

Internet Engineering Task Force (IETF)
Request for Comments: 5787
Category: Experimental
ISSN: 2070-1721

D. Papadimitriou
Alcatel-Lucent
March 2010

OSPFv2 Routing Protocols Extensions for Automatically Switched Optical Network (ASON) Routing

Abstract

The ITU-T has defined an architecture and requirements for operating an Automatically Switched Optical Network (ASON).

The Generalized Multiprotocol Label Switching (GMPLS) protocol suite is designed to provide a control plane for a range of network technologies including optical networks such as time division multiplexing (TDM) networks including SONET/SDH and Optical Transport Networks (OTNs), and lambda switching optical networks.

The requirements for GMPLS routing to satisfy the requirements of ASON routing, and an evaluation of existing GMPLS routing protocols are provided in other documents. This document defines extensions to the OSPFv2 Link State Routing Protocol to meet the requirements for routing in an ASON.

Note that this work is scoped to the requirements and evaluation expressed in RFC 4258 and RFC 4652 and the ITU-T Recommendations current when those documents were written. Future extensions of revisions of this work may be necessary if the ITU-T Recommendations are revised or if new requirements are introduced into a revision of RFC 4258.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for examination, experimental implementation, and evaluation.

This document defines an Experimental Protocol for the Internet community. This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <http://www.rfc-editor.org/info/rfc5787>.

Copyright Notice

Copyright (c) 2010 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	4
1.1. Conventions Used in This Document	5
2. Routing Areas, OSPF Areas, and Protocol Instances	5
3. Reachability	6
3.1. Node IPv4 Local Prefix Sub-TLV	6
3.2. Node IPv6 Local Prefix Sub-TLV	7
4. Link Attribute	8
4.1. Local Adaptation	8
4.2. Bandwidth Accounting	9
5. Routing Information Scope	9
5.1. Terminology and Identification	9
5.2. Link Advertisement (Local and Remote TE Router ID Sub-TLV)	10
5.3. Reachability Advertisement (Local TE Router ID sub-TLV) ...	11
6. Routing Information Dissemination	12
6.1. Import/Export Rules	13
6.2. Discovery and Selection	13
6.2.1. Upward Discovery and Selection	13
6.2.2. Downward Discovery and Selection	14
6.2.3. Router Information Experimental Capabilities TLV ...	16
6.3. Loop Prevention	16
6.3.1. Associated RA ID	17
6.3.2. Processing	18
6.4. Resiliency	19
6.5. Neighbor Relationship and Routing Adjacency	20
6.6. Reconfiguration	20
7. OSPFv2 Scalability	21
8. Security Considerations	21
9. Experimental Code Points	21
9.1. Sub-TLVs of the Link TLV	22
9.2. Sub-TLVs of the Node Attribute TLV	22
9.3. Sub-TLVs of the Router Address TLV	23
9.4. TLVs of the Router Information LSA	23
10. References	24
10.1. Normative References	24
10.2. Informative References	25
11. Acknowledgements	26
Appendix A. ASON Terminology	27
Appendix B. ASON Routing Terminology	28

1. Introduction

The Generalized Multiprotocol Label Switching (GMPLS) [RFC3945] protocol suite is designed to provide a control plane for a range of network technologies including optical networks such as time division multiplexing (TDM) networks including SONET/SDH and Optical Transport Networks (OTNs), and lambda switching optical networks.

The ITU-T defines the architecture of the Automatically Switched Optical Network (ASON) in [G.8080].

[RFC4258] details the routing requirements for the GMPLS suite of routing protocols to support the capabilities and functionality of ASON control planes identified in [G.7715] and in [G.7715.1].

[RFC4652] evaluates the IETF Link State routing protocols against the requirements identified in [RFC4258]. Section 7.1 of [RFC4652] summarizes the capabilities to be provided by OSPFv2 [RFC2328] in support of ASON routing. This document details the OSPFv2 specifics for ASON routing.

Multi-layer transport networks are constructed from multiple networks of different technologies operating in a client-server relationship. The ASON routing model includes the definition of routing levels that provide scaling and confidentiality benefits. In multi-level routing, domains called routing areas (RAs) are arranged in a hierarchical relationship. Note that as described in [RFC4652] there is no implied relationship between multi-layer transport networks and multi-level routing. The multi-level routing mechanisms described in this document work for both single-layer and multi-layer networks.

Implementations may support a hierarchical routing topology (multi-level) for multiple transport network layers and/or a hierarchical routing topology for a single transport network layer.

This document details the processing of the generic (technology-independent) link attributes that are defined in [RFC3630], [RFC4202], and [RFC4203] and that are extended in this document. As detailed in Section 4.2, technology-specific traffic engineering attributes (and their processing) may be defined in other documents that complement this document.

Note that this work is scoped to the requirements and evaluation expressed in [RFC4258] and [RFC4652] and the ITU-T Recommendations current when those documents were written. Future extensions of revisions of this work may be necessary if the ITU-T Recommendations are revised or if new requirements are introduced into a revision of [RFC4258].

This document is classified as Experimental. Significant changes to routing protocols are of concern to the stability of the Internet. The extensions described in this document are intended for cautious use in self-contained environments. The objective is to determine whether these extensions are stable and functional, whether there is a demand for implementation and deployment, and whether the extensions have any impact on existing routing protocol deployments.

1.1. 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 [RFC2119].

The reader is assumed to be familiar with the terminology and requirements developed in [RFC4258] and the evaluation outcomes detailed in [RFC4652].

General ASON terminology is provided in Appendix A. ASON routing terminology is described in Appendix B.

2. Routing Areas, OSPF Areas, and Protocol Instances

An ASON routing area (RA) represents a partition of the data plane, and its identifier is used within the control plane as the representation of this partition.

RAs are arranged in hierarchical levels such that any one RA may contain multiple other RAs, and is wholly contained by a single RA. Thus, an RA may contain smaller RAs inter-connected by links. The limit of the subdivision results in an RA that contains just two sub-networks interconnected by a single link.

An ASON RA can be mapped to an OSPF area, but the hierarchy of ASON RA levels does not map to the hierarchy of OSPF routing areas. Instead, successive hierarchical levels of RAs MUST be represented by separate instances of the protocol. Thus, inter-level routing information exchange (as described in Section 6) involves the export and import of routing information between protocol instances.

