Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 11, 2017

N. Gupta
A. Dogra
Cisco Systems, Inc.
C. Docherty

G. Mirsky
Ericsson
J. Tantsura
Individual

August 10, 2016

Fast failure detection in VRRP with BFD
draft-nitish-vrrp-bfd-04

Abstract

This document describes how Bidirectional Forwarding Detection (BFD) can be used to support sub-second detection of a Master Router failure in the Virtual Router Redundancy Protocol (VRRP).

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on February 11, 2017.

Copyright Notice

Table of Contents

1. Introduction

The Virtual Router Redundancy Protocol (VRRP) provides redundant Virtual gateways in the Local Area Network (LAN), which is typically the first point of failure for end-hosts sending traffic out of the LAN. Fast failure detection of VRRP Master is critical in supporting high availability of services and improved Quality of Experience to users. In VRRP [RFC5798] specification, Backup routers depend on VRRP packets generated at a regular interval by the Master router, to detect the health of the VRRP Master. Faster failure detection can be achieved within VRRP protocol by reducing the Advertisement Interval and hold down timers. However, aggressive timers can put extra load on CPU and the network bandwidth which may not be desirable.

Since the VRRP protocol depends on the availability of Layer 3 IPv4 or IPv6 connectivity between redundant peers, the VRRP protocol can interact with the Layer 3 variant of BFD as described in [RFC5881] or [I-D.draft-ietf-bfd-multipoint] to achieve a much faster failure detection of the VRRP Master on the LAN. BFD, as specified by the [RFC5880] or [I-D.draft-ietf-bfd-multipoint] can provide a much faster failure detection in the range of 150ms, if implemented in the part of a Network device which scales better than VRRP when aggressive timers are used.

2. Requirements Language

In this document, several words are used to signify the requirements of the specification. 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. [RFC2119]

3. Applicability of Single-hop BFD

BFD for IPv4 or IPv6 (Single Hop) [RFC5881] requires that in order for a BFD session to be formed both peers participating in a BFD session need to know its peer IPv4 or IPV6 address. This poses a unique problem with the definition of the VRRP protocol, that makes the use of BFD for IPv4 or IPv6 [RFC5881] more challenging. In VRRP it is only the Master router that sends Advert packets. This means that a Master router is not aware of any Backup routers, and Backup routers are only aware of the Master router. This also means that a Backup router is not aware of any other Backup routers in the Network.

Since BFD for IPv4 or IPv6 [RFC5881] requires that a session be formed by both peers using a full destination and source address, there needs to be some external means to provide this information to BFD on behalf of VRRP. Once the peer information is made available, VRRP can form BFD sessions with its peer Virtual Router. The BFD session for a given Virtual Router is identified as the Critical Path BFD Session, which is the session that forms between the current VRRP Master router, and the highest priority Backup router. When the Critical Path BFD Session identified by VRRP as having changed state from Up to Down, then this will be interpreted by the VRRP state machine on the highest priority Backup router as a Master Down event. A Master Down event means that the highest priority Backup peer will immediately become the new Master for the Virtual Router.

NOTE: At all times, the normal fail-over mechanism defined in the VRRP [RFC5798] will be unaffected, and the BFD fail-over mechanism will always resort to normal VRRP fail-over.

This draft defines the mechanism used by the VRRP protocol to build a peer table that will help in forming of BFD session and the detection of Critical Path BFD session. If the Critical Path BFD session were to go down, it will signal a Master Down event and make the most preferred Backup router as the VRRP Master router. This requires an extension to the VRRP protocol.

This can be achieved by defining a new type in the VRRP Advert packet, and allowing VRRP peers to build a peer table in any of the operational state, Master or Backup.

3.1. Extension to VRRP protocol

In this mode of operation VRRP peers learn the adjacent routers, and form BFD session between the learnt routers. In order to build the peer table, all routers send VRRP Advert packets whilst in any of the operational states (Master or Backup). Normally VRRP peers only send Advert packets whilst in the Master state, however in this mode VRRP Backup peers will also send Advert packets with the type field set to BACKUP ADVERTISEMENT type defined in Section 3.3 of this document. The VRRP Master router will still continue to send packets with the Advert type as ADVERTISEMENT as defined in the VRRP protocol. This is to maintain inter-operability with peers complying to VRRP protocol.

Additionally, Advert packets sent from Backup Peers must not use the Virtual router MAC address as the source address. Instead it must use the Interface MAC address as the source address from which the packet is sent from. This is because the source MAC override feature is used by the Master to send Advert packets from the Virtual Router MAC address, which is used to keep the bridging cache on LAN switches and bridging devices refreshed with the destination port for the Virtual Router MAC.

