Adaptive VLAN Assignment for Data Center RBridges
draft-zhang-trill-vlan-assign-04.txt

Abstract

   If RBridges are casually assigned as Appointed Forwarders for VLANs
   without considering the number of MAC addresses and traffic load of
   these VLANs, it may overload some of the RBridges while leave other
   RBridges lightly loaded, which reduces the scalability of a RBridge
   network and undermines its performance.

   A new mechanism is designed in this document to support the adaptive
   VLAN assignment (or Appointed Forwarder selection) based on the
   forwarders' reporting of their usage of MAC tables and available
   bandwidth.

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

The scales of Data Center Networks (DCNs) are expanding rapidly these years. In DCNs, Ethernet switches and bridges are abundantly used for the interconnection of servers. The plug-and-play feature and the simple management and configuration of Ethernet are appealing to DCN providers. A whole DCN can be a simple large layer 2 Ethernet which is either built on a real network or on a virtual platform.

Cloud Computing is growing up from DCNs which can be regarded as a virtual platform that provides the reuse of the network resources of DCNs. A lot of cloud applications have been developed by DCN providers, such as Amazon's Elastic Compute Cloud (EC2), Akamai's Application Delivery Network (ADN) and Microsoft's Azure. Cloud Computing clearly brings new challenges to the traditional Ethernet. The scales of the DCNs are becoming too large to be carried on the traditional Ethernet. The valuable MAC-tables of the bridges are running out of use for storing millions of MAC addresses. The broadcast of ARP messages consumes too much bandwidth and computing resources. The mobility of end stations brings dynamics to the network which can become a heavy burden if the management and configuration of the network involves too much manpower. The Spanning Tree Protocol used in the traditional Ethernet is becoming outdated since there is only a single viable path on the tree for a node pair and this path is not always the best path (e.g., shortest path).

RBridges are designed to improve the shortcomings of the traditional Ethernet. To make use of the rich connections, RBridges introduce multi-pathing to the Ethernet to break the single-path constraint of STP. Multiple points of attachment is a basic feature supported by RBridges and common for Data Center Bridges. This feature not only increases the "east-west" capacity but also greatly enhances the reliability of DCNs [VL2] [SAN]. If several RBridges are attached to a bridged LAN link at the same time, the DRB is responsible for the assignment of a VLAN to one of the RBridges as the Appointed forwarder. However, VLAN assignment is currently done in an one-way manner. The DRB casually assigns a VLAN to an RBridge attached to the local link without knowing its available MAC-table entries or bandwidth. The appointed forwarder does not feedback the utilization of its MAC-table or bandwidth either.

This document proposes to balance the load among VLANs. Two types of sub-TLVs are defined, with which a forwarder can report its MAC entries and traffic bit rate respectively. By gathering these report messages, the VLAN assignment can be done in a way that the usage of the MAC tables and bandwidth of the attached RBridges are balanced among VLANs.

[1.1](). Terminology

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](). Data Center RBridge

Data Center Networks grow rapidly. Ethernet is widely used in data centers because of its simple management and plug-and-play features. However, there are shortcomings of Ethernet. RBridges are designed to improve these shortcomings. In this section, we analyze the characteristics of DCNs that impact the design of RBridges and reveal why the adaptive VLAN assignment is important for RBridges to be used in DCNs.

[2.1](). Scalability

In the past years, a large DCN is typically composed of tens of thousands servers interconnected through switches. In the cloud computing era, there can be as many as millions of servers in one DCN. The management of the massive MAC addresses of the servers on the layer2 devices will become more and more complex. RBridges are used to replace the traditional bridges.

Valuable CAM-tables on RBridges can easily be used up if they are not used reasonably [[CAMtbl]()]. For RBridges to be widely used in DCNs, the VLANs should be assigned to the RBridges in a manner that MAC entries of VLANs on RBridges are balanced.

[2.2](). East-West Capacity Increase

The Spanning Tree Protocol (STP) in the traditional LAN blocks some ports of bridges for the purpose of loop avoidance. The side-effects of STP are obvious. The link bandwidth attached to the blocked ports are not utilized which greatly wastes the capacity of the network. On the tree topology, the communication between the bridges of the left branch and right branch must transit the single root bridge, which forms a "hair-pin turn".

