Interpretation request for 802.1ah:

L2 Gateway Protocol description

Mick Seaman

There appear to have been a number of editorial and technical mishaps in the development of 802.1ah's description of the L2GP extension to MSTP.

It appears that some changes were made to the wrong machines. Unless isL2gp is cleared (False) these errors will lead to indeterminate behavior as currently two machines are attempting to use and then clear the same variable (rcvdBPDU).

A consequence of the errors is that the 802.1ah state machines disagree with statements as to how L2GP is meant to operate. This note contains suggestions for corrections to the state machines and other normative descriptive elements in Clause 13 of P802.1ah, as it would amend IEEE Std 802.1Q. These are based on my review of the 802.1ah final text and P802.1ah ballot comments and related submissions (by others).

I would request that 802.1 determine whether these suggested changes are consistent with the way that 802.1ah should be interpreted.

A more difficult problem is a fundamental flaw in the 802.1ah logic, captured in checkBPDUConsistency(), for detecting a potential data loop due to some ports attached to the backbone being configured as L2GP ports while others are not. This procedure currently attempts to block a loop if **superior** information is received from the backbone. However it is the receipt of **inferior** information that would denote a loop from the L2 Gateway Port through a site (MSTP/RSTP being used within the site), out onto the backbone, and then back to the same L2GP port. Further, the way in which the L2GP port is meant to be blocked is through setting of the disputed variable, but the information is being received at a Root Port and the Port Role Transitions machine currently takes no action on disputed when selectedRole is RootPort. Note that the receipt of the backbone BPDU is not required to block the port if the L2GP priority were to be set too low resulting in the port becoming Designated, as that is already dealt with by the repeated generation of 'fake BPDUs' which would cause the Port Information Machine (PIM) to set disputed.

This interpretation request is not a complete set of changes to improve the 802.1ah L2GP description and related aspects of 802.1Q, but just highlights issues that could lead implementers to produce non-working implementations. I believe that the task of a thorough overhaul is best left to Q-REV or to an existing project that needs to make changes for its own consistency requirements. However I do suggest some name changes for future formal consideration, these would help point out the necessary changes.

The state machine corrections in this note do not contain any other functionality additions I have suggested in other notes, but are purely focused on repairing 802.1ah quickly. I make no claim that the corrected functionality is all that everyone might desire, merely that there are current inconsistencies in the P802.1ah that require interpretation, and indeed that the current L2GP description will result in indeterminate outcomes if isL2gp is True. I have not yet added L2GP functionality to my RSTP simulator, and it is unlikely that I will be able to do so for the forseeable future.

1. Summary

It appears that the general text of the P802.1ah changes to Clause 13 does not line up with the changes to the PIM and PRX state machines (2).

The Port Receive Pseudo Information machine has some curiosities that are also not reflect in the text. To make it possible to talk about the differences in interpretation, this note describes a drop in replacement machine —the L2GPRX machine (3).

The checkBPDUconsistency() procedure currently applies the wrong test. Connectivity should be prevented if the received BPDU contains worse information than pseudo-information. This note provides a replacement procedure (4). The way P802.1ah causes the checking to take effect is through setting the disputed variable, but this variable currently has no effect on a Root Port. Changes so that it has the desired effect are suggested in (5).

It might have been easier to spot these errors if the state machine overview figure had actually indicated what variables were being used and affected by the L2GP Port Receive Pseudo Information machine. I have produced an overview figure including my suggested replacement L2GPRX (4).

2. PIM and PRX changes

802.1ah makes a change to the Port Information Machine (PIM, 802.1Q Figure 13-14) preventing a transition to the RECEIVE state when isL2gp is True. This will defeat the discussed operation of the Port Receive Pseudo Information state machine, which is meant to set rcvdXstMsg to prompt this machine to process its regular pseudo messages. It appears that this state machine should have be left unchanged, as in Figure 2, and the Port Receive state machine (PRX) modified instead as in Figure 1. If the latter were not modified the operation of the state machines would be unpredictable as there would be a race between PRX and Port Receive Pseudo Information machine to clear rcvdBpdu.

