              Traffic Engineering (TE) Extensions to OSPF Version 2

Status of this Memo

Copyright Notice

Abstract

   This document describes extensions to the OSPF protocol version 2 to
   support intra-area Traffic Engineering (TE), using Opaque Link State
   Advertisements.

1. Introduction

   This document specifies a method of adding traffic engineering
   capabilities to OSPF Version 2 [1]. The architecture of traffic
   engineering is described in [5]. The semantic content of the
   extensions is essentially identical to the corresponding extensions
   to IS-IS [6]. It is expected that the traffic engineering extensions
   to OSPF will continue to mirror those in IS-IS.

   The extensions provide a way of describing the traffic engineering
   topology (including bandwidth and administrative constraints) and
   distributing this information within a given OSPF area. This
   topology does not necessarily match the regular routed topology,
   though this proposal depends on Network LSAs to describe multi-access
   links. This document purposely does not say how the mechanisms
   described here can be used for traffic engineering across multiple
   OSPF areas; that task is left to future documents. Furthermore, no
   changes have been made to the operation of OSPFv2 flooding; in

particular, if non-TE capable nodes exist in the topology, they MUST
flood TE LSAs as any other type 10 (area-local scope) Opaque LSAs
(see [3]).

1.1. Applicability

Many of the extensions specified in this document are in response to
the requirements stated in [5], and thus are referred to as "traffic
engineering extensions", and are also commonly associated with MPLS
Traffic Engineering. A more accurate (albeit bland) designation is
"extended link attributes", as the proposal is to simply add more
attributes to links in OSPF advertisements.

The information made available by these extensions can be used to
build an extended link state database just as router LSAs are used to
build a "regular" link state database; the difference is that the
extended link state database (referred to below as the traffic
engineering database) has additional link attributes. Uses of the
traffic engineering database include:

    o   monitoring the extended link attributes;
    o   local constraint-based source routing; and
    o   global traffic engineering.

For example, an OSPF-speaking device can participate in an OSPF area,
build a traffic engineering database, and thereby report on the
reservation state of links in that area.

In "local constraint-based source routing", a router R can compute a
path from a source node A to a destination node B; typically, A is R
itself, and B is specified by a "router address" (see below). This
path may be subject to various constraints on the attributes of the
links and nodes that the path traverses, e.g., use green links that
have unreserved bandwidth of at least 10Mbps. This path could then
be used to carry some subset of the traffic from A to B, forming a
simple but effective means of traffic engineering. How the subset of
traffic is determined, and how the path is instantiated, is beyond
the scope of this document; suffice it to say that one means of
defining the subset of traffic is "those packets whose IP
destinations were learned from B", and one means of instantiating
paths is using MPLS tunnels. As an aside, note that constraint-based
routing can be NP-hard, or even unsolvable, depending on the nature
of the attributes and constraints, and thus many implementations will
use heuristics. Consequently, we don't attempt to sketch an
algorithm here.

Finally, for "global traffic engineering", a device can build a
traffic engineering database, input a traffic matrix and an
optimization function, crunch on the information, and thus compute
optimal or near-optimal routing for the entire network. The device
can subsequently monitor the traffic engineering topology and react
to changes by recomputing the optimal routes.

1.2. Limitations

As mentioned above, this document specifies extensions and procedures
for intra-area distribution of Traffic Engineering information.
Methods for inter-area and inter-AS (Autonomous System) distribution
are not discussed here.

The extensions specified in this document capture the reservation
state of point-to-point links. The reservation state of multi-access
links may not be accurately reflected, except in the special case in
which there are only two devices in the multi-access subnetwork.
Operation over multi-access networks with more than two devices is
not specifically prohibited. A more accurate description of the
reservation state of multi-access networks is for further study.

This document also does not support unnumbered links. This
deficiency will be addressed in future documents; see also [7] and
[8].

1.3. Conventions

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 BCP 14, RFC 2119 [2].

2. LSA Format

2.1. LSA type

This extension makes use of the Opaque LSA [3].

Three types of Opaque LSAs exist, each of which has a different
flooding scope. This proposal uses only Type 10 LSAs, which have an
area flooding scope.

