

TRILL Working Group  
Internet Draft  
Intended status: Standards Track

Howard Yang  
Cisco Systems  
Ayan Banerjee  
Cisco Systems  
Donald Eastlake  
Huawei  
Radia Perlman  
Intel Labs  
July 17, 2015

Expires: January 17, 2016

TRILL: Parent Selection in Distribution Trees  
<[draft-yang-trill-parent-seletion-05.txt](#)>

#### Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#). This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

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

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/ietf/1id-abstracts.txt>

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

This Internet-Draft will expire on January 17, 2016.

#### Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

Yang, et al

Expires January 17, 2016

[Page 1]

---

Internet-Draft

Trill: Parent Selection

July 2015

This document is subject to [BCP 78](#) and the IETF Trust's Legal

Provisions Relating to IETF Documents  
(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Abstract

This document describes a protocol extension in TRILL IS-IS and a parent selection tiebreak algorithm in the calculation of distribution trees in TRILL. The proposal is to modify the current algorithm to improve the stability of the distribution trees when multiple equal cost parents are present. It also offers the capabilities of pinning down multi-destination traffic and re-shaping the distribution trees to improve the traffic load balancing.

## Table of Contents

<a href="#">1. Introduction</a>	<a href="#">3</a>
<a href="#">1.1. Conventions used in this document</a>	<a href="#">3</a>
<a href="#">2. Problem Definition</a>	<a href="#">3</a>
<a href="#">3. Explicit Parent Selection Algorithm</a>	<a href="#">5</a>
<a href="#">3.1. Implicit Selection vs. Explicit Selection</a>	<a href="#">5</a>
<a href="#">3.1.1. Explicit Selection is a Preference</a>	<a href="#">6</a>
<a href="#">3.1.2. Explicit Selection is a Local Decision</a>	<a href="#">7</a>
<a href="#">3.1.3. Explicit Selection is honoews only if viable</a>	<a href="#">7</a>
<a href="#">3.2. Migration and Capability Sub-TLV</a>	<a href="#">7</a>
<a href="#">3.2.1. A Second Approach in the Mixed Environment</a>	<a href="#">6</a>
<a href="#">4. Sub-TLV Extensions to IS-IS</a>	<a href="#">9</a>
<a href="#">4.1. Parent Selection Algorithm Version Sub-TLV</a>	<a href="#">9</a>
<a href="#">4.2. Explicit Parent Preference Sub-TLV</a>	<a href="#">10</a>
<a href="#">5. Operation of RBridge</a>	<a href="#">10</a>
<a href="#">6. Other Applications</a>	<a href="#">11</a>
<a href="#">7. Security Considerations</a>	<a href="#">12</a>
<a href="#">8. IANA Considerations</a>	<a href="#">12</a>
<a href="#">9. Conclusions</a>	<a href="#">12</a>
<a href="#">10. References</a>	<a href="#">12</a>
<a href="#">10.1. Normative References</a>	<a href="#">12</a>

<a href="#">11. Acknowledgments</a>	<a href="#">13</a>
-------------------------------------	--------------------

## [1. Introduction](#)

The IETF has standardized the TRILL protocol [[RFC6325](#)], which provides transparent Layer 2 forwarding using encapsulation with a hop count and link state routing. TRILL provides optimal pair-wise forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic as well as supporting VLANs. In [Section 4.5.1 of \[RFC6325\]](#), a tiebreak algorithm is described to select a parent node when there are multiple equal cost parents. It uses the tree number as a parameter in the algorithm.

While the algorithm described in [[RFC6325](#)] is simple and elegant to achieve the goal of traffic load splitting, it exhibits an undesired behavior that a link status change in one tree causes the re-selection of paths in other trees, which impacts the stability of the distribution trees in the event of a link status flap.

This document presents a solution to the above problem with the introduction of protocol extension in IS-IS and a modification of the parent selection tiebreak algorithm.

### [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](#)].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying [RFC-2119](#) significance.