An ASON RA may therefore be identified by the combination of its OSPF instance identifier and its OSPF area identifier. With proper and careful network-wide configuration, this can be achieved using just the OSPF area identifier, and this process is RECOMMENDED in this document. These concepts and the subsequent handling of network reconfiguration is discussed in Section 6.

3. Reachability

In order to advertise blocks of reachable address prefixes, a summarization mechanism is introduced that complements the techniques described in [RFC5786].

This extension takes the form of a network mask (a 32-bit number indicating the range of IP addresses residing on a single IP network/subnet). The set of local addresses is carried in an OSPFv2 TE LSA Node Attribute TLV (a specific sub-TLV is defined per address family, i.e., IPv4 and IPv6, used as network-unique identifiers).

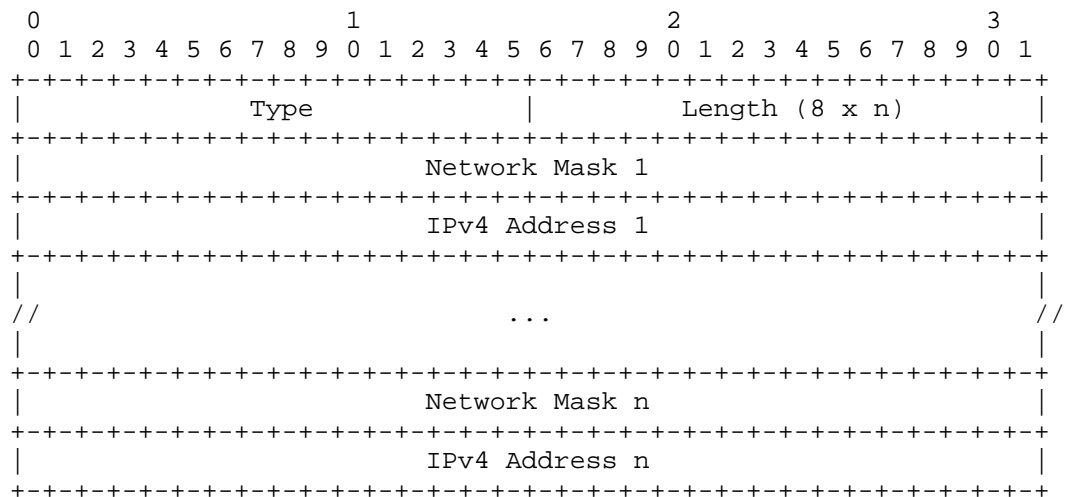
The proposed solution is to advertise the local address prefixes of a router as new sub-TLVs of the (OSPFv2 TE LSA) Node Attribute top-level TLV. This document defines the following sub-TLVs:

- Node IPv4 Local Prefix sub-TLV: Length: variable
- Node IPv6 Local Prefix sub-TLV: Length: variable

3.1. Node IPv4 Local Prefix Sub-TLV

The Type field of the Node IPv4 Local Prefix sub-TLV is assigned a value in the range 32768-32777 agreed to by all participants in the experiment. The Value field of this sub-TLV contains one or more local IPv4 prefixes. The Length is measured in bytes and, as defined in [RFC3630], reports the length in bytes of the Value part of the sub-TLV. It is set to $8 \times n$, where n is the number of local IPv4 prefixes included in the sub-TLV.

The Node IPv4 Local Prefix sub-TLV has the following format:



Network mask i : A 32-bit number indicating the IPv4 address mask for the i th advertised destination prefix.

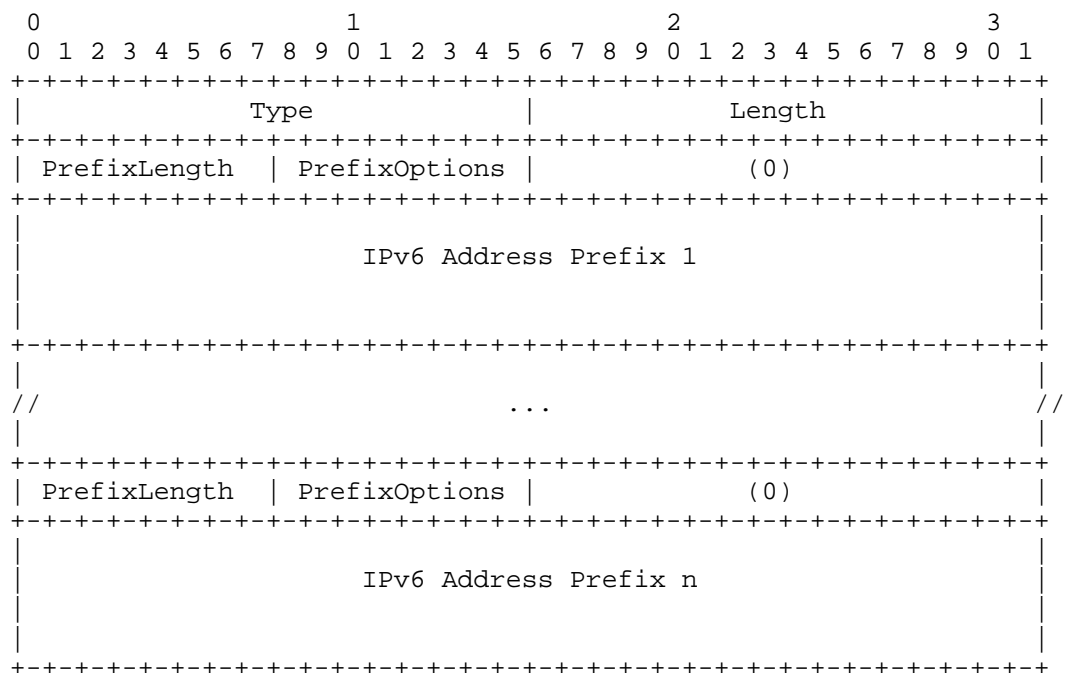
Each <Network mask, IPv4 Address> pair listed as part of this sub-TLV represents a reachable destination prefix hosted by the advertising Router ID.

The local addresses that can be learned from Opaque TE LSAs (that is, the router address and TE interface addresses) SHOULD NOT be advertised in the node IPv4 Local Prefix sub-TLV.

