Abstract

OSPFv3 is a candidate for deployments in environments where auto-
configuration is a requirement. One such environment is the IPv6
home network where users expect to simply plug in a router and have
it automatically use OSPFv3 for intra-domain routing. This document
describes the necessary mechanisms for OSPFv3 to be self-configuring.

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 10, 2013.

Copyright Notice

Copyright (c) 2012 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.
This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction .......................... 3
   1.1. Requirements notation ................ 3
   1.2. Acknowledgments ........................ 3
2. OSPFv3 Default Configuration ................ 4
3. OSPFv3 Router ID Selection ................ 5
4. OSPFv3 Adjacency Formation ................ 6
5. OSPFv3 Duplicate Router-ID Detection and Resolution ..... 7
   5.1. Duplicate Router-ID Detection for Neighbors ..... 7
   5.2. Duplicate Router-ID Detection for OSPFv3 Routers that are not Neighbors ............ 7
      5.2.1. OSPFv3 Router Auto-Configuration LSA ......... 7
      5.2.2. Router-Hardware-Fingerprint TLV ............... 9
   5.3. Duplicate Router-ID Resolution .............. 9
   5.4. Change to Received Self-Originated LSA Processing .... 10
6. Security Considerations .................... 11
7. Management Considerations .................. 12
8. IANA Considerations ........................ 13
9. Normative References ........................ 14
Authors’ Addresses .......................... 15
1. Introduction

OSPFv3 [OSPFV3] is a candidate for deployments in environments where auto-configuration is a requirement. Its operation is largely unchanged from the base OSPFv3 protocol specification [OSPFV3].

The following aspects of OSPFv3 auto-configuration are described:

1. Default OSPFv3 Configuration
2. Unique OSPFv3 Router-ID generation
3. OSPFv3 Adjacency Formation
4. Duplicate OSPFv3 Router-ID Resolution

1.1. Requirements notation

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-KEYWORDS].

1.2. Acknowledgments

This specification was inspired by the work presented in the Homenet working group meeting in October 2011 in Philadelphia, Pennsylvania. In particular, we would like to thank Fred Baker, Lorenzo Colitti, Ole Troan, Mark Townsley, and Michael Richardson.

Arthur Dimitrelis and Aidan Williams did prior work in OSPFv3 auto-configuration in the expired "Autoconfiguration of routers using a link state routing protocol" IETF Draft. There are many similarities between the concepts and techniques in this document.

Thanks for Abhay Roy and Manav Bhatia for comments regarding duplicate router-id processing.

Thanks for Alvaro Retana and Michael Barnes for comments regarding OSPFv3 Instance ID auto-configuration.

The RFC text was produced using Marshall Rose’s xml2rfc tool.
2. OSPFv3 Default Configuration

For complete auto-configuration, OSPFv3 will need to choose suitable configuration defaults. These include:

1. Area 0 Only - All auto-configured OSPFv3 interfaces MUST be in area 0.

2. OSPFv3 SHOULD be auto-configured on for IPv6 on all interfaces intended as general IPv6-capable routers. Optionally, an interface MAY be excluded if it is clear that running OSPFv3 on the interface is not required. For example, if manual configuration or an other condition indicates that an interface is connected to an Internet Service Provider (ISP), there is typically no need to employ OSPFv3. However, note that in many environments it can be useful to test whether an OSPFv3 adjacency can be established. In home networking environments, an interface where no OSPFv3 neighbors are found but a DHCP prefix can be acquired may be considered as an ISP interface.

3. OSPFv3 interfaces will be auto-configured to an interface type corresponding to their layer-2 capability. For example, Ethernet interfaces will be auto-configured as broadcast networks and Point-to-Point Protocol (PPP) interfaces will be auto-configured as Point-to-Point interfaces. Most extant OSPFv3 implementations do this already.

4. OSPFv3 interfaces MUST auto-configure the default HelloInterval and RouterDeadInterval as specified in [OSPFV3].

5. All OSPFv3 interfaces SHOULD be auto-configured to use an Interface Instance ID of 0 that corresponds to the base IPv6 unicast address family instance ID as defined in [OSPFV3-AF]. Similarly, if IPv4 unicast addresses are advertised in a separate auto-configured OSPFv3 instance, the base IPv4 unicast address family instance ID value, i.e., 64, SHOULD be auto-configured as the Interface Instance ID for all interfaces corresponding to the OSPFv3 instance [OSPFV3-AF].
3. OSPFv3 Router ID Selection

As OSPFv3 Router implementing this specification must select a unique Router-ID. A pseudo-random number SHOULD be used for the OSPFv3 Router-ID. The generation should be seeded with a variable that is likely to be unique in that environment. A good choice of seed would be some portion or hash of the Route-Hardware-Fingerprint as described in Section 5.2.2.

Since there is a possibility of a Router ID collision, duplicate Router ID detection and resolution are required as described in Section 5 and Section 5.3.
4. OSPFv3 Adjacency Formation

Since OSPFv3 uses IPv6 link-local addresses for all protocol messages other than message sent on virtual links (which are not applicable to auto-configuration), OSPFv3 adjacency formation can proceed as soon as a Router-ID has been selected and the IPv6 link-local address has completed Duplicate Address Detection (DAD) as specified in IPv6 Stateless Address Autoconfiguration [SLAAC]. Otherwise, there is no change to the OSPFv3 base specification except with respect to duplicate Router-ID detection and resolution as described in Section 5 and Section 5.3.
5. OSPFv3 Duplicate Router-ID Detection and Resolution

There are two cases of duplicate OSPFv3 Router-ID detection. One where the OSPFv3 router with the duplicate Router-ID is directly connected and one where it is not. In both cases, the resolution is for one of the routers with the duplicate OSPFv3 Router-ID to select a new one.

5.1. Duplicate Router-ID Detection for Neighbors

In this case, a duplicate Router-ID is detected if any valid OSPFv3 packet is received with the same OSPFv3 Router-ID but a different IPv6 link-local source address. Once that occurs, the OSPFv3 router with the numerically smaller IPv6 link-local address will need to select a new Router-ID as described in Section 5.3. Note that the fact that the OSPFv3 router is a neighbor on a non-virtual interface implies that the router is directly connected. An OSPFv3 router implementing this specification should assure that the inadvertent connection of multiple router interfaces to the same physical link in not misconstrued as detection of a different OSPFv3 router with a duplicate Router-ID.

5.2. Duplicate Router-ID Detection for OSPFv3 Routers that are not Neighbors

OSPFv3 Routers implementing auto-configuration, as specified herein, MUST originate an Auto-Config (AC) Link State Advertisement (LSA) including the Router-Hardware-Fingerprint Type-Length-Value (TLV). The Router-Hardware-Fingerprint TLV contains a variable length value that has a very high probability of uniquely identifying the advertising OSPFv3 router. An OSPFv3 router implementing this specification MUST compare a received self-originated Auto-Config LSA’s Router-Hardware-Fingerprint TLV against its own router hardware fingerprint. If the fingerprints are not equal, there is a Router-ID conflict and the OSPFv3 Router with the numerically smaller router hardware fingerprint MUST select a new Router-ID as described in Section 5.3.

This new LSA is designated for information related to OSPFv3 Auto-configuration and, in the future, could be used other auto-configuration information, e.g., global IPv6 prefixes. However, this is beyond the scope of this document.