With the rapid increase of the amount of servers in DCNs and their traffic demand, it is urgent to break the constraint of STP and enhance the "east-west" capacity of DCNs which are always richly connected. RBridges use the multi-path routing to set up the data plane of a TRILL campus. Multiple RBridges may be attached to a same LAN link, which offers multiple access points to the LAN link. The hosts on this LAN link is therefore multi-homed to a TRILL campus. All the attached RBridges can act as packet forwarders for VLANs

carried on this LAN link. In the worst case, all the VLANs are assigned to a single RBridge. Under this scenario, the ingress capacity on other RBridges is wasted. It is necessary to balance the traffic load of VLANs among these RBridges through the assignment of VLANs.

2.3. Virtualization

Virtualization is important for increasing the utilization of network resources in DCNs. For example, the VPNs can be used to separate the traffic from different services therefore they can be carried on the same pool of resources. When the VPNs is carried over a TRILL campus, RBridges can use a VLAN tag to identify each VPN. However, the use of VLANs multiplies the entries in the MAC table of RBridges.

Virtual Machines (VM) are widely used in DCNs. A physical host can support tens of VMs and each VM has to be identified by one MAC address that is need to be stored in MAC tables of RBridges. This seriously increases the number of MAC entries in RBridges. Moreover, the number of VMs in a VLAN is not necessarily equal to the number of physical hosts. VMs are spawned or destroyed based on the demand of applications. They can also migrate from one location to another, which may be either an in-service or out-of-service move. VMs bring about the volatility of the size of VLANs. It is hard to provide one static VLAN assignment for a TRILL campus based on the numbers of physical hosts of VLANs that is proper for all applications all the time.

3. MAC Entries Usage

A CAM-table on a switch is expensive, which is a major constraint on the scalability of Ethernet [CAMtable]. When an RBridge is used to connect lots of hosts in large Data Center Networks, the entries of the CAM-table can easily be used up. The network should be tactically interconnected and valuable MAC table entries should be used economically.

RBridges support multiple points of attachment [RFC6325]. When RBridges are used in a DCN to form a TRILL campus, a LAN link MAY have multiple access points to this campus. All these access RBridges are able to act as the packet forwarder of VLANs carried on this LAN link. The DRB of this LAN link is responsible to pick out one RBridge attached to this LAN link as the appointed forwarder for each VLAN-x. In other words, the DRB assigns VLAN-x to one of the RBridge. For an assigned VLAN, its forwarder is not only responsible for forwarding the packets for this VLAN but also need to store active MAC addresses of hosts on this VLAN.