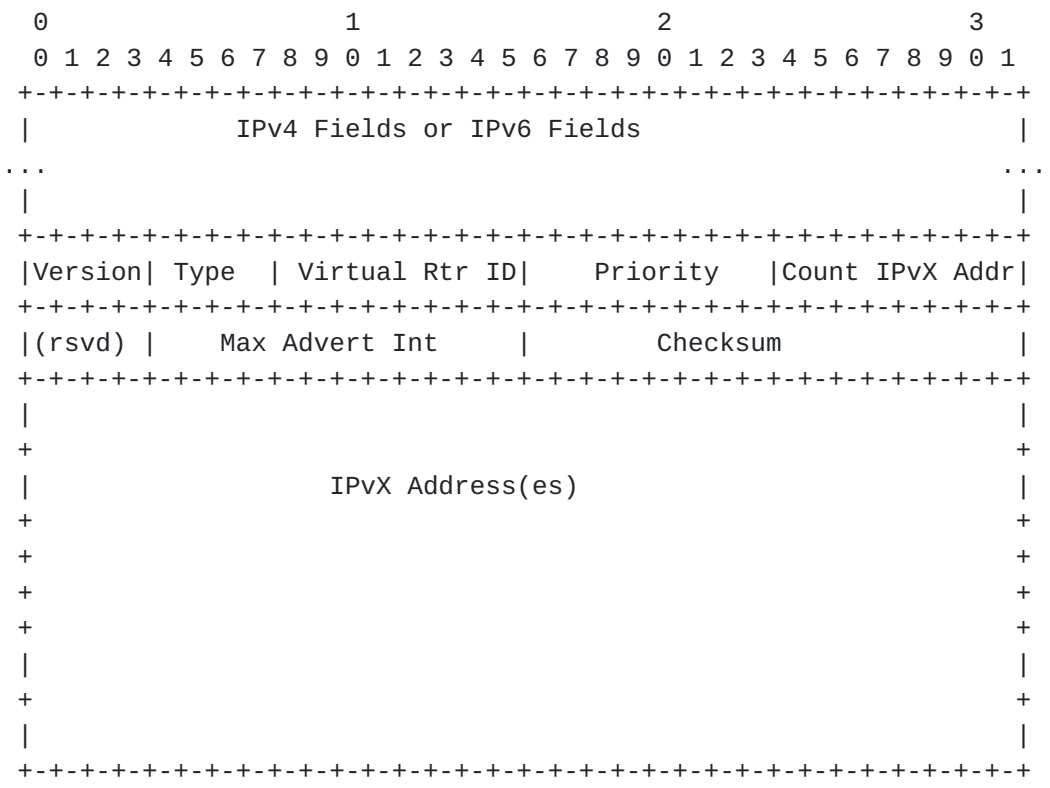
3.2. VRRP Peer Table

VRRP peers can now form the peer table by learning the source address in the ADVERTISEMENT or BACKUP ADVERTISEMENT packet sent by VRRP Master or Backup peers. This allows peers to create BFD sessions with other operational peers.

A peer entry should be removed from the peer table if Advert is not received from a peer for a period of (3 * the Advert interval).

3.3. VRRP BACKUP ADVERTISEMENT Packet Type

The following figure shows the VRRP packet as defined in VRRP [RFC5798] RFC.



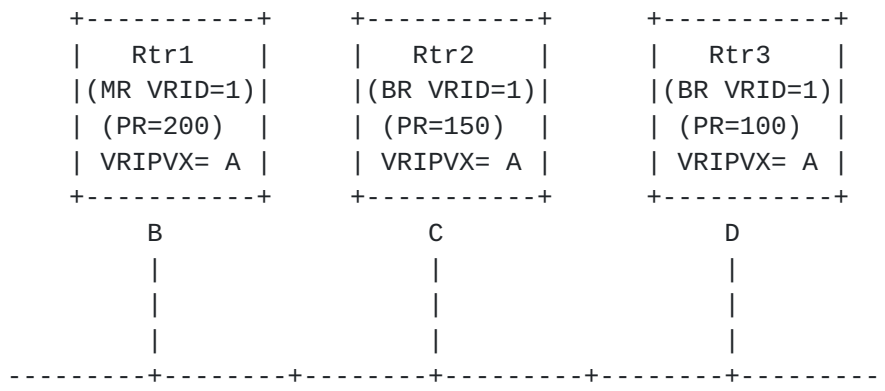
The type field specifies the type of this VRRP packet. The type field can have two values. Type 1 (ADVERTISEMENT) is used by the VRRP Master Router. Type 2 (BACKUP ADVERTISEMENT) is used by the VRRP Backup router. This is to distinguish the packets sent by the VRRP backup Router. VRRP Backup fills Backup_Advertisement_Interval in the Max Advert Int of BACKUP ADVERTISEMENT packet. Rest of the fields in Advert packet remain the same.

