Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 20, 2016

K. Raza
vIPtela
J. Cavanaugh
405Labs
A. Kulawiak
Morgan Stanley
P. Pillay-Esnault
F. Shamim
Cisco Systems
October 18, 2015

OSPF Stub Neighbors
draft-raza-ospf-stub-neighbor-02

Abstract

Open Shortest Path First stub neighbor is an enhancement to the protocol to support large scale of neighbors in some topologies with improved convergence behavior. It introduces limited changes protocol behavior to implement a scalable solution for hub and spoke topologies by limiting the functionality changes to the hub. The concepts are also applicable to a host running in a virtual machine environment.

Table of Contents

1. Introduction 2
2. Specification of Requirements 4
3. Incremental deployment 4
4. Link State Advertisement Filtering 4
  4.1. Area Border Router(ABR) Hub Routers 5
  4.2. Autonomous System Boundary Router (ASBR) Hub Routers . . 5
  4.3. Hub Routers which are neither ASBR or ABR 5
5. Proposed Changes 5
  5.1. Stub neighbor overview 5
  5.2. Local Adjacency 5
  5.3. Local Router LSA originated on the Hub Router 6
6. Hub Router Stub Neighbor Support Discovery and misconfiguration detection 8
7. Receiving and propagation of spoke routes 10
8. Demand Circuit 10
9. Benefits 10
10. Security Considerations 11
11. IANA Considerations 11
12. Acknowledgments 11
13. Normative References 11
Authors' Addresses 11

1. Introduction

With the growing size of an OSPF-network, most large networks are now deploying OSPF in large hub and spoke topologies. Also in lot of cases L3 routing would be extended to Top of rack or even to a host running virtual machines.

In any case these remote devices constitute a stub point in an OSPF network. These devices although being part of OSPF network will never be a transit point and thus do not need any topology information of the area nor do they require optimal routing calculations.

The spoke router in the case of a hub and spoke (or a host running OSPF) only need default route to the rest of the network, but they do need to send information about the connected network in the local site. In case of hosts they need to advertise routes in the virtual machines.

OSPF as network protocol was designed for an environment where routers were of similar capabilities. To protect the larger network, area hierarchy was introduced. Network was typically broken up into a backbone area and several subordinate areas. This breakup of the topology into areas serves multiple purposes

As OSPF has become pervasive protocol in the enterprise network it needs to evolve for large hub and spoke setups, these are typical retail environments. In a retail setup typical remote branch router does not have enough capacity to become part of a larger area, even if we break the network in large number of smaller areas. A remote router in one retail store does not need to have routes to all the router in other retail store that are part of its area setup.

Also increasing the number of areas on ABR can burden the ABR, this is due to the creation of large number of summary LSA. Although this can be handled by creating the areas as stub with no summary. Even by creating smaller sized areas with stub no summary, it does not completely eliminate the problem of having unnecessary information from the prospective of intra area.

With the advent of virtualized hosts, hosts are now advertising an increasing number of new virtual machine routes. These prefixes need to be advertised by a router that is connected to the host. Traditionally the host would be connected to the router via a shared link between the two (host and router). The host is often sourcing subnets that are not connected to the common subnet between the host and routers. However, the hosts (or spokes) themselves just need a default route from the router(or hub) to reach rest of the network. The solutions using current features of the protocol are not scalable. The overhead of protocol info and flooding of large number of unnecessary information to low-end routers caps the number of spokes on a hub.

This document describes extensions to OSPF to support very large Hub and spoke topologies more efficiently. Currently, the spoke router receives unnecessary information from the neighboring hub routers about all the other routers in the area. In most cases all a spoke router needs is IP reachability to hub routers which are the gateways to the rest of the network.

We presuppose familiarity with the contents of [RFC2328].

2. Specification of Requirements

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. Incremental deployment

For ease of deployment, the changes proposed in this document will be limited to the hub routers only.

By limiting changes only to the hub router the feature can be incrementally introduced without upgrading other routers in the network. Specifically, the spoke sites do not need to be upgraded.

It will be the responsibility of the hub router to mask the changes from the spoke as well as rest of the OSPF network such that the upgrading the network is simple from the point of interoperability and ease of deployment.