If VLANs on a LAN link are not appointed properly, some RBridges's
MAC tables are easily to be used up while other RBridges are left
idle. Take Figure 3.1 as an example, there are four VLANs carried on
the LAN link: w, x, y and z. There are two hosts in both VLAN-w and
VLAN-x and one host in both VLAN-y and VLAN-z. RB1 and RB2 are both
attached to this LAN link. RB1 is elected as the Designated RBridge
who is responsible to choose appointed forwarders for the above
VLANs. In the figure, VLAN-w,x are assigned to RB1 and VLAN-y,z are
assigned to RB2. Obviously, this assignment is not balanced, since
the MAC table of RB1 has four entries while the MAC table of RB2 only
has two entries. If the DRB can reassign VLAN-w to RB2 and reassign
VLAN-y to RB1, both RBridges will have three MAC entries, therefore a
more balanced assignment is achieved.

```
       MAC Entries
       +-----+
       |  w  |
       +-----+
       |  w  |                             MAC Entries
       +-----+ >---+                         +-----+
       |  x  |     |                         |  y  |
       +-----+     |                 +---< +-----+
       |  x  |     |                 |       |  z  |
       +-----+     |                 |       +-----+
                   |                 |
       DRB&AF:w,x  |                 |     AF:y,z
       +-----+     |                 |       +-----+
       | RB1 |-----+                 +-----| RB2 |
       +-----+                               +-----+
          |                                     |
       @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
        @                                         @
       @ +-------+ +-------+   +-------+ +-------+ @
       @ |[H] [H]| |[H] [H]|   | [H]  | | [H]  | @
       @ +-------+ +-------+   +-------+ +-------+ @
        @   VLAN-w    VLAN-x      VLAN-y    VLAN-z @
         @                                       @
       @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
                     LAN link
```
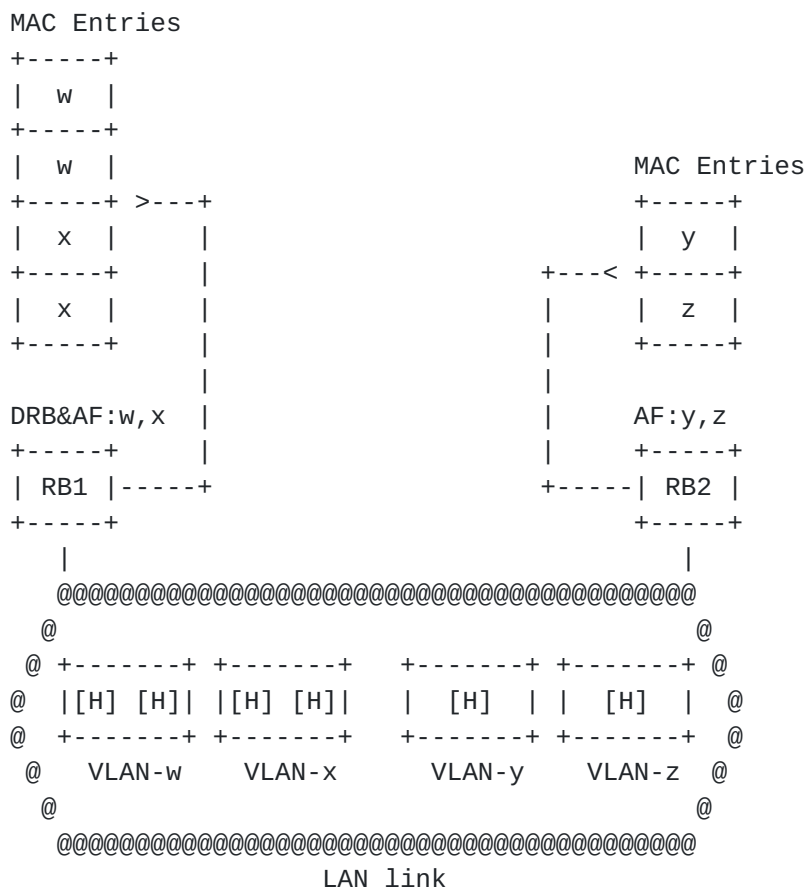
                Figure 3.1: Unbalanced assignment among VLANs

In order to assign the VLANs in a balanced way, the DRB need to know
the usage of MAC tables of its appointed forwarders. Since RBridges
only store active MAC addresses and a virtual machine can move from
one location to another, MAC entries that a VLAN occupies on an
RBridge varies from time to time. The assignment of VLANs cannot be
done once for all. It is necessary for the DRB to do the assignment

adaptively taking the usage of MAC tables of its appointed forwarders into consideration.

In Section 5.1, the MAC Entries Report sub-TLV is defined to deliver this kind of information from a forwarder to a DRB.

4. Load Distribution

The traffic from the TRILL campus to the local LAN link is the egressing traffic while the traffic from the local LAN link to the TRILL campus is the ingressing traffic. A forwarder RBridge acts as both the ingress and egress point of a VLAN's traffic. The assignment of the appointed forwarder for each VLAN affects both the egress and ingress traffic load distribution.

4.1. Egress Traffic

One RBridge MAY have multiple ports attached to the same local LAN link. These ports are called "port group" [RFC6325]. When a DRB assigns a VLAN to an RBridge, its total available egress bandwidth of the port group needs to be taken into consideration.

After VLAN-x has been assigned to an RBridge, the forwarding port assignment of one of the port group to VLAN-x as the forwarding port is entirely a local matter. Since a LAN link is a STP domain, more than one forwarding port for one VLAN will cause a loop. The forwarder MUST assign one and only one port for each VLAN. Load balancing among multiple ports on the same link can be realized through splitting the load among different VLANs as suggested in Section 4.4.4 of [RFC6325].

