IS-IS Flooding Reduction in MSDC

draft-xu-lsr-isis-flooding-reduction-in-msdc-02

Abstract

IS-IS is commonly used as an underlay routing protocol for MSDC (Massively Scalable Data Center) networks. For a given IS-IS router within the CLOS topology, it would receive multiple copies of exactly the same LSP from multiple IS-IS neighbors. In addition, two IS-IS neighbors may send each other the same LSP simultaneously. The unnecessary link-state information flooding wastes the precious process resource of IS-IS routers greatly due to the fact that there are too many IS-IS neighbors for each IS-IS router within the CLOS topology. This document proposes some extensions to IS-IS so as to reduce the IS-IS flooding within MSDC networks greatly. The reduction of the IS-IS flooding is much beneficial to improve the scalability of MSDC networks.

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 RFC 2119 [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 https://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 not supposed to be referenced in published informational documents.
Introduction

IS-IS is commonly used as an underlay routing protocol for Massively Scalable Data Center (MSDC) networks where CLOS is the most popular topology. For a given IS-IS router within the CLOS topology, it would receive multiple copies of exactly the same LSP from multiple IS-IS neighbors. In addition, two IS-IS neighbors may send each other the same LSP simultaneously. The unnecessary link-state information flooding wastes the precious process resource of IS-IS routers greatly and therefore IS-IS could not scale very well in MSDC networks.
To simplify the network management task, centralized controllers are becoming fundamental network elements in most MSDCs. One or more controllers are usually connected to all routers within the MSDC network via a Local Area Network (LAN) which is dedicated for network management purpose (called management LAN), as shown in Figure 1.

With the assistance of a controller acting as IS-IS Designated Intermediate System (DIS) for the management LAN, IS-IS routers within the MSDC network don’t need to exchange any IS-IS Protocol Datagram Units (PDUs) other than Hello packets among them. In order to obtain the full topology information (i.e., the fully synchronized link-state database) of the MSDC’s network, these IS-IS routers would exchange the link-state information with the controller being elected as IS-IS DIS for the management LAN instead.
To further suppress the flooding of multicast IS-IS PDUs originated from IS-IS routers over the management LAN, IS-IS routers would not send multicast IS-IS Hello packets over the management LAN. Instead, they just wait for IS-IS Hello packets originated from the controller being elected as IS-IS DIS initially. Once an IS-IS DIS for the management LAN has been discovered, they start to send IS-IS Hello packets directly (as unicasts) to the IS-IS DIS periodically. In addition, IS-IS routers would send IS-IS PDUs to the IS-IS DIS for the management LAN as unicasts as well. In contrast, the controller being elected as IS-IS DIS would send IS-IS PDUs as before. As a result, IS-IS routers would not receive IS-IS PDUs from one another unless these IS-IS PDUs are forwarded as unknown unicasts over the management LAN. Through the above modifications to the current IS-IS router behaviors, the IS-IS flooding is greatly reduced, which is much beneficial to improve the scalability of MSDC networks.

2. Terminology

This memo makes use of the terms defined in [RFC1195].

3. Modifications to Current IS-IS Behaviors

3.1. IS-IS Routers as Non-DIS

After the bidirectional exchange of IS-IS Hello packets among IS-IS routers, IS-IS routers would originate Link State PDUs (LSPs) accordingly. However, these self-originated LSPs need not to be exchanged directly among them anymore. Instead, these LSPs just need to be sent solely to the controller being elected as IS-IS DIS for the management LAN.

To further reduce the flood of multicast IS-IS PDUs over the management LAN, IS-IS routers SHOULD send IS-IS PDUs as unicasts. More specifically, IS-IS routers SHOULD send unicast IS-IS Hello packets periodically to the controller being elected as IS-IS DIS. In other words, IS-IS routers would not send any IS-IS Hello packet over the management LAN until they have found an IS-IS DIS for the management LAN. Note that IS-IS routers SHOULD NOT be elected as IS-IS DIS for the management LAN (This is done by setting the DIS Priority of those IS-IS routers to zero). As a result, IS-IS routers would not see each other over the management LAN. In other word, IS-IS routers would not establish adjacencies with one other. Furthermore, IS-IS routers SHOULD send all the types of IS-IS PDUs to the controller being elected as IS-IS DIS as unicasts as well.

To avoid the data traffic from being forwarded across the management LAN, the cost of all IS-IS routers’ interfaces to the management LAN SHOULD be set to the maximum value.
When a given IS-IS router lost its connection to the management LAN, it SHOULD actively establish adjacency with all of its IS-IS neighbors within the CLOS network. As such, it could obtain the full LSDB of the CLOS network while flooding its self-originated LSPs to the remaining part of the whole CLOS network through these IS-IS neighbor.

3.2. Controllers as DIS

The controller being elected as IS-IS DIS would send IS-IS PDUs as multicasts or unicast as before. And it SHOULD accept and process those unicast IS-IS PDUs originated from IS-IS routers. Upon receiving any new LSP from a given IS-IS router, the controller being elected as DIS MUST flood it immediately to the management LAN for two purposes: 1) implicitly acknowledging the receipt of that LSP; 2) synchronizing that LSP to all the other IS-IS routers.

Furthermore, to decrease the frequency of advertising Complete Sequence Number PDU (CSNP) on the controller being elected as DIS, it’s RECOMMENDED that IS-IS routers SHOULD send an explicit acknowledgement with a Partial Sequence Number PDU (PSNP) upon receiving a new LSP from the controller being elected as DIS.

4. Acknowledgements

The authors would like to thank Peter Lothberg and Erik Auerswald for his valuable comments and suggestions on this document.

5. IANA Considerations

TBD.

6. Security Considerations

TBD.

7. References

7.1. Normative References


7.2. Informative References


Authors' Addresses

Xiaohu Xu
Alibaba, Inc
Email: xiaohu.xxh@alibaba-inc.com

Luyuan Fang
Expedia, Inc
Email: luyuanf@gmail.com

Jeff Tantsura
Apstra, Inc.
Email: jefftant.ietf@gmail.com

Shaowen Ma
Juniper
Email: mashao@juniper.net