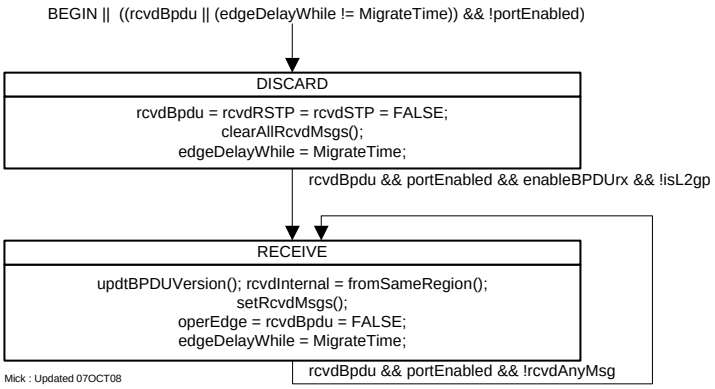


Figure 1—Port Receive state machine (with needed 802.1ah functionality added)

3. L2GPRX

802.1ah's Port Receive Pseudo Information machine has the following curiosities, which seem neither desired nor reflected in the text of 802.1ah:

a) It simulates two BPDU receptions when first enabled.

b) It simulates a pseudo-info reception when a BPDU is received even though that BPDU is not used, and its reception will have no effect apart from exercising the other state machines in the system with information they already have. This is true even if enableBPDUrx is False.

c) It lacks the capability to defend against receive message overrun, which PRX does by its use of !rcvdAnyMsg. Without this there could be unpredictable results if BPDUs arrived faster than the other machine can execute.

d) It feeds information to PIM even if the L2GP port is not enabled, therefore preventing use of redundant connections to the backbone, which seems contrary to the general intent.

In addition to the above the Pseudo Information machine got redrawn in P802.1ah D3.5 from a form that was completely consistent with the other state machines to one which largely ignored their diagrammatic conventions. Why I cannot determine.

Figure 3 is a Layer 2 Gateway Protocol state machine (L2GPRX) as a suggested replacement for the Port Receive Pseudo Information machine. Among other things it uses L2gpReceiveCheck() as a replacement for the current checkBPDUConsistency() procedure.