The hub router can be a normal router and there is no requirement for the hub to be a area border router or an autonomous system boundary router. Hub site is a sort of passive listener. It is there to receive routes from the spoke site, and to just provide exit towards rest of the network. A hub router SHOULD send a default or aggregated route towards the spoke and filters out all the information about rest of the network from the spoke.

4. Link State Advertisement Filtering

Routers establish adjacencies to flood topological information. The flooding process ensures all the information is consistent across the entire area and ensures the LSAs are delivered to all routers within the same area.

From the protocol prospective, topological information that is carried in the LSAs cannot be filtered, which it is essential to the loop free topology.

The topological information learned, by all routers within an area build the consistent graph of the network connections.

Vendors have implemented LSA filtering function on per neighbors basis specially for the purpose of scaling large full mesh environments. ISIS had the concept of mesh groups to avoid n2 flooding for a link failure and n3 flooding issue in case of node failure. LSA filtering gives the capability to filter information since it was done in the past in meshed topologies it was very

crucial that planning is done to make sure inconsistency does not happen inside of database thus causing loops.

Today prefix aggregation can only be achieved using summary type 3 or type 5 LSA. There is no way to limit or mask intra area information. The hub and spoke topologies or Data center cases, it would be beneficial to mask intra area information as it would not cause any loop.

4.1. Area Border Router(ABR) Hub Routers

In the case of hub routers being area border routers, aggregation can be achieved at the Hub router level using current features. The aggregation can be done by either using ranges or the default route injected as a type 3 LSA.

4.2. Autonomous System Boundary Router (ASBR) Hub Routers

In the case of hub routers being ASBR as well , aggregation can be achieved at the Hub router level using current features. The aggregation can be done by either using ranges or the default route injected as a type 5 or type 7 LSA.

4.3. Hub Routers which are neither ASBR or ABR

Currently there is no possibility of aggregating prefixes sent to the spoke routers and severely impact the scale.

5. Proposed Changes

5.1. Stub neighbor overview

We propose a new kind of adjacency for neighbors configured as stub. This adjacency will have a modified flooding content as the stub router only need a gateway through its neighbor. The hub router will send limited information to the remote spoke router without overwhelming the host with area topology. Another benefit is failures of the spoke node will be masked and would not impact the larger OSPF domain and other spoke nodes in the network. Spoke nodes SHOULD be considered a stub node when the remote site needs to send only prefixes to rest of the OSPF network without being considered a transit node.

5.2. Local Adjacency

The local adjacency concept in only present on a Hub router and it applies to those neighbors configured as stub neighbors. In this case, the hub router will maintain the adjacency to stub neighbors as

local only. Local adjacencies are not advertised in the normal router LSA flooded to other non-stub neighbors, thus masking the local adjacencies or stub nodes.

On the other hand, the hub router will flood a simplified router LSA to its local adjacencies so as to mask the area topology behind it. The Hub "Local" router LSA will contain only a p2p link to the stub neighbor when full adjacency is achieved and advertise one stub link with a configured range or the default prefix or both. The Hub router will effectively hide all the area topology including the prefixes behind it.

We are introducing a new type of default route with a local behavior. The current use of default route as type 3 or as type 5 cannot solve some of the use cases and more specifically in the Data center topologies.

The spoke router will function as normal advertising all its connected prefixes to Hub router.

5.3. Local Router LSA originated on the Hub Router

The local Router LSA MUST contain at least 2 links. One p2p link to the stub neighbor and a stub link to advertise the default prefix or a range defined per configuration.

```
Hub router-LSA for any area with default prefix
LS age = 0                     ;always true on origination
Options =          ;
LS type = 1                    ;indicates router-LSA
Link State ID = 192.0.2.1     ;Hub Router ID
Advertising Router = 192.0.2.1 ;Hub Router ID
bit E = 0                      ;not an AS boundary router
bit B = 0                      ;not area border router
#links = 2
Link ID = 192.0.2.2     ;Spoke Router ID.
Link Data = 192.0.2.1   ;Hub IP interface to net
Type = 1                ;connects to Point-to-point network
# TOS metrics = 0
metric = 1

Link ID = 0.0.0.0    ;Default prefix
Link Data = 0x0      ;Network mask
Type = 3                ;connects to stub network
# TOS metrics = 0
metric = 100
```

