Network Working Group   Henk Smit
INTERNET DRAFT   Cisco Systems
                                                                 Tony Li
                                                       Juniper Networks
                                                                May 1999

IS-IS extensions for Traffic Engineering

<draft-ietf-isis-traffic-01.txt>

Status

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 a 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.

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.

3.0 The router ID TLV

The 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.

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.

4.0 The extended IP reachability TLV

The extended IP reachability TLV is TLV type 135.

The existing IP reachability TLV is a single TLV that carries IP prefixes in a format that is analogous to the IS neighbor TLV. It carries 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) is a key problem that is not addressed by the existing IP reachability TLV. This problem occurs when an IP prefix is injected into a level one area, redistributed into level 2, subsequently redistributed into a second level one area, and then redistributed from the second level one area back into level two. This problem occurs because the path that the information can take forms a loop. The likely result is a forwarding loop.

To address these 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 bytes of metric information
        1 byte 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 bytes of IPv4 prefix
        0-250 optional octets of sub-TVLs, if present consisting of
                1 octet of length of sub-TLVs
                0-249 octets of sub-TLVs
```

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

The up/down bit shall be set to 0 when a prefix is first injected into IS-IS. If a prefix is redistributed 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 travelled down the hierarchy. Prefixes that have the up/down bit set to 1 must not be redistributed. If a prefix is redistributed from an area to another area at the same level, then the up/down bit shall be set to 1.

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 there are no sub-TLVs associated with this IP prefix, the bit indicating the presence of sub-TVLs 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. No sub-TLVs for the extended IP reachability TLV have been defined yet.

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 bytes for the given number of significant bits. This implies:

Significant bits	Bytes
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 an IP prefix is advertised with a metric larger then MAX_PATH_METRIC (0xFE000000, see below), this IP prefix should not be considered during the normal SPF computation. This will allow advertisment of an IP prefix for other purposes than building the normal IP routing table.

5.0 The extended IS reachability TLV

The extended IS reachability TLV is TLV type 22.

The existing IS reachability TLV is a single TLV that 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

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.

The metric octets are encoded as a 24-bit unsigned integer. 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 advertisment 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		Reserved for cisco specific extensions
255		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.

5.1 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'.

5.2 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.

5.3 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.

5.4 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.

5.5 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.

5.6 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 holding of 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.

5.7 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 TLV 22.

6.0 Security Considerations

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

7.0 Acknowledgments

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

8.0 References

[1] draft-ietf-mpls-traffic-eng-00.txt, "Requirements for Traffic Engineering Over MPLS", D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus, work in progress.

[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, Dec. 1990

9.0 Authors' Addresses

Henk Smit
Cisco Systems, Inc.
210 West Tasman Drive
San Jose, CA 95134
Email: hsmit@cisco.com
Voice: +31 20 342 3736

Tony Li
Juniper Networks, Inc.
385 Ravendale Dr.
Mountain View, CA 94043
Email: tli@juniper.net
Fax: +1 650 526 8001
Voice: +1 650 526 8006