INTERNET-DRAFT
Intended Status: Proposed Standard
Expires: March 25, 2013

Kesava Vijaya Krupakaran
Janardhanan Pathangi Narasimhan
Dell

September 21, 2012

Fair Share AF Load Share
draft-kvk-trill-fair-share-af-load-share-02

Abstract

In an access LAN of a TRILL campus, the DRB can choose to load share the AF responsibility among other RBridges in the LAN. This document throws light on one such approach where the AF appointment is fair share scheduled among the RBridges.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

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/1id-abstracts.html

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

Copyright and License Notice

Table of Contents

1 Introduction

In a shared access LAN, the appointed forwarder for a VLAN is responsible for encapsulating and decapsulating native traffic on that VLAN. Other non-AF RBridges in the LAN discard the native traffic for that VLAN.

The DRB can choose to be the AF for all VLANs or load share the AF responsibility among other RBRidges in the LAN. This ensures better utilization of resources like hardware tables and buffers. The VLAN partitioning scheme suggested in [RFC6439] section 2.2.1 is static and requires careful configuration. Another simple protocol would be to allocate VLANs in a round-robin fashion among all RBridges in the LAN. However this doesn't leave scope for schemes like retaining 50% of VLANs with the DRB and distribute only the rest among others.

Fair share scheduling of AF allows for the flexibility of assigning certain RBridges (say with higher switching capability) AF for higher proportion of VLANs than others.

1.1 Terminology

This document uses the acronyms defined in [RFC6439].

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].

2 Shares

Each RBridge is configured with certain quantity of shares. A share is the proportion of VLANs which would be allocated to the RBridge in comparison with other RBridges. The face value of the shares is a relative quantity and makes sense only when taken in conjunction with total shares allocated in the LAN.

These shares are advertised by each RBridge in its hello. The DRB load shares the AF among RBridges based on the relative value of shares.

For instance, let A, B and C be three RBridges with S(A) = 2, S(B) = 1 and S(C) = 1. Then A is assigned the AF for 1/2 of the VLANs while B and C the AF for 1/4th of the VLANs each.

Even when the number of VLANs for which the RBridge is to be AF calculates to a non integer value, it should be made sure that there

is only one AF for a VLAN in a multi-access LAN.

3 AF Affinity VLAN Set

Fair share scheduling distributes VLANs among RBridges according to proportion of shares allocated. This allows allocation of higher proportion of VLANs to certain RBridges (with higher switching capability). However, this does not guarantee that these RBridges would handle larger share of the native traffic.

Following the previous example, even though A is appointed AF for 50% of the VLANs while B only 25% of the VLANs, the traffic load of VLANs for which B is AF could be considerably higher that those in A.

In order to overcome this conundrum, each RBriged in access LAN is configured with an AF Affinity VLAN Set apart form the share proportion. This RBridge has AF affinity to the set of configured VLANs. Thus when the DRB appoints an RBridge AF for a set of VLANs, the members of the set are chosen from the AF Affinity VLAN Set advertised.

Expanding on the previous example, if X denotes an RBridge, let S(X) be the shares assigned to X, V(X) be the AF Affinity VLAN Set and AF(X) denote the set of VLANs for which X is assigned AF. Let the access LAN encompass ten shared VLANs [11, 20]. In this case the AF assignment with just the shares configured could be as in Table 1. If RBridge A has higher switching capability and VLANs [16, 20] are heavily loaded, this AF appointment defeats the purpose.

Table 1: AF appointment using fair share scheduling

X	S(X)	AF(X)
A	2	{11, 12, 13, 14, 15}
B	1	{16, 17, 18}
C	1	{19, 20}

By configuring AF Affinity VLAN set in each RBridge, this difficulty can be overcome. Such a configuration is shown in Table 2. How the AF Affinity VLAN set is arrived at is beyond the scope of this document. Long term traffic planning tools could be helpful in extrapolating a decent configuration.

