Workgroup: Network Working Group
Internet-Draft:
draft-ietf-lsr-isis-flood-reflection-05
Published: 26 November 2021
Intended Status: Experimental
Expires: 30 May 2022
Authors: A. Przygienda, Ed. (Juniper), C. Bowers (Juniper), Y. Lee (Comcast), A. Sharma (Comcast), R. White (Juniper)

IS-IS Flood Reflection

Abstract

This document describes a backwards compatible, optional IS-IS extension that allows the creation of IS-IS flood reflection topologies. Flood reflection allows topologies in which L1 areas provide transit forwarding for L2 using all available L1 nodes internally. It accomplishes this by creating L2 flood reflection adjacencies within each L1 area. Those adjacencies are used to flood L2 LSPDUs, and they are used in the L2 SPF computation. However, they are not used for forwarding within the flood reflection cluster. This arrangement gives the L2 topology significantly better scaling properties. As additional benefit, only those routers directly participating in flood reflection have to support the feature. This allows for the incremental deployment of scalable L1 transit areas in an existing network, without the necessity of upgrading other routers in the network.

Requirements Language

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].

Status of This Memo

This Internet-Draft will expire on 30 May 2022.

Copyright Notice

Table of Contents

1. Introduction

This section introduces the problem space and outlines the solution. Some of the terms may be unfamiliar to reader without extensive IS-IS background and in such case a glossary is provided in Section 2 and can be referenced.

Due to the inherent properties of link-state protocols the number of IS-IS routers within a flooding domain is limited by processing and flooding overhead on each node. While that number can be maximized by well written implementations and techniques such as exponential back-offs, IS-IS will still reach a saturation point where no further routers can be added to a single flooding domain. In some L2 backbone deployment scenarios, this limit presents a significant challenge.

The traditional approach to increasing the scale of an IS-IS deployement is to break it up into multiple L1 flooding domains and a single L2 backbone. This works well for designs where an L2 backbone connects L1 access topologies, but it is limiting where a large L2 is supposed to span large number of routers. In such scenarios, an alternative approach is to consider multiple L2 flooding domains connected together via L1 flooding domains. In other words, L2 flooding domains are connected by "L1/L2 lanes" through the L1 areas to form a single L2 backbone again. Unfortunately, in its simplest implementation, this requires the inclusion of most, or all, of the transit L1 routers as L1/L2 to allow traffic to flow along optimal paths through such transit areas. Consequently, this approach fails to reduce the number of L2 routers involved, so it fails to increase the scalability of the L2 backbone.

```
+----+  +-------+            +-------+               +-------+  +----+
| R1 |  |  R10  +------------+  R20  +---------------+  R30  |  | R4 |
| L2 +--+ L1/L2 |            |  L1   |               | L1/L2 +--+ L2 |
|    |  |       +--------+ +-+       |  +------------+       |  |    |
+----+  ++-+--+-+        | | +---+---+----------+    +-+--+-++  +----+
         | |  |          | | |   |   |  |       |      |  | |
         | |  |          | | |   |   |  |  +-----------+  | |
         | |  +-------+  | | |   |   |  |  |    |         | |
         | |          |  | | |   |   |  |  |    |  +------+ |
         | +------+ +--------+   |   +-------+  |  |        |
         |        | | |  | |     |      |  | |  |  |        |
+----+  ++------+---+ |  +---+---+---+--+  | +-------+------++  +----+
| R2 |  |  R11  | |   |    | |  R21  |     |    | | |  R31  |  | R5 |
| L2 +--+ L1/L2 +------------+  L1   +---------------+ L1/L2 +--+ L2 |
|    |  |       | |   |    | |       |     |    |  | |       |  |    |
+----+  ++------+---+ |    | +---+--++     | +-------+------++  +----+
         |        | | |    | |   |  |      | |  |  |        |
         | +---------------+ |   |  |      | |  |  |        |
         | |      | | |      |   |  |      | |  |  |        |
         | |  +--------------+   |  +-----------------+     |
         | |  |   | | |          |         | |  |  |  |     |
+----+  ++-+--+-+ | | +------+---+---+-----+ |  |  | ++-----++  +----+
| R3 |  |  R12  | +----------|  R22  |       |  +----+  R32  |  | R6 |
| L2 +--+ L1/L2 |   +--------|  L1   +-------+     | | L1/L2 +--+ L2 |
|    |  |       +------------+       |---------------+       |  |    |
+----+  +-------+            +-------+-------------+ +-------+  +----+
```

