PIM Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 14, 2013

H. Liu
Huawei Technologies
T. Tsou
Huawei Technologies (USA)

July 13, 2012

PIM MTU Hello Option for PIM Message Encapsulation
draft-lts-pim-hello-mtu-01

Abstract

This memo introduces a new PIM Hello MTU Option which is carried in PIM Hello messages. The MTU option enables interface MTU information to be exchanged among PIM neighbors, and PIM messages to be encapsulated in an efficient and consistent way.

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 January 14, 2013.

Copyright Notice

Table of Contents

1. Introduction

A PIM router often needs to preserve a great many (*,G) or (S,G) multicast forwarding states to enable traffic forwarding for large scale of multicast channels. These states are usually set up and kept alive by each downstream router periodically sending Join Messages carrying its own forwarding states to its upstream neighbor. For each round of assembling these states into a PIM message, multiple segments of packets might be generated due to the MTU limitation on the sending PIM interface.

Current implementation uses merely sending link MTU to calculate maximum PIM packet length without considering the receiving MTU of the neighbor(s). It has some drawbacks because if the MTU of the sending interface is larger than that of the receiving one, PIM protocol packets encapsulated according to the sending MTU will most possibly be discarded by the receiving router and the forwarding states cannot be properly established as a result. There are already faults being reported caused by inconsistent MTU configuration among PIM neighbors.

Even though the problem could be resolved by requiring each PIM downstream interface to take less or equal MTU value than its upstream interface, it is inflexible for operation and does not scale because the interface or link conditions across the network might be diverse in practice. As a remedy, this memo recommends exchanging link MTU information among PIM neighbors by using a new Hello MTU Option. The option is carried in periodical PIM Hello messages for a router to inform its receiving link MTU parameter on an interface to the connected neighbor(s), so that the MTU information could be referenced by the neighbor(s) when they are sending PIM protocol messages on this link.

PIM MTU Option can be applied to all variants of PIM protocols, i.e., PIM-SM, PIM-SSM, PIM-DM, and BIDIR-PIM, on both IPv4 and IPv6 networks. There is an exception for the processing of PIM-SM Register/Register-Stop Message, which should reference the MTU information on the entire path between source DR and RP, as described in 4.4.1 of [RFC4601].

It should be noted that PIM MTU Option extension is different from multicast PMTU discovery mentioned in [RFC1981] . Section 5.2 of RFC1981 describes that an implementation could maintain a single PMTU learned across the whole multicast distribution tree. This might result in using smaller packets than necessary for a lot of paths. And because the end to end paths can be very dynamic it could make the effort too complex. This PMTU is used in encapsulating a 'multicast data packet' to avoid fragmentation in multicast data

plane as the packet travels on all paths of the tree. Whereas PIM MTU option works in control plane and has a per-hop nature - it only functions between adjacent one-hop PIM neighbors to guide the sending of a 'PIM protocol message'.

The maintenance of MTU in control plane (by PIM Hello MTU Option) and data plane (by PMTU) are for different purposes and are run independently - the control plane makes sure that forwarding paths are setup even there exists asymmetric MTUs on different links, while the data plane is to make multicast delivery efficient by avoiding fragmenting/reassembling operation, which could be done by means of acquiring minimal MTU on all paths, and of applying it in generating a data packet on first-hop or head-end. Control plane cannot preclude fragmentation, but it is the premise of normal data forwarding - even if some data packets exceeding limitation of some points of the paths cannot be processed properly, other packets meeting the PMTU requirements will be normally forwarded and delivered.

2. Terminology

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].

3. MTU Option and its Operation Rule

To record the minimum usable sending MTU value on an interface, a new General Purpose non-group-specific state - Sending MTU state is introduced in PIM protocols (for General Purpose State referring to 4.1.1 of [RFC4601] and [RFC3973], and 3.1.1 of [RFC5015]). It is 32-bit long and is unique on an interface whether the link connected is point-to-point or multi-accessed. The initial value of the Sending MTU state should be set to the outbound MTU of the interface, taking either the configured MTU or the default MTU value (referring to 7.1 of [RFC1191] for common MTU for different link types).

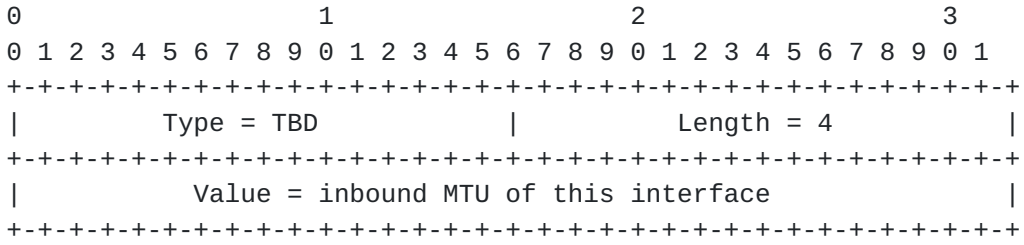
When an MTU Hello Option is received from a neighbor, a PIM router parses the MTU value in the option and decides whether or not it should accept the value and store it in the Sending MTU field. A router should not accept too small a value to prevent extreme fragmentation from deteriorating the router's performance. If the MTU value is valid from a legal neighbor, it compares the value with the MTU value currently stored in the Sending MTU field, and makes the replacement if the former is less than the latter.

Unlike other PIM Hello option, MTU Option is not required being supported simultaneously by all PIM neighbors connecting to a network. An MTU-capable router only considers the MTU of a trusty neighbor from which a valid MTU option is received. An MTU-capable PIM router should use MTU option in its Hello message, and should keep the Sending MTU state to the initial value if no neighbor reports a valid MTU Option. Finally, an MTU-incapable router should ignore an MTU option on reception.

The Sending MTU state should be checked before sending a multicast PIM message, to ensure the length of the message does not exceed the MTU limit of both the sending and receiving links. It should be noted that as a convention, the length calculation starts from the beginning of an IP header.

4. Option Format

A Hello MTU Option has the following format:



Type: to be assigned by IANA if this option is accepted. The field is 16-bit long.

Length: the length of the Value field. The field is 16-bit long.

Value: inbound MTU value for this interface. The field is 32-bit long.

5. IANA Considerations

The Type field should be allocated by IANA if MTU option is accepted.

6. Security Considerations

The potential security threat for MTU option should be the denial-of-service attack of extremely fragmenting PIM messages, by advertising much smaller MTU value than necessary. A remedy is to require a PIM router to check the validity of a neighbor's MTU value

before accepting it.

7. Acknowledgements

The authors would like to acknowledge Hou Yunlong, Mach Chen, Liu Yisong, Stig Venaas, Bill Fenner, Dino Farinacci, and Chiranjeevi Ramana Rao for their valuable comments and discussions on the work.

8. Normative References

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, November 1990.

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

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

[RFC3973] Adams, A., Nicholas, J., and W. Siadak, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)", RFC 3973, January 2005.

[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.

Authors' Addresses

Liu Hui
Huawei Technologies
Building Q14, No.156, Beiqing Rd.
Beijing 100095
China

Phone: 8610-60610012
Email: helen.liu@huawei.com

Tina Tsou
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara CA 95050
USA

Phone: +1 408 330 4424
Email: Tina.Tsou.Zouting@huawei.com