1 ADVERTISEMENT
2 BACKUP ADVERTISEMENT

A packet with unknown type MUST be discarded.

3.4. Sample configuration

The following figure shows a simple network with three VRRP routers implementing one virtual router.



Legend:

```
        ---+---+---+--    =  Ethernet, Token Ring, or FDDI
                    MR    =  Master Router
                    BR    =  Backup Router
                    PR    =  VRRP Router priority
                    VRID  =  VRRP Router ID
                    VRIPVX=  IPv4 or IPv6 address protected by
                             the VRRP Router
                    B,C,D =  Interface IPv4 or IPv6 address of
                             the Virtual Router
```

In the above configuration there are three routers on the LAN protecting an IPv4 or IPv6 address associated to a Virtual Router ID 1. Rtr1 is the Master router since it has the highest priority compared to Rtr2 and Rtr3. Now if peer learning extension is enabled on all the peers. Rtr1 will send the Advert packet with type field set to 1. While Rtr2 and Rtr3 will send the Advert packet with type field set to 2. In the above configuration the peer table built at each router is shown below:

Rtr1 Peer table

Peer Address	Priority
C	150
D	100

Rtr2 Peer table

Peer Address	Priority
B	200
D	100

Rtr3 Peer table

Peer Address	Priority
B	200
C	150

Once the peer tables are formed, VRRP on each router can form a BFD sessions with the learnt peers.

3.5. Critical BFD session

The Critical BFD Session is determined to be the session between the VRRP Master and the next best VRRP Backup. Failure of the Critical BFD session indicates that the Master is no longer available and the most preferred Backup will now become Master.

In the above example the Critical BFD session is shared between Rtr1 and Rtr2. If the BFD Session goes from Up to Down state, Rtr2 can treat it as a Master down event and immediately assume the role of VRRP Master router for VRID 1 and Rtr3 will become the critical Backup. If the priorities of two Backup routers are same then the primary IPvX Address of the sender is used to determine the highest priority Backup. Where higher IPvX address has higher priority.

3.6. Protocol State Machine

3.6.1. Parameters Per Virtual Router

Following parameters are added to the VRRP protocol to support this mode of operation.

Backup_Advertisement_Interval	Time interval between BACKUP ADVERTISEMENTS (centiseconds). Default is 100 centiseconds (1 second).
Backup_Adver_Interval	Advertisement interval contained in BACKUP ADVERTISEMENTS received from the Backup (centiseconds). This value is saved by virtual routers used, to compute Backup_Down_Interval.
Backup_Down_Interval	Time interval for VRRP instance to declare Backup down (centiseconds). Calculated as (3 * Backup_Adver_Interval) for each VRRP Backup.
Critical_Backup	Procedure outlined in section 3.4 of this document is used to determine the Critical_Backup at each VRRP Instance.
Critical_BFD_Session	The Critical BFD Session is the session between the VRRP Master and Critical_Backup.

3.6.2. Timers

Following timers are added to the VRRP protocol to support this mode of operation.

Backup_Down_Timer	Timer that fires when BACKUP ADVERTISEMENT has not been heard from a backup peer for Backup_Down_Interval.
Backup_Adver_Timer	Timer that fires to trigger sending of BACKUP ADVERTISEMENT based on Backup_Advertisement_Interval.

3.6.3. VRRP State Machine with BFD

Following State Machine replaces the state Machine outlined in section 6.4 of the VRRP protocol [RFC5798] to support this mode of operation. Please refer to the section 6.4 of [RFC5798] for State description.

3.6.3.1. Initialize

Following state machine replaces the state machine outlined in section 6.4.1 of [RFC5798]

(100) If a Startup event is received, then:

(105) - If the Priority = 255 (i.e., the router owns the IPvX address associated with the virtual router), then:

(110) + Send an ADVERTISEMENT

(115) + If the protected IPvX address is an IPv4 address, then:

(120) * Broadcast a gratuitous ARP request containing the virtual router MAC address for each IP address associated with the virtual router.

(125) + else // IPv6

(130) * For each IPv6 address associated with the virtual router, send an unsolicited ND Neighbor Advertisement with the Router Flag (R) set, the Solicited Flag (S) unset, the Override flag (O) set, the target address set to the IPv6 address of the virtual router, and the target link-layer address set to the virtual router MAC address.

(135) +endif // was protected addr IPv4?