5.2.1. OSPFv3 Router Auto-Configuration LSA

The OSPFv3 Auto-Configuration (AC) LSA has a function code of TBD and the S1/S2 bits set to B’01’ indicating Area Flooding Scope. The U bit will be set indicating that the OSPFv3 AC LSA should be flooded
even if it is not understood. The Link State ID (LSID) value will be a integer index used to discriminate between multiple AC LSAs originated by the same OSPF Router. This specification only describes the contents of an AC LSA with a Link State ID (LSID) of 0.

OSPFv3 Auto-Configuration (AC) LSA

The format of the TLVs within the body of an AC LSA is the same as the format used by the Traffic Engineering Extensions to OSPF [TE]. The LSA payload consists of one or more nested Type/Length/Value (TLV) triplets. The format of each TLV is:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Type             |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Value...                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

TLV Format

The Length field defines the length of the value portion in octets (thus a TLV with no value portion would have a length of 0). The TLV is padded to 4-octet alignment; padding is not included in the length.
field (so a 3-octet value would have a length of 3, but the total size of the TLV would be 8 octets). Nested TLVs are also 32-bit aligned. For example, a 1-byte value would have the length field set to 1, and 3 octets of padding would be added to the end of the value portion of the TLV. Unrecognized types are ignored.

The new LSA is designated for information related to OSPFv3 Auto-configuration and, in the future, can be used other auto-configuration information, e.g., global IPv6 prefixes.

5.2.2. Router-Hardware-Fingerprint TLV

The Router-Hardware-Fingerprint TLV is the first TLV defined for the OSPFv3 Auto-Configuration (AC) LSA. It will have type 1 and MUST be advertised in the LSID OSPFv3 AC LSA with an LSID of 0. It SHOULD occur, at most, once and the first instance of the TLV will take precedence over preceding TLV instances. The length of the Router-Hardware-Fingerprint is variable but must be 32 bytes or greater.