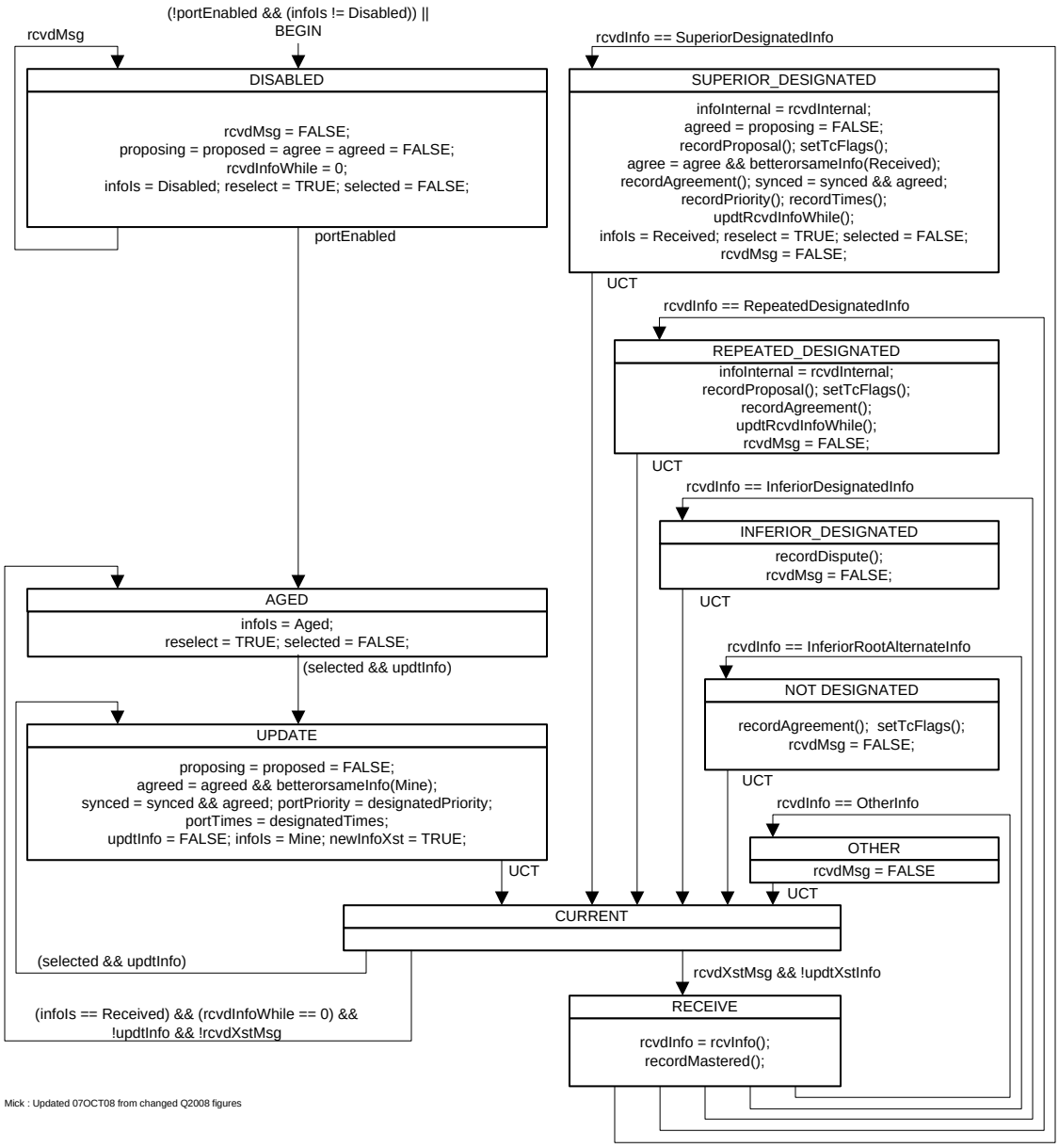


Figure 2—Port Information state machine (should not have been changed by 802.1ah)

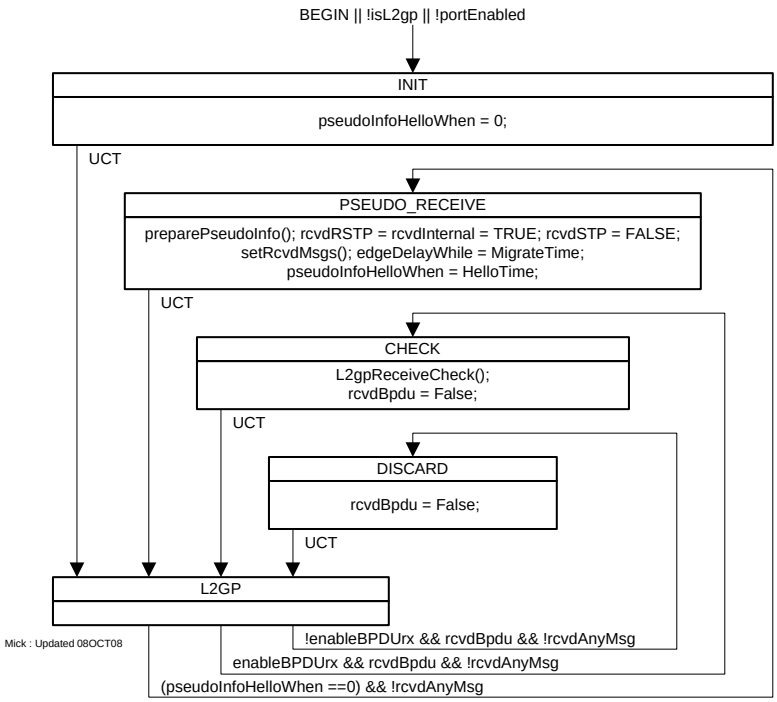


Figure 3—L2 Gateway Port Receive state machine

4. checkBPDUconsistency() problems

The following L2gpReceiveCheck() procedure is suggested as a replacement for the current checkBPDUConsistency(). The latter contains a construct of the form:

(a) and (b) or (c)

which is ambiguous, in addition to applying its check the wrong way round.