(140) + Set the Adver_Timer to Advertisement_Interval

(145) + Transition to the {Master} state

(150) - else // rtr does not own virt addr

(155) + Set Master_Adver_Interval to Advertisement_Interval

(160) + Set the Master_Down_Timer to Master_Down_Interval

(165) + Set Backup_Adver_Timer to Backup_Advertisement_Interval

(170) + Transition to the {Backup} state

(175) -endif // priority was not 255

(180) endif // startup event was recv

3.6.3.2. Backup

Following state machine replaces the state machine outlined in section 6.4.2 of [RFC5798]

(300) While in this state, a VRRP router MUST do the following:

(305) - If the protected IPvX address is an IPv4 address, then:

(310) + MUST NOT respond to ARP requests for the IPv4 address(es) associated with the virtual router.

(315) - else // protected addr is IPv6

(320) + MUST NOT respond to ND Neighbor Solicitation messages for the IPv6 address(es) associated with the virtual router.

(325) + MUST NOT send ND Router Advertisement messages for the virtual router.

(330) -endif // was protected addr IPv4?

(335) - MUST discard packets with a destination link-layer MAC address equal to the virtual router MAC address.

(340) - MUST NOT accept packets addressed to the IPvX address(es) associated with the virtual router.

(345) - If a Shutdown event is received, then:

(350) + Cancel the Master_Down_Timer.

(355) + Cancel the Backup_Adver_Timer.

(360) + Cancel Backup_Down_Timers.

(365) + Remove Peer table.

(370) + If Critical_BFD_Session Exists:

(375) * Tear down the Critical_BFD_Session.

(380) + endif // Critical_BFD_Session Exists?

(385) + Send a BACKUP ADVERTISEMENT with Priority = 0.

(390) + Transition to the {Initialize} state.

```
(395) -endif // shutdown recv

(400) - If the Master_Down_Timer fires or
        If Critical_BFD_Session transitions from UP to DOWN, then:

   (405) + Send an ADVERTISEMENT

   (415) + If the protected IPvX address is an IPv4 address, then:

      (420) * Broadcast a gratuitous ARP request on that interface
      containing the virtual router MAC address for each IPv4
      address associated with the virtual router.

   (425) + else // ipv6

      (430) * Compute and join the Solicited-Node multicast
      address [RFC4291] for the IPv6 address(es) associated with
      the virtual router.

      (435) * For each IPv6 address associated with the virtual
      router, send an unsolicited ND Neighbor Advertisement with
      the Router Flag (R) set, the Solicited Flag (S) unset, the
      Override flag (O) set, the target address set to the IPv6
      address of the virtual router, and the target link-layer
      address set to the virtual router MAC address.

   (440) +endif // was protected addr ipv4?

   (445) + Set the Adver_Timer to Advertisement_Interval.

   (450) + If the Critical_BFD_Session exists:

      (455) @ Tear Critical_BFD_Session.

    (460) + endif // Critical_BFD_Session exists

   (465) + Calculate the Critical_Backup.

   (470) + If the Critical_Backup exists:

      (475) * BootStrap Critical_BFD_Session with the
              Critical_Backup.

   (480) + endif //Critical_Backup exists?

   (485) + Transition to the {Master} state.

(490) -endif // Master_Down_Timer fired
```

(485) - If an ADVERTISEMENT is received, then:

(490) + If the Priority in the ADVERTISEMENT is zero, then:

(495) * Set the Master_Down_Timer to Skew_Time.

(500) * If the Critical_BFD_Session exists:

(505) * Tear Critical_BFD_Session with the Master.

(510) * endIf // Critical_BFD_Session exists

(515) + else // priority non-zero

(520) * If Preempt_Mode is False, or if the Priority in the ADVERTISEMENT is greater than or equal to the local Priority, then:

(525) @ Set Master_Adver_Interval to Adver Interval contained in the ADVERTISEMENT.

(530) @ Recompute the Master_Down_Interval.

(535) @ Reset the Master_Down_Timer to Master_Down_Interval.

(540) @ Determine Critical_Backup.

(545) @ If Critical_BFD_Session does not exists and this instance is the Critical_Backup:

(550) @+ BootStrap Critical_BFD_Session with Master.

(555) @ endif //Critical_BFD_Session exists check

(560) * else // preempt was true or priority was less

(565) @ Discard the ADVERTISEMENT.

(570) *endif // preempt test

(575) +endif // was priority zero?

(580) -endif // was advertisement recv?