The contents of the hardware fingerprint should be some combination of MAC addresses, CPU ID, or serial number(s) that provides an extremely high probability of uniqueness. It MUST be based on hardware attributes that will not change across hard and soft restarts.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              1                |             >32               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Router Hardware Fingerprint                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
0
```

Router-Hardware-Fingerprint TLV Format

5.3. Duplicate Router-ID Resolution

The OSPFv3 Router selected to resolve the duplicate OSPFv3 Router-ID condition must select a new OSPFv3 Router-ID. After selecting a new Router-ID, the Router-LSA with the prior duplicate Router-ID MUST be purged. all self-originated LSAs MUST be reoriginated, and any OSPFv3
neighbor adjacencies MUST be reestablished.

5.4. Change to Received Self-Originated LSA Processing

RFC 2328 [OSPFV2], Section 13.4, describes the processing of received self-originated LSAs. If the received LSA doesn’t exist, the receiving router will purge it from the OSPF routing domain. If the LSA is newer than the version in the Link State Database (LSDB), the receiving router will originate a newer version by advancing the LSA sequence number and reflooding. Since it is possible for an auto-configured OSPFv3 router to choose a duplicate OSPFv3 Router-ID, OSPFv3 routers implementing this specification should detect when multiple instances of the same self-originated LSA are purged or reoriginated since this is indicative of an OSPFv3 router with a duplicate Router-ID in the OSPFv3 routing domain. When this condition is detected, the OSPFv3 Router SHOULD delay self-originated LSA processing for LSAs that have recently been purged or reflooded. This specification recommends 10 seconds as the interval defining recent self-originated LSA processing and an exponential back off of 1 to 8 seconds for the processing delay.
6. Security Considerations

A unique OSPFv3 Interface Instance ID is used for auto-configuration to prevent inadvertent OSPFv3 adjacency formation, see Section 2.

The goals of security and complete OSPFv3 auto-configuration are somewhat contradictory. When no explicit security configuration takes place, auto-configuration implies that additional devices placed in the network are automatically adopted as a part of the network. However, auto-configuration can also be combined with password configuration (see below) or future extensions for automatic pairing between devices. These mechanisms can help provide an automatically configured, securely routed network.

It is RECOMMENDED that OSPFv3 routers supporting this specification also offer an option to explicitly configure a password for HMAC-SHA authentication as described in [OSPFV3-AUTH-TRAILER]. When configured, the password will be used on all auto-configured interfaces with the Security Association Identifier (SA ID) set to 1 and HMAC-SHA-256 will be used as the authentication algorithm.
7. Management Considerations

It is RECOMMENDED that OSPFv3 routers supporting this specification also allow explicit configuration of OSPFv3 parameters as specified in Appendix C of [OSPFV3]. This is in addition to the authentication key configuration recommended in Section 6. However, it is acknowledged that there may be some deployment scenarios where manual configuration is not required.
8. IANA Considerations

This specification allocates a new OSPFv3 LSA, OSPFv3 Auto-Configuration (AC) LSA, TBD, as described in Section 5.2.1.

This specification also creates a registry for OSPFv3 Auto-Configuration (AC) LSA TLVs. This registry should be placed in the existing OSPFv3 IANA registry, and new values can be allocated via IETF Consensus or IESG Approval.

Three initial values are allocated:

- 0 is marked as reserved.
- 1 is Router-Hardware-Fingerprint TLV (Section 5.2.2).
- 65535 is an Auto-configuration-Experiment-TLV, a common value that can be used for experimental purposes.
9. Normative References


Authors’ Addresses

Acee Lindem
Ericsson
102 Carric Bend Court
Cary, NC  27519
USA

Email: acee.lindem@ericsson.com

Jari Arkko
Ericsson
Jorvas, 02420
Finland

Email: jari.arkko@piuha.net
Abstract

This document presents a topology-transparent zone in a domain. A topology-transparent zone comprises a group of routers and a number of links connecting these routers. Any router outside of the zone is not aware of the zone. The information about the links and routers inside the zone is not distributed to any router outside of the zone. Any link state change such as a link down inside the zone is not seen by any router outside of the zone.

Status of this Memo

This Internet-Draft is submitted to IETF 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 April 26, 2012.

Copyright Notice

Copyright (c) 2011 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 ............................................. 3
2. Terminology ............................................. 4
3. Conventions Used in This Document .......................... 4
4. Topology-Transparent Zone .................................. 4
  4.1. TTZ as a Group of Edge Routers Connected ................ 7
  4.2. TTZ as a Single Router ................................ 7
5. Changes to OSPF Protocols in Router LSA .................... 8
  5.1. One Bit to Indicate an Internal TTZ Link ................ 8
  5.2. A New Field for TTZ ID to Indicate a TTZ Link .......... 10
6. Constructing and Maintaining Router LSA ..................... 11
  6.1. Router LSA for a Router in TTZ ....................... 11
  6.2. Router LSAs for TTZ as a Group of Edge Routers .......... 12
  6.3. Router LSA for TTZ as a Router ....................... 12
    6.3.1. Selection of TTZ-DR for TTZ ........................ 12
    6.3.2. Constructing Router LSA for TTZ as a Router ........ 13
    6.3.3. Maintaining Router LSA for TTZ as a Router .......... 15
7. Establishing Adjacencies .................................. 15
  7.1. Group of Edge Routers for TTZ ....................... 16
  7.2. Single Router for TTZ ................................ 16
8. Distribution of LSAs ..................................... 17
  8.1. Distribution of LSAs within TTZ ....................... 17
  8.2. Distribution of LSAs through TTZ ..................... 17
9. Computation of Routing Table ................................ 18
10. Security Considerations .................................. 19
11. IANA Considerations .................................... 19
12.Acknowledgement .......................................... 19
13. References ............................................. 19
  13.1. Normative References ............................... 19
  13.2. Informative References ............................. 19
Authors’ Addresses .......................................... 20
1. Introduction

The number of routers in an Autonomous System (AS) becomes larger and larger as the Internet traffic keeps growing. Thus the Open Shortest Path First (OSPF) Link State Database (LSDB) and OSPF routing table are bigger and bigger. Any link state change in an AS leads to a number of link state distributions to every router in the AS. This triggers every router in the AS to re-calculate its OSPF routes, update its Routing Information Base (RIB) and Forwarding Information Base (FIB). All these will consume network resource including network bandwidth and Central Process Unit (CPU) time.

RFC 2328 "OSPF Version 2" describes OSPF areas in an AS. Each area has a number of area border routers connected to the backbone area. Each area border router summarizes the topology of its attached non backbone areas for transmission on the backbone, and hence to all other area border routers.

A big AS may be divided into a number of OSPF areas. However, some policies need to be applied on area border routers for reducing the number of link states such as summary LSAs to be distributed to other routers in other areas. In this case, a router in an OSPF area can see all other routers in the same area. A link-state change anywhere in an OSPF area will be populated everywhere in the same area, and may even be distributed to other areas in the same AS indirectly. For example, all the routers and links in a Point-Of-Presence (POP) in an OSPF area will be seen by all the other routers in the same area. Any link state change in the POP will be distributed to all the other routers in the same area and may be distributed to routers in other areas indirectly.

A link state change in an area will lead to every router in the same area to re-calculate its OSPF routes, update its RIB and FIB. It may also lead to a number of link state distributions to other areas. This will trigger routers in other areas to re-calculate their OSPF routes, update their RIBs and FIBs.

This document presents a topology-transparent zone in a domain or an area and describes extensions to OSPFv2 for supporting the topology-transparent zone. A topology-transparent zone comprises a group of routers and a number of links connecting these routers. Any router outside of the zone is not aware of the zone. The information about the links and routers inside the zone is not distributed to any router outside of the zone. Any link state change such as a link down inside the zone is not seen by any router outside of the zone.
2. Terminology

This document uses terminologies defined in RFC 2328, and RFC 2740.

3. Conventions Used in This Document

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.

4. Topology-Transparent Zone

A Topology-Transparent Zone (TTZ) comprises an Identifier (ID), a group of routers and a number of links connecting the routers. A Topology-Transparent Zone is in an OSPF domain.

The ID of a Topology-Transparent Zone (TTZ) or TTZ ID is a number that is unique for identifying a node in an OSPF domain. It is not zero in general.

The figure below illustrates an example of a routing domain containing a topology-transparent zone: TTZ 600.
The routing domain comprises routers R15, R17, R23, R25, R29 and R31. It also contains a topology-transparent zone TTZ 600. The TTZ 600 comprises routers R61, R63, R65, R67, R71 and R73, and the links connecting them.

There are two types of routers in a Topology-Transparent Zone (TTZ): TTZ internal routers and TTZ edge routers. A TTZ internal router is a router inside the TTZ and every adjacent router of the TTZ internal router is a router inside the TTZ. A TTZ edge router is a router
inside the TTZ and has at least one adjacent router that is outside of the TTZ and at least one adjacent router that is inside the TTZ.

The TTZ in the figure above comprises four TTZ edge routers R61, R63, R65 and R67. Each TTZ edge router is connected to at least one router outside of the TTZ. For instance, router R61 is a TTZ edge router since it is connected to router R15, which is outside of the TTZ.

In addition, the TTZ comprises two TTZ internal routers R71 and R73. A TTZ internal router is not connected to any router outside of the TTZ. For instance, router R71 is a TTZ internal router since it is not connected to any router outside of the TTZ. It is just connected to routers R61, R63, R65, R67 and R73 inside the TTZ.