3.2. Node IPv6 Local Prefix Sub-TLV

The Type field of the Node IPv6 Local Prefix sub-TLV is assigned a value in the range 32768-32777 agreed to by all participants in the experiment. The Value field of this sub-TLV contains one or more local IPv6 prefixes. IPv6 Prefix representation uses [RFC5340], Section A.4.1.

The Node IPv6 Local Prefix sub-TLV has the following format:



Length reports the length of the Value part of the sub-TLV in bytes. It is set to the sum over all of the local prefixes included in the sub-TLV of $(4 + (\text{number of 32-bit words in the prefix}) * 4)$.

The encoding of each prefix potentially using fewer than four 32-bit words is described below.

PrefixLength: Length in bits of the prefix.

PrefixOptions: 8-bit field describing various capabilities associated with the prefix (see [RFC5340], Section A.4.2).

IPv6 Address Prefix i: The ith IPv6 address prefix in the list. Each prefix is encoded in an even multiple of 32-bit words using the fewest pairs of 32-bit words necessary to include the entire prefix. Thus, each prefix is encoded in either 64 or 128 bits with trailing zero bit padding as necessary.

The local addresses that can be learned from TE LSAs, i.e., router address and TE interface addresses, SHOULD NOT be advertised in the node IPv6 Local Prefix sub-TLV.

4. Link Attribute

[RFC4652] provides a map between link attributes and characteristics and their representation in sub-TLVs of the top-level Link TLV of the Opaque TE LSA [RFC3630] and [RFC4203], with the exception of the local adaptation (see below). Advertisement of this information SHOULD be supported on a per-layer basis, i.e., one Opaque TE LSA per switching capability (and per bandwidth granularity, e.g., low-order virtual container and high-order virtual container).

4.1. Local Adaptation

Local adaptation is defined as a TE link attribute (i.e., sub-TLV) that describes the cross/inter-layer relationships.

The Interface Switching Capability Descriptor (ISCD) TE Attribute [RFC4202] identifies the ability of the TE link to support cross-connection to another link within the same layer, and the ability to use a locally terminated connection that belongs to one layer as a data link for another layer (adaptation capability). However, the information associated with the ability to terminate connections within that layer (referred to as the termination capability) is embedded with the adaptation capability.

For instance, a link between two optical cross-connects will contain at least one ISCD attribute describing the lambda switching capable (LSC) switching capability; whereas a link between an optical cross-connect and an IP/MPLS LSR will contain at least two ISCD attributes: one for the description of the LSC termination capability and one for the packet switching capable (PSC) adaptation capability.

In OSPFv2, the Interface Switching Capability Descriptor (ISCD) is a sub-TLV (of type 15) of the top-level Link TLV (of type 2) [RFC4203].

The adaptation and termination capabilities are advertised using two separate ISCD sub-TLVs within the same top-level Link TLV.

Per [RFC4202] and [RFC4203], an interface MAY have more than one ISCD sub-TLV. Hence, the corresponding advertisements should not result in any compatibility issues.

Further refinement of the ISCD sub-TLV for multi-layer networks is outside the scope of this document.

4.2. Bandwidth Accounting

GMPLS routing defines an Interface Switching Capability Descriptor (ISCD) that delivers, among other things, information about the (maximum/minimum) bandwidth per priority that a Label Switched Path (LSP) can make use of. Per [RFC4202] and [RFC4203], one or more ISCD sub-TLVs can be associated with an interface. This information, combined with the Unreserved Bandwidth (sub-TLV defined in [RFC3630], Section 2.5.8), provides the basis for bandwidth accounting.

In the ASON context, additional information may be included when the representation and information in the other advertised fields are not sufficient for a specific technology (e.g., SDH). The definition of technology-specific information elements is beyond the scope of this document. Some technologies will not require additional information beyond what is already defined in [RFC3630], [RFC4202], and [RFC4203].

5. Routing Information Scope

5.1. Terminology and Identification

The definition of short-hand terminology introduced in [RFC4652] is repeated here for clarity.

- Pi is a physical (bearer/data/transport plane) node.

- Li is a logical control plane entity that is associated to a single data plane (abstract) node. Each Li is identified by a unique TE Router ID. The latter is a control plane identifier, defined as the Router Address top-level TLV of the Type 1 TE LSA [RFC3630].

Note: The Router Address top-level TLV definition, processing, and usage remain per [RFC3630]. This TLV specifies a stable IP address of the advertising router (Ri) that is always reachable if there is any IP connectivity to it (e.g., via the Data Communication Network). Moreover, each advertising router advertises a unique, reachable IP address for each Pi on behalf of which it makes advertisements.

- Ri is a logical control plane entity that is associated to a control plane "router". The latter is the source for topology information that it generates and shares with other control plane "routers". The Ri is identified by the (advertising) Router ID (32-bit) [RFC2328].

The Router ID, which is represented by Ri and which corresponds to the RC-ID [RFC4258], does not enter into the identification of the logical entities representing the data plane resources such as links. The Routing Database (RDB) is associated to the Ri.

Note: Aside from the Li/Pi mappings, these identifiers are not assumed to be in a particular entity relationship except that the Ri may have multiple Lis in its scope. The relationship between Ri and Li is simple at any moment in time: an Li may be advertised by only one Ri at any time. However, an Ri may advertise a set of one or more Lis. Hence, the OSPFv2 routing protocol must support a single Ri advertising on behalf of more than one Li.

5.2. Link Advertisement (Local and Remote TE Router ID Sub-TLV)

A Router ID (Ri) advertising on behalf multiple TE Router IDs (Lis) creates a 1:N relationship between the Router ID and the TE Router ID. As the link local and link remote (unnumbered) ID association is not unique per node (per Li unicity), the advertisement needs to indicate the remote Lj value and rely on the initial discovery process to retrieve the [Li;Lj] relationship. In brief, as unnumbered links have their ID defined on a per-Li basis, the remote Lj needs to be identified to scope the link remote ID to the local Li. Therefore, the routing protocol MUST be able to disambiguate the advertised TE links so that they can be associated with the correct TE Router ID.

For this purpose, a new sub-TLV of the (OSPFv2 TE LSA) top-level Link TLV is introduced that defines the Local and Remote TE Router ID.