Hub router-LSA for any area with default prefix

```
   Hub router-LSA for any area with configured ranges
   LS age = 0                     ;always true on origination
   Options =           ;
   LS type = 1                    ;indicates router-LSA
   Link State ID = 192.0.2.1     ;Hub Router ID
   Advertising Router = 192.0.2.1 ;Hub Router ID
   bit E = 0                      ;not an AS boundary router
   bit B = 0                      ;not area border router
   #links = 2
   Link ID = 192.0.2.2     ;Spoke Router ID.
   Link Data = 192.0.2.1   ;Hub interface to net
   Type = 1                ;connects to Point-to-point network
   # TOS metrics = 0
   metric = 1

   Link ID = 198.51.100.0    ;Aggregated prefix
   Link Data = 0xffffff00  ;Network mask
   Type = 3                ;connects to stub network
   # TOS metrics = 0
   metric = 100

            Hub router-LSA for any area with configured ranges


   Hub router-LSA for any area with configured ranges
   LS age = 0                     ;always true on origination
   Options =           ;
   LS type = 1                    ;indicates router-LSA
   Link State ID = 192.0.2.1     ;Hub Router ID
   Advertising Router = 192.0.2.1 ;Hub Router ID
   bit E = 0                      ;not an AS boundary router
   bit B = 0                      ;not area border router
   #links = 2
   Link ID = 192.0.2.2     ;Spoke Router ID.
   Link Data = 192.0.2.1   ;Hub interface to net
   Type = 1                ;connects to Point-to-point network
   # TOS metrics = 0
   metric = 1

   Link ID = 0.0.0.0    ;Default prefix
   Link Data = 0x0      ;Network mask
   Type = 3                ;connects to stub network
   # TOS metrics = 0
   metric = 100

            Hub router-LSA for any area with configured ranges
```

A spoke router is usually a leaf node or in some cases may be in a dual-homed topology with another hub. In these cases, both Hub routers MUST be configured to view the spoke as a stub neighbor. The Local Router LSA of a Hub will get flooded over the other ospf interfaces of a spoke router. The Hub routers SHOULD ignore local router LSAs from other Hub routers flooded by a stub neighbor.

```
                          Rest of OSPF area
                                |
                                |
                                |       Normal RTR LSA
                                |<-- Normal link
                                |
                                |
                               HUB
        (site-1)              /     \               (site-2)
        HOSTS ------SPOKE1 ----+/       \+------- SPOKE2 ------ HOSTS
                          ^            ^
                          |            |
                       Stub Neighbor links

            Simplified HUB Local RTR LSA contains only p2p link
             and a stub link with default or configured range


                       Hub and Spoke Example 1
```

6. Hub Router Stub Neighbor Support Discovery and misconfiguration detection

To avoid the possibility of any routing loops and misconfigurations due to partial deployments, this draft defines a new OSPF Router Functional Capability known as a Hub Router Stub Neighbor Support Capability. The value of this capability is a bit value to be assigned by IANA from OSPF Router Functional Capability Bits registry [I-D.ietf-ospf-rfc4970bis].

The Auto Discovery via announcement of the Hub Router Stub Neighbor Support Functional Capability ensures that the detection of Hub Routers configured with the feature and advertising both a normal and a modified router LSA to a spoke.

The deployment scenario assumes that all hubs will be upgraded with the new functionality and configure their link to their the spokes appropriately to prevent the modified router LSA of any hub to be flooded back over normal links.

A hub router receiving back its own modified local router LSA over one of its non-stub neighbors is an indication of misconfiguration and it SHOULD revert back to normal mode or log an error so the operation can intervene.

