Network Working Group
Request for Comments: 5120
Category: Standards Track

T. Przygienda
Z2 Sagl
N. Shen
Cisco Systems
N. Sheth
Juniper Networks

February 2008

M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)

Status of This Memo

Abstract

This document describes an optional mechanism within Intermediate System to Intermediate Systems (IS-ISs) used today by many ISPs for IGP routing within their clouds. This document describes how to run, within a single IS-IS domain, a set of independent IP topologies that we call Multi-Topologies (MTs). This MT extension can be used for a variety of purposes, such as an in-band management network "on top" of the original IGP topology, maintaining separate IGP routing domains for isolated multicast or IPv6 islands within the backbone, or forcing a subset of an address space to follow a different topology.

1. Introduction

Maintaining multiple MTs for IS-IS [ISO10589] [RFC1195] in a backwards-compatible manner necessitates several extensions to the packet encoding and additional Shortest Path First (SPF) procedures. The problem can be partitioned into the forming of adjacencies and advertising of prefixes and reachable intermediate systems within each topology. Having put all the necessary additional information in place, it must be properly used by MT capable SPF computation. The following sections describe each of the problems separately. To simplify the text, "standard" IS-IS topology is defined to be MT ID #0 (zero).

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].

1.2. Definitions of Terms Used in This Document

CSNP Complete Sequence Number Packet. Used to describe all the contents of a link state database of IS-IS.

DIS Designated Intermediate System. The intermediate system elected to advertise the pseudo-node for a broadcast network.

IIH IS-IS Hello. Packets that are used to discover adjacent intermediate systems.

LSP Link State Packet. Packet generated by an intermediate system and lists adjacent systems, prefixes, and other information.

PSNP Partial Sequence Number Packet. Used to request information from an adjacent intermediate system's link state database.

SPF Shortest Path First. An algorithm that takes a database of nodes within a domain and builds a tree of connectivity along the shortest paths through the entire network.

2. Maintaining MT Adjacencies

Each adjacency formed MUST be classified as belonging to a set of MTs on the interface. This is achieved by adding a new TLV into IIH packets that advertises to which topologies the interface belongs. If MT #0 is the only MT on the interface, it is optional to advertise it in the new TLV. Thus, not including such a TLV in the IIH implies MT ID #0 capability only. Through this exchange of MT capabilities, a router is able to advertise the IS TLVs in LSPs with common MT set over those adjacencies.

The case of adjacency contains multiple MTs on an interface, and if there exists an overlapping IP address space among the topologies, additional mechanisms MUST be used to resolve the topology identity of the incoming IP packets on the interface. See further discussion in Section 8.2.2 of this document.

2.1. Forming Adjacencies on Point-to-Point Interfaces

Adjacencies on point-to-point interfaces are formed as usual with IS-IS routers not implementing MT extensions. If a local router does not participate in certain MTs, it will not advertise those MT IDs in its IIHs and thus will not include that neighbor within its LSPs. On the other hand, if an MT ID is not detected in the remote side's IIHs, the local router MUST NOT include that neighbor within its LSPs. The local router SHOULD NOT form an adjacency if they don't have at least one common MT over the interface.

2.2. Forming Adjacencies on Broadcast Interfaces

On a LAN, all the routers on the LAN that implement the MT extension MAY advertise their MT capability TLV in their IIHs. If there is at least one adjacency on the LAN interface that belongs to this MT, the MT capable router MUST include the corresponding MT IS Reachable TLV in its LSP, otherwise it MAY include this MT IS Reachable TLV in its LSP if the LAN interface participates in this MT set.

Two routers on a LAN SHALL always establish adjacency, regardless of whether or not they have a common MT. This is to ensure all the routers on the LAN can correctly elect the same DIS. The IS SHOULD NOT include the MT IS TLV in its LSP if none of the adjacencies on the LAN contain this MT.

The DIS, CSNP, and PSNP functions are not changed by MT extension.

3. Advertising MT Reachable Intermediate Systems in LSPs

A router MUST include within its LSPs in the Reachable Intermediate Systems TLV-only adjacent nodes that are participating in the corresponding topology and advertise such TLVs only if it participates itself in the corresponding topology. The Standard Reachable Intermediate Systems TLV is acting here as MT ID #0, the equivalent of the newly introduced MT Reachable Intermediate Systems TLV. A router MUST announce the MT IS TLV when there is at least one adjacency on the interface that belongs to this MT, otherwise it MAY announce the MT IS TLV of an adjacency for a given MT if this interface participates in the LAN.

