Rapid Spanning Tree Migration

Mick Seaman

This note describes the BPDU formats and associated procedures that support plug and play migration from the original or 'legacy' spanning tree algorithm and protocol (STP) to the 'rapid' spanning tree protocol (RSTP). This description has been revised to take into account the conclusions reached at the Novmber 1999 802.1 meeting. The altenatives presented to that meeting are described in the 10/30/99 revision of this note.

Introduction

The proposed Rapid Spanning Tree Protocol was explicitly designed to be compatible with the legacy protocol specified in IEEE Std. 802.1D-1998 and prior revisions. Computation of the spanning tree is identical. Protocol changes are in the following areas:

a) Inclusion of the Port Roles (Root Port, Designated Port, Backup Port) in the computation of Port State (Blocking, Listening, Learning, Forwarding[1]). In particular a new Root Port transitions to forwarding rapidly.

b) Signaling to neighboring bridges of a bridge Port's desire to be Designated and Forwarding, and explicit acknowledgment by the neighboring bridge on a point-to-point link. This allows the transition of the Port State to Forwarding to complete without waiting for a timer expiry.

c) Acceptance of messages from a prior Designated Bridge even if they conveyed "inferior' information. Additionally a minimum increment to the Message Age is specified so that messages propagating in this way cannot 'go round in circles' for long.

d) Improvements in the propagation of topology change information so that that information does not have to be propagated all the way to the Root Bridge and back before unwanted learnt source address information is purged from forwarding databases.

e) Origination of Configuration BPDUs on a port by port basis, instead of transmission on Designated Ports following reception of information from the Root.

In addition to state machines described in P802.1w/D2, the following are required to support these changes:

1. Respecification of timer values to accommodate changed behavior in the cases where neighboring bridges do not implement the rapid algorithm, and the Forward Delay timers do actually run to completion. The default timers are believed to work well, but some care may be necessary in environments where timers have been tuned to their minimum values.
2. Detection of point-to-point links to allow selection of the procedures to indicate 'Designated wanting to become Forwarding' (referred to as 'designated indication' or 'txmt_di' and 'rcvd_di' in the state machines) and 'Yes, go ahead' ('designated confirmation' or 'txmt_dc' and 'rcvd_dc').
3. Specification of message formats to include designated indications and confirmations.

This note addresses this last message format question specifically.

The Message Format Problem

The initial proposal was to include two additional flags in the flags field of Configuration BPDUs and to increment the Protocol Version Identifier field. The intent of the 802.1D specification was that additions could be made to the BPDU so long as each BPDU contained at minimum all the fields of prior versions, and implementations conformant to any given version of the protocol effectively ignored the version identifier field for subsequent versions, processing them only according to their knowledge of the protocol. That is:

- a version 0 bridge would completely ignore the version identifier field and process all BPDUs as if they were version 0 BPDUs, ignoring additional fields
- a version 1 bridge would process version 0 BPDUs as version 0 BPDUs, version 1 BPDUs as version 1 BPDUs, and version 2 and higher BPDUs as version 1 BPDUs, igoring additional fields.

Unfortunately it appears that a conformance test house has widely disseminated a tool that checks that BPDUs of unknown version and unknown flags are discarded – thus ensuring an installed base of bridges that can not simply be resident in upgraded networks as originally envisioned.

The problem then, is to choose suitable message formats and possibly accompanying procedures to select a BPDU format on a specific link that provides forward migration with the minimum of fuss.

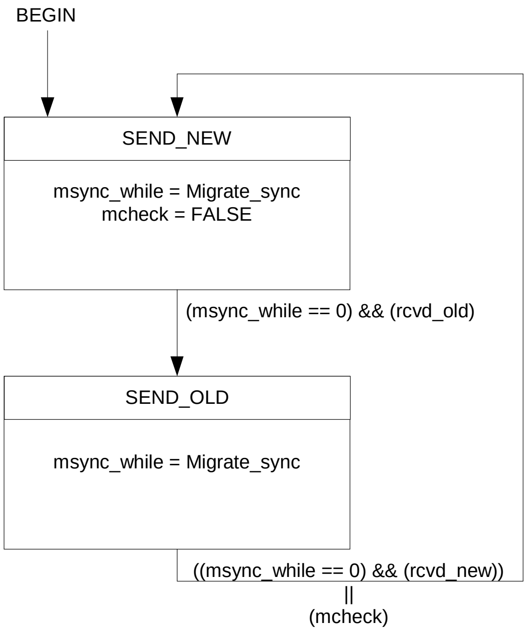
[1] Disabled for ports not participating in the algorithm

Message Format Selection

The proposed solution is as follows:

1. Use a new type of BPDU, referred to here as the 'RSTP BPDU' to carry the new parameters.
2. Retain the originally envisaged upgrade path for the the BPDU format, retaining the orginal 'reserved' multicast address and SAP values, but changing the version number of the protocol.
3. As part of the maintenance effort on 802.1D (P802.1t) a clearer and forceful statement of the originally envisaged migration strategy should be added.
4. Explicitly recognize and accommodate systems that conform to 802.1D-1998 in most respects, but that will discard these new BPDUs on the basis of their containing a different version number and/or unusual flags in the flags field.

