Network Working Group
Request for Comments: 3784
Category: Informational

H. Smit
Procket Networks
T. Li

June 2004

# Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)

## Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

## Copyright Notice

Copyright (C) The Internet Society (2004).

## Abstract

This document describes extensions to the Intermediate System to Intermediate System (IS-IS) protocol to support Traffic Engineering (TE). This document extends the IS-IS protocol by specifying new information that an Intermediate System (router) can place in Link State Protocol (LSP) Data Units. This information describes additional details regarding the state of the network that are useful for traffic engineering computations.

## 1. Introduction

The IS-IS protocol is specified in ISO 10589 [1], with extensions for supporting IPv4 specified in RFC 1195 [3]. Each Intermediate System (IS) (router) advertises one or more IS-IS Link State Protocol Data Units (LSPs) with routing information. Each 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.

This document contains the design of new TLVs to replace the existing IS Neighbor TLV, IP Reachability TLV, and to include additional information about the characteristics of a particular link to an IS-IS LSP. The characteristics described in this document are needed for Traffic Engineering [4] (TE). 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.

Mechanisms and procedures to migrate to the new TLVs are not discussed in this document.

## 2. 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 upon receipt.

The Sub-TLV type space is managed by the IETF IS-IS WG (http://www.ietf.org/html.charters/isis-charter.html). New type values are allocated following review on the IETF IS-IS mailing list. This will normally require publication of additional documentation describing how the new type is used. In the event that the IS-IS working group has disbanded the review shall be performed by a Designated Expert assigned by the responsible Area Director.

## 3. 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 [1]) 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 whether the metric is internal or external, and one bit that was originally unused, but which was later defined by RFC 2966 to be the up/down bit. The remaining 6 bits are used to store the actual metric, resulting in 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 indicating the pseudonode number. 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 of the value field of the sub-TLV
          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.

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 a traffic engineering 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. Therefore, we have chosen MAX_PATH_METRIC to be 4,261,412,864 (0xFE000000, 2^32 - 2^25).

If a link is advertised with the maximum link metric (2^24 - 1), this link MUST NOT be considered during the normal SPF computation. This will allow advertisement of a link for purposes other 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.

## 3.1. 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'.

This sub-TLV is OPTIONAL. This sub-TLV SHOULD appear once at most in each extended IS reachability TLV.

## 3.2. 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 MAY add or omit this sub-TLV from the description of an adjacency. If a router implements traffic engineering, it MUST include this sub-TLV.

### 3.3. 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 MAY add or omit this sub-TLV from the description of an adjacency. If a router implements traffic engineering, it MUST include this sub-TLV on point-to-point adjacencies.

### 3.4. 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.

This sub-TLV is optional. This sub-TLV SHOULD appear once at most in each extended IS reachability TLV.

### 3.5. 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.

This sub-TLV is optional. This sub-TLV SHOULD appear once at most in each extended IS reachability TLV.

### 3.6. Sub-TLV 11: Unreserved bandwidth

This sub-TLV contains the amount of bandwidth reservable in 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 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.

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

This sub-TLV is optional. This sub-TLV SHOULD appear once at most in each extended IS reachability TLV.

### 3.7. 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 a traffic engineering 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. Therefore, we have chosen MAX_PATH_METRIC to be 4,261,412,864 (0xFE000000, 2^32 - 2^25).

This sub-TLV is optional. This sub-TLV SHOULD appear once at most in each extended IS reachability TLV. 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.

## 4. 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 [1]. They carry four metrics, of which only the default metric is commonly used. The

default metric has a possible range of 0-63. We would like to remove this restriction.

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 presence 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 of the value field of the sub-TLV
               0-247 octets of value
```

This data structure can be replicated within the TLV, without exceeding 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 then MAX_PATH_METRIC (0xFE000000, see paragraph 3.0), this prefix MUST NOT be considered during the normal SPF computation. This allows advertisement of a prefix for purposes other than building the normal IP routing table.

### 4.1. The up/down Bit

If routers were allowed to redistribute IP prefixes freely in both directions between level 1 and level 2 without any additional mechanisms, those routers would not be able to determine looping of routing information. A problem occurs when a router learns a 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 2 to level 1), the bit MUST be set to 1, indicating 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 one area to another 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 [2].

### 4.2. 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 all sub-TLVs associated with this IPv4 prefix. 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 for the extended IP reachability TLV.

## 5. 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.

If a router does not implement traffic engineering, it MAY add or omit the Traffic Engineering router ID TLV. If a router implements traffic engineering, it MUST include this TLV in its LSP. This TLV SHOULD not be included more than once in an LSP.

If a router advertises the Traffic Engineering router ID TLV in its LSP, and if it advertises prefixes via the Border Gateway Protocol (BGP) with the BGP next hop attribute set to the BGP router ID, 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.

# 6. IANA Considerations

## 6.1. TLV Codepoint Allocations

This document defines the following new IS-IS TLV types, which have been reflected in the ISIS TLV code-point registry:

| Type | Description | IIH | LSP | SNP |
|---|---|---|---|---|
| 22 | The extended IS reachability TLV | n | y | n |
| 134 | The Traffic Engineering router ID TLV | n | y | n |
| 135 | The extended IP reachability TLV | n | y | n |

## 6.2. New Registries

IANA has created the following new registries.

### 6.2.1. Sub-TLVs for the Extended IS Reachability TLV

This registry contains codepoints for Sub-TLVs of TLV 22. The range of values is 0-255. Allocations within the registry require documentation of the proposed use of the allocated value and approval by the Designated Expert assigned by the IESG (see [5]).

Taking into consideration allocations specified in this document, the registry has been initialized as follows:

| Type | Description |
|---|---|
| 0-2 | unassigned |
| 3 | Administrative group (color) |
| 4-5 | unassigned |
| 6 | IPv4 interface address |
| 7 | unassigned |
| 8 | IPv4 neighbor address |
| 9 | Maximum link bandwidth |
| 10 | Reservable link bandwidth |
| 11 | Unreserved bandwidth |
| 12-17 | unassigned |
| 18 | TE Default metric |
| 19-254 | unassigned |
| 255 | Reserved for future expansion |

#### 6.2.2. Sub-TLVs for the Extended IP Reachability TLV

This registry contains codepoints for Sub-TLVs of TLV 135. The range of values is 0-255. Allocations within the registry require documentation of the use of the allocated value and approval by the Designated Expert assigned by the IESG (see [5]).

No codepoints are defined in this document.

## 7. References

### 7.1. Normative References

[1] ISO, "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)", International Standard 10589:2002, Second Edition

[2] Li, T., Przygienda, T. and H. Smit, "Domain-wide Prefix Distribution with Two-Level IS-IS", RFC 2966, October 2000.

### 7.2. Informative References

[3] Callon, R.W., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990

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

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

## 8. Security Considerations

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

## 9. Acknowledgments

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

## 10. Authors' Addresses

Henk Smit

EMail: hhwsmit@xs4all.nl

Tony Li

EMail: tony.li@tony.li

## 11. Full Copyright Statement

Copyright (C) The Internet Society (2004). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

### Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights 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; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat 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 implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

### Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.