(585) - If a BACKUP ADVERTISEMENT is received, then:

(590) + If the Priority in the BACKUP ADVERTISEMENT is zero,

```
                then:

          (595) * Cancel Backup_Down_Timer.

          (600) * Remove the Peer from Peer table.

       (605) + else // priority non-zero

          (610) * Update the peer table with peer information.

          (615) * Set Backup_Adver_Interval to Adver Interval
          contained in the BACKUP ADVERTISEMENT.

          (620) * Recompute the Backup_Down_Interval.

          (625) * Reset the Backup_Down_Timer to Backup_Down_Interval.

       (630) +endif // was priority zero?

       (635) + Recalculate Critical_Backup.

       (640) + If Critical_BFD_Session exists and this
       instance is not the Critical_Backup:

          (645) * Tear Down the Critical_BFD_Session.

       (650) + else If Critical_BFD_Session doesnot exists and this
       instance is the Critical_Backup:

          (655) * BootStrap Critical_BFD_Session with Master.

       (660) + endif // Critical_Backup change

    (665) -endif // was backup advertisement recv?

    (670) - If Backup_Down_Timer fires, then:

       (675) + Remove the Peer from Peer table.

       (680) + If Critical_BFD_Session does not exist:

          (685) @ Recalculate Critical_Backup.

          (690) @ If This instance is the Critical_Backup:

             (695) +@ BootStrap Critical_BFD_Session with Master.

          (700) @ endif // Critical_Backup change
```

(705) + endif // Critical_BFD_Session does not exist?

(710) -endif // Backup_Down_Timer fires?

(715) - If Backup_Adver_Timer fires, then:

(720) + Send a BACKUP ADVERTISEMENT.

(725) + Reset the Backup_Adver_Timer to Backup_Advertisement_Interval.

(730) -endif // Backup_Down_Timer fires?

(735) endwhile // Backup state

3.6.3.3. Master

Following state machine replaces the state machine outlined in section 6.4.3 of [RFC5798]

(800) While in this state, a VRRP router MUST do the following:

(805) - If the protected IPvX address is an IPv4 address, then:

(810) + MUST respond to ARP requests for the IPv4 address(es) associated with the virtual router.

(815) - else // ipv6

(820) + MUST be a member of the Solicited-Node multicast address for the IPv6 address(es) associated with the virtual router.

(825) + MUST respond to ND Neighbor Solicitation message for the IPv6 address(es) associated with the virtual router.

(830) + MUST send ND Router Advertisements for the virtual router.

(835) + If Accept_Mode is False: MUST NOT drop IPv6 Neighbor Solicitations and Neighbor Advertisements.

(840) -endif // ipv4?

(845) - MUST forward packets with a destination link-layer MAC address equal to the virtual router MAC address.

(850) - MUST accept 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. Otherwise, MUST NOT accept these packets.

(855) - If a Shutdown event is received, then:

(860) + Cancel the Adver_Timer.

(865) + Send an ADVERTISEMENT with Priority = 0,

(870) + Cancel Backup_Down_Timers.

(875) + Remove Peer table.

(880) + If Critical_BFD_Session Exists:

(885) * Tear down Critical_BFD_Session

(890) + endif // If Critical_BFD_Session Exists

(895) + Transition to the {Initialize} state.

(900) -endif // shutdown recv

(905) - If the Adver_Timer fires, then:

(910) + Send an ADVERTISEMENT.

(915) + Reset the Adver_Timer to Advertisement_Interval.

(920) -endif // advertisement timer fired

(925) - If an ADVERTISEMENT is received, then:

(930) -+ If the Priority in the ADVERTISEMENT is zero, then:

(935) -* Send an ADVERTISEMENT.

(940) -* Reset the Adver_Timer to Advertisement_Interval.

(945) -+ else // priority was non-zero

(950) -* If the Priority in the ADVERTISEMENT is greater than the local Priority,

