with the Ipv6 virtual router.
- Must respond to ND Neighbor Solicitation message for the IPv6 address(es) associated with the Ipv6 virtual router.
- Must send ND Router Advertisements for the virtual router.
- If `Accept_mode` is `false`: must not drop IPv6 Neighbor Solicitations and Neighbor Advertisements.
- Must forward packets with a destination link layer MAC address equal to the virtual router MAC address.

There are several options to implement the above. For example, we could choose to configure the virtual IP addresses over the VRRP VNIC only when the router becomes Master, and unconfigure those virtual IP addresses when the router becomes Backup. Thus the Backup router will not accept any packets destined to the virtual IP addresses, nor it will send the ND Router Advertisement messages for the virtual router.

The problem with the above approach is that the application will not be able to bind to the virtual IP addresses when the router is Backup, that would requires applications to be started up “cold” after a fail-over event.

Alternatively, we choose to always configure the virtual IP addresses over the VRRP VNIC, but only bring them up when the router becomes Master, and leaves the virtual IP addresses in place on the Backup router but keep them down. This allows the applications being brought up and left running, so that when fail-over occurs, the applications are immediately ready to accept new connections. In this case, VRRP will be the only authority that manages the `up/down` state of the virtual addresses, and any other applications will not be able to bring up/down the virtual IP addresses.

Once the `vrrpd` daemon has the list of the virtual IP addresses, it manages the virtual addresses based on the state of the VRRP router. `vrrpd` will bring up the virtual IP addresses when the VRRP router becomes Master , and bring down the addresses when the router becomes Backup. In another word, `vrrpd` has full management of the `up/down` state of the virtual IP addresses and no other applications and services are allowed to change the `up/down` state of those IP addresses.

A new MAC capability `MAC_CAPAB_VRRP` will be introduced and the special VRRP VNICs will own such capability and advertise a new `DL_CAPAB_VRRP DLPI` capability as part of the `DL_CAPABILITY_REQ/ACK` negotiation process. Note that this specific capability negotiation will be done when the VNIC is first plumbed. IP will then mark the corresponding `ill` as VRRP capable and set a new `IFF_VRRP` flag on such `ill` . This new flag will indicate all the addresses over such `ill` are VRRP virtual IP addresses.

A new `SO_VRRP` socket option will also be introduced to indicate a socket is a VRRP control socket. If the `IFF_VRRP` flag is set on a specific `ill`, the `IFF_UP` flag of the associated IP addresses will only be able to be changed over a VRRP control socket, and other attempts to change its `IFF_UP` flag will fail. The `priv_sys_ip_config` privilege is required to set the `SO_VRRP` socket option.

For now, the `vrrpd` daemon will be the only application to set the `SO_VRRP` socket option on the socket used to change the `IFF_UP` flag of the virtual IP addresses, and manages the `up/down` state of the virtual IP addresses.

5.5 Preempt Mode and Zero Priority

The `preempt_mode` controls whether a (just enabled or reenabled) higher priority Backup router preempts a lower priority Master router. If `preempt_mode` is `true`, then the preemption is allowed; otherwise, the preemption is prohibited. Default it `true`. The `preempt_mode` must be `true` if the VRRP router is the owner of the virtual IP addresses.

The priority value zero has special meaning indicating that the current Master has stopped participating in VRRP. This is used to trigger Backup routers to quickly transition to Master without having to wait for the current Master to timeout. When a Backup VRRP router receives the VRRP advertisement in which the priority is zero, the Backup router immediately starts a timer (whose interval is based on its priority: `((256 - priority) * Master_Adver_Interval) / 256)`), and if no advertisement is received from other router during that time, the Backup router declares the current Master is down and becomes the new Master.

When the current Master router is disabled, a zero priority VRRP advertisement packet will be sent before the Master's state transits to `Initialize` (see 5.7).

5.6 Accept Mode Implementation

According to the standard, if `accept_mode` is `true`, the Master will accept packets destined to the virtual IP addresses. If `accept_mode` is `false`, the Master must not accept those packets, although it must forward packets and respond to ARP requests or ND Neighbor Solicitations for the virtual IP addresses. The `accept_mode` must be `true` if the VRRP router is the owner of the virtual IP addresses.