One new LSA is defined, the Traffic Engineering LSA. This LSA
describes routers, point-to-point links, and connections to multi-
access networks (similar to a Router LSA). For traffic engineering
purposes, the existing Network LSA is sufficient for describing
multi-access links, so no additional LSA is defined for this purpose.

2.2. LSA ID

The LSA ID of an Opaque LSA is defined as having eight bits of type
data and 24 bits of type-specific data. The Traffic Engineering LSA
uses type 1. The remaining 24 bits are the Instance field, 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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |      1        |                   Instance                    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
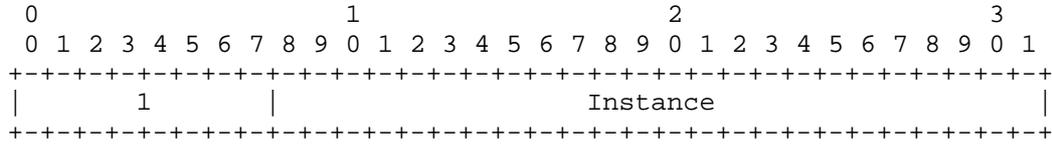
The Instance field is an arbitrary value used to maintain multiple
Traffic Engineering LSAs. A maximum of 16777216 Traffic Engineering
LSAs may be sourced by a single system. The LSA ID has no
topological significance.

2.3. LSA Format Overview

2.3.1. LSA Header

The Traffic Engineering LSA starts with the standard LSA header:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            LS age              |    Options    |      10       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |      1        |                   Instance                    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Advertising Router                      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       LS sequence number                      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         LS checksum           |             Length            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
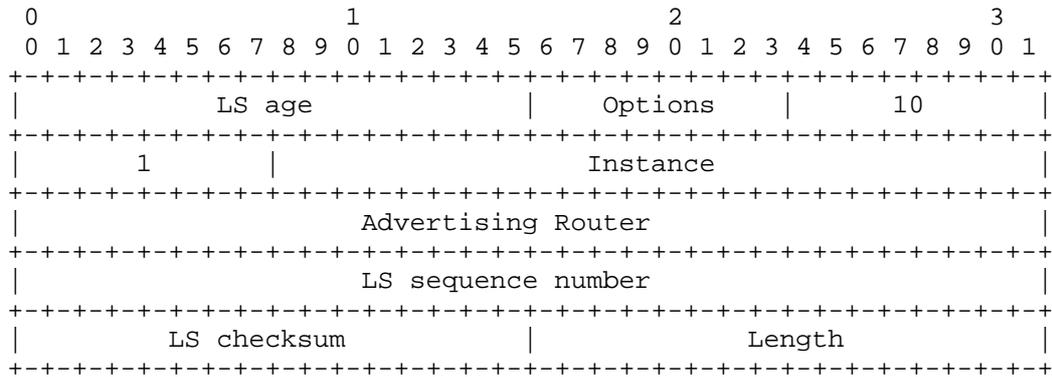
2.3.2. TLV Header

   The LSA payload consists of one or more nested Type/Length/Value
   (TLV) triplets for extensibility. 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 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |              Type             |             Length            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                            Value...                           |
   .                                                               .
   .                                                               .
   .                                                               .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
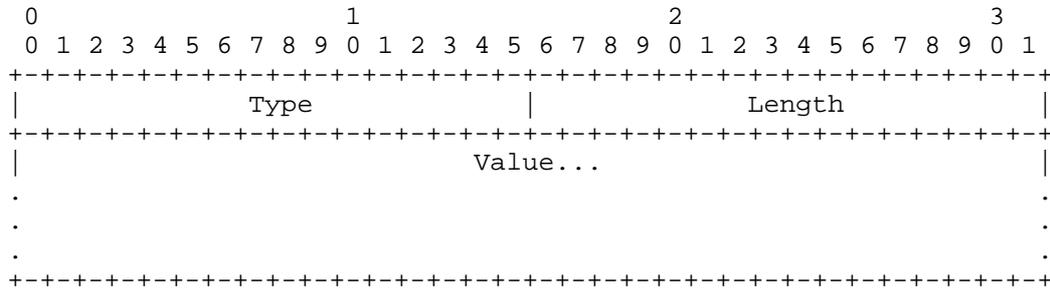
   The Length field defines the length of the value portion in octets
   (thus a TLV with no value portion would have a length of zero). The
   TLV is padded to four-octet alignment; padding is not included in the
   length field (so a three octet value would have a length of three,
   but the total size of the TLV would be eight octets). Nested TLVs
   are also 32-bit aligned. Unrecognized types are ignored.