(955) -* or

```
        (960) -* If the Priority in the ADVERTISEMENT is equal to
        the local Priority and the primary IPvX Address of the
        sender is greater than the local primary IPvX Address, then:

           (965) -@ Cancel Adver_Timer

           (970) -@ Set Master_Adver_Interval to Adver Interval
           contained in the ADVERTISEMENT

           (975) -@ Recompute the Skew_Time

           (980) @ Recompute the Master_Down_Interval

           (985) @ Set Master_Down_Timer to Master_Down_Interval

           (990) If Critical_BFD_Session Exists:

              (995) @+ Tear Critical_BFD_Session

           (960) @ endif //Critical_BFD_Session Exists?

           (965) @ Calculate Critical_Backup.

           (970) @ If this instance is Critical_Backup:

              (975) @+ BootStrap Critical_BFD_Session with new
                       Master.

           (980) @ endif // am i Critical_Backup?

           (985) @ Transition to the {Backup} state

        (990) * else // new Master logic

           (995) @ Discard ADVERTISEMENT

        (1000) *endif // new Master detected

     (1005) +endif // was priority zero?

  (1010) -endif // advert recv

  (1015) - If a BACKUP ADVERTISEMENT is received, then:

     (1020) + If the Priority in the BACKUP ADVERTISEMENT is
              zero, then:

        (1025) * Remove the Peer from peer table.
```

```
        (1030) + else: // priority non-zero

           (1035) * Update the Peer info in peer table.

           (1040) * Recompute the Backup_Down_Interval

           (1045) * Reset the Backup_Down_Timer to
                    Backup_Down_Interval

        (1050) + endif // priority in backup advert zero

        (1055) + Calculate the Critical_Backup

        (1060) + If Critical_BFD_Session doesnot exist:

           (1065) * BootStrap Critical_BFD_Session

        (1070) + else if Critical_BFD_Session exist and
                 Critical_Backup changes:

           (1075) + Tear Critical_BFD_Session with old Backup

           (1080) + BootStrap Critical_BFD_Session with Critical_Backup

        (1085) + endif //  Critical_BFD_Session check?

     (1090) - endif // backup advert recv

     (1095) - If Critical_BFD_Session transitions from UP to DOWN,
     then:
        (1100) + Cancel Backup_Down_Timer

        (1105) + Delete the Peer info from peer table

        (1200) + Calculate the Critical_Backup

        (1205) + BootStrap Critical_BFD_Session with Critical_Backup

     (1210) - endif // BFD session transition

  (1215) endwhile // in Master
```

4. Applicability of p2mp BFD

[I-D.draft-ietf-bfd-multipoint] describes extensions to the BFD protocol for its use in multipoint and multicast networks. With these extensions p2mp BFD can support sub-second failure detection of

the root by tail nodes. In fact, a redundancy group may be viewed as p2mp BFD session with its Master being the root and Backup routers being tail nodes. Once Master selection process is completed, the Master router starts transmitting BFD control packets with IPvX address associated with the VRID as source IPvX address. Backup router demultiplexes p2mp BFD tail sessions based on IPvX source address associated with the virtual router that it been configured with. Once Backup router accepts p2mp session from the new Master router, the Backup router MAY use My Discriminator from received p2mp BFD control packet to demultiplex p2mp BFD sessions. When a Backup router detects failure of the Master router it re-evaluates its role in the VRID. As result, the Backup router may become the Master router of the given VRID or continue as a Backup router. If the former is the case, then the new Master router MUST select My Discriminator and start transmitting p2mp BFD control packets using Master IPvX address as source IPvX address for p2mp BFD control packets. If the latter is the case, then the Backup router MUST wait for p2mp BFD control packet with source IPvX address set to IPvX address associated with the VRID.

4.1. VRRP State Machine with p2mp BFD

Following section outlines the interaction between VRRP protocol [RFC5798] state machine and p2mp BFD. Please refer to the section 6.4 of [RFC5798] for State description.

4.1.1. Initialize

Following state machine replaces the state machine outlined in section 6.4.1 of [RFC5798]

(100) If a Startup event is received, then:

(105) - If the Priority = 255 (i.e., the router owns the IPvX address associated with the virtual router), then:

(110) + Send an ADVERTISEMENT

(115) + If the protected IPvX address is an IPv4 address, then:

(120) * Broadcast a gratuitous ARP request containing the virtual router MAC address for each IP address associated with the virtual router.

(125) + else // IPv6

(130) * For each IPv6 address associated with the virtual router, send an unsolicited ND Neighbor Advertisement with the Router Flag (R) set, the Solicited Flag (S) unset, the Override flag (O) set, the target address set to the IPv6 address of the virtual router, and the target link-layer address set to the virtual router MAC address.

(135) +endif // was protected addr IPv4?

(140) + Set the Adver_Timer to Advertisement_Interval

(145) + Transition to the {Master} state

(150) - else // rtr does not own virt addr

(155) + Set Master_Adver_Interval to Advertisement_Interval

(160) + Set the Master_Down_Timer to Master_Down_Interval

(165) + BootStrap BFD MultipointTail Session

(170) + Transition to the {Backup} state

(175) -endif // priority was not 255

(180) endif // startup event was recv

4.1.2. Backup

Following state machine replaces the state machine outlined in section 6.4.2 of [RFC5798]

(300) While in this state, a VRRP router MUST do the following:

(305) - If the protected IPvX address is an IPv4 address, then:

(310) + MUST NOT respond to ARP requests for the IPv4 address(es) associated with the virtual router.