Since it is not possible to prevent a router that does not understand MT extensions from being responsible for the generation of the according pseudo-node, it is possible to neither introduce special TLVs in the pseudo-node LSPs, nor run distinct DIS elections per MT. Therefore, a generated pseudo-node LSP by DIS MUST contain

in its IS Reachable TLV all nodes on the LAN as usual, regardless of their MT capabilities. In other words, there is no change to the pseudo-node LSP construction.

4. MTs and Overload, Partition, and Attached Bits

For each of the MTs, a router could become potentially partitioned, overloaded, and attached independently. To prevent unnecessary complexity, MT extensions do not support MT based partition repair. The overload, partition, and attached bits in the LSP header only reflect the status of the default topology.

Attached bit and overload bit are part of the MT TLV being distributed within a node's LSP fragment zero. Since each adjacency can belong to different MTs, it is possible that some MTs are L2 attached, and others are not on the same router. The overload bit in the MT TLV can be used to signal the topology being overloaded. An MT-based system is considered overloaded if the overload bit in the MT is set.

Route leaking between the levels SHOULD only be performed within the same MT.

5. Advertising MT Specific IP Prefixes

Each of the MTs commands its own address space so a new TLV is necessary for prefixes stored in MTs other than MT ID #0. To make the encoding less confusing when same prefixes are present in multiple MTs and accelerate SPF per MT, rather than adding a sub-TLV in Traffic Engineered (TE) extensions, a new TLV is introduced for that purpose that closely follows TE encoding [RFC3784].

6. MT SPF Computation

Each MT MUST run its own instance of the decision process. The pseudo-node LSPs are used by all topologies during computation. Each non-default topology MAY have its attached bit and overload bit set in the MT TLV. A reverse-connectivity check within SPF MUST follow the according MT to assure the bi-directional reachability within the same MT.

The results of each computation SHOULD be stored in a separate Routing Information Base (RIB), in normal cases, otherwise overlapping addresses in different topologies could lead to undesirable routing behavior, such as forwarding loops. The forwarding logic and configuration need to ensure the same MT is traversed from the source to the destination for packets. The nexthops derived from the MT SPF MUST belong to the adjacencies

conforming to the same MT for correct forwarding. It is recommended for the administrators to ensure consistent configuration of all routers in the domain to prevent undesirable forwarding behavior.

No attempt is made in this document to allow one topology to calculate routes using the routing information from another topology inside SPF. Even though it is possible to redistribute and leak routes from another IS-IS topology or from external sources, the exact mechanism is beyond the scope of this document.

7. Packet Encoding

Four new TLVs are added to support MT extensions. One of them is common for the LSPs and IIHs. Encoding of Intermediate System TLV and IPv4 Reachable Prefixes is tied to traffic engineering extensions [RFC3784] to simplify the implementation effort. The main reasons we chose to use new TLVs instead of using sub-TLVs inside existing TLV type-22 and type-135 are:

1. In many cases, multi-topologies are non-congruent, using the sub-TLV approach will not save LSP space;

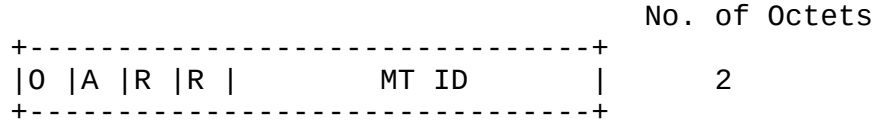
2. Many sub-TLVs are already being used in TLV type-22, and many more are being proposed while there is a maximum limit on the TLV size, from the existing TLVs;

3. If traffic engineering or some other applications are being applied per topology level later, the new TLVs can automatically inherit the same attributes already defined for the "standard" topology without going through long standard process to redefine them per topology.

7.1. Multi-Topology TLV

The TLV number of this TLV is 229. It contains one or more MTs; the router is participating in the following structure:

- CODE - 229
- LENGTH - total length of the value field, it SHOULD be 2 times the number of MT components.
- VALUE - one or more 2-byte MT components, structured as follows:



Bit O represents the OVERLOAD bit for the MT (only valid in LSP fragment zero for MTs other than ID #0, otherwise SHOULD be set to 0 on transmission and ignored on receipt).