4.2. Ingress Traffic

After packets enter a TRILL campus from the ingress RBridge, they are sent through the paths starting at this ingress point. Since the DRB knows the whole topology of the TRILL campus, it can figure out these paths as well. Therefore, the DRB should take the available bandwidth of these paths into consideration when assigning the appointed forwarder of a VLAN. Any assignment that is possible to congest an already busy ingress point or a path should be avoided.

Traffic Matrices are usually taken as the input to the traffic engineering methods [TE]. VLAN assignment actually changes the Traffic Matrix of a TRILL campus. Therefore, our adaptive VLAN assignment can work together with traffic engineering methods to achieve a balanced traffic distribution of the whole network. However, the design of this kind of cooperative mechanism is out the scope of this document.

With the TLV defined in Section 5.2, the egressing and ingressing
traffic of an appointed forwarder are reported to its DRB.

5. Feedback Sub-TLVs

The Appointed Forwarders TLV has already been defined in [RFC6326].
With this TLV, the DRB can appoint an RBridge on the local link to be
the forwarder for each VLAN. However, there is no feedback from the
appointed forwarder to notify the DRB whether the assignment is
reasonable. Two sub-TLVs are defined in this section to open the
passageway of feedback.

5.1. MAC Entries Report Sub-TLV

Appointed forwarders use MAC Entries Report sub-TLV to report the
usage of their MAC tables to the DRB. It has the following format:

```
+-+-+-+-+-+-+-+-+
|Type=MACEtrRep |                          (1 byte)
+-+-+-+-+-+-+-+-+
| Length        |                          (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  DRB Nickname              |              (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Maximum MAC Entries       |              (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Available MAC Entries     |              (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        MAC Entries of VLAN Range (1)    | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        ......                           | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        MAC Entries of VLAN Range (n)    | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where each "MAC Entries of VLAN Range" is of the form:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |    Start.VLAN      |              (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |    End.VLAN        |              (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| The Number of MAC Entries  |              (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o Type: MAC Entries Report sub-TLV.

o Length: 6+6n bytes, where n is the number of VLANs that the

appointed forwarder selects to report.

o DRB Nickname: The nickname of the Designated RBridge of the local
   link.

o Maximum MAC Entries: The maximum number of the entries of the MAC
   table of the appointed forwarder.

o Available MAC Entries: The number of available entries of the
   appointed forwarder's MAC table.

o RESV: These bits MUST be sent as zero and ignored on receipt.

o Start.VLAN, End.VLAN: These fields are the VLAN IDs of the
   appointment range, inclusive. To specify a single VLAN, the VLAN's
   ID appears as both the start and end VLAN [RBissib].

o The Number of MAC Entries: The number of MAC entries occupied by
   the given VLAN range. These MAC entries does not only contain the
   MAC addresses of hosts on the local link but also includes the MAC
   addresses on the remote link in the same VLAN range.

5.2. Traffic Bit Rate Report Sub-TLV

   Appointed forwarders use Traffic Bit Rate Report sub-TLV to report
   the bandwidth utilization of their ports (or port groups) to the DRB.
   This sub-TLV has the following format:

```
+-+-+-+-+-+-+-+-+
|Type=TrafficRep|                         (1 byte)
+-+-+-+-+-+-+-+-+
|  Length       |                         (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  DRB Nickname             |             (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Maximum Link Bandwidth      |          (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Available Link Bandwidth    |          (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Traffic Bit Rate of VLAN Range (1)      | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             ......                       | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Traffic Bit Rate of VLAN Range (n)      | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   where each "Traffic Bit Rate of VLAN Range" is of the form:

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | RESV|D|     Start.VLAN         |               (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | RESV  |     End.VLAN           |               (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  Traffic Bit Rate             |               (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   o Type: Traffic Bit Rate Report sub-TLV.

   o Length: 6+6n bytes, where n is the number of VLANs that the
      appointed forwarder selects to report.

   o DRB Nickname: The nickname of the Designated RBridge of the local
      link.

   o Maximum Link Bandwidth: The maximum bandwidth of the port (or port
      group) attached to the local link.

   o Available Link Bandwidth: The available bandwidth of the port (or
      port group) attached to the local link.

   o RESV: These bits MUST be sent as zero and ignored on receipt.

   o D: The direction of the traffic. Value "0" denotes the egressing
      traffic volume and value "1" denotes the ingressing traffic
      volume.

   o Start.VLAN, End.VLAN: These fields are the VLAN IDs of the
      appointment range, inclusive. To specify a single VLAN, the VLAN's
      ID appears as both the start and end VLAN [RBissib].

   o Traffic Bit Rate: The traffic bit rate of the given VLAN range
      from/to the local link through the attaching ports group of the
      appointed forwarder.