The defined behaviors is summarized as below:

	Accept packets	Forward packets	ND adv/ARP response
Master Accept	Y	Y	Y
Master No_accept	N	Y	Y
Backup	N	N	N

A new `IFF_NOACCEPT` flag will be introduced to support different `accept_mode` behavior. When a VRRP router becomes Master, if it's `accept_mode` is false, its `ill` will be set by `vrrpd` with the new `IFF_NOACCEPT` flag. IP will mark all the local `IREs` associated with the IP addresses over such `ill` to be "`no_accept`", and all the received unicast local packets will be dropped, with the exception of the Neighbor Solicitation packets and Neighbor Advertisement packets. This allows the Neighbor Unreachability Detection mechanism work as expected.

Note that the `IFF_NOACCEPT` flag can only be set on a `ill` if its `IFF_VRRP` flag is set. Further, the `IFF_NOACCEPT` flag over an `ill` can only be changed through a VRRP control socket (the `SO_VRRP` socket, see 5.4) .

5.7 VRRP Router Enabling and Disabling

According to the VRRP protocol, a VRRP router has three states (`Initialize`, `Backup` and `Master`). When a VRRP router is created, it will stay at the `Initialize` state until it is enabled. Once administrator enables a VRRP router, `vrrpd` will first check whether the VRRP VNIC exists, if not, the enabling operation will fail. The `vrrpd` daemon will then hold the VNIC and the data-link (the router is created on) open so that they will not be deleted. After that, `vrrpd` will determine the virtual IP addresses associated with the VRRP router and the primary IP address used to send the VRRP advertisement. If no virtual IP address or primary IP address can be found at that time, the VRRP router will keep staying at

the Initialize state. The vrrpd daemon will listen to the PF_ROUTE socket messages and keep track of the IP addresses configured on the system and update the virtual IP addresses list and select the primary IP address once one is brought up. When both are determined, the VRRP router will start the state machine transition and enter into either the Master state or the Backup state as defined by the VRRP protocol.

The vrrpd daemon will keep tracking of the state of all the IP addresses. When the primary IP addresses is brought down and no other primary IP addresses can be selected, or when none virtual IP addresses exists for the VRRP router, the state of the VRRP router will go back to Initialize.

The administrator can also disable a VRRP router by running the "vrrpadm disable-router" command. In that case, the state of the VRRP router will also change back to Initialize.

If the VRRP router was the Master, when it turns into the Initialize state because of any of the above reasons, it will stop sending the VRRP advertisement and the Backup VRRP router belongs to the same VRRP virtual router will eventually time out and declare Master_down, and transit to the Master state.

5.8 VRRP Sysevents

Some applications may want to get notified when there is a state transition for a specific VRRP router, so that they are able to respond accordingly. The sysevent s mechanism is used to serve this purpose. When a state transition occurs, a system event that contains the VRRP router name and the transition type (from what state to what state) will be delivered by the vrrpd daemon; interested programs can subscribe to these events and get notified.

The system event has the following properties:

- class: ES_VRRP
- subclass: ESC_VRRP_STATE_CHANGE
- vendor: SUNW
- publisher: vrrpd

An event has the following name-value pair attributes:

- vrrp_event_version (integer): "1" for current version
- vrrp_inst_name (string): the VRRP router name
- vrrp_state (integer): current state
- vrrp_prev_state (integer): previous state.

The value of the two 'state' attributes can be None, Initialized, Backup or Master. State change from None means the VRRP router is just created, and state change to None means the VRRP router is just deleted.

5.9 Router Advertisements

The VRRP protocol defines the Router Advertisement behaviors for VRRP IPv6 virtual IP addresses:

- The Master must send ND Router Advertisements for the virtual router while the Backup must not.
- When fail-over occurs, the new Master must take over the responsibility to send Router Advertisements using the same options as that of the original master.

The VRRP support itself does not assure sending the Router Advertisements for the virtual IPv6 addresses. Instead, we rely on the administrator to configure in.ndpd correctly: the RA options must be the same for all the VRRP VNICs participated in the same VRRP router.