Figure 1: Example Topology of L1 with L2 Borders

Figure 1 is an example of a network where a topologically rich L1 area is used to provide transit between six different L2-only routers (R1-R6). Note that the six L2-only routers do not have connectivity to one another over L2 links. To take advantage of the abundance of paths in the L1 transit area, all the intermediate systems could be placed into both L1 and L2, but this essentially combines the separate L2 flooding domains into a single one, triggering again maximum L2 scale limitation we try to address in first place.

A more effective solution would allow to reduce the number of links and routers exposed in L2, while still utilizing the full L1 topology when forwarding through the network.

[RFC8099] describes Topology Transparent Zones (TTZ) for OSPF. The TTZ mechanism represents a group of OSPF routers as a full mesh of adjacencies between the routers at the edge of the group. A similar mechanism could be applied to IS-IS as well. However, a full mesh of adjacencies between edge routers (or L1/L2 nodes) significantly limits the scale of the topology. The topology in Figure 1 has 6 L1/L2 nodes. Figure 2 illustrates a full mesh of L2 adjacencies between the 6 L1/L2 nodes, resulting in (5 * 6)/2 = 15 L2 adjacencies. In a somewhat larger topology containing 20 L1/L2 nodes, the number of L2 adjacencies in a full mesh rises to 190.

```
+----+  +-------+    +-------------------------------+-------+  +----+
| R1 |  |  R10  |    |                               |  R30  |  | R4 |
| L2 +--+ L1/L2 +------------------------------------+ L1/L2 +--+ L2 |
|    |  |       |    |                               |       |  |    |
+----+  ++-+-+--+-+  |                             +-+--+---++  +----+
         | | |    |  |                             |    |   |
         | +----------------------------------------------+ |
         |   |    |  |                             |    | | |
         |   +-----------------------------------+ |    | | |
         |        |  |                           | |    | | |
         |     +----------------------------------------+ | |
         |     |  |  |                           | |      | |
+----+  ++-----+- |  |                           | | -----+-++  +----+
| R2 |  |  R11  | |  |                           | | |  R31  |  | R5 |
| L2 +--+ L1/L2 +------------------------------------+ L1/L2 +--+ L2 |
|    |  |       | |  |                           | | |       |  |    |
+----+  ++------+------------------------------+ | | +----+-++  +----+
         |        |  |                         | | |      | |
         |        |  |                         | | |      | |
         |    +-------------------------------------------+ |
         |    |   |  |                         | | |        |
         |    |   |  |                         +----------+ |
         |    |   |  |                           | |      | |
         |    |   |  |                           +-----+  | |
         |    |   |  |                             |   |  | |
+----+  ++----+-+-+  |                             +-+-+--+-++  +----+
| R3 |  |  R12  |    |      L2 adjacency             |  R32  |  | R6 |
| L2 +--+ L1/L2 +------------------------------------+ L1/L2 +--+ L2 |
|    |  |       |    |                               |       |  |    |
+----+  +-------+----+                               +-------+  +----+
```

Figure 2: Example topology represented in L2 with a full mesh of L2 adjacencies between L1/L2 nodes

BGP, as specified in [RFC4271], faced a similar scaling problem, which has been solved in many networks by deploying BGP route reflectors [RFC4456]. We note that BGP route reflectors do not necessarily have to be in the forwarding path of the traffic. This incongruity of forwarding and control path for BGP route reflectors allows the control plane to scale independently of the forwarding plane.

