Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 22, 2011

D. Zhou, Ed.
Hangzhou H3C Tech. Co., Ltd.
H. Deng
China Mobile Research Institute
Y. Shi
Hangzhou H3C Tech. Co., Ltd.
H. Liu
Huawei Technologies Co., Ltd.
I. Bhattacharya
Cisco Systems

October 19, 2010

Unnecessary Multicast Flooding Problem Statement
draft-dizhou-pim-umf-problem-statement-01

Abstract

This document describes the unnecessary multicast stream flooding problem in the link layer switches between multicast source and PIM First Hop Router (FHR). The IGMP-Snooping Switch will forward multicast streams to router ports, and the PIM FHR must receive all multicast streams even if there is no request from receiver. This often leads to waste of switchs' cache and link bandwidth when the multicast streams are not actually required. This document details the problem and defines design goals for a generic mechanism to restrain the unnecessary multicast stream flooding.

1. Introduction

The Protocol Independent Multicast (PIM) is now the most popular multicast routing protocol in the world. The router on the edge of PIM routing domain is called PIM First Hop Router (FHR). PIM has four modes: Sparse Mode (SM), Dense Mode (DM), Bidirectional PIM (Bidir-PIM), Source-Specific Multicast (SSM). DM and Bidir-PIM suppose that the receivers are always existing. SM and SSM are designed for multicast streams to be transferred on demand.

The IGMP-Snooping, specified in RFC 4541 [RFC4541], is a link layer multicast streams forwarding control mechanism. It forwards all multicast streams to the router ports and selected multicast stream to membership ports. It also forwards IGMP Membership report messages to router ports and IGMP Query messages to all ports except the incoming port.

The PIM-Snooping is another link layer multicast streams forwarding control mechanism. It forwards the selected multicast streams to the requested router ports. But it can not run on the path between multicast source and PIM FHR because there is no PIM Join and PIM Prune messages.

In many typical deployment scenarios, some link layer switches are existing between multicast sources and PIM FHR. The receivers may be only exist between the source and PIM FHR, maybe exist in the network behind the PIM FHR, maybe even not exist temporarily. But if only the PIM FHR exists, the multicast streams are always transferred through these switches to PIM FHR, even if no receivers exist.

These unnecessary multicast streams will lead to waste of the switches' cache and link bandwidth. And the cache and link bandwidth are essential for application streams to transfer with less packet loss, latency, and jitter.

There are some attempts made to solve the problem of unnecessary multicast streams flooding on switches between the sources and PIM FHR in various ways. However, those solutions are either scenario-limited or deployment-limited.

This document provides a detailed description of protocol design goals for efficient PIM and PIM-Snooping based mechanism to solve this problem.

2. Terminology

In this document, several words are used to signify the requirements of the specification. These words are often capitalized. 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]

With respect to PIM, this document follows the terminology that has been defined in RFC 4601 [RFC4601].

3. Problem Statement

Under the existing model, the PIM FHR sends out PIM hello messages, as well as IGMP query messages if it is an IGMP querier in the network segment. The IGMP-Snooping switches between the sources and PIM FHR receive multicast streams from the sources and forward them to PIM FHR, even there is no request from receivers. If there are many sources, the multicast streams will flood to all PIM FHRs.

This will bring about some serious problems.

Firstly, the unnecessary multicast streams will seriously consume link bandwidth .

If there are many sources and PIM FHRs, the bandwidth between the source and PIM FHR will be seriously wasted. This problem will be more serious especially in such a type of application:

- The sources are much more than receivers.
- Each source may be requested simultaneously by some receivers.
- Most sources are NOT requested at most time.
- Receivers may be only exist between the source and PIM FHR, maybe exist in the network behind the PIM FHR, maybe even not exist temporarily.

In the application mentioned above, it is not acceptable to afford plenty of bandwidth to forward all multicast streams from the sources.

Secondly, the unnecessary multicast streams will consume outgress cache of switches.