If in.ndpd is configured to not sending Router advertisements for the virtual IPv6 addresses, the VRRP will work in the non-router mode. In that case, strictly speaking, the system no longer works as a router, but VRRP protocol can still provide the high availability for the protected virtual IP addresses.

5.10 Timers and Threading

Each VRRP router in the Master state has an advertisement timer that fires to send VRRP advertisements. Each VRRP instance in the Backup state has a Master_down_timer that is refreshed every time when an advertisement is received, and if the timer expires, the router claims Master down and starts the state transition.

To minimize the complexity, the libinetutil library is used to maintain the timers and the IO calls. The processing of received administrative requests and advertisement packets are serialized by the poll(2) call in libinetutil. Thus there will be only one single thread running in the vrrpd process and no locking and synchronizing is required.

5.11 Dynamic Reconfiguration (DR)

When a VRRP router is enabled, the underlyiny data-link and the router's VNIC will be held open to prevent them from being deleted/renamed. In order to make DR still possible with the existence of VRRP routers, a new RCM plugin will be implemented and disable/reenable the router when necessary.

5.12 Exclusive-IP Zone Support

In each exclusive-IP zone, the VRRP service (svc:/network/vrrp/default) will be enabled and will start a vrrpd daemon. The daemon will manages the VRRP router for that specific zone.

Unfortunately, VNIC cannot be created inside a non-global zone today. Thus an administrator must create the VRRP special VNIC in the global-zone first, then assign the VNIC to the non-global zone where the VRRP router resides. The VRRP router will be created and started in the non-global zone using the vrrpadm command.

Further, since sysevents are not support in an non-global zone, the VRRP sysevents will not be posted for the VRRP routers in the non-global zone.

5.13 Inter-operations with Other Network Features

VRRP service cannot work on IPMP interface. The reason is because VRRP requires specific VRRP MAC addresses while IPMP works completely in the IP layer.

Further, the VRRP virtual IP addresses can only be static configured and will not be able to be auto-configured by the two existing IP addresses auto-configuration tools - in.ndpd (IPv6 auto-configuration) and dhcpagent (DHCP configuration) . Because the Master and the Backup VRRP routers (VNICs) share the same MAC address, this will simply confuses in.ndpd and dhcpagent and eventually cause unexpected results. Therefore, IPv6 auto-configuration and DHCP configuration will not be supported over VRRP VNICs.

If an administrator configures either IPv6 auto-configuration or DHCP over a VRRP VNIC, since neither in.ndpd or dhcp client sets the SO_VRRP socket option (see section 5.4), the attempt to bring up the auto-configured IP address will fail, and fails the auto-configuration operation.

An ongoing project (ipadm) will make static IPv6 address configuration and IPv6 auto-configuration exclusive to each other on a specific IPv6 interface. Since VRRP configuration requires static IP address configuration, it will prevent in.ndpd from trying to auto-configure over a VRRP VNIC from the beginning.

5.14 Security considerations

5.14.1 Least Privilege of `vrrpd`

In the VRRP SMF manifest, the `vrrpd` will be set to be run by the "`root`" user, and its privilege property will be set to only include the "`basic`" privilege and the following privileges:

- `priv_sys_config`

  Required to post VRRP sysevents. Note that sysevents are not supported in the non-global zone, and this privilege is only needed in the global zone.

- `priv_net_rawaccess`

  Required to hold the physical data-link (which owns the primary IP address) and the VNIC (which owns the virtual IP addresses) open to prevent them from being deleted.

- `priv_net_icmpaccess`

  Required to open the `RAW` socket

- `priv_sys_ip_config`

  Required to bring up/down the virtual IP addresses and set the `SO_VRRP` socket option

5.14.2 Authorization Required by VRRP Administrators

A new `solaris.network.vrrp` authorization will be introduced and will be required to configure the VRRP service. Note that the "read-only" operation - "`vrrpadm show-router`" will not need this authorization.

The `solaris.network.vrrp` authorization will be added to the "`Network Management`" profile.

Further, when "`-f`", "`-P`" or "`-a`" is specified with the `create-router` subcommand,, `vrrpadm` will need correct privileges to create VNICs and configure primary and virtual IP addresses.