We propose here a similar solution for IS-IS. A simple example of what a flood reflector control plane approach would look like is shown in Figure 3, where router R21 plays the role of a flood reflector. Each L1/L2 ingress/egress router builds a tunnel to the flood reflector, and an L2 adjacency is built over each tunnel. In this solution, we need only 6 L2 adjacencies, instead of the 15 needed for a full mesh. In a somewhat larger topology containing 20 L1/L2 nodes, this solution requires only 20 L2 adjacencies, instead of the 190 need for a full mesh. Multiple flood reflectors can be used, allowing the network operator to balance between resilience, path utilization, and state in the control plane. The resulting L2 adjacency scale is R*n, where R is the number of flood reflectors used and n is the number of L1/L2 nodes. This compares quite favorably with n*(n-1)/2 L2 adjacencies required in a fully meshed L2 solution.

```
+----+  +-------+                                    +-------+  +----+
| R1 |  |  R10  |                                    |  R30  |  | R4 |
| L2 +--+ L1/L2 +--------------+   +-----------------+ L1/L2 +--+ L2 |
|    |  |       |  L2 adj      |   |      L2 adj     |       |  |    |
+----+  +-------+  over        |   |      over       +-------+  +----+
                   tunnel      |   |      tunnel
+----+  +-------+           +--+---+--+              +-------+  +----+
| R2 |  |  R11  |           |   R21   |              |  R31  |  | R5 |
| L2 +--+ L1/L2 +-----------+  L1/L2  +--------------+ L1/L2 +--+ L2 |
|    |  |       |  L2 adj   |  flood  |   L2 adj     |       |  |    |
+----+  +-------+  over     |reflector|   over       +-------+  +----+
                   tunnel   +--+---+--+   tunnel
+----+  +-------+              |   |                 +-------+  +----+
| R3 |  |  R12  +--------------+   +-----------------+  R32  |  | R6 |
| L2 +--+ L1/L2 |  L2 adj                 L2 adj     | L1/L2 +--+ L2 |
|    |  |       |  over                   over       |       |  |    |
+----+  +-------+  tunnel                 tunnel     +-------+  +----+
```

Figure 3: Example topology represented in L2 with L2 adjacencies from each L1/L2 node to a single flood reflector

As illustrated in Figure 3, when R21 plays the role of flood reflector, it provides L2 connectivity among all of the previously disconnected L2 islands by reflooding all L2 LSPDUs. At the same time, R20 and R22 in Figure 1 remain L1-only routers. L1-only routers and L1-only links are not visible in L2. In this manner, the flood reflector allows us provide L2 control plane connectivity in a scalable manner.

As described so far, the solution illustrated in Figure 3 relies only on currently standardized IS-IS functionality. Without new functionality, however, the data traffic will traverse only R21. This will unnecessarily create a bottleneck at R21 since there is still available capacity in the paths crossing the L1-only routers R20 and R22 in Figure 1.

Hence, some new functionality is necessary to allow the L1/L2 edge nodes (R10-12 and R30-32 in Figure 3) to recognize that the L2 adjacency to R21 should not be used for forwarding. The L1/L2 edge nodes should forward traffic that would normally be forwarded over the L2 adjacency to R21 over L1 links instead. This would allow the forwarding within the L1 area to use the L1-only nodes and links shown in Figure 1 as well. It allows networks to be built that use the entire forwarding capacity of the L1 areas, while at the same time introducing control plane scaling benefits provided by L2 flood reflectors.

This document defines all extensions necessary to support flood reflector deployment:

* A 'flood reflector adjacency' for all the adjacencies built for the purpose of reflecting flooding information. This allows these 'flood reflectors' to participate in the IS-IS control plane without being used in the forwarding plane. This is a purely local operation on the L1/L2 ingress; it does not require replacing or modifying any routers not involved in the reflection process. Deployment-wise, it is far less tricky to just upgrade the routers involved in flood reflection rather than have a flag day on the whole IS-IS domain.

* An (optional) full mesh of tunnels between the L1/L2 routers, ideally load-balancing across all available L1 links. This harnesses all forwarding paths between the L1/L2 edge nodes without injecting unneeded state into the L2 flooding domain or creating 'choke points' at the 'flood reflectors' themselves. The draft is agnostic as to the tunneling technology used but provides enough information for automatic establishment of such tunnels. The discussion of IS-IS adjacency formation and/or liveness discovery on such tunnels is outside the scope of this draft and is largely choice of the underlying implementation. A solution without tunnels is also possible by applying judicious scoping of reachability information between the levels as described in more details later.