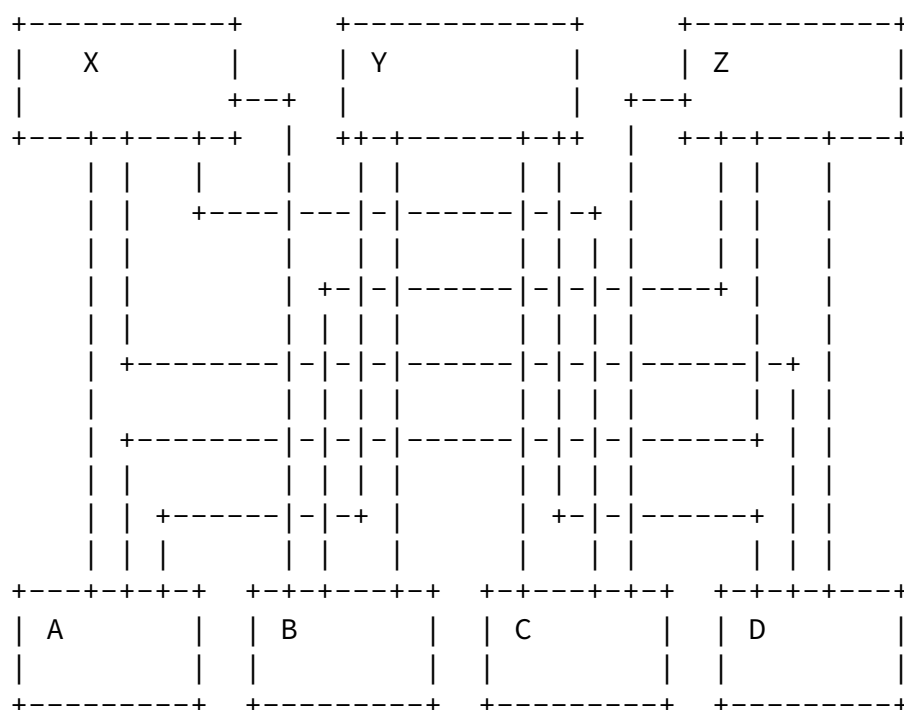
In this document, the characters ">>" preceding an indented line(s) indicates a compliance requirement statement using the key words listed above. This convention aids reviewers in quickly identifying or finding the explicit compliance requirements of this RFC.

## [2](#). Problem Definition

In the following example of TRILL network, there are two groups of RBridges or nodes: one group consists of RBridges X, Y and Z (Let's call it spine node group). The other group consists of RBridges A, B,

C, and D (Let's call it leaf node group). Each spine node has one link connecting to each leaf node. There are no links connecting any two spine nodes. There are no links between any two leaf nodes,

either. (In this document, `node` and `RBridge` are used interchangeably.)



Assume three multi-destination distribution trees are to be calculated, with Tree 1 rooted at X, Tree 2 at Y, and Tree 3 at Z. Assuming the 7-octet IS-IS IDs of A, B, C, D are in ascending order.

When a tree, for example, Tree 2 (rooted at Y), is calculated, and when node X is to be added to the tree, there are 4 equal cost possible parents: A, B, C, D. To pick one as the parent, a tiebreak algorithm is needed. [Section 4.5.1 of \[RFC6325\]](#) describes a parent selection tiebreak algorithm: For each node N, if N has p parents of equal cost, then order the parents in ascending order according to the 7-octet ISIS-ID and number them starting at zero. For Tree j, choose N's parent as choice (j-1) mod p. (See also [\[RFC7180\]](#) for the correction of the algorithm.) We call the algorithm "original tiebreak algorithm".

Applying this algorithm, on Tree 2, B is selected as the parent node for X, that is, link XB is on Tree 2. (In this document, link XB means the link between node X and node B.) Similarly, when Tree 3 is

calculated, C is selected as the parent for X. (On Tree 1, A is selected as the parent for Y and Z).

In the event when the link XA goes down, Tree 1 will select B as the parent node for Y and Z, as there are only 3 equal cost parents (B, C, and D) for Y or Z now. This change on Tree 1 is expected, as XA was on Tree 1.

But observe the changes on Tree 2 and 3: Tree 2 now changes to select C as the parent node for X, due to the failure of link XA. Similarly, Tree 3 changes to select D as the parent node for X.

	Parent of X on Tree 2	Parent of X on Tree 3
XA is up	B	C
XA is down	C	D

Although link XA is only on Tree 1, its link status change caused the re-selection of the parents on other trees. That is, the status change on one seemingly unrelated link causes the re-selection of the parent nodes and the changes of the tree paths on other trees.

This behavior affects the stability of the multi-destination distribution trees in TRILL.

We propose the following modifications to solve the problem.

In addition, network designers and administrators wish to "pin" the multi-destination traffic to the paths. They prefer to have the traffic follow trees with some deterministic in nature. The proposal here offers such a tool to "pin" down the traffic to the trees.

### [3. Explicit Parent Selection Algorithm](#)

#### [3.1. Implicit Selection vs. Explicit Selection](#)

In the current parent selection tiebreak algorithm, when there are  $p$  parent nodes, we list them in the ascending order and assign a number to them starting from 0. And Tree  $j$  selects the node  $(j-1) \bmod p$ , in the ordered list as its parent. We call this selection "implicit selection".