6. Adaptive VLAN Assignment

   Appointed forwarders MAY advertise the MAC Entries Report Sub-TLV and
   Traffic Bit Rate Report Sub-TLV included in Multi-Topology-Aware Port
   Capability TLV in LSPs to their DRBs. The number of VLANs in these
   TLVs is constrained by the maximum size of HELLO messages which is
   1470 octets [RFC6325]. If it is not big enough to hold the
   information for all VLANs, appointed forwarders choose VLANs in the
   order of most MAC entries first and highest load first to report to
   the DRB. An appointed forwarder SHOULD report these two sub-TLVs to
   its DRB each time the DRB assigns a VLAN to it. Its adjacency to the
   DRB MUST be in Report state [RFC6327]. In order to relieve the burden

of feedback, thresholds MAY be imposed on the report of these sub-
TLVs. For example, when the usage of MAC table increases by 10
percent, the Appointed Forwarder reports the MAC Entries Report Sub-
TLV to its DRB.

The usage of the MAC table and the utilization of bandwidth of this
appointed forwarder is calculated on the DRB. The DRB SHOULD choose
an appointed forwarder for the local link based on Least
Usage/Utilization First (LUF) policy. If an RBridge on the local link
is already overloaded, the DRB should refrain from assigning
additional VLANs to it.

An Appointed Forwarder may fail or get overloaded [RBclear]. The
usage of MAC tables and bandwidth of Appointed Forwarders attached to
a LAN link changes with time elapses and a balanced VLAN assignment
may become unbalanced. For example, the bandwidth of an Appointed
Forwarder may run out of use due to the increase of traffic demand.
The usage of the MAC table of an Appointed Forwarder may change
greatly due to host mobility. In such situations, the DRB need to
modify the appointment of VLANs on a LAN link. The change of
Appointed Forwarder of a VLAN brings service interruption to this
VLAN, therefore the reassignment of VLANs from existing Appointed
Forwarder to a new Appointed Forwarder should be limited.

It has been a common practice to collect MAC table usage in bridge
networks. Through the collection of the above two sub-TLVs (e.g.,
stored in MIB [RBmib].), operators will have a vision of the MAC
table usage and bandwidth utilization of the RBridges on the LAN
link. Based on this vision, operators can easily recognize the hot
spot of the TRILL campus, which facilitates the trouble shooting.

7. **Partial VLANs Appointment**

[RFC6349] indicates that the appointed VLAN list for one appointed
forwarder sent in an Appointed Forwarders Sub-TLV should be
contiguous. In particular, when the designated VLAN falls in the
range of the appointed VLANs list, the DRB should have two
appointments with contiguous appointed VLAN ranges for one appointed
forwarder. However, on a LAN link, a DRB can flexibly appoint any
VLAN to any appointed forwarder. This flexibility creates the
necessity to allow DRB to have discrete appointments to one appointed
forwarder. The following example shows this necessity and reveals the
issue caused by discrete VLAN appointment.

When enabled VLANs of two RBriges attached to the same link have
intersections, DRB SHOULD guarantee that the sets of VLANs appointed
to these two RBridges do not overlap. With the adaptive VLAN
assignment method defined in this document, the DRB is probably to

discretely assign VLANs of this intersection to achieve load balance
among these two RBridges. Suppose both RB1 and RB2 have enabled the
VLANs from 2 through 4094 on the local link. VLAN 1 is the Designated
VLAN for this link. Suppose RB1 gets all the odd numbered VLANs while
RB2 gets all the even numbered VLANs. The DRB has to use 2046
appointments for RB1 and 2047 appointments for RB2. Since one
appointment consumes 6 octets [RBisisb] and a TRILL-HELLO message
MUST NOT exceed 1470 octets [RFC6325]. Even these octets are all used
for VLAN appointments, a TRILL-HELLO at most contains 245
appointments. The 4093 appointments obviously exceeds the maximum
number of appointments that one TRILL-HELLO can have.