A TTZ may hide the information inside the TTZ from the outside. It may not distribute any internal information about the TTZ to a router outside of the TTZ.

For instance, the TTZ in the figure above does not send the information about TTZ internal router R71 to any router outside of the TTZ in the routing domain; it does not send the information about the link between TTZ router R61 and R65 to any router outside of the TTZ.

In order to create a Topology-Transparent Zone (TTZ), we must configure the same TTZ ID on every link that connects routers inside the TTZ and every router in the TTZ must support TTZ feature.

For example, the same TTZ ID is configured on the nine links below:

- the link between router R61 and R65,
- the link between router R65 and R67,
- the link between router R67 and R63,
- the link between router R63 and R61,
- the link between router R71 and R61,
- the link between router R71 and R63,
- the link between router R71 and R65,
- the link between router R71 and R67 and
The link between router R71 and R73.

Thus six routers R61, R63, R65, R67, R71 and R73, and nine links among these six routers form a topology-transparent zone TTZ 600 in the figure above.

For a router in a TTZ, there is a little bit of changes in the format of the OSPF protocol data. In addition, there are some corresponding changes in the related OSPF protocol procedures.

There are a number of ways to make some of these changes. A TTZ may be virtualized as a different object in a different way. In one way, a TTZ may be virtualized as a group of edge routers fully connected. In another way, a TTZ may be seen as a single router. These virtualizations of TTZ is described below.

4.1. TTZ as a Group of Edge Routers Connected

From a router outside of the TTZ, a TTZ is seen as a group of TTZ edge routers, which are fully connected. For instance, router R15 in the figure above, which is outside of TTZ 600, sees TTZ 600 as a group of TTZ edge routers: R61, R63, R65 and R67. These four TTZ edge routers are fully connected. The cost from one edge router to another edge router is the cost of the shortest path between these two routers.

In addition, a router outside of the TTZ sees TTZ edge routers having normal connections to the routers outside of the TTZ. For example, router R15 sees four TTZ edge routers R61, R63, R65 and R67, which have the normal connections to R15, R29, R17 and R23, R25 and R31 respectively.

4.2. TTZ as a Single Router

A TTZ is seen as a single router from a router outside of the TTZ. For instance, router R15 in the figure above, which is outside of TTZ 600 and connected to TTZ 600 through TTZ edge router R61, sees TTZ 600 as a single router.

A router outside of a TTZ sees a number of links connected to the TTZ as a single router, each of which is connected to a router outside of the TTZ. For instance, router R15 sees TTZ 600 as a single router with six links, connecting to router R15, R17, R23, R25, R29 and R31 respectively.

A TTZ as a special single router considers every connection between a router outside of the TTZ and an edge router of the TTZ as a link. There are a few of options for the router ID of this special router.
o One option is that the TTZ ID is used as the router ID of the special router.

o Another option is that the biggest router ID among the router IDs of routers in the TTZ is selected as the router ID of that special router.

o The third option is that the smallest router ID among the router IDs of routers in the TTZ is used as the router ID of the special router.

The first option "the TTZ ID is used as the router ID of the special router" is recommended.

5. Changes to OSPF Protocols in Router LSA

There are a number of ways to extend the existing OSPF protocols to support TTZ in a router LSA. This section describes a couple of them.

o One way is to use one bit to indicate that a link is a TTZ link or an internal link of a TTZ in a router LSA.

o A another option is that a new field for a TTZ ID may be added into a router link to indicate that the link described belongs to which TTZ.

The first option "use one bit to indicate an internal TTZ link" is preferred.

5.1. One Bit to Indicate an Internal TTZ Link

A router LSA contains the description of a number of router links. The existing format of a router LSA is illustrated as follows:
Figure 2: Format of Router LSA

For a router link, the value of an eight bit Type field indicates the kind of the link. The value of the Type field may be 1, 2, 3 or 4, which indicates that the kind of the link is a point-to-point connection to another router, a connection to a transit network, a connection to a stub network, or a virtual link respectively.

The existing eight bit Type field for a router link may be split into two fields as follows:
I bit flag:

1: This indicates that the router link is an internal link to a router inside the TTZ.

0: This indicates that the router link is an external link.

Type-1:

The kind of the link.

Figure 3: Bit to Indicate Internal TTZ Link

For a link inside a TTZ, the value of I bit flag is set to one, indicating that this link is an internal TTZ link. For a link connecting to a router outside of a TTZ from a TTZ edge router, the value of I bit flag is set to zero, indicating that this link is an external TTZ link.

The value of Type-1 field may have value 1, 2, 3, or 4, which indicates that the kind of a link being described is a point-to-point connection to another router, a connection to a transit network, a connection to a stub network, or a virtual link respectively.

5.2. A New Field for TTZ ID to Indicate a TTZ Link

A new field for a TTZ ID is added into a router link to indicate that the link described belongs to which TTZ in the figure below.
If the value of the new field (TTZ ID) is zero, it indicates that the corresponding link is a link connecting to a router outside of the TTZ. If the value of the new field is not zero and is a TTZ ID, it indicates that the link described belongs to the TTZ given by the TTZ ID in the new field.

6. Constructing and Maintaining Router LSA

Two types of router LSAs are generated. One is constructed by every router in a TTZ for the router to describe the links connecting to it. The other is generated by some routers in the TTZ for the TTZ to virtualize the TTZ as a group of edge routers connected or a single router.

6.1. Router LSA for a Router in TTZ

Every router in a TTZ constructs a router LSA for the router that comprises the router links connecting the routers inside the TTZ and the router links connecting to the routers outside of the TTZ. It sends this router LSA to its neighboring routers in the TTZ. For each of the router links in the router LSA, it can be represented in one of the ways described in the previous section.

For example, when "One Bit to Indicate an Internal TTZ Link" is used as an extension to the link type, for each of the router links in the router LSA, the value of I bit flag is set to one for an internal link inside the TTZ; and the value of I bit flag is set to zero for an external link connecting to a router outside of the TTZ; the value
of Type-1 field may have value 1, 2, 3, or 4, which indicates that the kind of a link being described is a point-to-point connection to another router, a connection to a transit network, a connection to a stub network, or a virtual link respectively.

When a router inside a TTZ receives a link state such as a router LSA from a neighboring router in the TTZ, it stores the link state and floods the link state to the other neighboring routers in the TTZ.

When a TTZ edge router receives a TTZ internal link state such as a router LSA for a router inside the TTZ from a neighboring router in the TTZ, it stores the link state and floods the link state to the other neighboring routers inside the TTZ. It does not flood the link state to any of its neighboring routers outside of the TTZ.

6.2. Router LSAs for TTZ as a Group of Edge Routers

For every edge router of a TTZ, in addition to generate a router LSA described above, it constructs a second router LSA and sends this second router LSA to its neighboring routers. The second router LSA comprises two groups of links.

The first group of links are the router links connecting the routers outside of the TTZ from this TTZ edge router. These router links are normal router links. There is a router link for every adjacency between this TTZ edge router and a router outside of the TTZ.

The second group of links are the "virtual" router links. For each of the other TTZ edge routers, there is a "virtual" router link to it from this TTZ edge router. The cost of the router link from this TTZ edge router to one of the other TTZ edge routers is the cost of the shortest path from this TTZ edge router to it.

6.3. Router LSA for TTZ as a Router

6.3.1. Selection of TTZ-DR for TTZ

Every TTZ has a TTZ Designated Router (TTZ-DR). The TTZ-DR originates LSAs for the TTZ.

The TTZ-DR for a TTZ is elected as follows: When a TTZ router first becomes functional, it checks to see whether there is currently a TTZ-DR for the TTZ. If there is, it accepts that TTZ-DR, regardless of its router ID. Otherwise, the router itself becomes TTZ-DR if it has the highest router ID among all the TTZ routers.

The procedure for selecting a TTZ-DR for a TTZ is described as follows: Call the TTZ router doing the selection Router X. The TTZ...
routers in the TTZ are considered.

1. Set TTZ-DR and TTZ-BDR to 0.

2. Set TTZ-BDR to the router ID of the TTZ router that has the highest router ID among all the TTZ routers and is not the TTZ-DR.

3. Set TTZ-DR to TTZ-BDR if there is not a TTZ-DR in the TTZ.

4. If router X becomes a newly elected TTZ-DR or TTZ-BDR, or is now no longer the TTZ-DR or the TTZ-BDR, then repeat step 2. and 3. above.

5. If router X becomes the TTZ-DR, then it originates LSAs for the TTZ.

The procedure for selecting the TTZ-DR is triggered by the events: a new TTZ router becomes functional or an existing TTZ router leaves the TTZ.

6.3.2. Constructing Router LSA for TTZ as a Router

For the TTZ-DR in a TTZ, in addition to generate a router LSA described above, it constructs a second router LSA or special router LSA for the TTZ as a special single router and sends this second router LSA to its neighboring routers.

The second router LSA comprises all the router links connecting the routers outside of the TTZ from any TTZ edge router. The Link State ID and Advertising Router of the router LSA is the ID of the special router for the TTZ.

When the TTZ-DR in the TTZ constructs and sends an OSPF packet to its neighboring routers, it sets the Router ID in the packet header of the packet to the ID of the special router for the TTZ.

A procedure for constructing all the router links of a Special Router LSA (SRL) on the TTZ-DR is described below in pseudo code. From the point of view of the router outside of the TTZ, this Special Router LSA (SRL) does not contain any TTZ specific information, it is just a normal router LSA containing router links from the router for the TTZ to the routers outside of the TTZ.
N = 0;
For each router LSA in the TTZ
{
    For each router link in the router LSA
    {
        If the router link is an external link
        {
            N = N + 1;
            Add the router link into router LSA SRL as a normal link;
        }
        Else If the router link is a stub link
