

IS-IS extensions for Traffic Engineering

<[draft-ietf-isis-traffic-03.txt](#)>

Status

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC 2026](#) except that the right to produce derivative works is not granted.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

1.0 Abstract

This document describes extensions to the IS-IS protocol to support Traffic Engineering [1]. The IS-IS protocol is specified in [2], with extensions for supporting IPv4 specified in [3].

This document extends the IS-IS protocol by specifying new information that an Intermediate System (IS) [router] can place in Link State Protocol Data Units (LSPs). This information describes additional information about the state of the network that is useful for traffic engineering computations.

INTERNET DRAFT

June 2001

[2.0](#) Introduction

An IS-IS LSP is composed of a fixed header and a number of tuples, each consisting of a Type, a Length, and a Value. Such tuples are commonly known as TLVs, and are a good way of encoding information in a flexible and extensible format.

The changes in this document include the design of new TLVs to replace the existing IS Neighbor TLV, IP Reachability TLV and add additional information. Mechanisms and procedures to migrate to the new TLVs are not discussed in this document.

The primary goal of these extensions is to add more information about the characteristics of a particular link to an IS-IS's LSP.

Secondary goals include increasing the dynamic range of the IS-IS metric and improving the encoding of IP prefixes. The router id is useful for traffic engineering purposes because it describes a single address that can always be used to reference a particular router.

This document is a publication of the IS-IS Working Group within the IETF, and is a contribution to ISO IEC JTC1/SC6, for eventual inclusion with ISO 10589 [\[2\]](#).

[3.0](#) Introducing Sub-TLVs

This document introduces a new way to encode routing information in IS-IS. The new object is called a sub-TLV. Sub-TLVs are similar to regular TLVs. They use the same concepts as regular TLVs. The difference is that TLVs exist inside IS-IS packets, while sub-TLVs exist inside TLVs. TLVs are used to add extra information to IS-IS packets. Sub-TLVs are used to add extra information to particular TLVs. Each sub-TLV consists of three fields. A one-octet Type field, a one-octet Length field, and zero or more octets of Value. The Type field indicates the type of items in the Value field. The Length field indicates the length of the Value field in octets. Each sub-TLV can potentially hold multiple items. The number of items in a sub-TLV can be computed from the length of the whole sub-TLV, when the length of each item is known. Unknown sub-TLVs are to be ignored and skipped on receipt.

INTERNET DRAFT

June 2001

[4.0](#) The extended IS reachability TLV

The extended IS reachability TLV is TLV type 22.

The existing IS reachability (TLV type 2, defined in ISO 10589 [\[2\]](#)) contains information about a series of IS neighbors. For each neighbor, there is a structure that contains the default metric, the delay, the monetary cost, the reliability, and the 7-octet ID of the adjacent neighbor. Of this information, the default metric is commonly used. The default metric is currently one octet, with one bit used to indicate that the metric is present and one bit used to indicate whether the metric is internal or external. The remaining 6 bits are used to store the actual metric, resulting a possible metric range of 0-63. This limitation is one of the restrictions that we would like to lift.

The remaining three metrics (delay, monetary cost, and reliability) are not commonly implemented and reflect unused overhead in the TLV. The neighbor is identified by its system Id (typically 6-octets), plus one octet to indicate the pseudonode number if the neighbor is on a LAN interface. Thus, the existing TLV consumes 11 octets per neighbor, with 4 octets for metric and 7 octets for neighbor identification. To indicate multiple adjacencies, this structure is repeated within the IS reachability TLV. Because the TLV is limited to 255 octets of content, a single TLV can describe up to 23 neighbors. The IS reachability TLV can be repeated within the LSP fragments to describe further neighbors.

