

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 15, 2015

G. Yan
Y. Liu
X. Zhang
Huawei
November 9, 2014

OSPF Synchronization Group
draft-yan-ospf-sync-group-01

Abstract

OSPF is a fundamental component for a routing system. It depends on the flooding mechanism to advertise and synchronize link-state database among distributed nodes in a network. As modern networks become larger and more complex, more and more nodes and adjacencies are involved. As a result, massive link-state information are generated and synchronized which are becoming an overhead of networks nowadays.

This document proposes a new design of OSPF database synchronization that is slightly different from the one stated in OSPF. This new design can help to alleviate the overhead by dividing OSPF routers into independent synchronization groups and limiting synchronization across the group border. Since less burden from synchronization, it is possible to accommodate more OSPF routers and adjacencies in a network.

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 <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 May 14, 2015.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) 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
2.	Terminology	3
3.	Problem Statement	3
4.	Proposed Solution	4
4.1.	Overview of a Synchronization Group	4
4.2.	LSA Synchronization in a Group	5
5.	Changes to the protocol	5
5.1.	Changes to the Flooding mechanism	5
5.2.	Route Calculation	6
5.3.	Protocol Extension	6
5.4.	Protocol Process	8
6.	Multi-homed SG consideration	8
6.1.	Problem Statement	8
6.2.	Proposed Solution	9
7.	Backward Compatibility	10
8.	IANA Considerations	10
9.	Security Considerations	10
10.	Acknowledgement	10
11.	References	10
11.1.	Normative References	10
11.2.	Informative References	11
	Authors' Addresses	11

[1.](#) Introduction

OSPF is a fundamental component for a routing system. It depends on the flooding mechanism to advertise and synchronize link-state database among distributed nodes in a network. As modern networks become larger and more complex, more and more nodes and adjacencies

are involved. As a result, massive link-state information are generated and synchronized which are becoming an overhead of networks nowadays.

This document proposes a new design of OSPF database synchronization that is slightly different from the one stated in [[RFC2328](#)]. This new design can help to alleviate the overhead by dividing OSPF routers into independent synchronization groups and limiting synchronization across the group border. Since less burden from synchronization, it is possible to accommodate more OSPF routers and adjacencies in a network.

In some scenarios, the routers in those networks suffer from limited CPU or storage resource which make them unqualified for large networks. With the help from this new design the situation can be improved.

2. Terminology

Synchronization Group (SG) : A sub-domain of one OSPF area in which the link-state database synchronization only happened among those routers in the same group.

Synchronization Group ID (SGID) : The identity of a Synchronization Group which MUST be unique in one OSPF network.

Synchronization Group Member (SGM) : One role of OSPF router which belongs to an unique Synchronization Group by carrying the SGID in its Hello packet. Adjacencies MUST NOT be established among SGMs from different SGs.

Synchronization Group Member Interface (SGMI) : The interface of a Synchronization Group Member.

Synchronization Group Director (SGD) : One role of OSPF router whose adjacencies MUST follow the standard procedure instead of affected by SGIDs.

Synchronization Group Director Interface (SGDI) : The interface of a Synchronization Group Director.

3. Problem Statement

As stated in [[RFC2328](#)], the flooding procedure supplied a reliable advertisement mechanism through which the link-state database is synchronized in an OSPF network. Forwarding loops or routing black-hole can be introduced if synchronization status is not achieved. There are some devices for which it is difficult to host the whole

link-state database since they may possess limited CPU or storage resource. Even for those devices which have enough resource, it is still an unneglectable overhead in a periodical manner.