Bit A represents the ATTACH bit for the MT (only valid in LSP fragment zero for MTs other than ID #0, otherwise SHOULD be set to 0 on transmission and ignored on receipt).

Bits R are reserved, SHOULD be set to 0 on transmission and ignored on receipt.

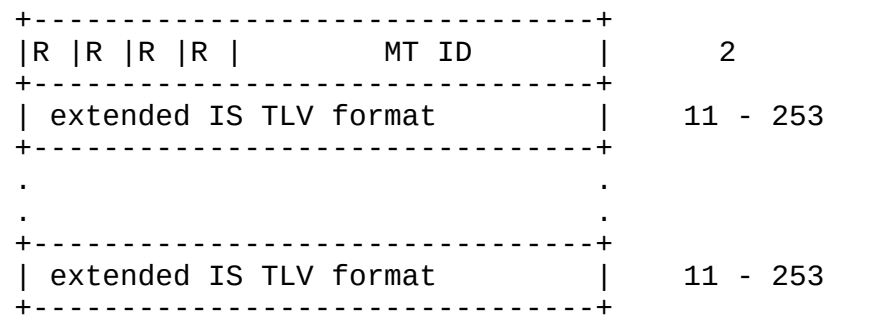
MT ID is a 12-bit field containing the ID of the topology being announced.

This MT TLV can advertise up to 127 MTs. It is announced in IIHs and LSP fragment 0, and can occur multiple times. The resulting MT set SHOULD be the union of all the MT TLV occurrences in the packet. Any other IS-IS PDU occurrence of this TLV MUST be ignored. Lack of MT TLV in hellos and fragment zero LSPs MUST be interpreted as participation of the advertising interface or router in MT ID #0 only. If a router advertises MT TLV, it has to advertise all the MTs it participates in, specifically including topology ID #0 also.

7.2. MT Intermediate Systems TLV

The TLV number of this TLV is 222. It is aligned with extended IS reachability TLV type 22 beside an additional two bytes in front at the beginning of the TLV.

- CODE - 222
- LENGTH - total length of the value field
- VALUE - 2-byte MT membership plus the format of extended IS reachability TLV, structured as follows:



Bits R are reserved, SHOULD be set to 0 on transmission and ignored on receipt.

MT ID is a 12-bit field containing the non-zero MT ID of the topology being announced. The TLV MUST be ignored if the ID is zero. This is to ensure the consistent view of the standard unicast topology.

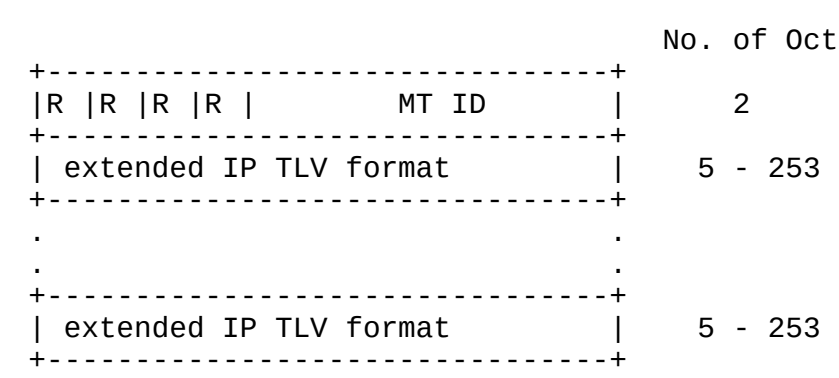
After the 2-byte MT membership format, the MT IS content is in the same format as extended IS TLV, type 22 [RFC3784]. It can contain up to 23 neighbors of the same MT if no sub-TLVs are used.

This TLV can occur multiple times.

7.3. Multi-Topology Reachable IPv4 Prefixes TLV

The TLV number of this TLV is 235. It is aligned with extended IP reachability TLV type 135 beside an additional two bytes in front.

- CODE - 235
- LENGTH - total length of the value field
- VALUE - 2-byte MT membership plus the format of extended IP reachability TLV, structured as follows:



Bits R are reserved, SHOULD be set to 0 on transmission and ignored on receipt.

MT ID is a 12-bit field containing the non-zero ID of the topology being announced. The TLV MUST be ignored if the ID is zero. This is to ensure the consistent view of the standard unicast topology.