The Type field of the Local and Remote TE Router ID sub-TLV is assigned a value in the range 32768-32777 agreed to by all participants in the experiment. The Length field takes the value 8. The Value field of this sub-TLV contains 4 octets of the Local TE Router Identifier followed by 4 octets of the Remote TE Router Identifier. The value of the Local and Remote TE Router Identifier SHOULD NOT be set to 0.

The format of the Local and Remote TE Router ID sub-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 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|               Type                 |               Length (8)         |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|               Local TE Router Identifier                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|               Remote TE Router Identifier                             |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This sub-TLV is only required to be included as part of the top-level Link TLV if the Router ID is advertising on behalf of more than one TE Router ID. In any other case, this sub-TLV SHOULD be omitted except if the operator plans to start off with 1 Li and progressively add more Lis (under the same Ri) such as to maintain consistency.

Note: The Link ID sub-TLV that identifies the other end of the link (i.e., Router ID of the neighbor for point-to-point links) MUST appear exactly once per Link TLV. This sub-TLV MUST be processed as defined in [RFC3630].

5.3. Reachability Advertisement (Local TE Router ID sub-TLV)

When the Router ID is advertised on behalf of multiple TE Router IDs (Lis), the routing protocol MUST be able to associate the advertised reachability information with the correct TE Router ID.

For this purpose, a new sub-TLV of the (OSPFv2 TE LSA) top-level Node Attribute TLV is introduced. This TLV associates the local prefixes (see above) to a given TE Router ID.

The Type field of the Local TE Router ID sub-TLV is assigned a value in the range 32768-32777 agreed to by all participants in the experiment. The Length field takes the value 4. The Value field of this sub-TLV contains the Local TE Router Identifier [RFC3630] encoded over 4 octets.

The format of the Local TE Router ID sub-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 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|               Type                 |               Length (4)          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|               Local TE Router Identifier                               |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This sub-TLV is only required to be included as part of the Node Attribute TLV if the Router ID is advertising on behalf of more than one TE Router ID. In any other case, this sub-TLV SHOULD be omitted.

6. Routing Information Dissemination

An ASON routing area (RA) represents a partition of the data plane, and its identifier is used within the control plane as the representation of this partition. An RA may contain smaller RAs inter-connected by links. The limit of the subdivision results is an RA that contains two sub-networks interconnected by a single link. ASON RA levels do not reflect routing protocol levels (such as OSPF areas).

Successive hierarchical levels of RAs can be represented by separate instances of the protocol.

Routing controllers (RCs) supporting RAs disseminate information downward and upward in this hierarchy. The vertical routing information dissemination mechanisms described in this section do not introduce or imply a new OSPF routing area hierarchy. RCs supporting RAs at multiple levels are structured as separate OSPF instances with routing information exchanges between levels described by import and export rules operating between OSPF instances.

The implication is that an RC that performs import/export of routing information as described in this document does not implement an Area Border Router (ABR) functionality.

6.1. Import/Export Rules

RCs supporting RAs disseminate information upward and downward in the hierarchy by importing/exporting routing information as Opaque TE LSAs (Opaque Type 1) of LS Type 10. The information that MAY be exchanged between adjacent levels includes the Router Address, Link, and Node Attribute top-level TLVs.

The Opaque TE LSA import/export rules are governed as follows:

- If the export target interface is associated with the same RA as is associated with the import interface, the Opaque LSA MUST NOT be imported.
- If a match is found between the advertising Router ID in the header of the received Opaque TE LSA and one of the Router IDs belonging to the RA of the export target interface, the Opaque LSA MUST NOT be imported.
- If these two conditions are not met, the Opaque TE LSA MAY be imported according to local policy. If imported, the LSA MAY be disseminated according to local policy. If disseminated, the normal OSPF flooding rules MUST be followed and the advertising Router ID MUST be set to the importing router's Router ID.

The imported/exported routing information content MAY be transformed, e.g., filtered or aggregated, as long as the resulting routing information is consistent. In particular, when more than one RC is bound to adjacent levels and both are allowed to import/export routing information, it is expected that these transformations are performed in a consistent manner. Definition of these policy-based mechanisms is outside the scope of this document.

In practice, and in order to avoid scalability and processing overhead, routing information imported/exported downward/upward in the hierarchy is expected to include reachability information (see Section 3) and, upon strict policy control, link topology information.

6.2 Discovery and Selection

6.2.1. Upward Discovery and Selection

In order to discover RCs that are capable of disseminating routing information up the routing hierarchy, the following capability descriptor bit is set in the OSPF Router Information Experimental Capabilities TLV (see Section 6.2.3) carried in the Router Information LSA ([RFC4970]).

- U bit: When set, this flag indicates that the RC is capable of disseminating routing information upward to the adjacent level.

In the case that multiple RCs are advertised from the same RA with their U bit set, the RC with the highest Router ID, among those RCs with the U bit set, SHOULD be selected as the RC for upward dissemination of routing information. The other RCs MUST NOT participate in the upward dissemination of routing information as long as the Opaque LSA information corresponding to the highest Router ID RC does not reach MaxAge. This mechanism prevents more than one RC advertising routing information upward in the routing hierarchy from the same RA.

Note that if the information to allow the selection of the RC that will be used to disseminate routing information up the hierarchy from a specific RA cannot be discovered automatically, it MUST be manually configured.

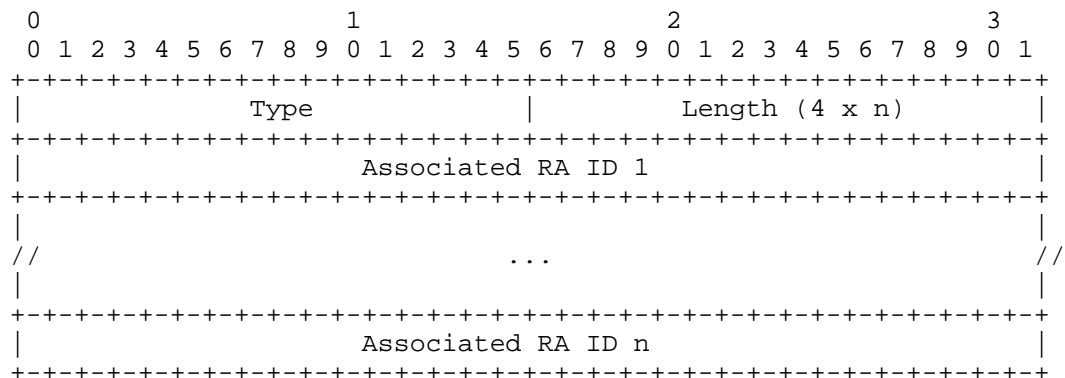
Once an RC has been selected, it remains unmodified even if an RC with a higher Router ID is introduced and advertises its capability to disseminate routing information upward the adjacent level (i.e., U bit set). This hysteresis mechanism prevents from disturbing the upward routing information dissemination process in case, e.g., of flapping.