The proposed extended IS reachability TLV contains a new data structure, consisting of

- 7 octets of system Id and pseudonode number

- 3 octets of default metric

- 1 octet of length of sub-TLVs

- 0-244 octets of sub-TLVs,

- where each sub-TLV consists of a sequence of

- 1 octet of sub-type

1 octet of length
0-242 octets of value

Thus, if no sub-TLVs are used, the new encoding requires 11 octets and can contain up to 23 neighbors. Please note that while the encoding allows for 255 octets of sub-TLVs, the maximum value cannot fit in the overall IS reachability TLV. The practical maximum is 255 octets minus the 11 octets described above, or 244 octets. Further, there is no defined mechanism for extending the sub-TLV space for a particular neighbor. Thus, wasting sub-TLV space is discouraged.

Expiration Date December 2001

[Page 3]

INTERNET DRAFT

June 2001

The metric octets are encoded as a 24-bit unsigned integer. Note that the metric field in the new extended IP reachability TLV is encoded as a 32-bit unsigned integer. These different sizes were chosen so that it is very unlikely that the cost of an intra-area route has to be chopped off to fit in the metric field of an inter-area route.

To preclude overflow within an SPF implementation, all metrics greater than or equal to MAX_PATH_METRIC shall be considered to have a metric of MAX_PATH_METRIC. It is easiest to select MAX_PATH_METRIC such that MAX_PATH_METRIC plus a single link metric does not overflow the number of bits for internal metric calculation. We assume that this is 32 bits. Thus, MAX_PATH_METRIC is 4,261,412,864 (0xFE000000, $2^{32} - 2^{25}$).

If a link is advertised with the maximum link metric ($2^{24} - 1$), this link should not be considered during the normal SPF computation. This will allow advertisement of a link for other purposes than building the normal Shortest Path Tree. An example is a link that is available for traffic engineering, but not for hop-by-hop routing.

Certain sub-TLVs are proposed here:

Sub-TLV type	Length (octets)	Name
3	4	Administrative group (color)
6	4	IPv4 interface address
8	4	IPv4 neighbor address
9	4	Maximum link bandwidth
10	4	Reservable link bandwidth
11	32	Unreserved bandwidth
18	3	TE Default metric

250-254
255

Reserved for cisco specific extensions
Reserved for future expansion

Each of these sub-TLVs is described below. Unless stated otherwise, multiple occurrences of the information are supported by multiple inclusions of the sub-TLV.

[4.1](#) Sub-TLV 3: Administrative group (color, resource class)

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.

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

Expiration Date December 2001

[Page 4]

INTERNET DRAFT

June 2001

[4.2](#) Sub-TLV 6: IPv4 interface address

This sub-TLV contains a 4-octet IPv4 address for the interface described by the (main) TLV. This sub-TLV can occur multiple times.

Implementations MUST NOT inject a /32 prefix for the interface address into their routing or forwarding table, because this can lead to forwarding loops when interacting with systems that do not support this sub-TLV.

If a router implements the basic TLV extensions in this document, it is free to add or omit this sub-TLV to the description of an adjacency. If a router implements traffic engineering, it must include this sub-TLV.

[4.3](#) Sub-TLV 8: IPv4 neighbor address

This sub-TLV contains a single IPv4 address for a neighboring router on this link. This sub-TLV can occur multiple times.

Implementations MUST NOT inject a /32 prefix for the neighbor address into their routing or forwarding table, because this can lead to forwarding loops when interacting with systems that do not support this sub-TLV.

If a router implements the basic TLV extensions in this document, it is free to add or omit this sub-TLV to the description of an adjacency. If a router implements traffic engineering, it must include this sub-TLV on point-to-point adjacencies.

[4.4](#) Sub-TLV 9: Maximum link bandwidth

This sub-TLV contains the maximum bandwidth that can be used on this link in this direction (from the system originating the LSP to its neighbors). This is useful for traffic engineering.

The maximum link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes (not bits!) per second.

[4.5](#) Sub-TLV 10: Maximum reservable link bandwidth

This sub-TLV contains the maximum amount of bandwidth that can be reserved in this direction on this link. Note that for oversubscription purposes, this can be greater than the bandwidth of the link.

The maximum reservable link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes (not bits!) per second.

[4.6](#) Sub-TLV 11: Unreserved bandwidth

This sub-TLV contains the amount of bandwidth reservable on this

direction on this link. Note that for oversubscription purposes, this can be greater than the bandwidth of the link.