In any network deployment, the switch between the sources and PIM FHR forwarding unnecessary multicast streams will also consume the outgress cache of switch including out-port specified cache and the cache shared by all ports. The cache is the key resource of switch to reduce streams' packet loss ratio, latency, and jitter.

Finally, the unnecessary multicast streams forwarding will increase the power consumption.

In summary, it is desirable to afford a mechanism to prohibit the switches between the source and PIM FHR from forwarding unnecessary multicast stream when it is not requested, and drive the switches to forward multicast stream in time when it is required.

4. Design Goals

The following are the goals and constraints in designing the mechanism for switch to restrain unnecessary multicast streams flooding:

- Switch SHALL forward the requested streams and SHALL NOT forward unrequired streams.
- Streams SHALL just be terminated at the exact switch.
- If a receiver appears, it MUST receive multicast streams in time.
- Deployment SHALL be flexible. The number and topology of switches between source and PIM FHR SHALL NOT be limited. The ip address deployment of multicast sources and receivers SHALL NOT be limited either. Sources and receivers may be in the same ip address segment, for example.
- The CPUs of switches SHALL receive no multicast stream data, but only protocol messages.

5. Use Cases and Related Work

In order to further clarify the items listed in scope of the proposed work, this section provides some background on related work and the use cases envisioned for the proposed work.

5.1. Source sending stream on demand

By adding a central controlling server, the multicast sources may be controlled to send streams on demand.

Note that once a receiver sends a request, the multicast stream will flow down toward switch's router ports, even if there is no other receivers behind the router ports.

5.2. Host simulation of PIM FHR

PIM FHR may be prohibited to send PIM hello messages and IGMP Query messages toward multicast sources. Instead, it can simulate host to send IGMP Membership Report and Leave messages if it receives PIM Join and Prune messages. So the switches between the sources and PIM FHRs would have no router ports.

But for PIM SM, the PIM FHR does not know at which port to send out IGMP messages, unless configured some information at the requested ports by network manager.

On the other hand, the switch will not forward IGMP Membership Report and Leave messages towards sources. It will only forward IGMP Membership Report and Leave messages to router ports.

5.3. IGMP Querier simulation of first-hop switch

For the second problem of PIM FHR host simulation, the switch directly connected to source can simulate IGMP Querier.

But when there are two or more switches simulating IGMP Queriers, the phenomenon of unnecessary multicast streams flooding still exists.

5.4. Replacing link layer switches with Routers

Replacing link layer switches directly connected to sources with Routers is not a perfect solution either. It will limit the flexibility of networking, and will further lead to waste of ip address and many ip address segments seriously if there are many sources. It will also bring about many ip address segments and then complicate network management.

6. some potential solutions

6.1. solution based on PIM and PIM-Snooping

The key points of it are as follows:

- When the PIM FHR receives a multicast stream, it creates an entry of (S,G) if the entry did not exist. And it judges whether the (S,G) entry has out interfaces. If the (S,G) has no out interface, the PIM router sends out an unicast PIM prune message towards the multicast source. The upstream neighbor address in the message is the source address.

- The switch between multicast sources and PIM FHR runs PIM snooping and IGMP snooping. When it intercepts the unicast PIM prune message by ip protocol field identification and finds out that the upstream neighbor address of the message is not in its PIM neighbor lists, it creates a (S,G) entry with a pruned port and an upstream port if not created before. The upstream port is found by looking up the unicast mac table. That (S,G) entry is punched with a specific sign which means that entry is different from traditional PIM-Snooping entry. The pruned port SHALL NOT forward multicast stream and has a lifetime which is 1/3 of that of PIM FHR's (S,G) entry, then converted to be a downstream port, so that the multicast stream will arrive at PIM FHR to refresh the (S,G) entry.

- Looking up IGMP-snooping entry and PIM-snooping entry, if the switch find there is no need to forward the multicast stream, it SHALL forward the unicast PIM prune message towards multicast source.

- When the switch receives an IGMP membership report, it shall forward the message through its router ports and upstream port.

