Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 12, 2011

IJ. Wijnands, Ed.
Y. Cai
Cisco Systems, Inc.

February 8, 2011

PIM neighbor reduction for transit LAN's. draft-wijnands-pim-neighbor-reduction-02

Abstract

PIM establishes a neighbor relationship with other routers directly connected to it on startup. Networks that are LANs or behave like a LAN, potentially create many PIM neighbors depending on how many routers are attached to it. If such a LAN is also a transit network (no directly connected source or receiver), many of the PIM procedures don't apply. This proposal describes a procedure to reduce the amount of neighbors established over a transit LAN.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on August 12, 2011.

Copyright Notice

Table of Contents

1. Introduction

PIM sends hello messages to discover other PIM enabled routers that are directly connected on a particular interface and form a PIM neighbor relationship with them. Various PIM procedures depend on having a PIM neighbor elected as Designated Router (DR), like for PIM register messages [RFC4601] and processing IGMP reports [RFC4604]. Most of these procedures are specific to either directly connected receivers or senders and do not apply to transit networks. Networks that are LANs or behave like a LAN (Mi-PMSI) [I-D.ietf-l3vpn-2547bis-mcast] create as many PIM neighbors as there are PIM enabled routers directly connected to that LAN. For networks where the sources and/or RPs are only in few locations, which is a very typical deployment, it's very likely that many of these PIM neighbors are never used as a target in any PIM J/P message. Combined with the fact that on transit networks there are no directly connected receivers or senders, having a PIM neighbor relationship with all the PIM routers over a transit LAN network seems unnecessary. It is however still useful to have a PIM neighbor relationship with PIM routers that are used as target in the PIM Join or Prune (J/P) messages. We'll discuss these later in this draft.

The proposal is to not form unessesary PIM neighbor relations by creating PIM neighbors dynamically on demand. Only PIM routers forwarding multicast data or on the path to the RP will be seen as a PIM neighbor. Other PIM routers on that LAN that act as receivers will stay passive and not form neighbor relationships. This will significantly reduce the number of PIM neighbors established over a LAN network where there are more passive receivers than there are senders. Networks that have directly connected senders and/or receivers are outside the scope of this draft.

1.1. Conventions used in this document

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].

1.2. Terminology

2. Reducing the number of PIM neighbors

PIM uses a unicast RIB to lookup the path to an upstream router for a particular Source or Rendezvous-Point (RP) [RFC4601]. The result of that lookup provides a directly connected next-hop that is used as the target in a PIM J/P message. [RFC4601] currently states that this next-hop also needs to be a PIM neighbor in order to send a PIM

J/P to it. However, whether you're not sending a PIM message because it is not a PIM neighbor or this upstream router is unable to parse the join, functionally does not make a big difference. The multicast tree can't be formed and traffic is interrupted. This draft proposes to send a PIM J/P to a target upstream router even if it is not a PIM neighbor. We also propose that a router accepts the PIM J/P and processes it as if it was received from a PIM neighbor. In most multicast deployments it is very likely that a next-hop for a source and an RP is also a PIM enabled router, so this is not considered to be a big issue. However, we do want to form a one-way PIM neighbor relationship with the target upstream router.

If a PIM router has a desire to send a PIM J/P to a non-PIM neighbor U, we propose to take one bit out of the PIM Join/Prune header 'Reserved' range and set it to 1 before we send the J/P packet. We call this bit the 'Hello Request' bit. A router that receives a PIM J/P with the 'Hello Request' bit on, sends a PIM hello out over the interface the PIM J/P was received on. The other routers on the LAN will receive the PIM hello and MAY form a one-way PIM neighbor relationship with U. A router that receives the Hello from U and has no interest in it MAY ignore the Hello to limit the amount of neighbor state. In the next PIM J/P the 'Hello Request' bit will be off because the PIM neighbor is known by the router sending the Join. Router U will continue to send periodic PIM Hello's out the interface as long as there is at least one downstream router joined over that interface for either a (*,G) or (S,G) state.

3. Bidir support

The support for PIM Bidir [RFC5015] on a LAN depends on the election of the Designated Forwarder (DF). The DF election mechanism has a few dependencies on PIM neighbors. [RFC5015] section 3.5.5 also describes a PIM Hello dependency on the DF election. For that reason routers that are bidir capable and a Candidate DF will send out a PIM Hello over that LAN. A PIM neighbor relationship will be established among the candidate DF routers. Note that a candidate DF router on a LAN is a router that has an RPF interface towards the RPA that is NOT on that same LAN. Please see [RFC5015] section 2.1 and 3.5.2 for details. It is expected that there are few Candidate DF routers and it's very likely these routers are already on the path to the RPA for the Sparse-Mode groups. We don't expect this procedure to add to the number of PIM neighbors that is etablished over that LAN.