6.2.2. Downward Discovery and Selection

The same discovery mechanism is used for selecting the RC responsible for dissemination of routing information downward in the hierarchy. However, an additional restriction MUST be applied such that the RC selection process takes into account that an upper level may be adjacent to one or more lower (RA) levels. For this purpose, a specific TLV indexing the (lower) RA ID to which the RCs are capable of disseminating routing information is needed.

The Downstream Associated RA ID TLV is carried in the OSPF Router Information LSA [RFC4970]. The Type field of the Downstream Associated RA ID TLV is assigned a value in the range 32768-32777 agreed to by all participants in the experiment. The Length of this TLV is $n \times 4$ octets. The Value field of this sub-TLV contains the list of Associated RA IDs. Each Associated RA ID value is encoded following the OSPF area ID (32 bits) encoding rules defined in [RFC2328].

The format of the Downstream Associated RA ID TLV is:



To discover RCs that are capable of disseminating routing information downward through the routing hierarchy, the following capability descriptor bit is set in the OSPF Router Information Experimental Capabilities TLV (see Section 6.2.3) carried in the Router Information LSA ([RFC4970]).

Note that the Downstream Associated RA ID TLV MUST be present when the D bit is set.

- D bit: when set, this flag indicates that the RC is capable of disseminating routing information downward to the adjacent levels.

If multiple RCs are advertised for the same Associated RA ID, the RC with the highest Router ID, among the RCs with the D bit set, MUST be selected as the RC for downward dissemination of routing information. The other RCs for the same Associated RA ID MUST NOT participate in the downward dissemination of routing information as long as the Opaque LSA information corresponding to the highest Router ID RC does not reach MaxAge. This mechanism prevents more than one RC from advertising routing information downward through the routing hierarchy.

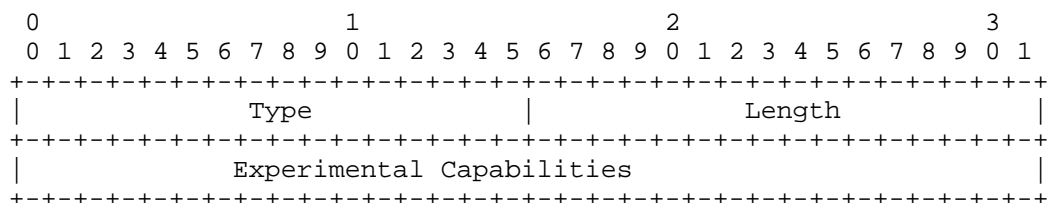
Note that if the information to allow the selection of the RC that will be used to disseminate routing information down the hierarchy to a specific RA cannot be discovered automatically, it MUST be manually configured.

The OSPF Router information Opaque LSA (Opaque type of 4, Opaque ID of 0) and its content, in particular the Router Information Capabilities TLV [RFC4970] and TE Node Capability Descriptor TLV [RFC5073], MUST NOT be re-originated.

6.2.3. Router Information Experimental Capabilities TLV

A new TLV is defined for inclusion in the Router Information LSA to carry experimental capabilities because the assignment policy for bits in the Router Information Capabilities TLV is "Standards Action" [RFC5226] prohibiting its use from Experimental documents.

The format of the Router Information Experimental Capabilities TLV is as follows:



Type A value in the range 32768-32777 agreed to by all participants in the experiment.

Length A 16-bit field that indicates the length of the value portion in octets and will be a multiple of 4 octets dependent on the number of capabilities advertised. Initially, the length will be 4, denoting 4 octets of informational capability bits.

Value A variable-length sequence of capability bits rounded to a multiple of 4 octets padded with undefined bits.

The following experimental capability bits are assigned:

Bit	Capabilities
0	The U bit (see Section 6.2.1)
1	The D bit (see Section 6.2.2)

6.3. Loop Prevention

When more than one RC is bound to an adjacent level of the hierarchy, and is configured or selected to redistribute routing information upward and downward, a specific mechanism is required to avoid looping of routing information. Looping is the re-introduction of routing information that has been advertised from the upper level back to the upper level. This specific case occurs, for example, when the RC advertising routing information downward in the hierarchy is not the same one that advertises routing upward in the hierarchy.

When these conditions are met, it is necessary to have a means by which an RC receiving an Opaque TE LSA imported/exported downward by an RC associated to the same RA does not import/export the content of this LSA back upward into the (same) upper level.

Note that configuration and operational simplification can be obtained when both functionalities are configured on a single RC (per pair of adjacent levels) fulfilling both roles. Figure 1 provides an example where such simplification applies.

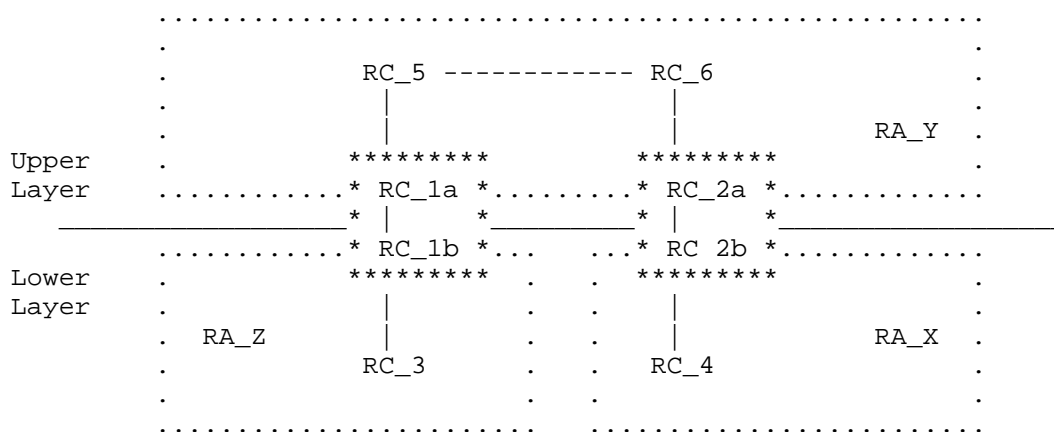


Figure 1. Hierarchical Environment (Example)

In this case, the procedure described in this section MAY be omitted, as long as these conditions are permanently guaranteed. In all other cases, without exception, the procedure described in this section MUST be applied.