The following two sub-TLVs are defined to enable the DRB to send out
a appointed VLANs list to its appointed forwarders in multiple TRILL-
HELLOs, which is similar as the TRILL Neighbor TLV [RBisisb].

7.1 Partial Appointed Forwarder Sub-TLV

This section defines a sub-TLV adapted from the Appointed Forwarders
sub-TLV [RBisisb]. This sub-TLV can appear in MT-PORT-CAP TLVs of
multiple TRILL-HELLOs.

```
+-+-+-+-+-+-+-+-+
|     Type      |                     (1 byte)
+-+-+-+-+-+-+-+-+
|     Length    |                     (1 byte)
+-+-+-+-+-+-+-+-+
|S|E| RESV      |                     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Appointment Information (1)       |  (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Appointment Information (2)       |  (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   ................                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Appointment Information (N)       |  (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where each appointment is of the form:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Appointee Nickname             |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |       Start.VLAN            |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |       End.VLAN              |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o Type: sub-TLV type, set to MT-PORT-CAP ParAppFwdr Sub-TLV TBD.

o Length: 1+6n bytes, where there are n appointments.

o Appointee Nickname: The nickname of the IS being appointed a
   forwarder.

o S: Start flag. If this bit is a one, then the first Appointment
   Information of this sub-TLV is the first appointment of the
   appointed VLANs list sent by the DRB.

o E: End flag. If this bit is a one, then the last Appointment
   Information of this sub-TLV is the last appointment of the
   appointed VLANs list sent by the DRB.

o R, RESV: These bits are reserved and MUST be sent as zero and
   ignored on receipt.

o Start.VLAN, End.VLAN: These fields are the VLAN IDs of the
   appointment range, inclusive. To specify a single VLAN, the VLAN's
   ID appears as both the start and end VLAN. As specified in
   [RFC6439], appointing an IS forwarder on a port for a VLAN not
   enabled on that port has no effect.

This sub-TLV enables the DRB to send VLAN appointment information in
multiple TRILL-HELLOs without the constraint of the size of a single
TRILL-HELLO. An IS's nickname can occur as an appointed forwarder for
multiple VLAN ranges by occurrences of this sub-TLV within the MT
Port Capability TLV [RBisisb]. However, an IS's nickname SHOULD only
occur as an appointed forwarder within the same TRILL-HELLO.

7.2 Partial VLANs Appointing Sub-TLV

This section defines another sub-TLV adapted from the VLANs Appointed
sub-TLV [RBisisb], which allows DRB to assign lots of VLANs to one
appointed forwarder in a bitmap format. This sub-TLV can appear in
MT-PORT-CAP sub-TLVs of multiple TRILL-HELLOs as well.

```
 +-+-+-+-+-+-+-+-+
 |    Type    |                        (1 byte)
 +-+-+-+-+-+-+-+-+
 |    Length    |                      (1 byte)
 +-+-+-+-+-+-+-+-+
 |S|E| RESV    |                       (1 byte)
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Appointee Nickname    |        (2 bytes)
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | RESV  |  Start VLAN ID      |       (2 bytes)
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | VLAN bit-map....
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o Type: sub-TLV type, set to MT-PORT-CAP ParVLANApp Sub-TLV TBD.

o Length: Variable, minimum 6.

o Appointee Nickname: The nickname of the IS being appointed a
   forwarder.

o S: Start flag. If this bit is a one, then this is the first
   appointment of the appointed VLANs list sent by the DRB.

o E: End flag. If this bit is a one, then this is the last
   appointment of the appointed VLANs list sent by the DRB.

o R, RESV: These bits are reserved and MUST be sent as zero and
   ignored on receipt.

o Start VLAN ID: The 12-bit VLAN ID that is represented by the high
   order bit of the first byte of the VLAN bit-map.

o VLAN bit-map: The highest order bit indicates the VLAN equal to
   the start VLAN ID, the next highest bit indicates the VLAN equal
   to start VLAN ID + 1, continuing to the end of the VLAN bit-map
   field.