        {
            N = N + 1;
            Add the router link into router LSA SRL and set cost to 0;
        }
    }
}
Set the value of Number of Links field in router LSA SRL to N;

Figure 5: Procedure for Constructing Router LSA for TTZ

In the procedure above, N is a variable for counting the number of external links that the TTZ has. Each router LSA in the TTZ is a router LSA that is generated by a router inside the TTZ and is sent to routers inside the TTZ.

When "One Bit to Indicate an Internal TTZ Link" is used as an extension to the link type, the value of I bit flag is set to one for an internal link inside the TTZ, and the value of I bit flag is set to zero for an external link connecting to a node outside of the TTZ. Thus the condition of the If statement is true if the I bit flag in the router link is zero.

When the value of the link type stays the same as its original definition, but a new field for a TTZ ID is added into a router link to indicate that the link described belongs to which TTZ (If the value of the new field is zero, it indicates that the corresponding link is a link connecting to a node outside of the TTZ. If the value of the new field is not zero and is a TTZ ID, it indicates that the link described belongs to the TTZ given by the TTZ ID in the new field), the condition of the If statement is true if the value of the new field for the TTZ ID in the router link is zero.

In the body of the If statement, the router link for the external link is added into the router LSA SRL as a normal link. The value of Type field for the router link is set to the kind of the link for the
external link, which may be 1, 2, 3, or 4, indicating the kind of the link being described is a point-to-point connection to another router, a connection to a transit network, a connection to a stub network, or a virtual link respectively.

6.3.3. Maintaining Router LSA for TTZ as a Router

The router LSA for a TTZ as a router is originated only by the DR in the TTZ. Since only this DR generates the router LSA for the TTZ, the sequence number of the LSA is also generated only by this DR. Thus all the router LSAs for the same TTZ as a router will have the same sequence number.

When this DR receives a router LSA for the TTZ with a sequence number that is greater than the sequence number of the router LSA for the TTZ in its link state database, it re-originates a new router LSA for the TTZ with a new sequence number equal to the sequence number of the router LSA received plus one, and floods the new router LSA to all its neighboring routers.

When any other router in the TTZ, including any other TTZ edge router, receives a router LSA for the TTZ with a sequence number that is greater than the sequence number of the router LSA for the TTZ in its link state database, it just accepts the router LSA into its link state database and floods the router LSA to its neighboring routers.

7. Establishing Adjacencies

A router in a TTZ forms an adjacency with another router in the TTZ in the same way as a normal router when these two routers have a connection.

For an edge router in a TTZ, in addition to establishing adjacencies with other routers in the TTZ that have connections with the edge router in the same way as a normal router, it forms an adjacency with any router outside of the TTZ that has a connection with the edge router.

When the edge router in the TTZ forms the adjacency with the router outside of the TTZ, there are two options. One option is that it acts as a TTZ edge router, which is one of the group of edge routers for TTZ; the other is that it acts as a special single router for the TTZ.
7.1. Group of Edge Routers for TTZ

An edge router of a TTZ, acting as one of the group of edge routers for the TTZ, forms an adjacency with a router outside of the TTZ in a way described below.

During and after the adjacency establishment, every OSPF protocol packet such as Database Description, which is sent to the router outside of the TTZ by the edge router, contains the edge router identifier (ID) as Router ID.

When the edge router synchronizes its link state database with the router outside of the TTZ, it sends the router outside of the TTZ the information about all the LSAs except for the LSAs belong to the TTZ that are hidden from any router outside of the TTZ.

At the end of the link state database synchronization, the edge router originates its own router LSA and sends this LSA to the router outside of the TTZ. This router LSA contains two groups of links.

The first group of links are the router links connecting to the routers outside of the TTZ from this TTZ edge router. The second group of links are the "virtual" router links connecting to the other TTZ edge routers from this TTZ edge router.

From the point of view of the router outside of the TTZ, it sees the other end as a normal router and forms the adjacency in the same way as a normal router. It is not aware of anything about its neighboring TTZ. From the LSAs related to the TTZ edge router in the other end, it knows that the TTZ edge router is connected to each of the other TTZ edge routers and some routers outside of the TTZ.

7.2. Single Router for TTZ

An edge router of a TTZ, acting as a special single router for the TTZ, forms an adjacency with a router outside of the TTZ in a way described below.

During and after the adjacency establishment, every OSPF protocol packet such as Database Description, which is sent to the router outside of the TTZ by the edge router, contains the special single router ID as Router ID.

When the edge router synchronizes its link state database with the router outside of the TTZ, it sends the router outside of the TTZ the information about all the LSAs except for the LSAs belong to the TTZ that are hidden from any router outside of the TTZ.
At the end of the link state database synchronization, the router LSA for the TTZ is originated and sent to the router outside of the TTZ. This router LSA contains the router links from every TTZ edge router to routers outside of the TTZ.

From the point of view of the router outside of the TTZ, it sees the other end as a normal single router and forms the adjacency in the same way as a normal router. It is not aware of anything about its neighboring TTZ. From the LSAs related to the special router in the other end, it knows that the special router for the TTZ is connected to the routers outside of the TTZ having connections to edge routers of the TTZ.

8. Distribution of LSAs

LSAs can be divided into two classes according to their distributions. One class of LSAs is distributed within a TTZ. The other is distributed through a TTZ.

8.1. Distribution of LSAs within TTZ

Any LSA about a link state in a TTZ is distributed within the TTZ. It will not be distributed to any router outside of the TTZ.

For example, any router LSA generated for a router in a TTZ is distributed within the TTZ. It will not be distributed to any router outside of the TTZ.

Any network LSA generated for a broadcast or NBMA network inside a TTZ is distributed within the TTZ. It will not be distributed to any router outside of the TTZ.

Any opaque LSA generated for a TTZ internal TE link is distributed within the TTZ. It will not be distributed to any router outside of the TTZ.

8.2. Distribution of LSAs through TTZ

Any LSA about a link state outside of a TTZ received by an edge router of the TTZ is distributed through the TTZ; and any LSA about a link state for the TTZ is distributed through the TTZ.

For example, when an edge router of a TTZ receives an LSA for a link state outside of the TTZ from a router outside of the TTZ, it floods it to its neighboring routers both inside the TTZ and outside of the TTZ. This LSA may be any LSA such as a router LSA and an opaque LSA that is distributed in a domain.
The routers in the TTZ continue to flood the LSA. When another edge router of the TTZ receives the LSA, it floods the LSA to its neighboring routers both outside of the TTZ and inside the TTZ.

In the case that a group of edge routers is for a TTZ, every edge router of the TTZ generates a router LSA for the TTZ. This LSA is distributes through the TTZ. That is that it is distributed to routers outside of the TTZ and to the routers inside the TTZ.

In the case that a single router is for a TTZ, the special router LSA generated for the TTZ is distributes through the TTZ.

9. Computation of Routing Table

The computation of the routing table on a router outside of a TTZ is the same as that described in RFC 2328. On a router inside the TTZ, it has the same procedure flow as that described in RFC 2328, but extends the meaning of a link and an association between two vertexes. In this section, we specify the extensions, and describe the the routing table computation on a router inside the TTZ.

A link between two vertexes can be a TTZ link. It can also be a normal link.

When examining the LSA associated with vertex V, for each link described in the LSA, supposing that vertex W is the other end of the link,

- if it is a normal link, then vertex W is an adjacent vertex of vertex V;
- if it is an internal TTZ link and the LSA is generated by a router in a TTZ, then vertex W can be considered as an adjacent vertex of vertex V;
- if it is an external TTZ link and the LSA is generated by a TTZ edge router, then vertex W, which is the other end of the external TTZ link and outside of the TTZ, can be considered as an adjacent vertex of vertex V.

When a TTZ is virtualized as a group of TTZ edge routers fully connected, the routing table on a router inside the TTZ is computed in two phases.

In the first phase, the shortest path to every destination inside the TTZ is computed through using the internal TTZ links (i.e., the links inside the TTZ). Then in the second phase, the shortest path to
every destination outside of the TTZ is computed over all the links including the links inside the TTZ and the links outside of the TTZ.

When a TTZ is virtualized as a special single router, the routing table on a router inside the TTZ is computed in the following way.

The cost/metric of a link (including external TTZ link) outside of the TTZ is considered as a special type of metrics. This type of metrics is an order of magnitude larger than that of metrics of a link inside the TTZ. That is that any metric of this special type is considered greater than the cost of any path internal to the TTZ. The path to every destination is computed through constructing a shortest path tree from the router in the TTZ to every destination.

10. Security Considerations

The mechanism described in this document does not raise any new security issues for the OSPF protocols.

11. IANA Considerations

12. Acknowledgement

The author would like to thank Acee Lindem, Dean Cheng, Lin Han and Yang Yu for their valuable comments on this draft.

13. References

13.1. Normative References


13.2. Informative References


Authors’ Addresses

Huaimo Chen
Huawei Technologies
Boston, MA
USA

Email: Huaimochen@huawei.com

Renwei Li
Huawei Technologies
2330 Central expressway
Santa Clara, CA
USA

Email: Renweili@huawei.com

Gregory CAUCHIE
France Telecom
38-40 avenue du General LECLERC
Issy-les-Moulineaux 92130,
FRANCE

Email: gregory_pro@orange.fr
Service Distribution using OSPF
draft-pillay-esnault-ospf-service-distribution-00