Because of the need for priority and preemption, each head end needs to know the amount of reserved bandwidth at each priority level. Thus, this sub-TLV contains eight 32 bit IEEE floating point numbers. The units are bytes (not bits!) per second. The values correspond to the bandwidth that can be reserved with a setup priority 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.

For stability reasons, rapid changes in the values in this sub-TLV should not cause rapid generation of LSPs.

[4.7](#) Sub-TLV 18: Traffic Engineering Default metric

This sub-TLV contains a 24-bit unsigned integer. This metric is administratively assigned and can be used to present a differently weighted topology to traffic engineering SPF calculations.

To preclude overflow within an SPF implementation, all metrics greater than or equal to MAX_PATH_METRIC shall be considered to have a metric of MAX_PATH_METRIC. It is easiest to select MAX_PATH_METRIC such that MAX_PATH_METRIC plus a single link metric does not overflow the number of bits for internal metric calculation. We assume that this is 32 bits. Thus, MAX_PATH_METRIC is 4,261,412,864 (0xFE000000, $2^{32} - 2^{25}$).

If a link is advertised without this sub-TLV, traffic engineering SPF calculations must use the normal default metric of this link, which is advertised in the fixed part of the extended IS reachability TLV.

[5.0](#) The extended IP reachability TLV

The extended IP reachability TLV is TLV type 135. The existing IP reachability TLVs (TLV type 128 and TLV type 130, defined in [RFC 1195](#) [3]) carry IP prefixes in a format that is

analogous to the IS neighbor TLV from ISO 10589 [2]. They carry four metrics, of which only the default metric is commonly used. Of this, the default metric has a possible range of 0-63. This limitation is one of the restrictions that we would like to lift.

In addition, route redistribution (a.k.a. route leaking) has a key problem that was not fully addressed by the existing IP reachability TLVs. RFC 1195 [3] allows a router to advertise prefixes upwards in the level hierarchy. Unfortunately there were no mechanisms defined to advertise prefixes downwards in the level hierarchy.

To address these two issues, the proposed extended IP reachability TLV provides for a 32 bit metric and adds one bit to indicate that a prefix has been redistributed 'down' in the hierarchy.

The proposed extended IP reachability TLV contains a new data structure, consisting of:

- 4 octets of metric information
- 1 octet of control information, consisting of
 - 1 bit of up/down information
 - 1 bit indicating the existence of sub-TLVs
 - 6 bits of prefix length
- 0-4 octet of IPv4 prefix
- 0-250 optional octets of sub-TLVs, if present consisting of
 - 1 octet of length of sub-TLVs
 - 0-249 octets of sub-TLVs,
 - where each sub-TLV consists of a sequence of
 - 1 octet of sub-type
 - 1 octet of length
 - 0-247 octets of value

This data structure can be replicated within the TLV, not to exceed the maximum length of the TLV.

The 6 bits of prefix length can have the values 0-32 and indicate the number of significant bits in the prefix. The prefix is encoded in the minimal number of octets for the given number of significant bits. This implies:

Significant bits	Octets
0	0
1-8	1
9-16	2
17-24	3
25-32	4

The remaining bits of prefix are transmitted as zero and ignored upon receipt.

If a prefix is advertised with a metric larger than MAX_PATH_METRIC (0xFE000000, see paragraph 4.0), this prefix should not be considered during the normal SPF computation. This will allow advertisement of a prefix for other purposes than building the normal IP routing table.

5.1 The up/down bit

Without any additional mechanisms, if routers were allowed to redistribute IP prefixes freely in both directions between level 1 and level 2, those routers can not determine looping of routing information. A problem occurs when a router learns an prefix via level 2 routing and advertises that prefix down into a level 1 area, where another router might pick up the route and advertise the prefix back up into the level 2 backbone. If the original source withdraws the prefix, those two routers might end up having a routing loop between them, where part of the looped path is via level 1 routing and the other part of the looped path is via level 2 routing. The solution that [RFC 1195](#) [3] poses is to allow only advertising prefixes upward in the level hierarchy, and to disallow the advertising of prefixes downward in the hierarchy.