Table 2: Fair share scheduling with AF Affinity VLAN set			
X	S(X)	V(X)	AF(X)
A	2	{16, 17, 18, 19, 20}	{16, 17, 18, 19, 20}
B	1	{11, 12, 13, 14}	{11, 12, 13}
C	1	{12, 13, 14}	{14, 15}

4 AF Affinity VLAN Set Overlap

If the AF Affinity VLAN sets advertised by the RBridges overlap, the RBridge with higher share has priority over the affinity of common VLANs. In case the RBridges advertise same share with conflicting AF Affinity VLAN sets, then the one with higher system ID gets more AF affinity over the common VLANs.

Table 3: Fair share scheduling with AF Affinity VLAN set overlap				
X	ID(X)	S(X)	V(X)	AF(X)
A	0000.0000.000a	2	{15, 16, 17, 18, 19, 20}	{15, 16, 17, 18, 19}
B	0000.0000.000b	1	{11, 12, 13, 14, 15}	{14, 20}
C	0000.0000.000c	1	{11, 12, 13, 14}	{11, 12, 13}

Consider the previous examples with the LAN comprising of three RBridges A, B and C coloured for VLANs [11, 20]. As shown in table 3, the AF Affinity VLAN sets overlap in RBridges {A, C} as well as {B, C}. A, having the highest share has the most affinity over the VLANs configured there. In this example, A has higher AF affinity to VLAN 15 than C. Similarly, C has greater the AF affinity of VLANs [11, 14] than B on virtue of its higher system ID.

It is possible to calculate a better AF distribution by examining common VLANs in AF Affinity VLAN sets when they overlap. Such algorithms have been avoided to keep the computation at DRB simple.

5 AF Distribution Among Heterogeneous RBridges

An access LAN could constitute motley set of RBridges with some that support fair share AF scheduling and some that doesn't. If the DRB doesn't support fair share AF scheduling, it ignores the sub-TLVs advertised by other RBridges and continue to distribute AF as it did previously.

If the DRB does support fair share AF scheduling and it receives hello form an RBridge without Fair Share Sub-TLV, it is assigned a default share equal to average of all shares advertised in the LAN during AF computation. If the AF Affinity VLAN Set Sub-TLV was not advertised, it is taken to be a NULL set. In case an RBridge advertises AF Affinity Sub-TLV without saying the shares, such TLV is ignored and the behaviour follows as though it had not advertised AF Affinity VLAN set.

In particular, if DRB is the only RBridge supporting the feature, all the RBridges get equal shares (equal to the one configured at DRB, consistent with the average rule discussed).

For instance, in an access LAN with RBridges A, B and C where S(A) = 7, S(B) = 3, and C doesn't support fair share AF scheduling, the DRB assigns it a default of 5 shares.

As a special case, if DRB supports fair share AF load share and none of the RBridges advertise any share and no share is configured in DRB, then DRB assigns a share value of 1 to all RBridges and load shares VLANs equally among all the RBridges.

6 AF Computation at DRB

DRB runs through all RBridges, in descending order of shares configured and assigns the AF based on Affinity VLAN set. If the shares advertised are equal, then RBridges are ordered based on system ID. If RBridges don't advertise shares, they are assigned default shares and are placed below RBridges who advertise shares in the ordered list of RBridges. If there are multiple such non congruous RBridges, they are again ordered based on system ID.

The DRB also monitors hellos for any change from previously advertised shares or AF Affinity VLAN set. If it detects a change, the AF assignment is recomputed for all RBridges. Any addition or deletion of adjacency also triggers fresh AF assignment. This simplifies the computation at DRB.

7 AF and VLAN Mapping

If the DRB detects VLAN mapping, it appoints one RBridge (possibly itself) as the AF for all VLANs as suggested in [RFC6439] section 2.4 to prevent loops.

8 AF and Multiple ports on a link

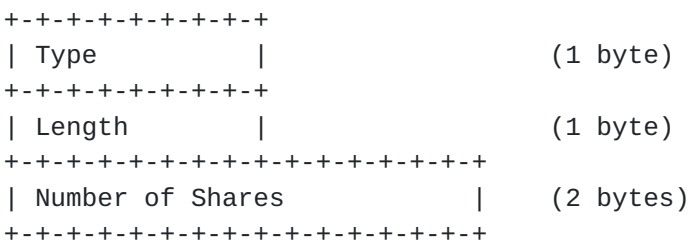
The shares configured represents the whole RBridge's proportion of AF sought. Further load sharing of AF among multiple ports on same link in an RBridge is a local decision.

