Independent Submission
Request for Comments: 7910
Category: Informational
ISSN: 2070-1721

W. Zhou
Cisco Systems

June 2016

Interoperability between the Virtual Router Redundancy Protocol and PIM

Abstract

This document introduces VRRP-aware PIM, a redundancy mechanism for the Protocol Independent Multicast (PIM) to interoperate with the Virtual Router Redundancy Protocol (VRRP). It allows PIM to track VRRP state and to preserve multicast traffic upon failover in a redundant network with virtual routing groups enabled. The mechanism described in this document is based on Cisco IOS software implementation.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This is a contribution to the RFC Series, independently of any other RFC stream. The RFC Editor has chosen to publish this document at its discretion and makes no statement about its value for implementation or deployment. Documents approved for publication by the RFC Editor are not a candidate for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7910.

Table of Contents

1. Introduction

The Virtual Router Redundancy Protocol (VRRP) [RFC5798] is a redundancy protocol for establishing a fault-tolerant default router. The protocol establishes a framework between network devices in order to achieve default device failover if the primary devices become inaccessible.

Protocol Independent Multicast (PIM) [RFC7761] has no inherent redundancy capabilities and its operation is completely independent of VRRP group states. As a result, IP multicast traffic is not necessarily forwarded by the same device that is elected by VRRP. The VRRP-aware PIM feature provides consistent IP multicast forwarding in a redundant network with virtual routing groups enabled.

In a multi-access segment (such as LAN), the elected PIM designated router (DR) is unaware of the redundancy configuration, and the elected DR and VRRP master router (MR) may not be the same. In order to ensure that the PIM DR is always able to forward a PIM Join/Prune (J/P) message towards Rendezvous Point (RP) or first-hop router, the VRRP MR becomes the PIM DR (if there is only one VRRP group). PIM is responsible for adjusting the DR priority based on the group state. When a failover occurs, multicast states are created on the new MR elected by the VRRP group and the MR assumes responsibility for the routing and forwarding of all the traffic addressed to the VRRP virtual IP address (vIP). This ensures that the PIM DR runs on the same router as the VRRP MR and maintains multicast routing (mroute) states. It enables multicast traffic to be forwarded through the VRRP MR, allowing PIM to leverage VRRP redundancy, avoid potential duplicate traffic, and enable failover, depending on the VRRP states in the router.

This mechanism can be safely deployed into a PIM network without changing the behavior of other routers. When only a specific set of routers enable this feature, a user can configure PIM interfaces to track state-change events of desired VRRP group(s). When a router becomes the VRRP MR, the PIM component applies the user-defined DR priority value to the interface in order to make it PIM DR. Other routers will not break the functionality of this feature, as long as their configured DR priority does not conflict with the participating routers. When deployed in a PIM transit network, downstream routers should configure the static route to use vIP as the next-hop address for PIM J/P messages in order to take advantage of this feature. If dynamic routing is used and the next-hop address is selected by unicast routing information as described in [RFC5294], then these routes cannot leverage the VRRP redundancy and failover mechanism. These downstream routers, however, do not have to support this new feature and there is no additional configuration or coordination required from a manageability point of view. This mechanism does not change any bit on the wire, and it has been implemented on Cisco IOS software.

2. Tracking and Failover

Without the mechanisms described in this document, a PIM component will send PIM J/P messages with the DR's IP address to upstream routers. A GenID (Generation Identifier) in a PIM Hello message is randomly selected when the router boots and remains the same as long as the router is up. A PIM neighbor reboot can easily be detected if its GenID is different from before; in this case, the PIM J/P and RP-Set information can be redistributed to the rebooted neighbor. With the VRRP-aware PIM mechanism enabled, the PIM component listens to the state-change notifications from VRRP and automatically adjusts the priority of the PIM DR based on the VRRP state and ensures the VRRP MR (if there is only one VRRP group) becomes the DR of the LAN. If there are multiple VRRP groups, the DR is determined by the user-configured priority value.

Upon failover, the PIM component triggers communication between upstream and downstream routers in order to create mroute states on the new VRRP MR. The PIM component sends an additional PIM Hello message using the VRRP vIP as the source address for each active VRRP group when a router becomes the VRRP MR. The PIM Hello message with a new GenID will trigger other routers to respond to the VRRP failover event in the same way as the PIM neighbor reboot event as described in [RFC5294]. Specifically, when a downstream router receives this PIM Hello message, it will add the source IP address (in this case the VRRP vIP) into its PIM neighbor list and immediately send triggered PIM J/P messages towards vIP. Upstream routers will process PIM J/P messages based on the VRRP group state.