6.3.1. Associated RA ID

We need some way of filtering the downward/upward re-originated Opaque TE LSA. Per [RFC5250], the information contained in Opaque LSAs may be used directly by OSPF. By adding the RA ID associated with the incoming routing information, the loop prevention problem can be solved.

This additional information, referred to as the Associated RA ID, MAY be carried in Opaque LSAs that include any of the following top-level TLVs:

- Router Address top-level TLV
- Link top-level TLV
- Node Attribute top-level TLV

The Associated RA ID reflects the identifier of the area from which the routing information is received. For example, for a multi-level hierarchy, this identifier does not reflect the originating RA ID; it will reflect the RA from which the routing information is imported.

The Type field of the Associated RA ID sub-TLV is assigned a value in the range 32768-32777 agreed to by all participants in the experiment. The same value MUST be used for the Type regardless of which TLV the sub-TLV appears in.

The Length of the Associated RA ID TLV is 4 octets. The Value field of this sub-TLV contains the Associated RA ID. The Associated RA ID value is encoded following the OSPF area ID (32 bits) encoding rules defined in [RFC2328].

The format of the Associated RA ID TLV is defined as follows:

```

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 (4)          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |
|               Associated RA ID     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

6.3.2. Processing

When fulfilling the rules detailed in Section 6.1, a given Opaque LSA is imported/exported downward or upward the routing hierarchy, and the Associated RA ID TLV is added to the received Opaque LSA list of TLVs such as to identify the area from which this routing information has been received.

When the RC adjacent to the lower or upper routing level receives this Opaque LSA, the following rule is applied (in addition to the rule governing the import/export of Opaque LSAs as detailed in Section 6.1).

- If a match is found between the Associated RA ID of the received Opaque TE LSA and the RA ID belonging to the area of the export target interface, the Opaque TE LSA MUST NOT be imported.
- Otherwise, this Opaque LSA MAY be imported and disseminated downward or upward the routing hierarchy following the OSPF flooding rules.

This mechanism ensures that no race condition occurs when the conditions depicted in Figure 2 are met.

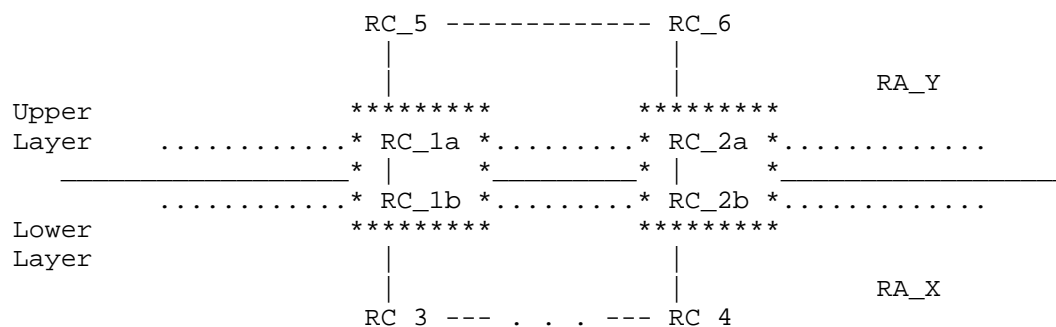


Figure 2. Race Condition Prevention (Example)

Assume that RC_1b is configured for exporting routing information upward toward RA_Y (upward the routing hierarchy) and that RC_2a is configured for exporting routing information toward RA_X (downward the routing hierarchy).

Assume that routing information advertised by RC_3 would reach RC_4 faster across RA_Y through hierarchy.

If RC_2b is not able to prevent from importing that information, RC_4 may receive that information before the same advertisement would propagate in RA_X (from RC_3) to RC_4. For this purpose, RC_1a inserts the Associated RA X to the imported routing information from RA_X. Because RC_2b finds a match between the Associated RA ID (X) of the received Opaque TE LSA and the ID (X) of the RA of the export target interface, this LSA MUST NOT be imported.

6.4. Resiliency

OSPF creates adjacencies between neighboring routers for the purpose of exchanging routing information. After a neighbor has been discovered, bidirectional communication is ensured, and a routing adjacency is formed between RCs, loss of communication may result in partitioned OSPF areas and so in partitioned RAs.

Consider for instance (see Figure 2) the case where RC_1a and RC_1b are configured for exchanging routing information downward and upward RA_Y, respectively, and that RC_2a and RC_2b are not configured for exchanging any routing information toward RA_X. If the communication between RC_1a and RC_2a is broken (due, e.g., to RC_5 - RC_6 communication failure), RA_Y could be partitioned.

In these conditions, it is RECOMMENDED that RC_2a be re-configurable such as to allow for exchanging routing information downward to RA_X. This reconfiguration MAY be performed manually or automatically. In

the latter cases, automatic reconfiguration uses the mechanism described in Section 6.2 (forcing MaxAge of the corresponding opaque LSA information in case the originating RC becomes unreachable). Manual reconfiguration MUST be supported.

6.5. Neighbor Relationship and Routing Adjacency

It is assumed that (point-to-point) IP control channels are provisioned/configured between RCs belonging to the same routing level. Provisioning/configuration techniques are outside the scope of this document.

Once established, the OSPF Hello protocol is responsible for establishing and maintaining neighbor relationships. This protocol also ensures that communication between neighbors is bidirectional. Routing adjacency can subsequently be formed between RCs following mechanisms defined in [RFC2328].

6.6 Reconfiguration

This section details the RA ID reconfiguration steps.

Reconfiguration of the RA ID occurs when the RA ID is modified, e.g., from value Z to value X or Y (see Figure 2).

The process of reconfiguring the RA ID involves:

- Disable the import/export of routing information from the upper and lower levels (to prevent any LS information update).
- Change the RA ID of the local level RA from, e.g., Z to X or Y. Perform a Link State Database (LSDB) checksum on all routers to verify that LSDBs are consistent.
- Enable import of upstream and downstream routing information such as to re-synchronize local-level LSDBs from any LS information that may have occurred in an upper or a lower routing level.
- Enable export of routing information downstream such as to re-sync the downstream level with the newly reconfigured RA ID (as part of the re-advertised Opaque TE LSA).
- Enable export of routing information upstream such as to re-sync the upstream level with the newly reconfigured RA ID (as part of the re-advertised Opaque TE LSA).

Note that the re-sync operation needs to be carried out only between the directly adjacent upper and lower routing levels.

7. OSPFv2 Scalability

- Routing information exchange upward/downward in the hierarchy between adjacent RAs SHOULD by default be limited to reachability information. In addition, several transformations such as prefix aggregation are RECOMMENDED when allowing the amount of information imported/exported by a given RC to be decreased without impacting consistency.
- Routing information exchange upward/downward in the hierarchy involving TE attributes MUST be under strict policy control. Pacing and min/max thresholds for triggered updates are strongly RECOMMENDED.