* Some way to support reflector redundancy, and potentially some way to auto-discover and advertise such adjacencies as flood reflector adjacencies. Such advertisements may allow L2 nodes outside the L1 to perform optimizations in the future based on this information.

2. Glossary

This section is introduced with the intention of allowing quick reference in the more detailed parts of the document to terms used

Flood Reflector:
Node configured to connect L2 only to flood reflector clients and reflect (reflood) IS-IS L2 LSPs amongst them.

Flood Reflector Client:
Node configured to build flood reflector adjacencies and normal L2 nodes.

Flood Reflector Adjacency:
IS-IS L2 adjacency limited by one end being client and the other reflector and agreeing on the same Flood Reflector Cluster ID.

Flood Reflector Cluster:
Collection of clients and flood reflectors configured with the same cluster identifier. Cluster ID value of 0 SHOULD NOT be used since it may be used in the future for special purposes.

Tunnel Deployment:
Deployment where flood reflector clients build a full mesh of tunnels in L1 to "shortcut" forwarding of L2 traffic through the cluster.

No Tunnel Deployment:
Deployment where flood reflector clients redistribute L2 reachability into L1 to allow forwarding through the cluster without underlying tunnels.

3. Further Details

Several considerations should be noted in relation to such a flood reflection mechanism.

First, this allows multi-area IS-IS deployments to scale without any major modifications in the IS-IS implementation on most of the nodes deployed in the network. Unmodified (traditional) L2 routers will compute reachability across the transit L1 area using the flood reflector adjacencies.

Second, the flood reflectors are not required to participate in forwarding traffic through the L1 transit area. These flood reflectors can be hosted on virtual devices outside the forwarding topology.

Third, astute readers will realize that flooding reflection may cause the use of suboptimal paths. This is similar to the BGP route reflection suboptimal routing problem described in [ID.draft-ietf-idr-bgp-optimal-route-reflection-28]. The L2 computation determines the egress L1/L2 and with that can create illusions of ECMP where there is none. And in certain scenarios lead to an L1/L2 egress which is not globally optimal. This represents a straightforward instance of the trade-off between the amount of control plane state and the optimal use of paths through the network often encountered when aggregating routing information.

One possible solution to this problem is to expose additional topology information into the L2 flooding domains. In the example network given, links from router 01 to router 02 can be exposed into L2 even when 01 and 02 are participating in flood reflection. This information would allow the L2 nodes to build 'shortcuts' when the L2 flood reflected part of the topology looks more expensive to cross distance wise.

Another possible variation is for an implementation to approximate with the tunnel cost the cost of the underlying topology.

Redundancy can be achieved by building multiple flood reflectors in a L1 area. Multiple flood reflectors do not need any synchronization mechanisms amongst themselves, except standard IS-IS flooding and database maintenance procedures.