If a hub router receive both a normal router LSA over a normal link and a modified router LSA with aggregation of a known Hub Router with stub neighbor support over a stub link then

1. It should acknowledge the LSA to form the adjacency but not flood it over its normal links

2. It MUST ignore the Local Router LSA and use the normal router LSA in its own SPF calculations

Any hub router receiving back a modified local router LSA over one of its non-stub neighbors is an indication of misconfiguration and it SHOULD log an error so the operation can intervene.

```
                       Rest of OSPF Area
                      |                |
                   n  |      n         | n
                    HUB1 ----------- HUB2
                       |             /   |
(site-1)               | s        s /    | s               (site-2)
HOSTS ------SPOKE1 ----+           /     +------- SPOKE2 ------ HOSTS
              |                   /
              |                  /
              |-----------------/
```

(s) Stub neighbor links and (n) normal links.

Hub and Spoke Example 2

The dual homed spoke1 will flood the Local RTR LSAs to the hub. The hubs will not propagate the flooding of local rtr lsa on any normal links. If one of the hubs is misconfigured then the originator can detect the misconfiguration if its receives the local router lsa over a normal link.

Implementations are encouraged to provide a knob to manually override and enforcement the functionality in partial deployment scenarios for cases where the topology guarantees that the router supporting the stub neighbor will not cause routing loops.

7. Receiving and propagation of spoke routes

Hub router upon receiving the route from the spoke SHOULD NOT treat that route as an intra area route. For interoperability reason rest of the network does not have to have any knowledge of this new adjacency.

A hub router that acts as an ABR just converts the entire stub neighbor routes as if they were part of an area. Since in case of OSPF area, id is not carried and only the Hub router understands that it is connected to stub neighbor it can convert all the stub neighbor and treat them as part of single area. Since the hub router is filtering all the LSA it is well aware of all the neighbors being part of the same area.

Hub router will be able to summarize at the area boundary. That way all the spokes could be summarized into a single route.

8. Demand Circuit

Sections 4.1, 4.2 described how to reduce the amount of information flooded and increase scalability. The use of Demand Circuit capability can further enhance the scalability for some use cases.

By making the spoke neighbors as demand circuit we will be able to suppress the refresh of all the routes we have learned from spoke sites. Only incremental changes are flooded in the network. Most networks have large number of spoke sites, in some large network there could be around 18-20K spoke sites each sending up to 3-5 subnets. Have to refresh these large number of LSAs can have unnecessary information flooded throughout large OSPF domain.

Second type of spoke sites that are emerging are running over long distance wireless networks. Sending periodic hellos for neighbor detection is not desired behavior in long distance wireless network. We do understand this can have convergence impact for the spoke that is dual homed.

9. Benefits

By making hub router define a stub neighbor we would be able to run OSPF in a true hub and spoke setup. Where the router that connects to the network and has local routes that needs to be advertising to rest of the network does not have to participate in the larger OSPF topology. Also the core network does not get destabilize due to flaps on the spoke churns causing impact on core convergence.

10. Security Considerations

This memo does not introduce any new security concerns or take any directed action towards improving the security of OSPF deployments in general. However, since all links in between OSPF neighbors do not add to router link states it could be considered as a security improvement by protecting an adjacency that can have larger network impact.

11. IANA Considerations

There are no IANA considerations.

12. Acknowledgments

This document was produced using Marshall Rose's xml2rfc tool.

13. Normative References

[I-D.ietf-ospf-rfc4970bis]
Lindem, A., Shen, N., Vasseur, J., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", draft-ietf-ospf-rfc4970bis-07 (work in progress), October 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <http://www.rfc-editor.org/info/rfc2328>.

Authors' Addresses

Khalid Raza
vIPtela
1735 Technology Drive
San Jose, CA 95110
USA

EMail: khalid.raza@viptela.com

John Cavanaugh
405Labs
6285 Lusk Blvd
San Diego, CA 92121
USA

EMail: John@405labs.com

Andrew Kulawiak
Morgan Stanley
1 New York Plaza
New York, NY 10004
USA

EMail: andrew.kulawiak@bankofamerica.com

Padma Pillay-Esnault
Cisco Systems
510 McCarty Blvd
Milpitas, CA 95035
USA

EMail: ppe@cisco.com

Faraz Shamim
Cisco Systems
2200 President George Bush TPKE
Richardson, TX 75082
USA

EMail: sshamim@cisco.com