To prevent this looping of prefixes between levels, a new bit of information is defined in the new extended IP reachability TLV. This bit is called the up/down bit. The up/down bit shall be set to 0 when a prefix is first injected into IS-IS. If a prefix is advertised from a higher level to a lower level (e.g. level two to level one), the bit shall be set to 1, to indicate that the prefix has traveled down the hierarchy. Prefixes that have the up/down bit set to 1 may only be advertised down the hierarchy, i.e. to lower levels.

These semantics apply even if IS-IS is extended in the future to have additional levels. By insuring that prefixes follow only the IS-IS hierarchy, we have insured that the information does not loop, thereby insuring that there are no persistent forwarding loops.

If a prefix is advertised from an area to another area at the same level, then the up/down bit shall be set to 1. This situation can arise when a router implements multiple virtual routers at the same level, but in different areas.

The semantics of the up/down bit in the new extended IP reachability TLV are identical to the semantics of the up/down bit defined in [RFC 2966](#) [4].

[5.2](#) Expandability of the extended IP reachability TLV with sub-TLVs

The extended IP reachability TLV can hold sub-TLVs that apply to a particular prefix. This allows for easy future extensions. If there are no sub-TLVs associated with a prefix, the bit indicating the presence of sub-TLVs shall be set to 0. If this bit is set to 1, the first octet after the prefix will be interpreted as the length of sub-TLVs. Please note that while the encoding allows for 255 octets of sub-TLVs, the maximum value cannot fit in the overall extended IP reachability TLV. The practical maximum is 255 octets minus the 5-9 octets described above, or 250 octets.

This document does not define any sub-TLVs yet for the extended IP reachability TLV.

[6.0](#) The Traffic Engineering router ID TLV

The Traffic Engineering router ID TLV is TLV type 134.

The router ID TLV contains the 4-octet router ID of the router originating the LSP. This is useful in several regards:

For traffic engineering, it guarantees that we have a single stable address that can always be referenced in a path that will be reachable from multiple hops away, regardless of the state of the node's interfaces.

If OSPF is also active in the domain, traffic engineering can compute

the mapping between the OSPF and IS-IS topologies.

INTERNET DRAFT

June 2001

If a router advertises the Traffic Engineering router ID TLV in its LSP, and if it advertises BGP routes with the BGP next hop attribute set to the BGP router ID, in that case the Traffic Engineering router ID should be the same as the BGP router ID.

Implementations MUST NOT inject a /32 prefix for the router ID into their forwarding table, because this can lead to forwarding loops when interacting with systems that do not support this TLV.

[7.0](#) Security Considerations

This document raises no new security issues for IS-IS.

[8.0](#) Acknowledgments

The authors would like to thank Yakov Rekhter and Dave Katz for their comments on this work.

[9.0](#) References

- [1] [RFC 2702](#), "Requirements for Traffic Engineering Over MPLS," D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus, September 1999.
- [2] ISO 10589, "Intermediate System to Intermediate System Intra-Domain Routeing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)" [Also republished as [RFC 1142](#)]
- [3] [RFC 1195](#), "Use of OSI IS-IS for routing in TCP/IP and dual environments", R.W. Callon, December 1990
- [4] [RFC 2966](#), "Domain-wide Prefix Distribution with Two-Level IS-IS", T. Li, T. Przygienda, H. Smit, October 2000.

INTERNET DRAFT

June 2001

[10.0](#) Authors' Addresses

Tony Li
Procket Networks, Inc.
1100 Cadillac Court
Milpitas, CA 95035
Email: tli@procket.com
Voice: +1 408 6357900
Fax: +1 408 6356166

Henk Smit
Procket Networks, Inc.
1100 Cadillac Court
Milpitas, CA 95035
Email: henk@procket.com
Voice: +1 408 6357900
Fax: +1 408 6356166

Expiration Date December 2001

[Page 11]