

Protocol Independent Multicast
Internet-Draft
Updates: [5015](#) (if approved)
Intended status: Standards Track
Expires: April 19, 2014

Z. Zhang
K. Windisch
J. A. Gralak
Juniper Networks, Inc.
October 16, 2013

PIM-Bidir RPL Resiliency
draft-zzhang-pim-bidir-rpl-resiliency-00.txt

Abstract

With PIM-Bidir, the RPA does not have to be associated with a router. Rather, it only needs to be a routable address on a RPL (typically a multi-access network). Such a scenario is commonly referred as Phantom RPA. This achieves RP resiliency to some extent, because the "RP" will not fail. However, if the RPL itself partitions, traffic converged to one partition will not be able to reach other parts of the network where joins converge to the other partitions of the RPL.

This document proposes simple procedures, which does not require signaling extensions, to achieve RPL resiliency.

Requirements Language

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 [RFC2119](#).

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 April 19, 2014.

Copyright Notice

Internet-Draft

pim-bidir-rpl-resiliency

October 2013

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](http://trustee.ietf.org/license-info) 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
1.1.	Problem Description	2
1.2.	Motivations	3
1.3.	Proposed Solutions	4
2.	Operations	5
2.1.	Modified PIM-Bidir Procedures	5
2.2.	Detect partitioning and elect active partition	6
2.2.1.	Using Host Routes advertised by any protocol	6
2.2.2.	Using Link State Routing protocol	6
2.2.3.	Comparison between the two detection and election methods	8
3.	IANA Considerations	8
4.	Security Considerations	8
5.	Contributors	8
6.	Acknowledgements	9
7.	References	9
7.1.	Normative References	9
7.2.	Informative References	9