- The number of routing levels MUST be maintained under strict policy control.

8. Security Considerations

This document specifies the contents and processing of Opaque LSAs in OSPFv2 [RFC2328]. Opaque TE and RI LSAs defined in this document are not used for SPF computation, and so have no direct effect on IP routing. Additionally, ASON routing domains are delimited by the usual administrative domain boundaries.

Any mechanisms used for securing the exchange of normal OSPF LSAs can be applied equally to all Opaque TE and RI LSAs used in the ASON context. Authentication of OSPFv2 LSA exchanges (such as OSPF cryptographic authentication [RFC2328] and [RFC5709]) can be used to secure against passive attacks and provide significant protection against active attacks. [RFC5709] defines a mechanism for authenticating OSPF packets by making use of the HMAC algorithm in conjunction with the SHA family of cryptographic hash functions.

[RFC2154] adds 1) digital signatures to authenticate OSPF LSA data, 2) a certification mechanism for distribution of routing information, and 3) a neighbor-to-neighbor authentication algorithm to protect local OSPFv2 protocol exchanges.

9. Experimental Code Points

This document is classified as Experimental. It defines new TLVs and sub-TLVs for inclusion in OSPF LSAs. According to the assignment policies for the registries of code points for these TLVs and sub-TLVs, values must be assigned from the experimental ranges and must not be recorded by IANA or mentioned in this document.

The following sections summarize the TLVs and sub-TLVs concerned.

9.1. Sub-TLVs of the Link TLV

This document defines the following sub-TLVs of the Link TLV carried in the OSPF TE LSA:

- Local and Remote TE Router ID sub-TLV
- Associated RA ID sub-TLV

The defining text for code point assignment for sub-TLVs of the OSPF TE Link TLV says ([RFC3630]):

- o Types in the range 10-32767 are to be assigned via Standards Action.
- o Types in the range 32768-32777 are for experimental use; these will not be registered with IANA, and MUST NOT be mentioned by RFCs.
- o Types in the range 32778-65535 are not to be assigned at this time.

That means that the new sub-TLVs must be assigned type values from the range 32768-32777. It is a matter for experimental implementations to assign their own code points, and to agree with cooperating implementations participating in the same experiments what values to use.

Note that the same value for the Associated RA ID sub-TLV MUST be used when it appears in the Link TLV, the Node Attribute TLV, and the Router Address TLV.

9.2. Sub-TLVs of the Node Attribute TLV

This document defines the following sub-TLVs of the Node Attribute TLV carried in the OSPF TE LSA.

- Node IPv4 Local Prefix sub-TLV
- Node IPv6 Local Prefix sub-TLV
- Local TE Router ID sub-TLV
- Associated RA ID sub-TLV

The defining text for code point assignment for sub-TLVs of the OSPF Node Attribute TLV says ([RFC5786]):

- o Types in the range 3-32767 are to be assigned via Standards Action.

- o Types in the range 32768-32777 are for experimental use; these will not be registered with IANA, and MUST NOT be mentioned by RFCs.
- o Types in the range 32778-65535 are not to be assigned at this time. Before any assignments can be made in this range, there MUST be a Standards Track RFC that specifies IANA Considerations that covers the range being assigned.

That means that the new sub-TLVs must be assigned type values from the range 32768-32777. It is a matter for experimental implementations to assign their own code points, and to agree with cooperating implementations participating in the same experiments what values to use.

Note that the same value for the Associated RA ID sub-TLV MUST be used when it appears in the Link TLV, the Node Attribute TLV, and the Router Address TLV.

9.3. Sub-TLVs of the Router Address TLV

The OSPF Router Address TLV is defined in [RFC3630]. No sub-TLVs are defined in that document and there is no registry or allocation policy for sub-TLVs of the Router Address TLV.

This document defines the following new sub-TLV for inclusion in the OSPF Router Address TLV:

- Associated RA ID sub-TLV

Note that the same value for the Associated RA ID sub-TLV MUST be used when it appears in the Link TLV, the Node Attribute TLV, and the Router Address TLV. This is consistent with potential for a future definition of a registry with policies that match the other existing registries.

9.4. TLVs of the Router Information LSA

This document defines two new TLVs to be carried in the Router Information LSA.

- Downstream Associated RA ID TLV
- Router Information Experimental Capabilities TLV

The defining text for code point assignment for TLVs of the OSPF Router Information LSA says ([RFC4970]):

- o 1-32767 Standards Action.

- o Types in the range 32768-32777 are for experimental use; these will not be registered with IANA and MUST NOT be mentioned by RFCs.
- o Types in the range 32778-65535 are reserved and are not to be assigned at this time. Before any assignments can be made in this range, there MUST be a Standards Track RFC that specifies IANA Considerations that covers the range being assigned.

That means that the new TLVs must be assigned type values from the range 32768-32777. It is a matter for experimental implementations to assign their own code points, and to agree with cooperating implementations participating in the same experiments what values to use.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2154] Murphy, S., Badger, M., and B. Wellington, "OSPF with Digital Signatures", RFC 2154, June 1997.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC3945] Mannie, E., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4202] Kompella, K., Ed., and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4970] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.

- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.
- [RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.
- [RFC5786] Aggarwal, R. and K. Kompella, "Advertising a Router's Local Addresses in OSPF TE Extensions", RFC 5786, March 2010.

10.2. Informative References

- [RFC4258] Brungard, D., Ed., "Requirements for Generalized Multi-Protocol Label Switching (GMPLS) Routing for the Automatically Switched Optical Network (ASON)", RFC 4258, November 2005.
- [RFC4652] Papadimitriou, D., Ed., Ong, L., Sadler, J., Shew, S., and D. Ward, "Evaluation of Existing Routing Protocols against Automatic Switched Optical Network (ASON) Routing Requirements", RFC 4652, October 2006.
- [RFC5073] Vasseur, J., Ed., and J. Le Roux, Ed., "IGP Routing Protocol Extensions for Discovery of Traffic Engineering Node Capabilities", RFC 5073, December 2007.
