

PIM WG
Internet-Draft
Intended status: Standards Track
Expires: December 8, 2016

Zheng. Zhang
Fangwei. Hu
BenChong. Xu
ZTE Corporation
June 6, 2016

PIM DR Improvement
draft-ietf-pim-dr-improvement-01.txt

Abstract

PIM is worldly deployed multicast protocol. This document will improve the stability of PIM protocol, decrease the lost of multicast packets when the PIM DR (Designed Router) is down.

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 December 8, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 2

2. Terminology 3

3. PIM hello message format 3

 3.1. DR Address Option format 3

 3.2. BDR Address Option format 4

4. The Protocol Treatment 4

 4.1. Election Algorithm 5

 4.2. Sending Hello Messages 6

 4.3. Receiving Hello Messages 6

 4.4. The treatment 7

5. Deployment suggestion 7

6. Security Considerations 7

7. IANA Considerations 7

8. Normative References 7

Authors' Addresses 8

1. Introduction

Multicast technology is used widely. Many modern technology use PIM technology, such as IPTV, Net-Meeting, and so on. There are many events that will influence the quality of multicast services. The change of unicast routes will cause the lost of multicast packets. The change of DR cause the lost of multicast packets too.

When a DR on a share-media LAN is down, other routers will elect a new DR until the expiration of Hello-Holdtime. The default value of Hello-Holdtime is 105 seconds. Although the value of Hello-Holdtime can be changed by manual, when the DR is down, there are still many multicast packets will be lost. The quality of IPTV and Net-Meeting will be influenced.

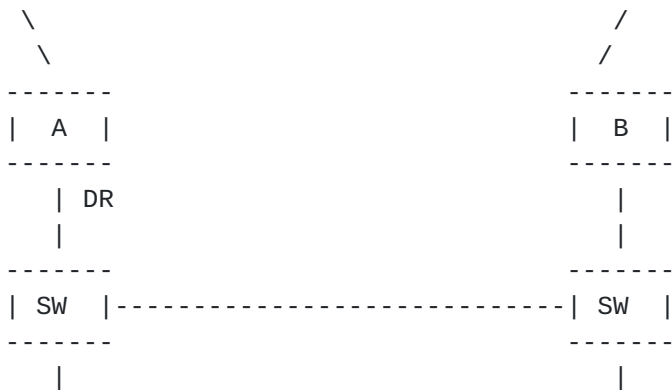


Figure 1: An example of multicast network

For example, there were two routers on one Ethernet. RouterA was elected to DR. When RouterA is down, the multicast packets are

discarded until the RouterB is elected to DR and RouterB imports the multicast flows successfully.

We suppose that there is only a RouterA in the Ethernet at first in Figure 1. RouterA is the DR who is responsible for forwarding multicast flows. When RouterB connects the Ethernet, RouterB will be elected to DR because a higher priority. So RouterA will stop forwarding multicast packets. The multicast flows will not recover until RouterB joins the multicast group after it is elected to DR.

2. Terminology

Backup Designated Router (BDR): A shared-media LAN like Ethernet may have multiple PIM-SM routers connected to it. Except for DR, a router who will act on behalf of directly connected hosts with respect to the PIM-SM protocol. But BDR will not forward the flows. When DR is down, the BDR will forward multicast flows immediately. A single BDR is elected per interface like the DR.

3. PIM hello message format

In [RFC4601] and [RFC7761], the PIM hello message format was defined. In this document, we define two new option values which are including Type, Length, and Value.