   This memo defines Types 1 and 2. See the IANA Considerations section
   for allocation of new Types.

2.4. LSA payload details

   An LSA contains one top-level TLV.

   There are two top-level TLVs defined:

      1 - Router Address
      2 - Link

2.4.1. Router Address TLV

   The Router Address TLV specifies a stable IP address of the
   advertising router that is always reachable if there is any
   connectivity to it; this is typically implemented as a "loopback
   address". The key attribute is that the address does not become
   unusable if an interface is down. In other protocols, this is known
   as the "router ID," but for obvious reasons this nomenclature is
   avoided here. If a router advertises BGP routes with the BGP next
   hop attribute set to the BGP router ID, then the Router Address
   SHOULD be the same as the BGP router ID.

If IS-IS is also active in the domain, this address can also be used
to compute the mapping between the OSPF and IS-IS topologies. For
example, suppose a router R is advertising both IS-IS and OSPF
Traffic Engineering LSAs, and suppose further that some router S is
building a single Traffic Engineering Database (TED) based on both
IS-IS and OSPF TE information. R may then appear as two separate
nodes in S's TED. However, if both the IS-IS and OSPF LSAs generated
by R contain the same Router Address, then S can determine that the
IS-IS TE LSA and the OSPF TE LSA from R are indeed from a single
router.

The router address TLV is type 1, has a length of 4, and a value that
is the four octet IP address. It must appear in exactly one Traffic
Engineering LSA originated by a router.

2.4.2. Link TLV

The Link TLV describes a single link. It is constructed of a set of
sub-TLVs. There are no ordering requirements for the sub-TLVs.

Only one Link TLV shall be carried in each LSA, allowing for fine
granularity changes in topology.

The Link TLV is type 2, and the length is variable.

The following sub-TLVs of the Link TLV are defined:

    1 - Link type (1 octet)
    2 - Link ID (4 octets)
    3 - Local interface IP address (4 octets)
    4 - Remote interface IP address (4 octets)
    5 - Traffic engineering metric (4 octets)
    6 - Maximum bandwidth (4 octets)
    7 - Maximum reservable bandwidth (4 octets)
    8 - Unreserved bandwidth (32 octets)
    9 - Administrative group (4 octets)

This memo defines sub-Types 1 through 9. See the IANA Considerations
section for allocation of new sub-Types.

The Link Type and Link ID sub-TLVs are mandatory, i.e., must appear
exactly once. All other sub-TLVs defined here may occur at most
once. These restrictions need not apply to future sub-TLVs.
Unrecognized sub-TLVs are ignored.

Various values below use the (32 bit) IEEE Floating Point format.
For quick reference, this format is 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|   Exponent    |                   Fraction                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
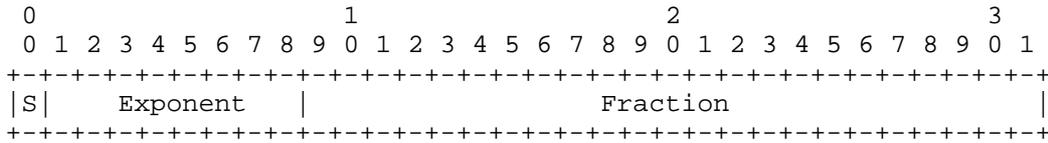
S is the sign, Exponent is the exponent base 2 in "excess 127"
notation, and Fraction is the mantissa - 1, with an implied binary
point in front of it. Thus, the above represents the value:

$$(-1)^{**}(S) * 2^{**}(Exponent-127) * (1 + Fraction)$$

For more details, refer to [4].

2.5. Sub-TLV Details

2.5.1. Link Type

The Link Type sub-TLV defines the type of the link:

     1 - Point-to-point
     2 - Multi-access

The Link Type sub-TLV is TLV type 1, and is one octet in length.