In the previous example, when the link XA is up, in the calculation of Tree 2, when node X is being added to the tree, B is chosen as the parent. This is implicit selection. In the proposed algorithm, this implicit selection is noted down on X. That is, on X, (Tree 2, B) is

saved in memory to be used later on. Similarly, on X, (Tree 3, C) is also noted down in the parent selection database.

When link XA goes down, the new implicit selections change for X on Tree 2 and 3. The new implicit selections are (Tree 2, C) and (Tree 3, D). Now we propose to change the algorithm to override this.

After the link XA goes down, X detects such link status change, and checks against its parent selection database. X then makes a conscious decision and advertises, in its LSP, the previously saved selections (Tree 2, B) and (Tree 3, C). Such advertised selections are called "explicit selections".

The essence of the proposed algorithm is: instead of relying on the well-known algorithm for everyone to pick C as the parent for X on Tree 2 and D on Tree 3, X advertises the explicit selection preference and suggests everybody that B should continue to be the parent for X on Tree 2, and C on Tree 3.

The way to advertise such explicit selections is to put them in the TLVs of X's LSP, which are propagated to all the nodes in the TRILL network.

The new tiebreak algorithm is then modified in such a way that when node X is to be considered to be added to a tree, the RBridge running the algorithm looks for explicit "parent selection preference" TLVs in X's LSP. If no such TLV is present, the current algorithm is used and the implicit selection is chosen. If there is an explicit selection in LSP, this parent selection preference advertisement must be honored, if it is a viable option, meaning it is among the list of currently available equal cost shortest path parent choices.

In our example, since X advertises (Tree 2, B), and B is a viable option, B is therefore chosen as the parent for X on Tree 2. Similarly, C is chosen as parent for X on Tree 3.

This achieves our goal of keeping the parent selections unchanged on Tree 2 and Tree 3 when link XA goes down.

#### [3.1.1.](#) Explicit Selection is a Preference

When a node advertises the explicit parent selection, such advertisement is a preference, which may or may not be honored. The explicit selection is honored only when it is among the equal cost parent choices.

In the previous example, if Links XA and YB both go down the same

time, when calculation Tree 2, and when X is added to the tree path, X's advertisement of (Tree 2, B) cannot be honored, as B is not viable. X advertised that it preferred to choose B as the parent on Tree 2, but such preference is not honored. In such case, the original tiebreak algorithm is applied and C is chosen as the parent.

It is worthwhile to point out that in the above case when both XA and YB go down, with the proposed algorithm when X advertises (Tree 2, B) and (Tree 3, C), and Y advertises (Tree 1, A) and (Tree 3, C), Tree 3 is not affected by the link status changes.

### [3.1.2](#). Explicit Selection is a Local Decision

It is up to an RBridge to decide whether or not to advertise any explicit selection preference. For example, in the above example, X can choose not to advertise any explicit selection even when link XA goes down, and allow all the nodes in the network to continue to select the implicit selections (which of course causes changes to Trees 2 and 3). X can even advertise a different node, for example, (Tree 2, D).

Whether or not to advertise or which node to advertise is a local decision. This characteristic makes it possible to have a smooth system software upgrade and migration in the TRILL network.

It is also the responsibility of the local node to decide when to stop any preference advertisement.

### [3.1.3](#). Explicit Selection is honored only if viable

Before the explicit selection advertisement is honored, a "viability check" must be performed. If the explicit advertisement is among the choices of equal cost shortest path parents, then it is viable. Only viable explicit advertisements are to be honored. This is to ensure that the multi-destination trees resulting from the new algorithm are still the SPF trees (short path trees).

## [3.2](#). Migration and Capability Sub-TLV

It is important that all the RBridges in the TRILL network runs the same SPF algorithm (including tiebreak algorithm) for the distribution tree calculation. Therefore, in the software upgrade and migration case, until all the RBridges are upgraded with the capability to run the new explicit selection tiebreak algorithm, any RBridge must continue to run the original tiebreak algorithm described in [Section 4.5.1 of \[RFC6325\]](#).

To achieve the migration goal, a new router-capability sub-TLV will be introduced, which will indicate whether or not an RBridge is capable to run the new algorithm. The absence of such sub-TLV in the LSP implies that the RBridge is not capable to run the new algorithm.

RBridge must inspect the LSPs of all the reachable nodes before the tree calculation. When an RBridge detects that there is at least one RBridge which does not advertise this capability sub-TLV, it should not advertise any explicit selections, and the implicit selections MUST be chosen, and any explicit selection advertisement is ignored.