Further the current check does not deal with the fact that the received BPDU could be internal to the MST Region, and thus have a learning flag for each tree. Looking at ballot comments and past submissions it is clear that the use of the CIST flags was intended. Given the current restrictions as to what can be in the information delivered to PIM by L2GPRX that is sufficient, but does not support maximum per VLAN connectivity between multiple sites with configuration error protection.

The current text of checkBPDUConsistency() also disagrees with clause 14.4 Validation of received BPDUs and that needs to be fixed for the document to be self consistent.

L2gpReceiveCheck()

This procedure compares the CIST message priority vector of BPDUs received on an L2GP port with the port's own priority vector. Each received BPDU is identified as an STP, RSTP, or MSTP BPDU as specified in clause 14.4. The disputed flag is set and agreed is cleared for the CIST and all MSTIs if the received CIST message priority vector is inferior to the CIST port priority vector; and

a) The BPDU is an STP BPDU; or

b) The BPDU is an RST or MST BPDU with a CIST Port Role of Designated and has the Learning flag set for the CIST.

5. PRT changes

Figure 4 is a suggested set of changes to the Root Port states of the Port Role Transition machine (PRT) that cause the port to transition to Discarding if disputed is set, and continues to be set at intervals. I believe that this fixes the identified problem without unwanted side effects, but cannot be absolutely sure.

6. State machine overview

Figure 5 is a revised state machine overview, i.e. a proposed replacement for 1Q Figure 13-9. It provides information missing from the equivalent figure in 802.1ah.