(315) - else // protected addr is IPv6

(320) + MUST NOT respond to ND Neighbor Solicitation messages for the IPv6 address(es) associated with the virtual router.

(325) + MUST NOT send ND Router Advertisement messages for the virtual router.

(330) -endif // was protected addr IPv4?

(335) - MUST discard packets with a destination link-layer MAC address equal to the virtual router MAC address.

(340) - MUST NOT accept packets addressed to the IPvX address(es) associated with the virtual router.

(345) - If a Shutdown event is received, then:

(350) + Cancel the Master_Down_Timer

(355) + Transition to the {Initialize} state

(360) -endif // shutdown recv

(365) - If the Master_Down_Timer fires or
BFD MultipointTail session transitions from UP to DOWN,
then:

(370) + Send an ADVERTISEMENT

(375) + If the protected IPvX address is an IPv4 address, then:

(380) * Broadcast a gratuitous ARP request on that interface containing the virtual router MAC address for each IPv4 address associated with the virtual router.

(385) + else // ipv6

(390) * Compute and join the Solicited-Node multicast address [RFC4291] for the IPv6 address(es) associated with the virtual router.

(395) * For each IPv6 address associated with the virtual

router, send an unsolicited ND Neighbor Advertisement with the Router Flag (R) set, the Solicited Flag (S) unset, the Override flag (O) set, the target address set to the IPv6 address of the virtual router, and the target link-layer address set to the virtual router MAC address.

(400) +endif // was protected addr ipv4?

(405) + Set the Adver_Timer to Advertisement_Interval

(410) + Tear down BFD MultipointTail Session

(415) + BootStrap BFD MultipointHead Session

(420) + Transition to the {Master} state

(425) -endif // Master_Down_Timer fired

(430) - If an ADVERTISEMENT is received, then:

(435) + If the Priority in the ADVERTISEMENT is zero, then:

(440) * Set the Master_Down_Timer to Skew_Time

(445) + else // priority non-zero

(450) * If Preempt_Mode is False, or if the Priority in the ADVERTISEMENT is greater than or equal to the local Priority, then:

(455) @ Set Master_Adver_Interval to Adver Interval contained in the ADVERTISEMENT

(460) @ Recompute the Master_Down_Interval

(465) @ Reset the Master_Down_Timer to Master_Down_Interval

(470) * else // preempt was true or priority was less

(475) @ Discard the ADVERTISEMENT

(480) *endif // preempt test

(485) +endif // was priority zero?

(490) -endif // was advertisement recv?

(495) endwhile // Backup state

4.1.3. Master

Following state machine replaces the state machine outlined in section 6.4.3 of [RFC5798]

(600) While in this state, a VRRP router MUST do the following:
(605) - If the protected IPvX address is an IPv4 address, then:

(610) + MUST respond to ARP requests for the IPv4 address(es) associated with the virtual router.

(615) - else // ipv6

(620) + MUST be a member of the Solicited-Node multicast address for the IPv6 address(es) associated with the virtual router.

(625) + MUST respond to ND Neighbor Solicitation message for the IPv6 address(es) associated with the virtual router.

(630) ++ MUST send ND Router Advertisements for the virtual router.

(635) ++ If Accept_Mode is False: MUST NOT drop IPv6 Neighbor Solicitations and Neighbor Advertisements.

(640) +-endif // ipv4?

(645) - MUST forward packets with a destination link-layer MAC address equal to the virtual router MAC address.

(650) - MUST accept 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. Otherwise, MUST NOT accept these packets.

(655) - If a Shutdown event is received, then:

(660) + Cancel the Adver_Timer

(665) + Send an ADVERTISEMENT with Priority = 0

(670) + Tear down BFD MultipointHead Session

(675) + Transition to the {Initialize} state

(680) -endif // shutdown recv

(685) - If the Adver_Timer fires, then:

(690) + Send an ADVERTISEMENT

(695) + Reset the Adver_Timer to Advertisement_Interval

(700) -endif // advertisement timer fired

(705) - If an ADVERTISEMENT is received, then:

(710) -+ If the Priority in the ADVERTISEMENT is zero, then:

(715) -* Send an ADVERTISEMENT

(720) -* Reset the Adver_Timer to Advertisement_Interval

(725) -+ else // priority was non-zero

(730) -* If the Priority in the ADVERTISEMENT is greater than the local Priority,

(735) -* or

(740) -* If the Priority in the ADVERTISEMENT is equal to the local Priority and the primary IPvX Address of the sender is greater than the local primary IPvX Address, then:

(745) -@ Cancel Adver_Timer

(750) -@ Set Master_Adver_Interval to Adver Interval contained in the ADVERTISEMENT

(755) -@ Recompute the Skew_Time

(760) @ Recompute the Master_Down_Interval

(765) @ Set Master_Down_Timer to Master_Down_Interval

(770) + Tear down BFD MultipointHead Session

(775) + BootStrap BFD MultipointTail Session

(780) @ Transition to the {Backup} state

(785) * else // new Master logic

(790) @ Discard ADVERTISEMENT

(795) *endif // new Master detected

(800) +endif // was priority zero?