In a mixed environment where the network consists of RBs running both existing algorithm and the proposed enhanced algorithm, the above approach prevents loops when multi-destination trees are calculated. Although, with this approach, until all the RBs are upgraded to run the new algorithm, none of the RBs can take advantage of the new algorithm.

### [3.2.1](#). A Second Approach in the mixed environment

A second proposal is to introduce a new flag in the highest priority RB's LSP. This flag signals to all the RBs in the network which algorithm to run, when highest priority RB's LSP is propagated.

Let's call the existing parent selection algorithm "V0", and let's call the new proposed parent selection algorithm "V1".

If the new flag in the highest priority RB indicates to run V0, then all the RBs continue to run the existing algorithm, and should ignore any explicit parent advertisements, if they exist.

If the new flag in the highest priority RB indicates to run V1, then all the RBs which understand this flag MUST run the new algorithm. Of course, all those RBs which have not been upgraded will not understand this flag and will continue to run the existing algorithm.

We also need to introduce a new TLV or sub-TLV in the ISIS hellos to indicate whether an RB supports the new enhanced algorithm. And an RB running V1 must not form ISIS adjacency with RBs running V0.

With this approach, RBs can be turned on to run the new algorithm as they are upgraded, and do not have to wait until all the RBs are upgraded. The drawbacks of this approach are that it requires an additional configuration (to set the flag on highest priority RB to signal which version to run), and the V1 and V0 RBs are partitioning the network (as they do not form adjacency with each other).

Summary of this approach: Each RBridge announces, in its Hellos (and possibly also in its LSP), whether it can use the new tree algorithm. The highest priority RBridge R1 announces, in its LSP, whether to run the new or old algorithm. R1 makes the decision based solely on configuration. If R1 announces using the new algorithm, and there are still old RBridges that do not support the new algorithm, then new RBridges will refuse to form an adjacency with them.

#### [4.](#) Sub-TLV Extensions to IS-IS

Two new sub-TLVs are introduced in below sub-sections. Both sub-TLVs can be carried in Router Capability TLV. The Router Capability TLV is defined in [\[RFC4971\]](#).

##### [4.1.](#) Parent Selection Algorithm Version Sub-TLV

The Parent Selection Algorithm Version sub-TLV indicates the maximum version of the algorithm supported by an RBridge. By implication, lower versions are also supported. If this sub-TLV is missing, the originating RBridge only supports the based version, which is the original algorithm described in [Section 4.5.1 of \[RFC6325\]](#).

This sub-TLV can also be used in ISIS hellos in port capability TLV if the second approach is used in the mixed environment ([Section 3.2.1](#)).

```

+---+---+---+---+---+
| Type                |                (1 byte)
+---+---+---+---+---+
| Length              |                (1 byte)
+---+---+---+---+---+
| Max-version         |                (1 byte)
+---+---+---+---+---+

```

- o Type: Router Capability sub-TLV type, set to PARENT-SELECT-VER.
- o Length: 1.
- o Max-version: Set to maximum version supported. 0: base version. 1: Explicit Parent Selection algorithm.

## 4.2. Explicit Parent Preference Sub-TLV

The Explicit Parent Preference Sub-TLV is a list of pairs of tree number and IS-IS ID. It includes the explicit parent selection advertisements which indicate which nodes the originating RBridge prefers in the equal cost multiple parent case for the trees.

```
+---+---+---+---+---+
| Type                | (1 byte)
+---+---+---+---+---+
| Length              | (1 byte)
+---+---+---+---+---+
| Tree Number a      | (2 bytes)
+---+---+---+---+---+
| IS-IS ID           | (7 bytes)
+---+---+---+---+---+
| Tree Number b      | (2 bytes)
+---+---+---+---+---+
| IS-IS ID           | (7 bytes)
+---+---+---+---+---+
| Tree Number (...)  | (2 bytes)
+---+---+---+---+---+
| IS-IS ID (...)     | (7 bytes)
+---+---+---+---+---+
```

- o Type: Router Capability sub-TLV type, set to PARENT-PREFERENCE.
- o Length:  $9 \times n$ , where  $n$  is the number of trees listed.
- o Tree Number: This is the tree number on which the explicit parent node will be specified in the next IS-IS ID field.
- o IS-IS ID: The 7-octet IS-IS ID of the parent node for the originating node on the tree whose number is specified in the previous "Tree Number" field.

## 5. Operation of RBridge

An RBridge MUST examine the LSPs of all the reachable nodes before each calculation of multi-destination distribution trees. Depending on the examination result, the RBridge can be in one of the following two states:

- o "Ineligible" state: if there is at least one node which does not

include the "Parent Selection Algorithm Version" sub-TLV, or it includes a lower version than "Explicit Parent Selection" version.