If the PIM J/P next-hop address matches the VRRP vIP, only the current VRRP MR will process the PIM J/P messages. This allows all PIM J/P messages to reach the VRRP group vIP and minimizes changes and configurations at the downstream routers.

Alternatively, the implementation can choose to have all VRRP passive routers maintain mroute states and record the GenID of the current MR. When a passive router becomes the MR, it uses the existing mroute states and the recorded MR GenID in its PIM Hello message. This will avoid resending PIM J/P messages upon failover and will eliminate the requirement of an additional PIM Hello with vIP. There is no change in on-the-wire behavior or in the PIM and VRRP message format.

3. PIM Assert Metric Auto-Adjustment

It is possible that, after the VRRP MR switches from router A to B, A would still forward multicast traffic, which will result in duplicate traffic. The PIM Assert mechanism will kick in because PIM Assert with redundancy is enabled.

- If there is only one VRRP group, passive routers will send an arbitrary penalty metric preference (PIM_ASSERT_INFINITY - 1) and make MR the Assert winner.

- If there are multiples VRRP groups configured on an interface, the Assert metric preference will be (PIM_ASSERT_INFINITY - 1) if and only if all VRRP groups are in Passive state.

- If there is at least one VRRP group in Active state, then the original Assert metric preference will be used. That is, the winner will be selected between routers using their real Assert metric preference with at least one active VRRP Group, as if no VRRP is involved.

4. DF Election for BiDir Group

Change to Designated Forwarder (DF) offer/winner metric is handled similarly to PIM Assert handling with VRRP.

- If there is only one VRRP group, passive routers will send a large penalty metric preference in an offer (PIM_BIDIR_INFINITY_PREF- 1) and make MR the DF winner.

- If there are multiples VRRP groups configured on an interface, the offer metric preference will be (PIM_BIDIR_INFINITY_PREF- 1) if and only if all VRRP groups are in Passive state.

- o If there is at least one VRRP group in Active state, then the original offer metric preference to RP will be used. That is, the winner will be selected between routers using their real offer metric, as if no VRRP is involved.

5. Tracking Multiple VRRP Groups on an Interface

A user can configure a PIM component to track more than one VRRP groups on an interface. This allows other applications to exploit PIM/VRRP interoperability to achieve various goals (e.g., load balancing). Since each VRRP group that is configured on an interface could be in different states at any moment, the DR priority is adjusted. The PIM Assert metric and PIM BiDir DF metric should be adjusted if and only if all VRRP groups that are configured on an interface are in Passive (non-Active) states to ensure that interfaces with all-passive VRRP groups do not win DR, Assert, and DF election. In other words, the DR, Assert, and DF winners will be elected among the interfaces with at least one active VRRP group.

6. Support of HSRP

Although there are differences between VRRP and the Hot Standby Router Protocol (HSRP) [RFC2281] -- including the number of backup (standby) routers, virtual IP address, and timer intervals -- the proposed scheme can also enable HSRP-aware PIM with similar failover and the tracking mechanism described in this document.

7. Security Considerations

The proposed tracking mechanism does not discuss adding authentication to the protocols and introduces no new negative impact or threats on security to PIM in either SSM (Source-Specific Multicast) or ASM (Any-Source Multicast) mode. Note that VRRP messages from malicious nodes could cause unexpected behaviors such as multiple MRs and PIM DRs, which are associated with VRRP-specific security issues. To mitigate the vulnerability of frequent VRRP and PIM DR state change from malicious attack, an implementation can choose to enable VRRP preemption such that a higher-priority VRRP backup router does not take over for a lower-priority MR; this will reduce the state-change notifications to a PIM component and subsequent mroute state changes. Detailed analysis of PIM and VRRP security is provided in [RFC5294] and [RFC5798].

8. Informative References

[RFC2281] Li, T., Cole, B., Morton, P., and D. Li, "Cisco Hot Standby Router Protocol (HSRP)", RFC 2281, DOI 10.17487/RFC2281, March 1998, <http://www.rfc-editor.org/info/rfc2281>.

[RFC5294] Savola, P. and J. Lingard, "Host Threats to Protocol Independent Multicast (PIM)", RFC 5294, DOI 10.17487/RFC5294, August 2008, <http://www.rfc-editor.org/info/rfc5294>.

[RFC5798] Nadas, S., Ed., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, DOI 10.17487/RFC5798, March 2010, <http://www.rfc-editor.org/info/rfc5798>.

[RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March 2016, <http://www.rfc-editor.org/info/rfc7761>.

Acknowledgments

I would like to give a special thank you and appreciation to Stig Venaas for his ideas and comments in this document.

Author’s Address

Wei Zhou
Cisco Systems
Tasman Drive
San Jose, CA 95134
United States

Email: zhouweiisu@gmail.com