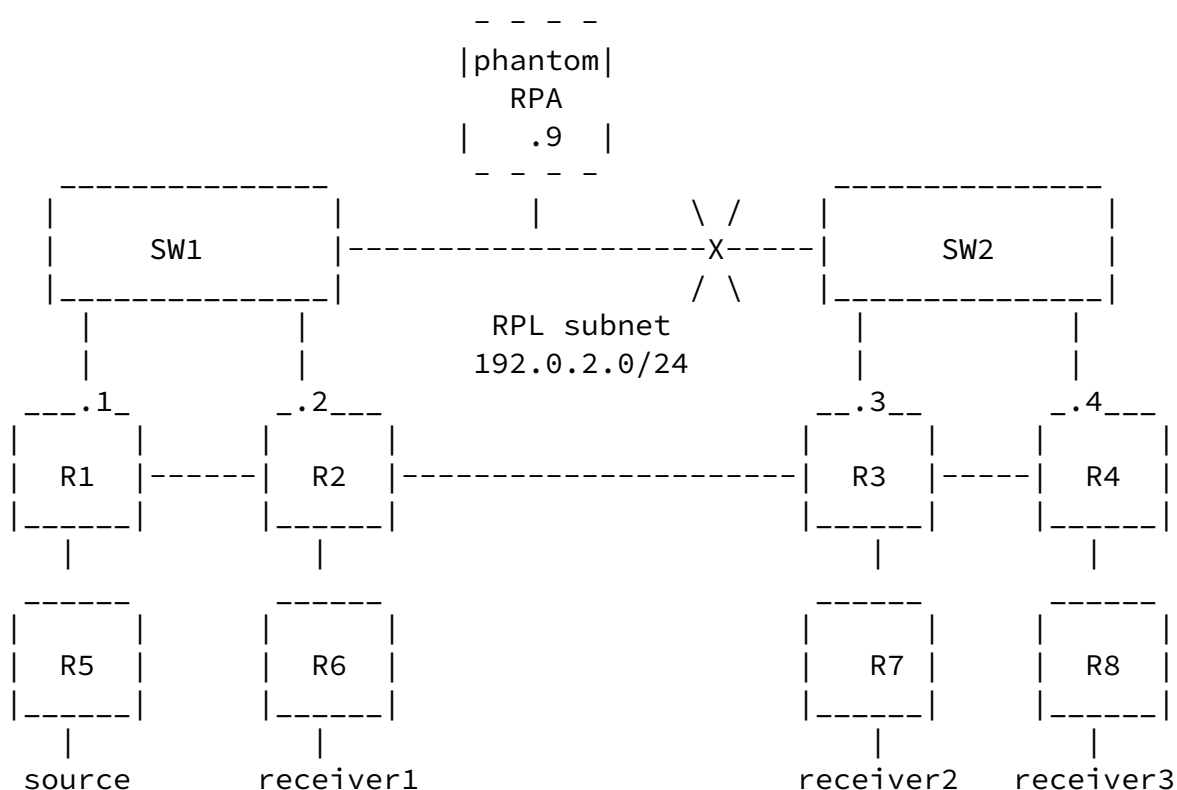
[1.](#) Introduction

[1.1.](#) Problem Description

The problem with partitioned RPL is that routers on the RPL still expect traffic to be exchanged over the RPL to reach other parts of the network, even though that won't happen across the RPL partitions.

This can be illustrated by Figure 1. The RPL is served by two

interconnected switches and if the link between the switches breaks, R1~R4 will all continue to treat the link as RPL, and terminate the joins. R3~4 continue to expect traffic injected by R5 to arrive on the RPL link, instead of sending joins to R2.



RPL partition caused by the inter-switch link failure.

Figure 1

1.2. Motivations

The importance of ensuring traffic reachability in spite of RPL partitioning is obvious. Additionally, [\[I-D.wijnands-pim-source-discovery-bsr\]](#) provides a perfect example of PIM-Bidir as a solution once the partitioning problem is solved.

[\[I-D.wijnands-pim-source-discovery-bsr\]](#) proposes to extend BSR to flood source information so that routers connecting to receivers can send (s,g) SPT joins, bypassing the RTP->SPT switch. It points out

that the solution is not suitable "for applications with strong dependency on the initial packet(s)" and PIM-Bidir [[RFC5015](#)] should be used for that. However, PIM-Bidir is not suitable where high resiliency is required, unless the partitioning problem is resolved.

[I-D.wijnands-pim-source-discovery-bsr] also raises a question whether BSR should be extended to a generic flooding mechanism for opaque information. Due to the way BSR flooding is done, while it is acceptable to flood group-to-rp mapping, it becomes inefficient to flood large amount of data. PIM-Bidir can be used as a generic protocol for efficient many-to-many data distribution and solving the partitioning problem enables the same level of resiliency as BSR flooding.

1.3. Proposed Solutions

This problem can be solved as follows:

- o Routers on the RPL detect RPL partitioning, elect an active partition to continue function as RPL, and stop treating the inactive partitions as RPL.
- o All routers route joins and traffic towards the active partition.

For the first task, this document specifies two methods to detect RPL partitioning and elect an active partition. For the second task, a host route to the RPA can be announced by the routers on the active partition.

This solution not only addresses RPL partitioning, it can also be used to mitigate the impact of network partitioning (where a part of network may be completely separated from the rest) by intentionally placing RPL segments into different parts of the network, as illustrated on Figure 2. This is called Anycast RPL in this

document, because the segments will all have the same subnet. With that, only one segment will be active and treated as RPL before the network partitions.