- o "Eligible" state: if all the reachable nodes includes the "Parent Selection Algorithm Version" sub-TLV, and the version is greater or equal to "Explicit Parent Selection" Version (version 1)

An RBridge must follow the following rules.

An RBridge can include the "Parent Selection Algorithm Version" sub-TLV in Router-Capability TLV, although such include is not mandatory.

In "Ineligible" state, the RBridge should not include "Explicit Parent Preference" sub-TLV in its LSP. It MUST ignore any "Explicit Parent Preference" in other nodes' LSPs, if there is any. And it MUST run the original parent selection algorithm ([[RFC6325](#)] as modified by [[RFC7180](#)]).

In "Eligible" state, the RBridge can include "Explicit Parent Preference" sub-TLV in its LSP. It MUST run the modified tiebreak algorithm in the distribution tree calculation: when a node X is added into the tree j, and when there are p number of equal cost parent choices, the RBridge MUST exam the "Explicit Parent Preference" sub-TLV in X's LSP, and MUST take one of the following actions:

1. If the Explicit Parent Preference is among the equal cost parent choices, the explicit parent preference MUST be honored.
2. Otherwise, the Explicit Parent Preference MUST be ignored. Select the choice of  $(j-1) \bmod p$ , as described in the original tiebreak algorithm.

## [6.](#) Other Applications

The proposed algorithm provides a tool to influence the tree parent node selection, and further makes it possible to re-shape the distribution trees. It is demonstrated how to improve the stability of the distribution trees with the tool.

With this tool, network engineers can also redirect multi-destination traffic and improve traffic load sharing. For example, in our previous example, while X can advertise (Tree 2, B) as explicit parent preference, but X can also advertise (Tree 2, D) instead. This makes it possible to utilize link XD, which in some cases may be desirable for the purposes of traffic engineering.

## 7. Security Considerations

No additional security risk is introduced by using the mechanisms proposed in this document.

For TRILL general Security Considerations, see [[RFC6325](#)].

## 8. IANA Considerations

This document introduces two new router-capability sub-TLVs that require code point assignment:

- o Parent selection tiebreak algorithm version, to be assigned from the IANA "IS-IS TLV Codepoints Registry".
- o Explicit parent selection, to be assigned from the IANA "IS-IS TLV Codepoints Registry".

## 9. Conclusions

This document presents a parent selection tiebreak algorithm in the calculation of distribution trees in TRILL. A protocol extension in IS-IS is defined, and a network migration strategy is also designed. The proposed algorithm provides a tool to influence the parent node selection, which can improve the stability and traffic load sharing of the distribution trees when multiple equal cost parents are present. It also offers a tool to network administrators to pin down the multi-destination traffic to trees in more deterministic way.

## 10. References

### 10.1. Normative References

- [RFC6325] Perlman, R., Eastlake, D., Dutt, D., Gai, S., and A. Ghanwani, "RBridges: Base Protocol Specification", [RFC 6325](#), June 2011.
- [RFC7180] D. Eastlake, M. Zhang, A. Ghanwani, V. Manral, A. Banerjee, "TRILL: Clarifications, Corrections, and Updates", May 2014.
- [RFC4971] Vasseur, JP. and N. Shen, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Information", 2007.

## [11](#). Acknowledgments

The authors would like to thank Abhay Roy, Erik Nordmark, Varun Shah, Ramkumar Parameswaran and Deepak Sreekantan for review and suggestions.

This document was prepared using 2-Word-v2.0.template.dot.

Yang, et al

Expires January 17, 2016

[Page 13]

Internet-Draft

Trill: Parent Selection

July 2015

### Authors' Addresses

Howard Yang  
Cisco Systems  
170 West Tasman Drive, San Jose, CA 95134

Email: [howardy@cisco.com](mailto:howardy@cisco.com)

Ayan Banerjee  
Cisco Systems  
170 West Tasman Drive, San Jose, CA 95134

Email: [ayabaner@gmail.com](mailto:ayabaner@gmail.com)

Donald Eastlake  
Huawei R&D USA  
155 Beaver Street, Milford, MA 01757

Phone: 1-508-333-2270  
Email: [d3e3e3@gmail.com](mailto:d3e3e3@gmail.com)

Radia Perlman  
Intel Labs  
2200 Mission College Blvd., Santa Clara, CA 95054-1549

Phone: 1-408-765-8080  
Email: [radia@alum.mit.edu](mailto:radia@alum.mit.edu)