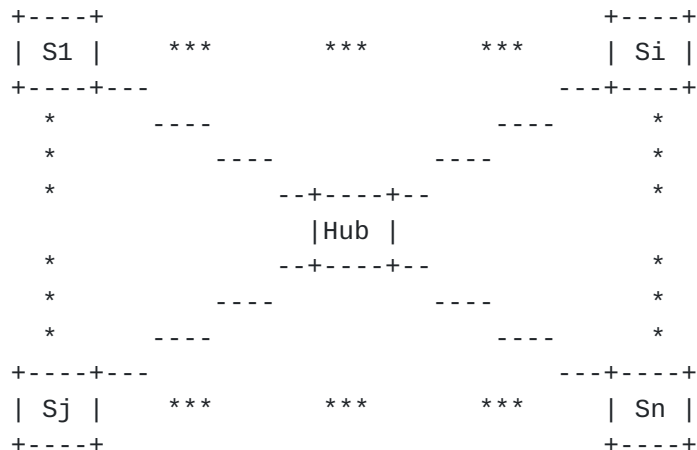


Figure 1 Hub and Spoke scenario

As showed in Figure 1, the hub sub-network established OSPF adjacencies with many spoke sub-networks indexed from S1 to Sn separately. Every LSAs generated by a single spoke have to be flooded to the rest of spokes through hub and vice versa. Let's assume there are m LSAs originated by each spoke then the total number of LSAs advertised among hub and spokes can be roughly counted as $m * n$, excluding the number of retransmission. What is worse, these LSA copies have to be refreshed every LSRefreshTime. This advertisement is indeed an unnecessary burden for devices with limited resources and even those devices with enough resources since all routes in one spoke share the same next hop which is the hub.

4. Proposed Solution

This document introduces a new mechanism which can solve the issue stated above through limiting synchronization scope inside a Synchronization Group instead of an area. The solution metioned here should be effective primarily in the hub-and-spoke scenario.

4.1. Overview of a Synchronization Group

A Synchronization Group (SG) is a sub-domain of one OSPF area in which the link-state database synchronization only happened among those routers in the same group. Each SG is identified uniquely by an identification number which is called SGID.

There are two roles involved into one Synchronization Group: Synchronization Group Member (SGM) and Synchronization Group Director

(SGD). The same SGD may be involved into several SGs simultaneously. Different SGDs are REQUIRED to interconnect with each other without passing through SGMs. The interfaces SGM and SGD used to form adjacencies are inherently called SGMI and SGDI. A SGMI or SGDI MUST belong to a single SG.

4.2. LSA Synchronization in a Group

Link-state database synchronization among SGDs follows the same procedure stated in [[RFC2328](#)]. They maintain the complete database of the area they belong to. This database is used to advertise among SGDs and consumed in the SPF calculation.

On the other side, SGMs only possess those LSAs that are learned from other SDMs and several LSAs leaked by their corresponding SGDs. SGMs advertise and use their LSDB in the manner as the standard document specified.

When OSPF adjacencies built between a SGD and a SGM, the synchronization between them SHOULD follow the specification defined in this document. In order to decrease the size of SGM's LSDB, a SGD only advertise necessary LSAs to its adjacent SGMs. Those LSAs in necessity include the Router-LSAs of SGDs in the same SG, the Network-LSAs if some of SGDs are DR for their corresponding networks and some Extended Prefix Opaque LSAs[I-D.ietf-ospf-prefix-link-attr] originated by SGDs to serve for limited reachability for SGMs.

5. Changes to the protocol

This document introduced some changes to OSPF[RFC2328] which is necessary to support SG.

5.1. Changes to the Flooding mechanism

SGDI and SGMI SHOULD be used to send and receive the LSAs updating in one SG. The LSA's SG belonging is identified by its originator's SGID. If MaxAge LSA is received, it SHOULD be processed as described in [section 13](#) of OSPF[RFC2328]. If a LSA is received from a neighbor that does not support SG, it SHOULD be processed as standardized since SG feature is ineffective between them.

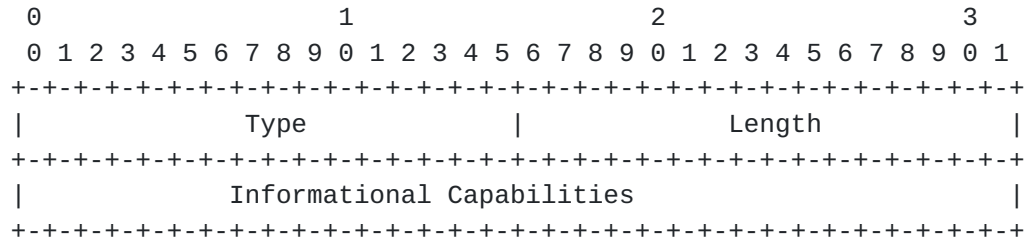
When LSDB synchronization happened between SGDs and SGMs, only limited LSDB SHOULD be flooded from SGDs to SGMs. As stated above, instead of flooding all LSAs to SGMs, only Router-LSAs and Network-LSAs in the same Group SHOULD be flooded to SGMs. On the contrary, SGMs SHOULD synchronize their whole LSDB to SGDs as standandized.

5.2. Route Calculation

No change introduced for route calculation in this document.

5.3. Protocol Extension

One new bit is introduced into Router Informational Capabilities TLV to indicate its originator supporting SG capability or not.