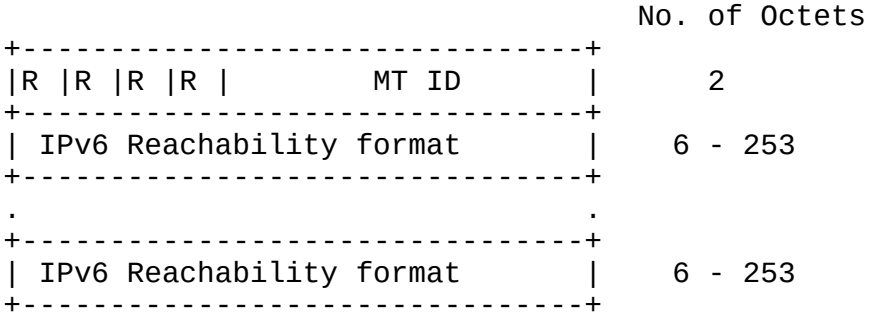
After the 2-byte MT membership format, the MT IPv4 content is in the same format as extended IP reachability TLV, type 135 [RFC3784].

This TLV can occur multiple times.

7.4. Multi-Topology Reachable IPv6 Prefixes TLV

The TLV number of this TLV is 237. It is aligned with IPv6 Reachability TLV type 236 beside an additional two bytes in front.

- x CODE - 237
- x LENGTH - total length of the value field
- x VALUE - 2-byte MT membership plus the format of IPv6 Reachability TLV, structured as follows:



Bits R are reserved, SHOULD be set to 0 on transmission and ignored on receipt.

MT ID is a 12-bit field containing the ID of the topology being announced. The TLV MUST be ignored if the ID is zero.

After the 2-byte MT membership format, the MT IPv6 context is in the same format as IPv6 Reachability TLV, type 236 [H01].

This TLV can occur multiple times.

7.5. Reserved MT ID Values

Certain MT topologies are assigned to serve predetermined purposes:

- MT ID #0: Equivalent to the "standard" topology.
- MT ID #1: Reserved for IPv4 in-band management purposes.
- MT ID #2: Reserved for IPv6 routing topology.
- MT ID #3: Reserved for IPv4 multicast routing topology.
- MT ID #4: Reserved for IPv6 multicast routing topology.
- MT ID #5: Reserved for IPv6 in-band management purposes.
- MT ID #6-#3995: Reserved for IETF consensus.
- MT ID #3996-#4095: Reserved for development, experimental and proprietary features [RFC3692].

8. MT IP Forwarding Considerations

Using MT extension for IS-IS routing can result in multiple RIBs on the system. In this section, we list some of the known considerations for IP forwarding in various MT scenarios. Certain deployment scenarios presented here imply different trade-offs in terms of deployment difficulties and advantages obtained.

8.1. Each MT Belongs to a Distinct Address Family

In this case, each MT related route is installed into a separate RIB. Multiple topologies can share the same IS-IS interface on detecting the incoming packet address family. As an example, IPv4 and IPv6 can share the same interface without any further considerations under MT ISIS.

8.2. Some MTs Belong to the Same Address Family

8.2.1. Each Interface Belongs to One and Only One MT

In this case, MTs can be used to forward packets from the same address family, even with overlapping addresses, since the MTs have their dedicated interfaces, and those interfaces can be associated with certain MT RIBs and FIBs.

8.2.2. Multiple MTs Share an Interface with Overlapping Addresses

Some additional mechanism is needed to select the correct RIBs for the incoming IP packets to determine the correct RIB to make a forwarding decision. For example, if the topologies are Quality of Service (QoS) partitioned, then the Differentiated Services Code Point (DSCP) bits in the IP packet header can be utilized to make the decision. Some IP headers, or even packet data information, MAY be checked to make the forwarding table selection, for example, the source IP address in the header can be used to determine the desired forwarding behavior.

This topic is not unique to IS-IS or even to Multi-topology, it is a local policy and configuration decision to make sure the inbound traffic uses the correct forwarding tables. For example, preferred customer packets are sent through a Layer 2 Tunneling Protocol (L2TP) towards the high-bandwidth upstream provider, and other packets are sent through a different L2TP to a normal-bandwidth provider. Those mechanisms are not part of the L2TP protocol specifications.

The generic approach of packet to multiple MT RIB mapping over the same inbound interface is outside the scope of this document.