4. Encodings

4.1. Flood Reflection TLV

The Flood Reflection TLV is a new top-level TLV that MAY appear in L2 IIHs. The Flood Reflection TLV indicates the flood reflector cluster (based on Flood Reflection Cluster ID) that a given router is configured to participate in. It also indicates whether the router is configured to play the role of either flood reflector or flood reflector client. The Flood Reflection Cluster ID and flood reflector roles advertised in the IIHs are used to ensure that flood reflector adjacencies are only formed between a flood reflector and flood reflector client, and that the Flood Reflection Cluster IDs match. The Flood Reflection TLV has the following format:

```
 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     |C|  RESERVED   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Flood Reflection Cluster ID                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Sub-TLVs ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD

Length: The length, in octets, of the following fields.

C (Client): This bit is set to indicate that the router acts as a flood reflector client. When this bit is NOT set, the router acts as a flood reflector. On a given router, the same value of the C-bit MUST be advertised across all interfaces advertising the Flood Reflection TLV in IIHs.

RESERVED: This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Flood Reflection Cluster ID: Flood Reflection Cluster Identifier. These same 32-bit value MUST be assigned to all of the flood reflectors and flood reflector clients in the same L1 area. The value MUST be unique across different L1 areas within the IGP domain. In case of violation of those rules multiple L1 areas may become a single cluster or a single area may split in flood reflection sense and several mechanisms such as auto-discovery of tunnels may not work correctly. On a given router, the same value of the Flood Reflection Cluster ID MUST be advertised across all interfaces advertising the Flood Reflection TLV in IIHs. When a router discovers that a node is using multiple Cluster IDs based on its advertised TLVs and IIHs, the node MAY adequately log such violations subject to rate limiting. This implies that a flood reflector MUST NOT participate in more than a single L1 area. In case of Cluster ID value of 0, the TLV containing it MUST be ignored.

Sub-TLVs: Optional sub-TLVs. For future extensibility, the format of the Flood Reflection TLV allows for the possibility of including optional sub-TLVs. No sub-TLVs of the Flood Reflection TLV are defined in this document.

The Flood Reflection TLV SHOULD NOT appear more than once in an IIH. A router receiving multiple Flood Reflection TLVs in the same IIH

MUST use the values in the first TLV and it SHOULD adequately log such violations subject to rate limiting.

4.2. Flood Reflection Discovery Sub-TLV

Flood Reflection Discovery sub-TLV is advertised as a sub-TLV of the IS-IS Router Capability TLV-242, defined in [RFC7981]. The Flood Reflection Discovery sub-TLV is advertised in L1 and L2 LSPs with area flooding scope in order to enable the auto-discovery of flood reflection capabilities. The Flood Reflection Discovery sub-TLV has the following format:

```
 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     |C|  Reserved   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Flood Reflection Cluster ID                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD

Length: The length, in octets, of the following fields.

C (Client): This bit is set to indicate that the router acts as a flood reflector client. When this bit is NOT set, the router acts as a flood reflector.

RESERVED: This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Flood Reflection Cluster ID: The Flood Reflection Cluster Identifier is the same as that defined in the Flood Reflection TLV and obeys the same rules.

The Flood Reflection Discovery sub-TLV SHOULD NOT appear more than once in TLV 242. A router receiving multiple Flood Reflection Discovery sub-TLVs in TLV 242 MUST use the values in the first sub-TLV and it SHOULD adequately log such violations subject to rate limiting.

4.3. Flood Reflection Discovery Tunnel Type Sub-Sub-TLV

Flood Reflection Discovery Tunnel Type sub-sub-TLV is advertised optionally as a sub-sub-TLV of the Flood Reflection Discovery Sub-TLV, defined in Section 4.2. It allows the automatic creation of L2 tunnels to be used as flood reflector adjacencies and L1 shortcut tunnels. The Flood Reflection Tunnel Type sub-sub-TLV has the following format:

```
 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    | Reserved    |F|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Tunnel Encapsulation Attribute                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD

Length: The length, in octets, of zero or more of the following fields.

Reserved: SHOULD be 0 on transmission and ignored on reception.

F Flag: When set indicates flood reflection tunnel endpoint, when clear, indicates possible L1 shortcut tunnel endpoint.

Tunnel Encapsulation Attribute: Carries encapsulation type and further attributes necessary for tunnel establishment as defined in [RFC9012]. Protocol type sub-TLV as defined in [RFC9012] MAY be included but MUST when F flag is set include according type that allows carrying of encapsulated IS-IS frames. Such tunnel type MUST provide according mechanisms to carry up to `originatingL2LSPBufferSize` sized IS-IS frames across.

A flood reflector receiving multiple Flood Reflection Discovery Tunnel Type sub-sub-TLVs in Flood Reflection Discovery sub-TLV with F flag set SHOULD use one or more of the specified tunnel endpoints to automatically establish one or more tunnels that will serve as flood reflection adjacency(-ies).

A flood reflection client receiving multiple Flood Reflection Discovery Tunnel Type sub-sub-TLVs in Flood Reflection Discovery sub-TLV with F flag clear from other leaves MAY use one or more of the specified tunnel endpoints to automatically establish one or more tunnels that will serve as L1 tunnel shortcuts.

Optional address validation procedures as defined in [RFC9012] MUST be disregarded.

4.4. Flood Reflection Adjacency Sub-TLV

The Flood Reflection Adjacency sub-TLV is advertised as a sub-TLV of TLVs 22, 23, 25, 141, 222, and 223. Its presence indicates that a given adjacency is a flood reflector adjacency. It is included in L2 area scope flooded LSPs. Flood Reflection Adjacency sub-TLV has the following format:

```
 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     |C|  Reserved   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Flood Reflection Cluster ID                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD

Length: The length, in octets, of the following fields.

C (Client): This bit is set to indicate that the router advertising this adjacency is a flood reflector client. When this bit is NOT set, the router advertising this adjacency is a flood reflector.

RESERVED: This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Flood Reflection Cluster ID: The Flood Reflection Cluster Identifier is the same as that defined in the Flood Reflection TLV and obeys the same rules.

The Flood Reflection Adjacency sub-TLV SHOULD NOT appear more than once in a given TLV. A router receiving multiple Flood Reflection Adjacency sub-TLVs in a TLV MUST use the values in the first sub-TLV and it SHOULD adequately log such violations subject to rate limiting.

4.5. Flood Reflection Discovery

A router participating in flood reflection as client or reflector MUST be configured as an L1/L2 router. It SHOULD originate the Flood Reflection Discovery sub-TLV with area flooding scope in L1 and L2. Normally, all routers on the edge of the L1 area (those having traditional L2 adjacencies) will advertise themselves as route reflector clients. Therefore, a flood reflector client will have both traditional L2 adjacencies and flood reflector L2 adjacencies.

A router acting as a flood reflector MUST NOT have any traditional L2 adjacencies. It will be an L1/L2 router only by virtue of having flood reflector L2 adjacencies. A router desiring to act as a flood reflector SHOULD advertise itself as such using the Flood Reflection Discovery sub-TLV in L1 and L2.

A given flood reflector or flood reflector client can only participate in a single cluster, as determined by the value of its Flood Reflection Cluster ID and should disregard other routers' TLVs for flood reflection purposes if the cluster ID is not matching.

Upon reception of Flood Reflection Discovery sub-TLVs, a router acting as flood reflector client SHOULD initiate a tunnel towards each flood reflector with which it shares an Flood Reflection Cluster ID using one or more of the tunnel encapsulations provided with F flag being set. The L2 adjacencies formed over such tunnels MUST be marked as flood reflector adjacencies. If the client or reflector has a direct L2 adjacency with the according remote side it SHOULD use it instead of instantiating a new tunnel.

In absence of auto-discovery an implementation MAY use statically configured tunnels to create flood reflection adjacencies.

The IS-IS metrics for all flood reflection adjacencies in a cluster SHOULD be uniform.

Upon reception of Flood Reflection Discover TLVs, a router acting as a flood reflector client MAY initiate tunnels with L1-only adjacencies towards any of the other flood reflector clients with lower router IDs in its cluster using encapsulations with F flag clear. These tunnels MAY be used for forwarding to improve the load-balancing characteristics of the L1 area. If the clients have a direct L2 adjacency they SHOULD use it instead of instantiating a new tunnel.

4.6. Flood Reflection Adjacency Formation

In order to simplify both implementations and network deployments, this draft does not allow the formation of complex hierarchies of flood reflectors and clients or allow multiple clusters in a single L1 area. Consequently, all flood reflectors and flood reflector clients in the same L1 area MUST share the same Flood Reflector Cluster ID. Deployment of multiple cluster IDs in the same L1 area are outside the scope of this document.

A flood reflector MUST only form flood reflection adjacencies with flood reflector clients with matching Cluster ID. A flood reflector MUST NOT form any traditional L2 adjacencies.

Flood reflector clients MUST only form flood reflection adjacencies with flood reflectors with matching Cluster ID.

Flood reflector clients MAY form traditional L2 adjacencies with flood reflector clients or nodes not participating in flood reflection. When two clients form traditional L2 adjacency Cluster ID is disregarded.