In [RBisisb], an appointed forwarder sends the VLANs Appointed Sub-
TLV to its neighbors including the DRB. In this document, the
Partial VLANs Appointing sub-TLV is used by the DRB to assign
multiple discrete VLANs to an appointed forwarder.

A DRB can alternatively pick one of the above two sub-TLVs to send
appointment information to a specific appointed forwarder. However,
it is recommended that the DRB choose the option consuming less
octets in the TRILL-HELLO. In the above example, if the DRB sends out
the appointment using the Partial Appointed Forwarder Sub-TLV, RB1

has 2047 appointments which consumes 2047*6 = 12276 octets. Instead,
if the DRB uses the Partial VLANs Appointing Sub-TLV, it only
consumes 512 octets.

An appointed forwarder begins to accumulate VLAN appointments when it
receives either of the above two sub-TLVs whose first Appointment
Information sets the Start flag to a one. This appointed forwarder
will wait its Holding Time for either a Partial Appointed Forwarders
Sub-TLV or a Partial VLANs Appointing Sub-TLV whose End flag is set
to a one. If that sub-TLV is received before the Holding Timer
expires, the appointed forwarder will issue the accumulated VLAN
appointments. If the sub-TLV whose End flag is a one is not received
before the Holding Timer expires, all these accumulated VLAN
appointments will be discarded.

8. Security Considerations

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

9. IANA Considerations

This document suggests to specify four new Sub-TLVs from existing
sub-TLV sequences -- namely MACEtrRep, TrafficRep, ParAppFwdr and
ParVLANApp.

	Section	Sub-TLV#	MT Port Capabil.	Router Capabil.	Extended IS Reach	NUM
MACEtrRep	5.1	TBD	X	-	-	*
TrafficRep	5.2	TBD	X	-	-	*
ParAppFwdr	7.1	TBD	X	-	-	*
ParVLANApp	7.1	TBD	X	-	-	*

10. Acknowledgements

The authors would like to thank Somnath Chatterjee and Weiguo Hao for
their comments.

11. References

11.1. Normative References

[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.

   [RBisisb] D. Eastlake, A. Banerjee, et al, "Transparent
             Interconnection of Lots of Links (TRILL) Use of IS-IS",
             draft-eastlake-isis-rfc6326bis-01.txt, working in progress.

   [RBmib]   A. Rijhsinghani, K. Zebrose, "Definitions of Managed
             Objects for RBridges", draft-ietf-trill-rbridge-mib-06.txt,
             working in progress.

   [RFC6327] D. Eastlake, R. Perlman, et al, "Routing Bridges
             (RBridges): Adjacency", RFC 6327, July 2011.

   [RBclear] D. Eastlake, M. Zhang, et al, "TRILL: Clarifications,
             Corrections, and Updates", draft-ietf-trill-clear-correct-
             00.txt, working in progress.

11.2. Informative References

   [CAMtbl] B. Hedlund, "Evolving Data Center Switching",
             http://internetworkexpert.s3.amazonaws.com/2010/trill1/
             TRILL-intro-part1.pdf

   [SAN]   "Configuring an iSCSI Storage Area Network Using Brocade
             FCX Switches", Brocade CONFIGURATION GUIDE, 2010.

   [VL2]   A. Greenberg, J.R. Hamilton, et al, "VL2: A scalable and
             flexible data center network", in Proceedings of ACM
             SIGCOMM, 2009.

   [TE]   M. Roughan, M. Throup, and Y. Zhang, "Traffic Engineering
             with Estimated Traffic Matrices" , in Proceedings of ACM
             IMC, 2003.

Author's Addresses

    Mingui Zhang
    Huawei Technologies Co.,Ltd
    Huawei Building, No.156 Beiqing Rd.
    Z-park ,Shi-Chuang-Ke-Ji-Shi-Fan-Yuan,Hai-Dian District,
    Beijing 100095 P.R. China

    Email: zhangmingui@huawei.com

    Dacheng Zhang
    Huawei Technologies Co.,Ltd
    Huawei Building, No.156 Beiqing Rd.
    Z-park ,Shi-Chuang-Ke-Ji-Shi-Fan-Yuan,Hai-Dian District,
    Beijing 100095 P.R. China

    Email: zhangdacheng@huawei.com