Abstract

The Open Shortest Path First (OSPF) protocol is used to carry data on behalf of other services using the Opaque Link State Advertisements. The protocol’s flooding mechanism is well suited to cover the data propagation requirements of services such as Traffic Engineering. The current mechanism cannot scale for a large number of services nor satisfy some of their new set of requirements. This document describes a new mechanism in OSPF to support service and data distribution for a large number of services, computation of preferred service access points and a controlled service data exchange.

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 September 7, 2012.

Copyright Notice

Copyright (c) 2012 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 ................................................. 3
2. Specification of Requirements ............................... 3
3. Requirements for service data propagation .................. 3
4. Typical Scenario for Services Distribution Router .......... 4
5. OSPF Service Distribution Router ............................ 5
6. Storage Of Service Data ..................................... 6
7. Mechanics of the OSPF Service Information Distribution Implementation ............................................. 7
   7.1. Advertising and Signaling of SDR Capability ............. 7
   7.2. Advertising the Service Distribution Router and its address mapping ......................................... 8
   7.3. Advertising the Directory of Producers and Consumers .... 9
   7.4. Service Routing Capable Router Operations ............... 12
       7.4.1. Operation due to Producer changes ................... 13
       7.4.2. Operation due to Consumer changes ................... 13
8. Calculation of Optimal Producer .............................. 14
9. Service Router Data Operations .............................. 14
   9.1. Implementation of SDDA .................................. 15
   9.2. SDR Message Formats .................................... 15
   9.3. SDR Message Header Format ................................ 16
   9.4. OPEN Message Format .................................... 16
   9.5. CONFIRM Message Format .................................. 18
   9.6. UPDATE Message Format ................................... 18
   9.7. NOTIFICATION Message Format ............................. 20
       9.7.1. Message Header Error ................................. 21
       9.7.2. OPEN Message Error .................................. 21
       9.7.3. CONFIRM Message Error ............................... 22
       9.7.4. UPDATE Message Error ............................... 22
       9.7.5. Cease .............................................. 22
10. Security Considerations ..................................... 22
11. IANA Considerations ....................................... 23
12. Contributors ................................................. 23
13. Acknowledgments ............................................. 23
14. Normative References ....................................... 23
1. Introduction

Originally, routing protocols were designed to propagate routing related information only. With the advent of Traffic Engineering, the IGP s started to be used as a transport mechanism. Most of the applications using IGP s as transport are still very much limited and confined to routing applications with similar requirements.

Today, OSPF can carry data for applications using Opaque LSAs. These Opaque LSAs are an integral part of the OSPF database and will be flooded, synchronized and updated just as any other LSA. However, they do not contribute directly to any routes or trigger an SPF.

Opaque LSAs will need to be flooded across all the OSPF area or domain and neighbor adjacencies to FULL state will depend on successful exchange of these LSAs.

The Link State IGP s are limited on the size of payload information they can carry as it will be flooded and stored in every single router all across their areas or domain regardless whether it is of interest or not.

This document describes a new mechanism in OSPF to support service and data distribution for a large number of services, computation of preferred service access points and controlled distribution of service data.

We presuppose familiarity with the contents of [RFC4970], [RFC5250], [RFC2328] and [RFC5340].

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. Requirements for service data propagation

Services requirements differ from the traditional routing information dissemination model. The service data may be unrelated to routing or be of interest only to some routers. The new set of requirements for using OSPF as Service Distribution Router (SDR) is as follows

Scale to a large number of services

No assumption regarding size, format or nature of the data
No assumption regarding topology
Routing and service data separated and independent
Must support cases where minimal number of routers only may be upgraded
Must support dynamic events
Routers only store and process data of interest
Ability to compute the shortest path to a producer or consumer of a service per IGP metrics or service metrics.
There is no assumptions regarding producers and consumers of services, their location or uniqueness.
Secured data may reside only on some routers.
In addition the routing requirements for OSPF as Service data distribution are
Backward compatible with Open Standard OSPF
Minimal/No impact on routing convergence and performance

4. Typical Scenario for Services Distribution Router

A SDR is typically reachable by multiple consumers or producers of data. The router itself may not be connected directly to any other router with Service Distribution Capability. The intermediate routers may have limited storage capability or cannot store the data for security reasons.

The SDR is aware of the topological information of the other service routers and can compute paths to the preferred Producer SDR (PSDR) or the Consumers SDR (CSDR) of a service. The SDR will implement tables of producers and consumers for services.

The SDR ensures that interested subscribers to a service are notified with the latest updates.

Producers or consumers can join or leave a service at any time using APIs. The SDR receiving the notification of "registration" or "de-registration" flood the change of state to all the known SDRs in its topology. Therefore, all SDR have the same view of the producers/consumers topology.
5. OSPF Service Distribution Router

A SDR leverages OSPF’s capability to store and flood the topology and other attributes of SDR capable routers. SDRs form an overlay and do not require to be directly connected to each other. SDRs do not need to maintain adjacency between them other than the normal OSPF adjacency for routing purposes. The SDRs rely on the OSPF underlying network for reachability to other SDR routers.

SDRs advertise a directory of producers and consumers of services and are capable to compute preferred producers. The SDRs delegate data exchange processing to remote SDRs to an external agent. This agent is described in detail in section 9 of this document.

The OSPF Opaque LSAs is used to carry relevant and interesting information for reachability and nature of SDR capable routers.

In order to limit the service data dissemination costs (storage, bandwidth, security, ..), SDRs may store only data of interest.
The OSPF Service Distribution Router

The following sections describe the extensions in OSPF protocol to support this capability.

6. Storage Of Service Data

The service data can be stored in an independent Service Data Database(SDD). There is no assumption made here on the size, format or nature of data. The data can even be stored on the disk of the router and accessible by APIs to OSPF and other applications for
query and update. A SDD is not part of OSPF and does not participate in the bringing up of adjacencies.

It is desirable that the service data database have a very flexible format to cater for a broad range of applications. A possible solution is that the database records be defined as container objects which themselves contain service metadata.

7. Mechanics of the OSPF Service Information Distribution Implementation

7.1. Advertising and Signaling of SDR Capability

The OSPF SDR router will identify itself to the rest of the domain by advertising its capability and a routable ip address. For example, this address MAY be a loopback interface configured to uniquely identify an OSPF SDR router. A new bit for SDR capability is reserved in the Router Information Capabilities TLV of the Router Information LSA, as defined in section 2.1 of [RFC4970].

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |     Options   |  11           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       4       |                    0                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             LS sequence number                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

The OSPF Router Information LSA

Flooding scope for AS 11

The format of the Router Informational Capabilities TLV is defined in 2.3 and 2.4 of [RFC4970]
The Router Informational Capabilities TLV

A new informational capability bit is defined for Service Distribution Routers

Bit Capabilities

6 Service Distribution Router Capability

7.2. Advertising the Service Distribution Router and its address mapping

A new TLV is defined in the Router Information LSA is used to advertise a routable address to reach the router.

Service router TLV and address Mapping

Type: A 16-bit field set to 2 representing the Service Distribution Router Address Mapping This TLV is applicable both to OSPFv2 and OSPFv3.
Length: A 16-bit field that indicates the length of the value portion in octets.

Address Format: A 16-bit field that indicates the length of the value portion in octets.

Possible values

1 : IPv4 Address
2 : IPv6 Address

Address Length: A 16-bit field that indicates the length of the value portion in octets.

Address : Routable IPv4/IPv6 address mapped to SDR

SDR metric: A 16-bit field that indicates SDR metric greater than 0.

Type of metric:

0 : None defined - Ignore SDR Metric
1 : SDR metric overrides the IGP metric
2 : Computed metric is composite of IGP metric + SDR metric

7.3. Advertising the Directory of Producers and Consumers

Opaque LSAs with autonomous system flooding scope as described in [RFC5250], are used to describe the services reachable through this router using TLVs.

Definition of TLV
Example of TLV for Directory of Producers and Subscribers

Type: A 16-bit field set to 3 representing the Directory of the Service Distribution Router. This TLV is applicable both to OSPFv2 and OSPFv3.

Length: A 16-bit field that indicates the length of the value portion in octets.

Services are described in a unique sub-TLV. The sub-TLV should contain a Service Identifier which uniquely identifies the service with network wide significance. The sub-TLV format should be flexible and it MAY be used to advertise a preference metric for the service.
Service Description Sub-TLV

Type: A 16-bit field set to 1 representing the Service Description Sub-TLV This TLV is applicable both to OSPFv2 and OSPFv3.

Length: A 16-bit field that indicates the length of the value portion in octets.