The Flood Reflector Cluster ID and flood reflector roles advertised in the Flood Reflection TLVs in IIHs are used to ensure that flood reflection adjacencies that are established meet the above criteria.

On change in either flood reflection role or cluster ID on IIH on the local or remote side the adjacency has to be reset and re-established if possible.

Once a flood reflection adjacency is established, the flood reflector and the flood reflector client MUST advertise the adjacency by including the Flood Reflection Adjacency Sub-TLV in the Extended IS reachability TLV or MT-ISN TLV.

5. Route Computation

To ensure loop-free routing, the route reflection client MUST follow the normal L2 computation to determine L2 routes. This is because nodes outside the L1 area will generally not be aware that flood reflection is being performed. The flood reflection clients need to produce the same result for the L2 route computation as a router not participating in flood reflection.

5.1. Tunnel Based Deployment

In tunnel based option the reflection client, after L2 and L1 computation, MUST examine all L2 routes and replace all flood reflector adjacencies with the correct underlying tunnel next-hop to the egress.

5.2. No Tunnel Deployment

In case of deployment without underlying tunnels, the necessary L2 routes are distributed into the area, normally as L2->L1 routes. Due to the rules in Section 4.6 the computation in the resulting topology is relatively simple, the L2 SPF from a flood reflector client is guaranteed to reach within a hop the Flood Reflector and in the following hop the L2 egress to which it has a forwarding tunnel again. All the flood reflector tunnel nexthops in the according L2 route can hence be removed and if the L2 route has no other ECMP L2 nexthops, the L2 route MUST be suppressed in the RIB by some means to allow the less preferred L2->L1 route to be used to forward traffic towards the advertising egress.

In the particular case the client has L2 routes which are not route reflected, those will be naturally preferred (such routes normally "hot-potato" route of the L1 area). However in the case the L2 route through the flood reflector egress is "shorter" than such present non flood reflected L2 routes, the node SHOULD ensure that such routes are suppressed so the L2->L1 towards the egress still takes preference. Observe that operationally this can be resolved in a relatively simple way by configuring flood reflector adjacencies to have a high metric, i.e. the flood reflector topology becomes "last resort" and the leaves will try to "hot-potato" out the area as fast as possible which is normally the desirable behavior.

In deployment scenarios where tunnels are not used, all L1/L2 edge nodes MUST be ultimately flood reflector clients except during during transition phase.

6. Redistribution of Prefixes

When L2 prefixes need to be redistributed into L1 by the route reflector clients a client that does not have any L2 flood reflector adjacencies MUST NOT redistribute those routes into the area in case of application of Section 5.2. The L2 prefixes advertisements redistributed into L1 with flood reflectors SHOULD be normally limited to L2 intra-area routes (as defined in [RFC7775]), if the information exists to distinguish them from other other L2 prefix advertisements.

On the other hand, in topologies that make use of flood reflection to hide the structure of L1 areas while still providing transit forwarding across them using tunnels, we generally do not need to redistribute L1 prefixes advertisements into L2.

7. Special Considerations

In pathological cases setting the overload bit in L1 (but not in L2) can partition L1 forwarding, while allowing L2 reachability through flood reflector adjacencies to exist. In such a case a node cannot replace a route through a flood reflector adjacency with a L1 shortcut and the client can use the L2 tunnel to the flood reflector for forwarding while it MUST initiate an alarm and declare misconfiguration.

A flood reflector with directly L2 attached prefixes should advertise those in L1 as well since based on preference of L1 routes the clients will not try to use the L2 flood reflector adjacency to route the packet towards them. A very, very corner case is when the flood reflector is reachable via L2 flood reflector adjacency (due to underlying L1 partition) only in which case the client can use the L2 tunnel to the flood reflector for forwarding towards those prefixes while it MUST initiate an alarm and declare misconfiguration.

A flood reflector SHOULD NOT set the attached bit on its LSPs.

Instead of modifying the computation procedures one could imagine a flood reflector solution where the Flood Reflector would re-advertise the L2 prefixes with a 'third-party' next-hop but that would have less desirable convergence properties than the solution proposed and force a fork-lift of all L2 routers to make sure they disregard such prefixes unless in the same L1 domain as the Flood Reflector.