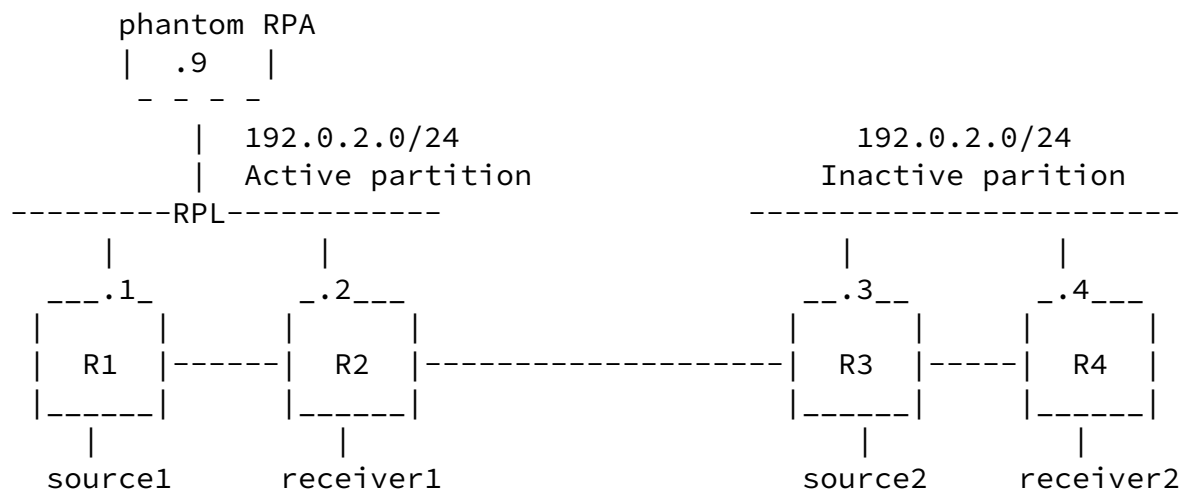


Figure 2

When the network separates into completely disjoint partitions, see figure Figure 3, each partition may have their own active RPL so intra-partition traffic will continue to flow. In the extreme case, all routers can be put onto RPL segments, making the network extremely resilient from PIM-Bidir point of view.

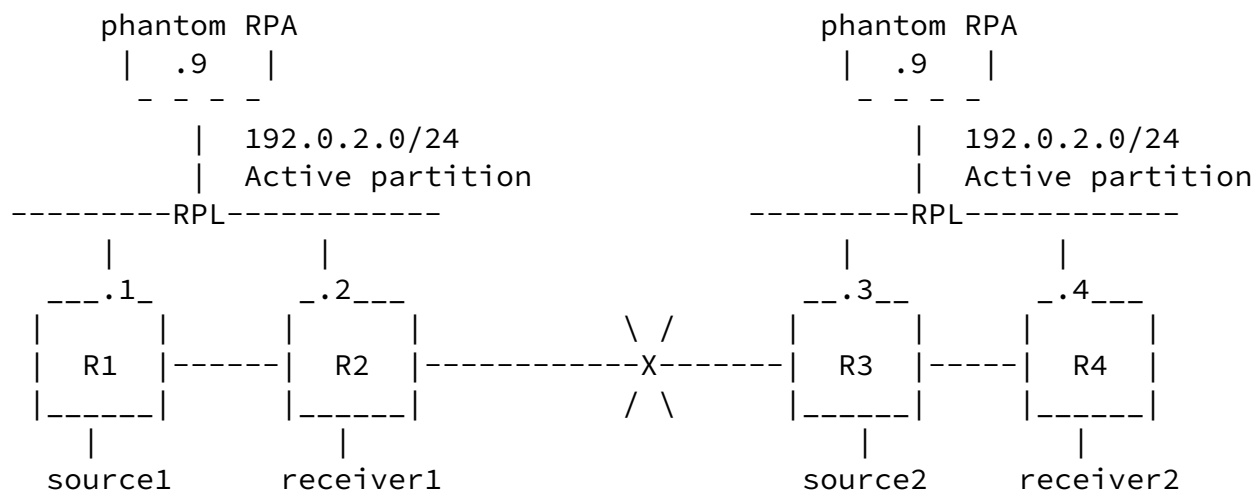


Figure 3

For simplicity and practicality, this document assumes that the RPA does not belong to any router on the RPL. Such a scenario is commonly referred as phantom RPA. These procedures MUST NOT be used when the RPA is an address belonging to a router.