- [RFC5709] Bhatia, M., Manral, V., Fanto, M., White, R., Barnes, M., Li, T., and R. Atkinson, "OSPFv2 HMAC-SHA Cryptographic Authentication", RFC 5709, October 2009.

For information on the availability of ITU Documents, please see <http://www.itu.int>.

- [G.7715] ITU-T Rec. G.7715/Y.1306, "Architecture and Requirements for the Automatically Switched Optical Network (ASON)", June 2002.
- [G.7715.1] ITU-T Draft Rec. G.7715.1/Y.1706.1, "ASON Routing Architecture and Requirements for Link State Protocols", November 2003.
- [G.805] ITU-T Rec. G.805, "Generic functional architecture of transport networks)", March 2000.

- [G.8080] ITU-T Rec. G.8080/Y.1304, "Architecture for the Automatically Switched Optical Network (ASON)," November 2001 (and Revision, January 2003).

11. Acknowledgements

The author would like to thank Dean Cheng, Acee Lindem, Pandian Vijay, Alan Davey, Adrian Farrel, Deborah Brungard, and Ben Campbell for their useful comments and suggestions.

Lisa Dusseault and Jari Arkko provided useful comments during IESG review.

Question 14 of Study Group 15 of the ITU-T provided useful and constructive input.

Appendix A. ASON Terminology

This document makes use of the following terms:

Administrative domain: (See Recommendation [G.805].) For the purposes of [G7715.1], an administrative domain represents the extent of resources that belong to a single player such as a network operator, a service provider, or an end-user. Administrative domains of different players do not overlap amongst themselves.

Control plane: performs the call control and connection control functions. Through signaling, the control plane sets up and releases connections, and may restore a connection in case of a failure.

(Control) Domain: represents a collection of (control) entities that are grouped for a particular purpose. The control plane is subdivided into domains matching administrative domains. Within an administrative domain, further subdivisions of the control plane are recursively applied. A routing control domain is an abstract entity that hides the details of the RC distribution.

External NNI (E-NNI): interfaces are located between protocol controllers between control domains.

Internal NNI (I-NNI): interfaces are located between protocol controllers within control domains.

Link: (See Recommendation G.805.) A "topological component" that describes a fixed relationship between a "subnetwork" or "access group" and another "subnetwork" or "access group". Links are not limited to being provided by a single server trail.

Management plane: performs management functions for the transport plane, the control plane, and the system as a whole. It also provides coordination between all the planes. The following management functional areas are performed in the management plane: performance, fault, configuration, accounting, and security management.

Management domain: (See Recommendation G.805.) A management domain defines a collection of managed objects that are grouped to meet organizational requirements according to geography, technology, policy, or other structure, and for a number of functional areas such as configuration, security, (FCAPS), for the purpose of providing control in a consistent manner. Management domains can be disjoint, contained, or overlapping. As such, the resources

within an administrative domain can be distributed into several possible overlapping management domains. The same resource can therefore belong to several management domains simultaneously, but a management domain shall not cross the border of an administrative domain.

Subnetwork Point (SNP): The SNP is a control plane abstraction that represents an actual or potential transport plane resource. SNPs (in different subnetwork partitions) may represent the same transport resource. A one-to-one correspondence should not be assumed.

Subnetwork Point Pool (SNPP): A set of SNPs that are grouped together for the purposes of routing.

Termination Connection Point (TCP): A TCP represents the output of a Trail Termination function or the input to a Trail Termination Sink function.

Transport plane: provides bidirectional or unidirectional transfer of user information, from one location to another. It can also provide transfer of some control and network management information. The transport plane is layered; it is equivalent to the Transport Network defined in Recommendation G.805.

User Network Interface (UNI): interfaces are located between protocol controllers between a user and a control domain. Note: There is no routing function associated with a UNI reference point.

Appendix B. ASON Routing Terminology

This document makes use of the following terms:

Routing Area (RA): an RA represents a partition of the data plane, and its identifier is used within the control plane as the representation of this partition. Per [G.8080], an RA is defined by a set of sub-networks, the links that interconnect them, and the interfaces representing the ends of the links exiting that RA. An RA may contain smaller RAs inter-connected by links. The limit of subdivision results in an RA that contains two sub-networks interconnected by a single link.

Routing Database (RDB): a repository for the local topology, network topology, reachability, and other routing information that is updated as part of the routing information exchange and may additionally contain information that is configured. The RDB may contain routing information for more than one routing area (RA).

Routing Components: ASON routing architecture functions. These functions can be classified as protocol independent (Link Resource Manager or LRM, Routing Controller or RC) or protocol specific (Protocol Controller or PC).

Routing Controller (RC): handles (abstract) information needed for routing and the routing information exchange with peering RCs by operating on the RDB. The RC has access to a view of the RDB. The RC is protocol independent.

Note: Since the RDB may contain routing information pertaining to multiple RAs (and possibly to multiple layer networks), the RCs accessing the RDB may share the routing information.

Link Resource Manager (LRM): supplies all the relevant component and TE link information to the RC. It informs the RC about any state changes of the link resources it controls.

Protocol Controller (PC): handles protocol-specific message exchanges according to the reference point over which the information is exchanged (e.g., E-NNI, I-NNI), and internal exchanges with the RC. The PC function is protocol dependent.

Author's Address

Dimitri Papadimitriou
Alcatel-Lucent Bell
Copernicuslaan 50
B-2018 Antwerpen
Belgium
Phone: +32 3 2408491
EMail: dimitri.papadimitriou@alcatel-lucent.be