Depending on pseudo-node choice in case of a broadcast domain with multiple flood reflectors attached this can lead to a partitioned LAN and hence a router discovering such a condition MUST initiate an alarm and declare misconfiguration.

8. IANA Considerations

This document requests allocation for the following IS-IS TLVs and Sub-TLVs.

8.1. New IS-IS TLV Codepoint

This document requests the following IS-IS TLV:

Value	Name	IIH	LSP	SNP	Purge
TBD1	Flood Reflection	y	n	n	n

Suggested value for TBD1 is 161.

8.2. Sub TLVs for TLV 242

This document request the following registration in the "sub-TLVs for TLV 242" registry.

Type	Description
TBD2	Flood Reflection Discovery

Suggested value for TBD2 is 161.

8.3. Sub-sub TLVs for Flood Reflection Discovery sub-TLV

This document request the following registration in the "sub-sub-TLVs for Flood Reflection Discovery sub-TLV" registry.

Type	Description
TBD3	Flood Reflection Discovery Tunnel Encapsulation Attribute

Suggested value for TBD3 is 161.

8.4. Sub TLVs for TLV 22, 23, 25, 141, 222, and 223

This document requests the following registration in the "sub-TLVs for TLV 22, 23, 25, 141, 222, and 223" registry.

Type	Description	22	23	25	141	222	223
TBD4	Flood Reflector Adjacency	y	y	n	y	y	y

Suggested value for TBD4 is 161.

9. Security Considerations

This document introduces tunnels carrying IS-IS control traffic via tunnels. In case of statically configured tunnels a deployment SHOULD provide enough security protection to prevent malicious attackers from using the tunnel endpoints. For information used to form dynamically discovered tunnels, it SHOULD be protected by the the deployed IS-IS security mechanism preventing malicious nodes from spoofing rogue information on behalf of other members.

10. Acknowledgements

The authors thank Shraddha Hegde, Peter Psenak, Acee Lindem and Les Ginsberg for their thorough review and detailed discussions. Thanks are also extended to Michael Richardson for an excellent routing directorate review.

11. References

11.1. Informative References

[ID.draft-ietf-idr-bgp-optimal-route-reflection-28]
Raszuk et al., R., "BGP Optimal Route Reflection", July 2019, <https://www.ietf.org/id/draft-ietf-idr-bgp-optimal-route-reflection-28.txt>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <https://www.rfc-editor.org/info/rfc4456>.

[RFC8099] Chen, H., Li, R., Retana, A., Yang, Y., and Z. Liu, "OSPF Topology-Transparent Zone", RFC 8099, DOI 10.17487/

RFC8099, February 2017, <https://www.rfc-editor.org/info/rfc8099>.

11.2. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC7775] Ginsberg, L., Litkowski, S., and S. Previdi, "IS-IS Route Preference for Extended IP and IPv6 Reachability", RFC 7775, DOI 10.17487/RFC7775, February 2016, <https://www.rfc-editor.org/info/rfc7775>.

[RFC7981] Ginsberg, L., Previdi, S., and M. Chen, "IS-IS Extensions for Advertising Router Information", RFC 7981, DOI 10.17487/RFC7981, October 2016, <https://www.rfc-editor.org/info/rfc7981>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

Authors' Addresses

Tony Przygienda (editor)
Juniper
1137 Innovation Way
Sunnyvale, CA
United States of America

Email: prz@juniper.net

Chris Bowers
Juniper
1137 Innovation Way
Sunnyvale, CA
United States of America

Email: cbowers@juniper.net

Yiu Lee
Comcast
1800 Bishops Gate Blvd
Mount Laurel, NJ 08054
United States of America

Email: Yiu_Lee@comcast.com

Alankar Sharma
Comcast
1800 Bishops Gate Blvd
Mount Laurel, NJ 08054
United States of America

Email: Alankar_Sharma@comcast.com

Russ White
Juniper
1137 Innovation Way
Sunnyvale, CA
United States of America

Email: russw@juniper.net