Service ID: A 32-bit field representing the Service Identifier. This TLV is applicable both to OSPFv2 and OSPFv3.

Service Metric: A 16-bit field that indicates the metric associated with the service. A metric of 0 would represent undefined. An unreachable or oversubscribed service has a metric of 0xFFFFFFFF.

Type of metric:

0 : None defined - Ignore Service Metric

1 : Service metric overrides the IGP/SDR metric

2 : Computed metric is composite of IGP metric + SDR metric + Service metric

The Services of interest (Consumers exist) are described in a unique sub-TLV. The sub-TLV should contain a Service Identifier which uniquely identifies the service with network wide significance. The sub-TLV format should be flexible and MUST contained the preferred SDR ID. If no producer exists yet for the service then the Preferred SDR ID should be set to 0.
Service Subscription Sub-TLV

Type: A 16-bit field set to 2 representing the Service Subscription Sub-TLV. This TLV is applicable both to OSPFv2 and OSPFv3.

Length: A 16-bit field that indicates the length of the value portion in octets.

Service ID: A 32-bit field representing the Service Identifier. A Service Identifier may only be defined in a unique sub-tlv. This TLV is applicable both to OSPFv2 and OSPFv3.

PSDR ID: A 32-bit field that indicates the PSDR for data exchange. Set to 0 if there is no producer for the service.

The topological view and characteristics of the OSPF Overlay Service Distribution Routers can be used to compute preferred producer independent of IGP metrics. It is possible to have multiple LSAs for large directories however a service must be described in a unique sub-tlv for the SDR.

7.4. Service Routing Capable Router Operations

The additional requirements are

No assumption on topology

Multiple producers may exists

Multiple consumers can all have different service interest

Producers/Consumers may join and leave at anytime

Consumers and Producers have access to the Service Data database

The SDR capable routers advertise the consumers who subscribe to a
service. The producers may connect to the SDR router to update the services/data in the Service Data database. The SDR router then builds the Opaque LSA describing the producer services which are reachable through it as well as the services its consumers are interested in.

When the router has a full neighbor relationship, it now has the topological view of all SDR capable routers in the domain as well as the services they offer and are interested in.

Leveraging the fact that the OSPF has already run its SPF, the reachability of overlay SDR capable routers and services offered. It is possible to calculate the preferred Producer SDR for a service by using a composite of the IGP metric, the SDR metric and the service metric. The list of preferred producers for a service can then be evaluated at each SDR.

The list of Consumer SDRs interested in service can also easily be computed from the directory of consumers.

7.4.1. Operation due to Producer changes

The producer service operations are

- New producer advertises a service
- Existing Producer start advertising a new service
- Existing Producer stops advertising a service.

The router will be notified by the application regarding the new producer and the services offered. The router will then either update or create an Opaque LSA to advertise this new information and flood it to all SR routers.

Upon receiving this information, remote SDR routers can recalculate the preferred PPSR. It may also need to perform some operations if it have consumers for this new service.

7.4.2. Operation due to Consumer changes

The consumer service operations are

- A new consumer join and add subscription
- An existing consumer stops subscriptions
An existing consumer adds subscriptions

The router will be notified by the application regarding the new consumer and the services it is interested in. The router will then either update or create an Opaque LSA to advertise this new information and flood it to all SDR routers.

Upon receipt of the new Opaque LSA the remote SDR routers can then update the list of CSDRs interested in their services per latest information.

8. Calculation of Optimal Producer

Leveraging OSPF capability to store and compute paths on a topology, the same mechanisms can be used to compute the Optimal PSDRs using the SPF for SDR reachable address using IGP metrics, SDR metric and the service metric. The Optimal PSDR is used in the consumer subtlv.

9. Service Router Data Operations

OSPF SDR delegates the task of SDD distribution to the Data Exchange Agent. This text defines an implementation of such an agent and named it the Service Data Distribution Agent (SDDA). OSPF SDR provides SDDA information about which Consumer SDR is interested which service provided by this OSPF (Producer) SDR. The SDDA makes use of such information to setup distribution channel for SDD distribution from this OSPF Producer SDR to other OSPF Consumer SDRs.

For each OSPF Consumer SDR which subscribes to at least one service provided by this OSPF Producer SDR, there will be a different distribution channel created.

The distribution channel is setup when the OSPF Consumer SDR has subscribed to its first service provided by this OSPF Producer SDR. When the OSPF Consumer SDR subscribes to additional service provided by this OSPF Producer SDR, service data for the new service will be carried over the existing distribution channel. In order words, the same distribution channel can carry service data for different services. The services carried are said to be bound to the distribution channel.

When a distribution channel is first setup for a service or a new service is bound to the channel, the SDDA will notify the SDD. In turn, the SDD will send the latest data for that service to the SDDA for distribution over that channel.

On the other hand, whenever the SDD has new version of data for a service, the SDD will send those data to the SDDA, which will
distribute the new data to all the distribution channels which carry the service.

9.1. Implementation of SDDA

The SDDA uses TCP [RFC793] as its transport protocol for service data distribution. This eliminates the need to implement explicit update fragmentation, retransmission, acknowledgement, and sequencing. SDDA listens on TCP Port 1001. The error notification mechanism used in SDDA assumes that TCP supports a "graceful" close (i.e., that all outstanding data will be delivered before the connection is closed).

The SDDA MUST listen on TCP Port 1001 for incoming connections.

Connection is unidirectional in respect of service data distribution. At most two connections exist between any two SDRs, one for each direction. The same connection is used to distribute data of multiple services from the PSDR to the CSDR.

Connection is initiated and teardown by the PSDR. The connection is referred as the outgoing connection on the PSDR and as incoming connection on the CSDR.

Connection is initiated when the PSDR discover, through the producer and consumer directory, that the CSDR has chosen it as the provider of a service. If PSDR already have an outgoing connection to the same CSDR, no new connection is made and the existing outgoing connection will be used to distribute data for the service.

Connection is teardown when the PSDR discover, through the producer and consumer directory, the CSDR no longer use it as provider for any services.

Connection is teardown when the CSDR discover, through the producer and consumer directory, the PSDR no longer support any services it is interested in.

Connection is also teardown when error occurs, as indicated by the NOTIFICATION message.

9.2. SDR Message Formats

This section describes message formats used by SDDA over the TCP connection.

SDR messages are sent over TCP connections. A message is processed only after it is entirely received. The maximum message size is 65535, including the message header.
All multi-octet fields are in network byte order.

TCP AO [RFC5925] is used for authentication.

9.3. SDR Message Header Format

Each message has a fixed-size header. There may or may not be a data portion following the header, depending on the message type. The layout of these fields is shown below:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type                 |         Length                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

SDR Message Header Format

Type: This 2-octet unsigned integer indicates the type code of the message. This document defines the following type codes:

1 - OPEN
2 - CONFIRM
3 - UPDATE
4 - NOTIFICATION

Length: This 2-octet unsigned integer indicates the total length of the message, including the header in octets. The maximum message length is 65535, and MAY be further constrained, depending on the message type. The value must be at least 4.

9.4. OPEN Message Format

After a TCP connection is established, the first message sent by both side is an OPEN message. If the OPEN message is accepted and processed successfully, a CONFIRM message confirming the OPEN is sent back.

Following the message header, the OPEN message contains the following fields:
Version: This 1-octet unsigned integer indicates the protocol version number of the message. The current OSPF SDR version number is 1.

OSPF Producer SDR Identifier: This 4-octet unsigned integer indicates the OSPF router ID of the producer SDR of the service.

OSPF Consumer SDR Identifier: This 4-octet unsigned integer indicates the OSPF router ID of the consumer SDR of the service.

Length of TLVs: This 2-octet indicates the total length of TLVs information in the OPEN message. Could be 0.

TLVs: Zero or more Type-Length-Value tuple of the following format.

No TLV for OPEN message is defined at the moment.

If the OSPF Producer SDR Identifier or the OSPF Consumer SDR
Identifier does not match information learned through the producer and consumer directory, the NOTIFICATION is sent back to reject the connection.

The subTLVs could define further constraint that must be met before the connection is OPEN message is considered succeed.

The NOTIFICATION message will be sent to indicate if the OPEN is accept or not.

If OPEN is rejected, the TCP connection MUST be teardown after sending the NOTIFICATION. The TCP connection should be retried if the condition that cause the failure is corrected.

9.5. CONFIRM Message Format

CONFIRM messages are used to confirm the OPEN message is accepted.

It has no additional data following the header.

The OSPF Producer SDR MUST NOT send UPDATE message until a CONFIRM message is received.

The OSPF Consumer SDR MUST NOT accept UPDATE message until a CONFIRM message is received. If the OSPF Consumer receive UPDATE message before the CONFIRM message, it should send a NOTIFICATION and close the connection.

9.6. UPDATE Message Format

UPDATE messages are used to push the latest service data from SDD from OSPF producer SDR to OSPF consumer SDR.

Following the message header, the UPDATE message contains the following fields:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---------------------------------------------------------------+
<table>
<thead>
<tr>
<th>Length of TLVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLVs</td>
</tr>
</tbody>
</table>
+---------------------------------------------------------------+
```