2.5.2. Link ID

The Link ID sub-TLV identifies the other end of the link. For
point-to-point links, this is the Router ID of the neighbor. For
multi-access links, this is the interface address of the designated
router. The Link ID is identical to the contents of the Link ID
field in the Router LSA for these link types.

The Link ID sub-TLV is TLV type 2, and is four octets in length.

2.5.3. Local Interface IP Address

The Local Interface IP Address sub-TLV specifies the IP address(es)
of the interface corresponding to this link. If there are multiple
local addresses on the link, they are all listed in this sub-TLV.

The Local Interface IP Address sub-TLV is TLV type 3, and is 4N
octets in length, where N is the number of local addresses.

2.5.4. Remote Interface IP Address

   The Remote Interface IP Address sub-TLV specifies the IP address(es)
   of the neighbor's interface corresponding to this link. This and the
   local address are used to discern multiple parallel links between
   systems. If the Link Type of the link is Multi-access, the Remote
   Interface IP Address is set to 0.0.0.0; alternatively, an
   implementation MAY choose not to send this sub-TLV.

   The Remote Interface IP Address sub-TLV is TLV type 4, and is 4N
   octets in length, where N is the number of neighbor addresses.

2.5.5. Traffic Engineering Metric

   The Traffic Engineering Metric sub-TLV specifies the link metric for
   traffic engineering purposes. This metric may be different than the
   standard OSPF link metric. Typically, this metric is assigned by a
   network administrator.

   The Traffic Engineering Metric sub-TLV is TLV type 5, and is four
   octets in length.

2.5.6. Maximum Bandwidth

   The Maximum Bandwidth sub-TLV specifies the maximum bandwidth that
   can be used on this link, in this direction (from the system
   originating the LSA to its neighbor), in IEEE floating point format.
   This is the true link capacity. The units are bytes per second.

   The Maximum Bandwidth sub-TLV is TLV type 6, and is four octets in
   length.

2.5.7. Maximum Reservable Bandwidth

   The Maximum Reservable Bandwidth sub-TLV specifies the maximum
   bandwidth that may be reserved on this link, in this direction, in
   IEEE floating point format. Note that this may be greater than the
   maximum bandwidth (in which case the link may be oversubscribed).
   This SHOULD be user-configurable; the default value should be the
   Maximum Bandwidth. The units are bytes per second.

   The Maximum Reservable Bandwidth sub-TLV is TLV type 7, and is four
   octets in length.

2.5.8. Unreserved Bandwidth

   The Unreserved Bandwidth sub-TLV specifies the amount of bandwidth
   not yet reserved at each of the eight priority levels in IEEE
   floating point format. The values correspond to the bandwidth that
   can be reserved with a setup priority of 0 through 7, arranged in
   increasing order with priority 0 occurring at the start of the sub-
   TLV, and priority 7 at the end of the sub-TLV. The initial values
   (before any bandwidth is reserved) are all set to the Maximum
   Reservable Bandwidth. Each value will be less than or equal to the
   Maximum Reservable Bandwidth. The units are bytes per second.

   The Unreserved Bandwidth sub-TLV is TLV type 8, and is 32 octets in
   length.

2.5.9. Administrative Group

   The Administrative Group sub-TLV contains a 4-octet bit mask assigned
   by the network administrator. Each set bit corresponds to one
   administrative group assigned to the interface. A link may belong to
   multiple groups.

   By convention, the least significant bit is referred to as 'group 0',
   and the most significant bit is referred to as 'group 31'.

   The Administrative Group is also called Resource Class/Color [5].

   The Administrative Group sub-TLV is TLV type 9, and is four octets in
   length.

3. Elements of Procedure

   Routers shall originate Traffic Engineering LSAs whenever the LSA
   contents change, and whenever otherwise required by OSPF (an LSA
   refresh, for example). Note that this does not mean that every
   change must be flooded immediately; an implementation MAY set
   thresholds (for example, a bandwidth change threshold) that trigger
   immediate flooding, and initiate flooding of other changes after a
   short time interval. In any case, the origination of Traffic
   Engineering LSAs SHOULD be rate-limited to at most one every
   MinLSInterval [1].