To do this, systems that implement RSTP should be sensitive to the presence of legacy bridges that discard RSTP BPDUs. Such bridges will naturally attempt to become Designated on LAN segments that would otherwise only have RSTP BPDUs transmitted on them. If such legacy systems are detectedthe new systems should use old protocol formats (Config and TCN BPDUs). Naturally this will mean the loss of certain new capabilities, though the key element of rapidly transitioning new root ports to Forwarding is retained.


BEGIN
SEND_NEW
msync_while = Migrate_sync
mcheck = FALSE
(msync_while == 0) && (rcvd_old)
SEND_OLD
msync_while = Migrate_sync
((msync_while == 0) && (rcvd_new))
||
(mcheck)

A suitable choices for the timer value Migrate_sync is 3 seconds. mcheck is a flag set by management that forces one of the systems attached to the LAN to use the new RSTP BPDUs, and hence to test the hypothesis that all legacy systems that do not understand the new BPDU formats have been removed from the LAN. There is one uinstance of this machine per RSTP capable port.

The proposed BPDU formats are not sensitive tpo the details of this migration machine. It is sufficient for the purposes of this paper to note that such a machine can be specified.

This machine controls the choice of BPDU format on transmission as follows. If a new bridge is added to a shared LAN it will start by sending a new style BPDU. It will receive and process BPDUs in any format for a 3 second period, but receipt of any old style BPDU will not cause it to change the format that it uses for transmission (transmission only occurs as required by the algorithm). If all the bridges attached to the LAN are new style bridges then they will see the new style BPDU and send new style BPDUs themselves (if they need to transmit). However if an old style bridge is present it will persist in sending old style BPDUs after 3 seconds has elapsed. After the initial three second period in the SEND_NEW state, any old style BPDU will cause a transition to the SEND_OLD state. In this state any transmissions required are sent as old style BPDUs, and the state is not changed for 3 seconds. If after 3 seconds a new style BPDU is received, then the machine reverts to the initial SEND_NEW state. It also reverts to the initial state when the port is first initialized, and on an explicit management request.

A likely scenario is that the remaining old style bridge port(s) are Root Ports or Alternate Ports. In this case when a new style Designated Port checks to see if they have been removed from the LAN, they will be silent for a period until they time out the existing Root at which time they will attempt to become Designated. However this will drive the new bridge to send old style BPDUs, and the would be Designated port will be forced back to the Blocking port state well before it enters the Learning or Forwarding states.

Note one subtlety. In case the legacy system discards BPDUs based only on an analysis and validation of the 'Flags' field, the static 'In_sync' case in RSTP BPDUs sets a new flag. This is prefereable to discovering that BPDUs are discarded only during times of significant network change!

Note: the above determination of BPDU format is made independently for each Bridge Port, any given Bridge may have some ports using new style BPDUs and some old style.

In the all new state the initial value of fd_while assigned in the DBT and DLT states (see the state machines proposed in P802.1w/D2) can be cut to Hello_time.

RSTP BPDU Format

Protocol Identifier 0000 0000 0000 0000
Version = 2
BPDU Type = 0
Flags
Reserved
Root Identifier
Root Path Cost
Bridge Identifier
Port Identifier
Message Age
Max Age
Hello Time
Forward Delay
Version 1 Length 0000 0000 0000 0000

The Protocol Identifier is the same as for legacy STP. The Version number is 2, since Version number 1 was reserved for 802.1G. The Version 1 Length is as required by 802.1G for versions of 1 or higher.
The flags field contains the following information:

Bit1 : Topology Change Flag

Bit2 : Topology Change Notification Flag

Bits 3 (less significant) and 4:

> Encode the following port roles:
>
> 0 Unknown
>
> 1 Alternate Port (or Backup)
>
> 2 Root Port
>
> 3 Designated Port

Bit 5 : Learning

Bit 6 : Forwarding

Bit 7 : In Sync

> (operational state matches administrative state, prior root ports retired or confirmed)

Bit 8 : Topology Change Acknowledge Flag.