UPDATE Message Format
Length of TLVs: This 2-octets unsigned integer indicates the total length of TLV information contained in the message.

TLVs: Zero or more Type-Length-Value tuple of the following format.

```
+----------------+-----------------+-----------------+
| Type           | Length          |
|                |                 |
|                |                 |
+----------------+-----------------+-----------------+
```

TLV Format

The following TLV is defined for the UPDATE message.

Services Update TLV:

```
+----------------+-----------------+-----------------+
| 1              | Length          |
|                |                 |
|                |                 |
+----------------+-----------------+-----------------+
```

Service Update TLV Format

Service ID: The 4-Octets indicates the service this update is originated from.

Service Data: Various length of service data. Its format is opaque to OSPF SDR and is defined by the service itself.
9.7. NOTIFICATION Message Format

A NOTIFICATION message is sent when an error condition is detected. The TCP connection is closed immediately after it is sent.

Following the message header, the NOTIFICATION message contains the following fields:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Error code    | Error subcode |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Length of TLVs            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                            TLVs                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

NOTIFICATION Message Format

Error Code: This 1-octet unsigned integer indicates the type of NOTIFICATION.

The following Error Codes have been defined:

1 - Message header error
2 - OPEN message error
3 - CONFIRM message error
4 - UPDATE message error
5 - Cease

Error subcode: This 1-octet unsigned integer provides more specific information about the nature of the reported error. The value 0 means unspecific error.

Length of TLVs: This 2-octet indicates the total length of TLVs information in the NOTIFICATION message. May be 0 to indicate no Sub-tlvs.
SubTLVs: Zero or more Type-Length-Value tuple of the following format.

```
+-----------------+-------------------+
| Type            | Length            |
| Value           |                   |
+-----------------+-------------------+
```

NOTIFICATION Message Format

No TLV for NOTIFICATION message is defined at the moment.

9.7.1. Message Header Error

The error code 1 is used for error detected at the common message header.

Error subcodes:

0 - Unknown
1 - Bad Message Length
2 - Bad Message Type

9.7.2. OPEN Message Error

The error code 2 is used for error detected while processing the OPEN message.

Error subcodes:

0 - Unknown
1 - Unsupported Version Number.
2 - Bad Producer SDR Identifier.
3 - Bad Consumer SDR Identifier.
4 - No Producer
5 - No Consumer
6 - Malformed TLV

9.7.3. CONFIRM Message Error

The error code 3 is used for error detected while processing the UPDATE message.

Error subcodes:
0 - Unknown
1 - Bad Message Sequence

9.7.4. UPDATE Message Error

The error code 4 is used for error detected while processing the UPDATE message.

Error subcodes:
0 - Unknown
1 - Bad Message Sequence
2 - Malformed TLV

9.7.5. Cease

The error code 5 is used when the OSPF SDR choose to close the connection due to reason other than message processing.

Error subcodes:
0 - Unknown
1 - No Producer
2 - No Consumer

10. Security Considerations

The new extensions defined in this document do not introduce any new security concerns other than those already defined in Section 6 of [RFC2328] and [RFC5340].
11. IANA Considerations

IANA is requested to allocate the value 1001 to "Service Data Distribution" within the "Port Numbers" registry.

12. Contributors

The authors would like to acknowledge the contributions of

Mike Dubroskiy
Rashmi Shrivastava
Jean-Michel Esnault

13. Acknowledgments

The authors would like to thank Les Ginsberg, Keyur Patel and many others who participated in numerous discussions.

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

14. Normative References


Authors’ Addresses

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

EMail: ppe@cisco.com

Burjiz Pithawala
Cisco Systems
510 McCarty Blvd
Milpitas, CA  95035
USA

EMail: bpithaw@cisco.com

Derek Yeung
Cisco Systems
510 McCarty Blvd
Milpitas, CA  95035
USA

EMail: myeung@cisco.com