Bit TBD: OSPF Synchronization Group capable

Figure 2 SG-bit in Informational Capabilities TLV

A new TLV called Synchronization Group TLV is defined to be included in Router Information Opaque LSA[RFC4970]. Every router that support SG feature MUST contain this TLV in its RI Opaque-LSA. SGID and role of SGM or SGD can be learnt by parsing this TLV and act accordingly.

As shown below, when link between SGD2 and SG1A failed, the best path to reach SG1A is SGD2->SGD4->SG2A->SGD3->SGD1->SG1A. Since SG2A only have default route originated by its SGs, saying SGD3 and SGD4, forwarding loops can be observed. Even special handling was taken on SGD4, such as avoiding traveling through SDMs, traffic black hole could happen on SGD4 since SGD2 will insist on its choice. What is worse, assuming SGD2 generated the same prefix as SG1A did but with shorter prefix length, since SGD4 should ignore the link between SGD4 and SG2A that will cause transversal traffic, this shorter prefix will be the best match for the original destination, so forwarding loop can be observed between SGD2 and SGD4.

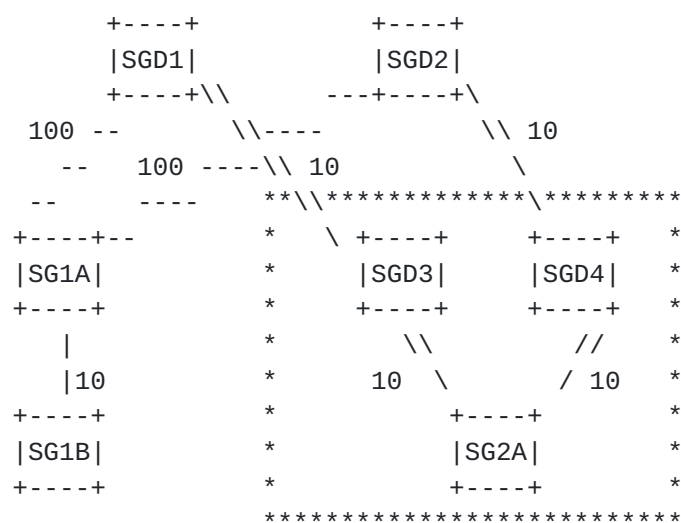


Figure 5 Multi-homed SG scenario

6.2. Proposed Solution

The root cause for the issue above is the inconsistent status of LSDB between SGs and SGMs. In order to solve this flaw, we may simply add one restriction to SGs that SG sub-networks can't be passed through to reach another SGD. With this restriction, inconsistent routing-table can be observed between SGs and the rest of networks in the same area, like SGD2 and SGD4 did above. Two solutions proposed here.

Solution I: SGs in the same SG are REQUIRED to automatically interconnect with each other using certain tunnels. The tunnel can be created when the SGD Router-LSA in the same SG is received. The traffic SHOULD be redirected into tunnels when the SGD finds the next hop points to one SGM. The exact tunnel type used here is out of the scope of this document.

Solution II: When a SG is multi-homed to multiple SGDs, SGDs and SGMs in the same SG SHOULD elect one Designated SGD (DSGD) from those candidate SGDs. Adjacencies SHOULD NOT be built between the non-designated SGDs and SDMs. A new DSGD SHOULD be elected among left candidates when the current DSGD failed.

With one of the solutions above, forwarding loops and traffic black hole are believed to be prevented.

7. Backward Compatibility

It is RECOMMENDED that SG feature is deployed all over the network at the same time. Otherwise It will work in the standardized manner without harm introduced into current network if partial deployment is used.

8. IANA Considerations

This document requests that IANA allocate from the OSPF TLV Codepoints Registry for a new TLV, referred to as the "Synchronization Group TLV".

9. Security Considerations

This document does not introduce any new security concerns to OSPF or any other specifications referenced in this document.

10. Acknowledgement

The authors would like to thank Eric Wu for his valuable suggestion on this draft.

11. References

11.1. Normative References

- [I-D.ietf-ospf-prefix-link-attr]
Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", [draft-ietf-ospf-prefix-link-attr-01](#) (work in progress), September 2014.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#), April 1998.

11.2. Informative References

- [RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", [RFC 4970](#), July 2007.
- [RFC5613] Zinin, A., Roy, A., Nguyen, L., Friedman, B., and D. Yeung, "OSPF Link-Local Signaling", [RFC 5613](#), August 2009.

Authors' Addresses

Gang Yan
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: yangang@huawei.com

Yuanjiao Liu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: liuyuanjiao@huawei.com

Xudong Zhang
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: zhangxudong@huawei.com