- When PIM FHR creates an out interface for a (S,G) entry that had no out interface before, it shall send unicast PIM join message towards multicast source. The upstream neighbor address of the message is the source address.

- When the switch receives the unicast PIM join message and finds out that the upstream neighbor address of the message is not in its PIM neighbor lists, it will convert the pruned port to be downstream port. When the (S,G) entry with specific sign has no pruned ports, it should be deleted in order to save the entry space.

- By the information from IGMP-snooping entry and PIM-snooping entry, the switch can decide whether it shall forward the unicast PIM join message towards multicast source.

- The role of membership port is prior than that of pruned port, and the role of pruned port is prior than that of router port or downstream port.

- If two or more switches or PIM FHRs are connected by one port directly, or through HUB or normal switch, some query mechanism shall be implemented.

6.2. solution based on IGMP and IGMP-Snooping

The key points of it are as follows:

- When the PIM FHR receives a multicast stream, it creates an entry of (S,G) if the entry did not exist. And it judges whether the (S,G) entry has out interfaces. If the (S,G) has no out interface, the PIM router sends out an unicast IGMP prune message towards multicast source.

- The switch between multicast sources and PIM FHR runs IGMP snooping. When it intercepts the unicast IGMP prune message by ip protocol field identification, it creates a IGMP-Snooping entry with a pruned port and an source port. The source port is found by looking up the unicast mac table. The pruned port has a lifetime which is 1/3 of the lifetime of PIM FHR's (S,G) entry, so that the multicast stream will arrive at PIM FHR before its (S,G) entry dies out.

- By the information from IGMP-snooping entry, the switch can decide whether it shall forward the unicast IGMP prune message towards multicast source.

- When the switch receives an IGMP membership report, it shall forward the message through its router ports and source port.

- When PIM FHR creates an out interface for a (S,G) entry that had no out interface before, it shall send unicast IGMP graft message towards multicast source.

- When the switch receives the unicast IGMP graft message, it will change the pruned port to be router port. When the IGMP-Snooping entry has only router ports and source ports, it should be deleted in order to save the entry space.

- By the information from IGMP-snooping entry, the switch can decide whether it shall forward the unicast IGMP graft message towards multicast source.

- The role of membership port is prior than that of pruned port, and the role of pruned port is prior than that of router port.

- If two or more switches or PIM FHRs are connected by one port directly, or through HUB or normal switch, some query mechanism shall be implemented.

The first solution is more simple than the second one because the PIM-snooping has afforded some essential information and there is no need to add some new messages. In the first solution the switches must run PIM-snooping besides IGMP-snooping.

Any advice is welcome.

7. Security Considerations

8. Contributors

9. Acknowledgements

10. Normative References

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

[RFC4541] Christensen, M., Kimball, K., and F. Solensky, "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches", RFC 4541, May 2006.

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

Authors' Addresses

Di Zhou (editor)
Hangzhou H3C Tech. Co., Ltd.
310 Liuhe Road
Hangzhou, Zhejiang
China(310053)

Phone: +86-571-86761327
Email: zhoudi@h3c.com

Hui Deng
China Mobile Research Institute
Unit2,28 Xuanwumenxi Ave,Xuanwu District
Beijing, Beijing
China(100053)

Phone: +86-010-15801696688-3314
Email: denghui@chinamobile.com

Yang Shi
Hangzhou H3C Tech. Co., Ltd.
Beijing R&D Center of H3C, Digital Technology Plaza,
NO.9 Shangdi 9th Street,Haidian District,
Beijing
China(100085)

Phone: +86 010 82775276
Email: young@h3c.com

Hui Liu
Huawei Technologies Co., Ltd.
Huawei Bld., No.3 Xinxi Rd.
Shang-Di Information Industry Base, Hai-Dian Distinct,
Beijing
China(100085)

Email: Liuhui47967@huawei.com

Indranil Bhattacharya
Cisco Systems
India(560037)

Email: myselfindranil@gmail.com