4. Generation ID Hello option

PIM routers may use the generation ID in a PIM hello to make

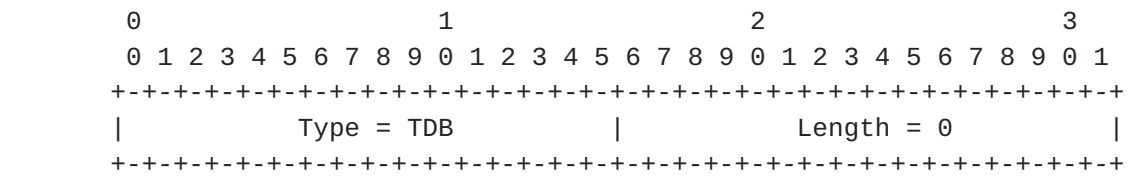
downstream router trigger PIM J/P's to it. This feature is used for upstream router High Avalability (HA) and when a router or interface becomes active. Using this feature there is no need to wait for the next periodic PIM J/P interval to (re)populate the forwarding state on the upstream router. If a Router or LAN becomes active, it is allowed to send a PIM hello on that LAN interface to speed up convergence, but it SHOULD not continue to send hello's periodically. Note, it's up to the downstream router(s) to either respond to this PIM Hello or ignore it if there is no interest in this PIM neighbor.

5. Hello suppression and options

PIM includes options in its Hello packets. We can group these options in two categories, options that are significant per neighbor or per LAN. For example, the GenID option is significant to the neighbor originating it, the Bidir option is significant to the LAN. Options that are significant per neighbor are learned as soon as a node has any interest that the neighbor. For these we don't need any special procedures. However, options relevant to the LAN, like Bidir capable or DR priority may not be learned because nodes on the LAN may suppress their Hello's using the procedures described in this draft. Its not important to know which nodes on the LAN support it or not, is good enough to know that at least one node does not support it. In order to discover the LAN specific options without creating PIM hello neighbor relationship between all the nodes we introduce the procedure below.

5.1. PIM suppress Hello option

We introduce a new PIM hello option called the PIM suppress option that is included in Hello's sent on the LAN. A PIM node on the LAN that receives this option in the PIM Hello (and supports it) will suppress its Hello if the set of included options match the options of this node. If this node has no interest in the sender of the Hello, no PIM neighbor relationship is created. The option set that this PIM neighbor advertised will be stored with an expire timer set to the advertised PIM hello holdtime. If this option set did already exist, only the option set expire timer is updated. The PIM hello periodic interval timer is started at the PIM hello interval time plus a random delay between 0 and 3 seconds. After the timer expires a PIM Hello is originated, unless a PIM Hello with the same set of options was received before the timer expired. This is similar to how PIM Join suppression works. With these procedures we are suppressing PIM hello's that share the same option set. Its likely that the PIM nodes on the LAN have the same option set, or at least have a limited set of option combinations. Below is the proposed PIM Hello suppress option encoding;



Type indicates PIM Hello suppression is supported.

5.2. Backwards compatibility

PIM nodes on the LAN that don't understand the suppress capability will obviously not suppress their Hello. They will just ignore the capability and create a PIM neighbor relation with the sender. This node does not expect other nodes to suppress their Hello so will assume that an upstream neighbor is not enabled with PIM. This may prevent PIM from sending PIM Join/Prunes. How this situation should be handled depends on the PIM implementation. Some implementations deployed in the field already ignore PIM neighbors for sending PIM Join/Prunes. For these implementations no special procedures are needed. Implementations that depend on PIM neighbors may only apply Hello suppression if all the PIM nodes on that LAN support the PIM suppress option. We propose the following two options to be supported;

As soon as one PIM node on the LAN does not support the suppress option all routers on the LAN will default back to sending periodic PIM hello's. Routers on the LAN continue to include the suppress option. As soon as all the routers on the LAN support the suppress option, PIM Hello suppression will be activated.

PIM hello suppression is always one and will not fall back to sending periodic PIM hello's.

6. Security Considerations

For securing PIM J/P messages please see the security section in [RFC4601].

7. IANA considerations

This document requests the reservation of a bit from the PIM Join/ Prune header reserved field. This bit field is called 'Hello Request' bit.

8. Acknowledgments

Thanks to Stig Venaas, Eric Rosen and Maria Napierala for their comments on the draft.

9. Contributing authors

Below is a list of the contributing authors in alphabetical order:

Yiqun Cai
Cisco Systems, Inc.
170 Tasman Drive
San Jose, CA, 95134
USA
E-mail: ycai@cisco.com

IJsbrand Wijnands
Cisco Systems, Inc.
De kleetlaan 6a
1831 Diegem
Belgium
E-mail: ice@cisco.com

10. References

10.1. Normative References

[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.

[RFC4604] Holbrook, H., Cain, B., and B. Haberman, "Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast", RFC 4604, August 2006.

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

[RFC5384] Boers, A., Wijnands, I., and E. Rosen, "The Protocol Independent Multicast (PIM) Join Attribute Format",

RFC 5384, November 2008.

10.2. Informative References

[I-D.ietf-l3vpn-2547bis-mcast]
Aggarwal, R., Bandi, S., Cai, Y., Morin, T., Rekhter, Y., Rosen, E., Wijnands, I., and S. Yasukawa, "Multicast in MPLS/BGP IP VPNs", draft-ietf-l3vpn-2547bis-mcast-10 (work in progress), January 2010.

Authors' Addresses

IJsbrand Wijnands (editor)
Cisco Systems, Inc.
De kleetlaan 6a
Diegem 1831
Belgium

Email: ice@cisco.com

Yiqun Cai
Cisco Systems, Inc.
170 Tasman Drive
San Jose CA, 95134
USA

Email: ycai@cisco.com