8.2.3. Multiple MTs Share an Interface with Non-Overlapping Addresses

When there is no overlap in the address space among all the MTs, strictly speaking, the destination address space classifies the topology to which a packet belongs. It is possible to install routes from different MTs into a shared RIB. As an example of such a deployment, a special IS-IS topology can be set up for certain External Border Gateway Protocol (eBGP) nexthop addresses.

8.3. Some MTs Are Not Used for Forwarding Purposes

MT in IS-IS MAY be used even if the resulting RIB is not used for forwarding purposes. As an example, multicast Reverse Path Forwarding (RPF) check can be performed on a different RIB than the standard unicast RIB, albeit an entirely different RIB is used for the multicast forwarding. However, an incoming packet MUST still be clearly identified as belonging to a unique topology.

9. MT Network Management Considerations

When multiple IS-IS topologies exist within a domain, some of the routers can be configured to participate in a subset of the MTs in the network. This section discusses some of the options we have to enable operations or the network management stations to access those routers.

9.1. Create Dedicated Management Topology to Include All the Nodes

This approach is to set up a dedicated management topology or 'in-band' management topology. This 'mgmt' topology will include all the routers need to be managed. The computed routes in the topology will be installed into the 'mgmt' RIB. In the condition that the 'mgmt' topology uses a set of non-overlapping address space with the default topology, those 'mgmt' routes can also be optionally installed into the default RIB. The advantages of duplicate 'mgmt' routes in both RIBs include: the network management utilities on the system does not have to be modified to use a specific RIB other than the default RIB; the 'mgmt' topology can share the same link with the default topology if so designed.

9.2. Extend the Default Topology to All the Nodes

Even in the case that default topology is not used on some of the nodes in the IP forwarding, we MAY want to extend the default topology to those nodes for the purpose of network management. Operators SHOULD set high costs on the links that belong to the extended portion of the default topology. This way, the IP data traffic will not be forwarded through those nodes during network topology changes.

10. Acknowledgments

The authors would like to thank Andrew Partan, Dino Farinacci, Derek Yeung, Alex Zinin, Stefano Previdi, Heidi Ou, Steve Luong, Pekka Savola, Mike Shand, Shankar Vemulapalli, and Les Ginsberg for the discussion, their review, comments, and contributions to this document.

11. Security Considerations

IS-IS security applies to the work presented. No specific security issues with the proposed solutions are known. The authentication procedure for IS-IS PDUs is the same regardless of MT information inside the IS-IS PDUs.

Note that an authentication mechanism, such as the one defined in [RFC3567], SHOULD be applied if there is high risk resulting from modification of multi-topology information.

As described in Section 8.2.2, multiple topologies share an interface in the same address space, some mechanism beyond IS-IS needs to be used to select the right forwarding table for an inbound packet. A misconfiguration on the system or a packet with a spoofed source address, for example, can lead to packet loss or unauthorized use of premium network resource.

12. IANA Considerations

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

Name	Value
MT-ISN	222
M-Topologies	229
MT IP. Reach	235
MT IPv6 IP. Reach	237

IANA has created a new registry, "IS-IS Multi-Topology Parameters", with the assignments listed in Section 7.5 of this document and registration policies [RFC2434] for future assignments. The MT ID values range 6-3995 are allocated through Expert Review; values in the range of 3996-4095 are reserved for Private Use. In all cases, assigned values are to be registered with IANA.

13. References

13.1. Normative References

[ISO10589] ISO. Intermediate System to Intermediate System Routing Exchange Protocol for Use in Conjunction with the Protocol for Providing the Connectionless-Mode Network Service. ISO 10589, 1992.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3692] Narten, T., "Assigning Experimental and Testing Numbers Considered Useful", BCP 82, RFC 3692, January 2004.

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

13.2. Informative References

[RFC3567] Li, T. and R. Atkinson, "Intermediate System to Intermediate System (IS-IS) Cryptographic Authentication", RFC 3567, July 2003.

[RFC3784] Smit, H. and T. Li, "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", RFC 3784, June 2004.

[H01] C. Hopps, "Routing IPv6 with IS-IS", Work in Progress.

Authors' Addresses

Tony Przygienda
Z2 Sagl
Via Rovello 32
CH-6942 Savosa
EMail: prz@net4u.ch

Naiming Shen
Cisco Systems
225 West Tasman Drive
San Jose, CA, 95134 USA
EMail: naiming@cisco.com

Nischal Sheth
Juniper Networks
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
EMail: nsheth@juniper.net

Full Copyright Statement

Intellectual Property