9 Multi-Topology-Aware Port Capability Sub-TLVs

Two new Multi-Topology-Aware Port Capability Sub-TLVs are required for the purpose of fair share AF appointment - Fair Share Sub-TLV and AF Affinity VLAN Set Sub-TLV.

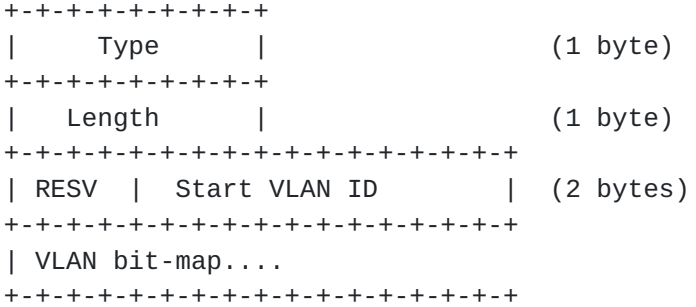
9.1 Fair Share Sub-TLV

Fair Share Sub-TLV is used to advertise the number of shares configured in the RBridge. Number of shares is a two octet value. When an RBridge advertises zero shares, it is not assigned any AF.



9.2 AF Affinity VLAN Set Sub-TLV

AF Affinity VLAN Set Sub-TLV is used to advertise the AF Affinity VLAN set configured in an RBridge. It is a facsimile of the Enabled-VLANs sub-TLV.



9.3 Partial VLANs Appointing Sub-TLV

As discussed in [RFC6439] section 2.2.3, the size of hello imposes a limit on the distribution of AF info in AF Sub-TLV by the DRB. The nature of the algorithm means that the AF appointment information could be disjoint. If the number of VLANs on a shared link is too high, all AF appointments cannot be accommodated in a single hello using the start end mechanism of AF Sub-TLV. In such case, the DRB should appoint one RBridge (possibly itself) as AF for all VLANs.

Alternatively, the AF information can be sent in a bitmap rather than start-end mechanism as suggested in AF Sub-TLV. For this purpose the Partial VLANs Appointing Sub-TLV suggested in Adaptive VLAN Assignment draft [VlanAsn] can be used.

10 Security Considerations

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

11 IANA Considerations

This document suggests two additional Sub-TLV to Multi-Topology-Aware Port Capability TLV apart from the reuse of Partial VLANs Appointing Sub-TLV from Adaptive VLAN Assignment draft.

- Fair Share Sub-TLV
- AF Affinity VLAN Set Sub-TLV

12 References

12.1 Normative References

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

[RFC6325] R. Perlman, D. Eastlake, et al, "RBridges: Base Protocol Specification", RFC 6325, July 2011.

[RFC6326] D. Eastlake, A. Banerjee, et al, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 6326, July 2011.

[RFC6439] D. Eastlake, R. Perlman, et al, "Routing Bridges (RBridges): Appointed Forwarders", RFC 6439, November 2011.

[RBisisb] D. Eastlake, A. Banerjee, et al, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", draft-eastlake-isis-rfc6326bis-09.txt, work in progress.

12.2 Informative References

[VlanAsn] M.Zhang and D.Zhang, "Adaptive VLAN Assignment for Data Center RBridges", draft-zhang-trill-vlan-assign-04.txt, work in progress.

Authors' Addresses

Kesava Vijaya Krupakaran
Dell
Olympia Technology Park,
Guindy Chennai 600 032

Phone: +91 44 4220 8496
Email: Kesava_Vijaya_Krupak@Dell.com

Janardhanan Pathangi
Dell
Olympia Technology Park,
Guindy Chennai 600 032

Phone: +91 44 4220 8459
Email: Pathangi_Janardhanan@Dell.com