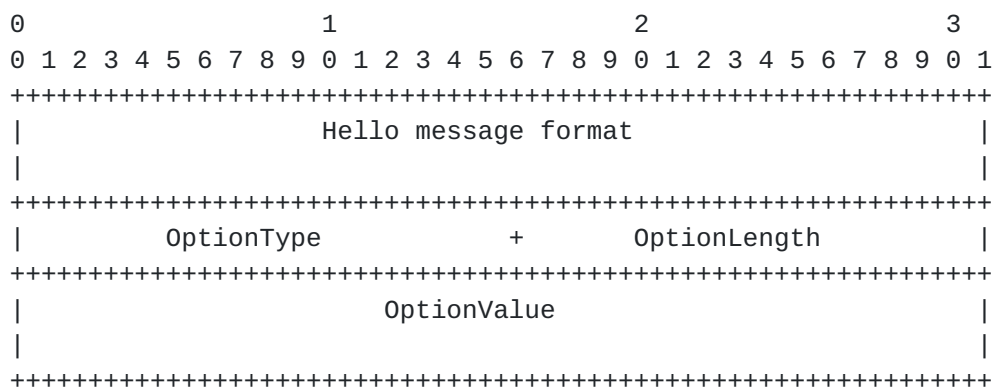


Figure 2: Hello message format

3.1. DR Address Option format

- o OptionType : The value is TBD.
- o OptionLength: If the network is support IPv4, the OptionLength is 4 octets. If the network is support IPv6, the OptionLength is 16 octets.

- o OptionValue: The OptionValue is IP address of DR. If the network is support IPv4, the value is IPv4 address of DR. If the network is support IPv6, the value is IPv6 address of DR.

3.2. BDR Address Option format

- o OptionType : The value is TBD.
- o OptionLength: If the network is support IPv4, the OptionLength is 4 octets. If the network is support IPv6, the OptionLength is 16 octets.
- o OptionValue: The OptionValue is IP address of BDR. If the network is support IPv4, the value is IPv4 address of BDR. If the network is support IPv6, the value is IPv6 address of BDR.

4. The Protocol Treatment

A new router starts to send hello messages with the values of DR and BDR are all set to 0 after its interface is enabled in PIM on a share-media LAN. When the router receive hello messages from other routers on the same share-media LAN, the router will check if the value of DR is filled. If the value of DR is filled with IP address of router who is sending hello messages, the router will store the IP address as the DR address of this interface.

Then the new router compare the priority and IP address itself to the stored IP address of DR and BDR according to the algorithm of [\[RFC4601\]](#) and [\[RFC7761\]](#). If the new router notices that it is better to be DR than the existed DR or BDR. The router will make itself the BDR, and send new hello messages with its IP address as BDR and existed DR. If the router notices that the existed DR is most priority in the share-media LAN, but the existed BDR is set to 0x0 in the received hello messages, or the existed BDR is not better than the new router, the router will elect itself to BDR. If the router notices that it is not better to be DR than existed DR and BDR, the router will respect the PIM protocol.

When the new router becomes the new BDR, the router will join the existed multicast groups, import multicast flows from upstream routers. But the BDR MUST not forward the multicast flows to avoid the duplicate multicast packets in the share-media LAN. The new router will monitor the DR. When the DR becomes unavailable because of the down or other reasons, the BDR will forward multicast flows immediately.

4.1. Election Algorithm

The DR and BDR election is according the rules defined below, the algorithm is similar to the DR election definition in [[RFC2328](#)].

- (1) Note the current values for the network's Designated Router and Backup Designated Router. This is used later for comparison purposes.
- (2) Calculate the new Backup Designated Router for the network as follows. Those routers that have not declared themselves to be Designated Router are eligible to become Backup Designated Router. The one who have the highest priority will be chosen to be Backup Designated Router. In case of a tie, the one having the highest Router ID is chosen.
- (3) Calculate the new Designated Router for the network as follows. If one or more of the routers have declared themselves Designated Router (i.e., they are currently listing themselves as Designated Router in their Hello Packets) the one having highest Router Priority is declared to be Designated Router. In case of a tie, the one having the highest Router ID is chosen. If no routers have declared themselves Designated Router, assign the Designated Router to be the same as the newly elected Backup Designated Router.
- (4) If Router X is now newly the Designated Router or newly the Backup Designated Router, or is now no longer the Designated Router or no longer the Backup Designated Router, repeat steps 2 and 3, and then proceed to step 5. For example, if Router X is now the Designated Router, when step 2 is repeated X will no longer be eligible for Backup Designated Router election. Among other things, this will ensure that no router will declare itself both Backup Designated Router and Designated Router.
- (5) As a result of these calculations, the router itself may now be Designated Router or Backup Designated Router.

The reason behind the election algorithm's complexity is the desire for an orderly transition from Backup Designated Router to Designated Router, when the current Designated Router fails. This orderly transition is ensured through the introduction of hysteresis: no new Backup Designated Router can be chosen until the old Backup accepts its new Designated Router responsibilities.