   Upon receipt of a changed Traffic Engineering LSA or Network LSA
   (since these are used in traffic engineering calculations), the
   router should update its traffic engineering database. No Shortest
   Path First (SPF) or other route calculations are necessary.

4. Compatibility Issues

   There should be no interoperability issues with routers that do not
   implement these extensions, as the Opaque LSAs will be silently
   ignored.

   The result of having routers that do not implement these extensions
   is that the traffic engineering topology will be missing pieces.
   However, if the topology is connected, TE paths can still be
   calculated and ought to work.

5. Security Considerations

   This document specifies the contents of Opaque LSAs in OSPFv2. As
   Opaque LSAs are not used for SPF computation or normal routing, the
   extensions specified here have no affect on IP routing. However,
   tampering with TE LSAs may have an effect on traffic engineering
   computations, and it is suggested that any mechanisms used for
   securing the transmission of normal OSPF LSAs be applied equally to
   all Opaque LSAs, including the TE LSAs specified here.

   Note that the mechanisms in [1] and [9] apply to Opaque LSAs. It is
   suggested that any future mechanisms proposed to secure/authenticate
   OSPFv2 LSA exchanges be made general enough to be used with Opaque
   LSAs.

6. IANA Considerations

   The top level Types in a TE LSA, as well as Types for sub-TLVs for
   each top level Type, have been registered with IANA, except as noted.

   Here are the guidelines (using terms defined in [10]) for the
   assignment of top level Types in TE LSAs:

   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.

The guidelines for the assignment of types for sub-TLVs in a TE LSA
are as follows:

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. 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.

7. Intellectual Property Rights Statement

   The IETF takes no position regarding the validity or scope of any
   intellectual property or other rights that might be claimed to
   pertain to the implementation or use of the technology described in
   this document or the extent to which any license under such rights
   might or might not be available; neither does it represent that it
   has made any effort to identify any such rights. Information on the
   IETF's procedures with respect to rights in standards-track and
   standards-related documentation can be found in BCP-11. Copies of
   claims of rights made available for publication and any assurances of
   licenses to be made available, or the result of an attempt made to
   obtain a general license or permission for the use of such
   proprietary rights by implementors or users of this specification can
   be obtained from the IETF Secretariat.

   The IETF invites any interested party to bring to its attention any
   copyrights, patents or patent applications, or other proprietary
   rights which may cover technology that may be required to practice
   this standard. Please address the information to the IETF Executive
   Director.

8. References

8.1. Normative References

   [1] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

   [2] Bradner, S., "Key words for use in RFCs to Indicate Requirement
        Levels", BCP 14, RFC 2119, March 1997.

   [3] Coltun, R., "The OSPF Opaque LSA Option", RFC 2370, July 1998.

   [4] IEEE, "IEEE Standard for Binary Floating-Point Arithmetic",
        Standard 754-1985, 1985 (ISBN 1-5593-7653-8).

8.2. Informative References

   [5] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M. and J.
        McManus, "Requirements for Traffic Engineering Over MPLS", RFC
        2702, September 1999.

   [6] Smit, H. and T. Li, "ISIS Extensions for Traffic Engineering",
        work in progress.

   [7] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in
        Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)",
        RFC 3477, January 2003.

   [8] Kompella, K., Rekhter, Y. and A. Kullberg, "Signalling
        Unnumbered Links in CR-LDP (Constraint-Routing Label
        Distribution Protocol)", RFC 3480, February 2003.

   [9] Murphy, S., Badger, M. and B. Wellington, "OSPF with Digital
        Signatures", RFC 2154, June 1997.

   [10] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA
        Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

9. Authors' Addresses

   Dave Katz
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA 94089 USA

   Phone: +1 408 745 2000
   EMail: dkatz@juniper.net

   Derek M. Yeung
   Procket Networks, Inc.
   1100 Cadillac Court
   Milpitas, CA 95035 USA

   Phone: +1 408 635-7900
   EMail: myeung@procket.com

   Kireeti Kompella
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA 94089 USA

   Phone: +1 408 745 2000
   EMail: kireeti@juniper.net

10. Full Copyright Statement

Acknowledgement