[2.](#) Operations

[2.1.](#) Modified PIM-Bidir Procedures

A PIM router treats a link as RPL when the following two conditions are all met:

- o [existing] The route towards a RPA is directly over the link
- o [new] The router is in the elected active partition

Note that the active partition could be the one and only "partition" (when there is no RPL partitioning).

A router MUST advertise a host route to the RPA if and only if it treats a link as RPL. It MUST start the DF election on the link and treat it as a regular link when it stops treating a link as RPL. When it starts treating a link as RPL, it MUST stop the DF election.

The following sections specify two methods to detect partitioning and elect an active partition. Each elected active partition is

identified by one of the routers, and other routers determine if they are in the active partition by checking their neighborhood with the identifying router.

The neighborhood check can be done via either IGP mechanism (e.g. OSPF Hello) or PIM Hello (if used). In either case, fast neighborhood change detection SHOULD be used, e.g., via BFD or short Hello interval.

[2.2.](#) Detect partitioning and elect active partition

For the detection and election, each partition needs to be represented by one or more identifiers. This can be done by two methods.

[2.2.1.](#) Using Host Routes advertised by any protocol

In each partition, routers learn of each other by way of PIM Hellos. Of all the neighbors, the one with the lowest routable unicast interface address on the subnet MUST advertise a host route to the address itself, e.g. via a Stub Link in the OSPF Router LSA or a BGP NLRI. Optionally, to speed up convergence and facilitate make-before-break process, the one with the second lowest address or even all may do the same.

The host routes represent all partitions, potentially with N:1 mapping.

Routers on the RPL subnet find all the host routes that fall into the RPL subnet range, and select the one with the lowest address which itself not RPA address. That address identifies the active partition. Whenever such a host route is added or deleted, the election process is rerun.

[2.2.2.](#) Using Link State Routing protocol

When a Link State Routing protocol is used, the link states for the RPL subnet can be used. For example, with OSPF, each partition may have its own Network LSA for the same subnet, or in case of no Network LSA (there may be no DR or adjacency between the DR and a non-DR), each router on the partition will advertise a stub link in its Router LSA for the RPL subnet. Routers on the RPL subnet check all the reachable Network LSAs for the subnet and reachable Router LSAs that have a stub link for the subnet. The Network LSA with the lowest Advertising Router among all those Network LSAs, or in case of no Network LSAs the Router LSA with the lowest Advertising Router is selected to identify the active partition. If a Network LSA is selected, then a router is on the active partition if and only if it

originated the Network LSA, or it is a neighbor on the subnet with the Advertising Router. If a Router LSA is selected, then only the Advertising Router itself is on the active partition.

Whenever a corresponding Network LSA or stub link for the RPL subnet is added/deleted or its reachability changes, the election process is rerun.

The above procedure does not need any PIM/IGP signaling extensions, but only works if all the partitions are in the same area. That is sufficient to address RPL partitioning, but if it is desired to put Anycast RPLs in different areas, then IGP signaling extension is needed. Again using OSPF as an example:

- o When an Area Border Router (ABR) advertises a Type 3 Summary LSA into the backbone area B from a non-backbone area A for a RPL subnet that it learns in area A, the Summary LSA MUST carry, in a TLV according to [[I-D.acee-ospfv3-lsa-extend](#)](details TBD), the lowest Advertising Router of the reachable Network LSAs for the RPL Subnet, or in case of no Network LSAs, the lowest Advertising Router of reachable Router LSAs that have a stub link for the RPL subnet, plus the LSA Type. If the ABR belongs to multiple non-backbone areas and the RPL subnet is reachable in more than one of the areas, a single Summary LSA is originated. In that case, Advertising Router and LSA Type in the TLV is set according to the selected LSA in the area with the lowest Area ID.
- o When an ABR advertises a Type 3 Summary LSA into the non-backbone area A for a RPL subnet that it learns in the backbone area B, the Summary LSA MUST carry in a TLV according to [[I-D.acee-ospfv3-lsa-extend](#)] the Advertising Router of the LSA that identifies the elected active partition. If the LSA is a Network LSA or Router LSA, the LSA Type in the TLV is set accordingly. If it is a Summary LSA, the LSA Type is copied from the Summary LSA's TLV. The LSA Type is not really used, but included for consistency.
- o For the election process in the backbone area, Advertising Routers in the following ordered groups are compared, and the lowest Advertising Router in the first non-empty group is elected to identify the active partition.
 - * Of reachable Network LSAs for the RPL subnet
 - * Of reachable Router LSAs with stub link for the RPL subnet
 - * Carried in the above mentioned TLV of reachable Summary LSAs for the RPL subnet, with Network LSA type in the TLV

- * Carried in the above mentioned TLV of reachable Summary LSAs

for the RPL subnet with Router LSA type in the TLV

- o In a non-backbone area, the following group order is used instead, so that the partitions in the backbone area are always preferred.
 - * Carried in the above mentioned TLV of reachable Summary LSAs for the RPL subnet
 - * Of reachable local Network LSAs for the RPL subnet
 - * Of reachable Router LSAs with stub link for the RPL subnet
- o To determine if a router is on the active partition, the router checks if the active partition is identified by a Summary LSA. If yes, the Advertising Router from the TLV is used. Otherwise, the Advertising Router of the identifying Network LSA or Router LSA is used. Then, the same neighborship checking as in single area case is done to determine if the router is on the active partition.

[2.2.3.](#) Comparison between the two detection and election methods

The Host Route method universally works with any routing protocol w/o any signaling changes, and can work across AS boundaries. However, it requires advertising additional host routes, and the election is purely based on address comparison.

The other method works only with Link State Routing protocols and only works intra-AS. It needs IGP signaling extensions if multiple RPL segments need to be intentionally placed in different areas. For OSPF, currently the extension is only considered for OSPFv3. On the other hand, it only needs a little additional signaling for the intentional inter-area Anycast RPL deployment, and the election prefers the RPL segments in the backbone area, which may be desired.

[3.](#) IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

[4.](#) Security Considerations

This document does not introduce new security risks.

[5.](#) Contributors

[6.](#) Acknowledgements

[7.](#) References

[7.1.](#) Normative References

[I-D.acee-ospfv3-lsa-extend]

Lindem, A., Mirtorabi, S., Roy, A., and F. Baker, "OSPFv3 LSA Extendibility", [draft-acee-ospfv3-lsa-extend-02](#) (work in progress), September 2013.

[RFC2119] Bradner, S. ., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

[RFC4601] Fenner, B. ., Handley, M. ., Holbrook, H. ., and I. . Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", [RFC 4601](#), August 2006.

[RFC5015] Handley, M. ., Kouvelas, I. ., Speakman, T. ., and L. . Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", [RFC 5015](#), October 2007.

[7.2.](#) Informative References

[I-D.wijnands-pim-source-discovery-bsr]

Wijnands, I., Venaas, S., and M. Brig, "PIM flooding mechanism and source discovery", [draft-wijnands-pim-source-discovery-bsr-03](#) (work in progress), July 2013.

Authors' Addresses

Zhaohui (Jeffrey) Zhang
Juniper Networks, Inc.
10 Technology Park Drive
Westford, MA 01886

EMail: zzhang@juniper.net

Kurt Windisch
Juniper Networks, Inc.

EMail: kurtw@juniper.net

Internet-Draft

pim-bidir-rpl-resiliency

October 2013

Jaroslav Adam Gralak
Juniper Networks, Inc.

EMail: jgralak@juniper.net