The above procedure may elect the same router to be both Designated Router and Backup Designated Router, although that router will never be the calculating router (Router X) itself. The elected Designated Router may not be the router having the highest Router Priority. If

Router X is not itself eligible to become Designated Router, it is possible that neither a Backup Designated Router nor a Designated Router will be selected in the above procedure. Note also that if Router X is the only attached router that is eligible to become Designated Router, it will select itself as Designated Router and there will be no Backup Designated Router for the network.

4.2. Sending Hello Messages

According to [Section 4.3.1 in \[RFC4601\]](#) and [\[RFC7761\]](#), when a new router's interface is enabled in PIM protocol, the router send hello messages with the values of DR and BDR are filled with 0x0. Then the interface is in waiting state and start the hold-timer which is like the neighbor hold-timer. When the hold-timer is expired, the interface will elect the DR and BDR according to the DR election rules.

When a new router sets itself BDR after receive hello messages from other routers, the router send hello messages with the value of DR is set to the IP address of existed DR and the value of BDR is set to the IP address of the router itself.

When a existed router sets itself non DR and non BDR after receive hello messages from other routers, the router will send hello messages with the value of DR is set to existed DR and the value of BDR is set to new BDR.

4.3. Receiving Hello Messages

When the values of DR and BDR which are carried by hello messages are received is all set to 0x0, the router MUST elect the DR due to the algorithm of [\[RFC4601\]](#) and [\[RFC7761\]](#). And elect a new BDR which are the best choice except DR.

When the value of DR which is carried by received hello messages is not 0x0, and the value of BDR is set to 0x0, the router will elect itself to BDR.

When the values of DR and BDR that carried by received hello messages are all larger than 0x0. The router will mark the existed DR, and compare itself and the BDR in message. When the router notice that it is better to be DR than existed BDR. The router will elect itself to the BDR.

When a router receives a new hello message with the values of DR and BDR are set to 0x0. The router will compare the new router with itself. If the router noticed that the new router is better to be DR than itself, the router will set the BDR to the new router.

4.4. The treatment

When all the routers on a shared-media LAN are start to work on the same time, the election result of DR is same as [[RFC4601](#)] and [[RFC7761](#)]. And all the routers will elect a BDR which is suboptimum to DR. The routers in the network will store the DR and BDR. The hello messages sent by all the routers are same with the value of DR and BDR are all set.

When a new router start to work on a shared-media LAN and receive hello messages from other routers that the value of DR is set. The new router will not change the existed DR even if it is superior to the existed DR. If the new router is superior to existed BDR, the new router will replace the existed BDR.

When the routers receive hello message from a new router, the routers will compare the new router and all the other routers on the LAN. If the new router is superior to existed BDR, the new router will be new BDR. Then the old BDR will send prune message to upstream routers.

As a result, the BDR is the one who has the highest priority except DR. Once the DR is elected, the DR will not change until it fails. After the DR and BDR are elected, the routers in the network will store the address of DR and BDR.

5. Deployment suggestion

If there are two and more routers on a share-media LAN, and the multicast services is sensitive to the lost of multicast packets, the function of DR and BDR defined in this document should be deployed.

6. Security Considerations

For general PIM Security Considerations.

7. IANA Considerations

IANA is requested to allocate OptionTypes in TLVs of hello message. Include DR and BDR.

8. Normative References

[RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#), DOI 10.17487/RFC2328, April 1998, <<http://www.rfc-editor.org/info/rfc2328>>.

- [RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", [RFC 4601](#), DOI 10.17487/RFC4601, August 2006, <<http://www.rfc-editor.org/info/rfc4601>>.
- [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>>.

Authors' Addresses

Zheng(Sandy) Zhang
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: zhang.zheng@zte.com.cn

Fangwei Hu
ZTE Corporation
No.889 Bibo Rd
Shanghai
China

Email: hu.fangwei@zte.com.cn

BenChong Xu
ZTE Corporation
No. 68 Zijinghua Road, Yuhuatai Distinct
Nanjing
China

Email: xu.benchong@zte.com.cn