(805) -endif // advert recv

(810) endwhile // in Master

5. Scalability Considerations

To reduce the number of packets generated at a regular interval, Backup Advert packets may be sent at a reduced rate as compared to Advert packets sent by the VRRP Master.

In a Data Centre with VXLAN extending the Layer 2 network, when implementing Section 4 of this document, inherently multicast traffic is flooded or replicated to all the Virtual Tunneling End Points by means of multicast traffic in the underlay network. The amount of replication or flooding depends on the number of Virtual Tunneling End Points connected to the VXLAN network. VRRP is typically deployed on the Virtual Tunneling End Points. If Multipoint BFD is used for tracking the state of VRRP Master Router the Multipoint BFD packets will get carried over the Layer 2 Overlay, this can lead to a lot of traffic getting flooded on the overlay as the rate at which BFD packets are generated will be typically in sub second range. Which is the problem if VRRP is configured with sub second timers. So in such scenarios where flooding of Multicast traffic is a concern, it is recommended to use Point to Point BFD sessions to avoid inherent flooding of Multicast traffic and configure VRRP to default or slow timers.

6. Operational Considerations

A VRRP peer that forms a member of this Virtual Router, but does not support this feature or extension must be configured with the lowest priority, and will only operate as the Router of last resort on failure of all other VRRP routers supporting this functionality.

It is recommended that mechanism defined by this draft, to interface VRRP with BFD should be used when BFD can support more aggressive monitoring timers than VRRP. Otherwise it is desirable not to interface VRRP with BFD for determining the health of VRRP Master.

This Draft does not preclude the possibility of the peer table being populated by means of manual configuration, instead of using the BACKUP ADVERTISEMENT as defined by the Draft.

7. IANA Considerations

This document requests IANA to create a new name space that is to be managed by IANA. The document defines a new VRRP Packet Type. The VRRP Packet Types are discussed below.

a) Type 1 (ADVERTISEMENT) defined in section 5.2.2 of [RFC5798]
b) Type 2 (BACKUP ADVERTISEMENT) defined in section 3.3 of this document

7.1. A New Name Space for VRRP Packet Types

This document defines in Section 3.3 a "BACKUP ADVERTISEMENT" VRRP Packet Type. The new name space has to be created by the IANA and they will maintain this new name space. The field for this namespace is 4-Bits, and IANA guidelines for assignments for this field are as follows:

ADVERTISEMENT	1
BACKUP ADVERTISEMENT	2

Future allocations of values in this name space are to be assigned by IANA using the "Specification Required" policy defined in [IANA-CONS]

8. Security Considerations

Security considerations discussed in [RFC5798], [RFC5880] and [I-D.draft-ietf-bfd-multipoint], apply to this document. There are no additional security considerations identified by this draft.

9. Acknowledgements

The authors gratefully acknowledge the contributions of Gerry Meyer, and Mouli Chandramouli, for their contributions to the draft. The authors will also like to thank Jeffrey Haas, Maik Pfeil, Chris Bowers and Vengada Prasad Govindan for their comments and suggestions.

10. Normative References

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, 2010.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, 1997.

[RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, 2010.

[RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, 2010.

[I-D.draft-ietf-bfd-multipoint] Katz, D., Ward, D., and S. Pallagatti, "BFD for Multipoint Networks", Work in Progress draft-ietf-bfd-multipoint-07, 2015.

[IANA-CONS] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 2434, 1998.

Authors' Addresses

Nitish Gupta
Cisco Systems, Inc.
Sarjapur Outer Ring Road
Bangalore 560103
India

Phone: +91 80 4429 2530
Email: nitisgup@cisco.com
URI: http://www.cisco.com/

Aditya Dogra
Cisco Systems, Inc.
Sarjapur Outer Ring Road
Bangalore 560103
India

Phone: +91 80 4429 2166
Email: addogra@cisco.com
URI: http://www.cisco.com/

Colin Docherty
25 George Grieve Way
Tranent
East Lothian, Scotland EH332QT
United Kingdom

Email: colin@doch.org.uk

Greg Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com

Jeff Tantsura
Individual

Email: jefftant.ietf@gmail.com