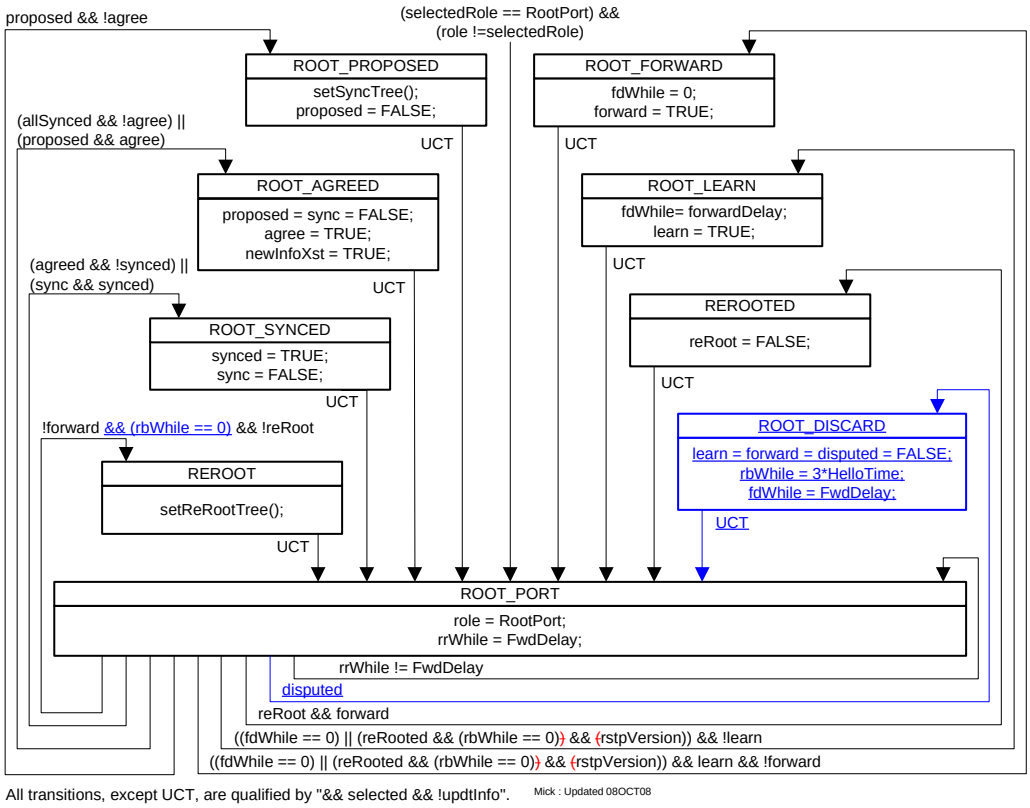
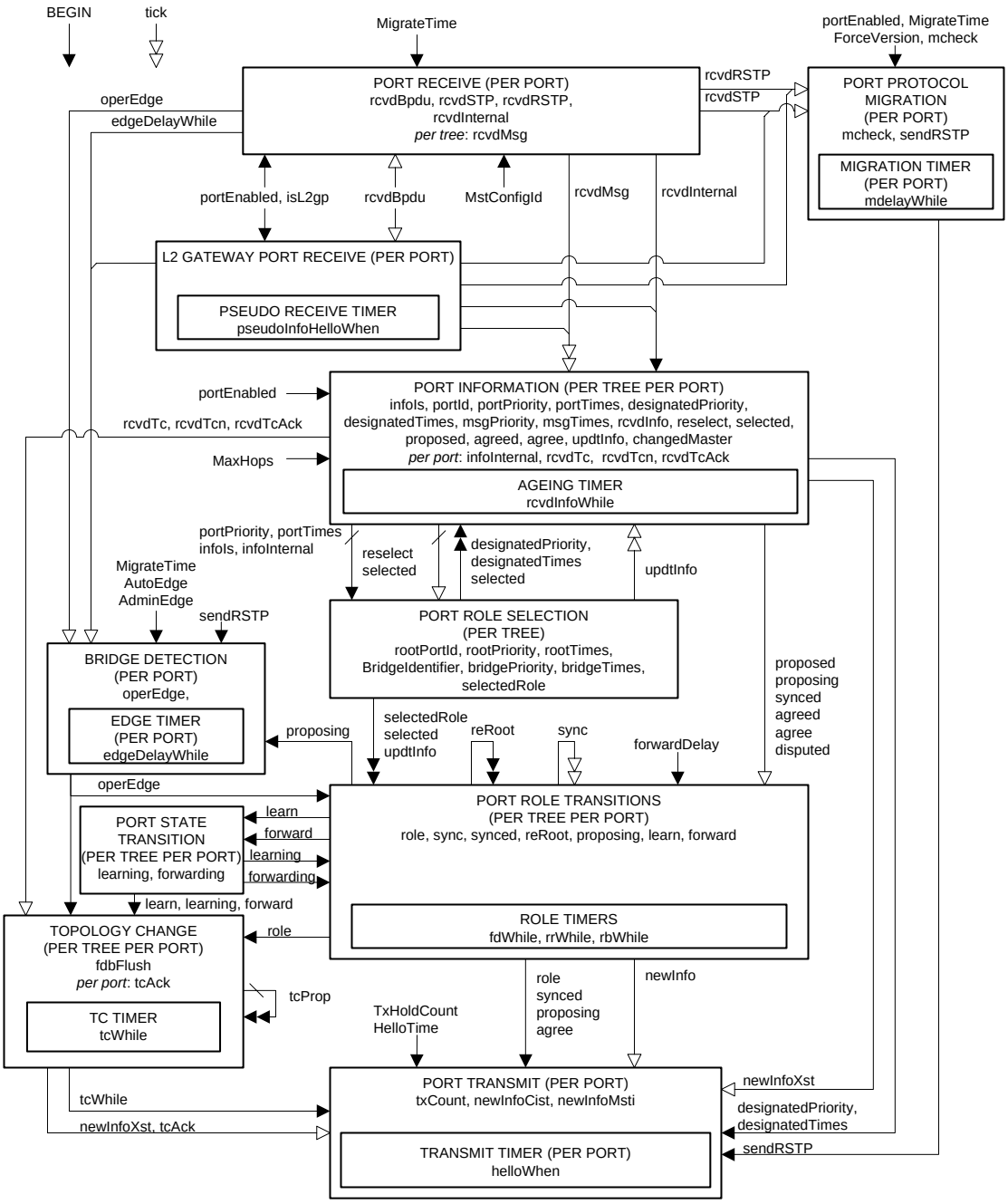


Figure 4—Changed Port Role Transitions state machine (change from 802.1Q-2006/2008)



NOTE: For convenience all timers are collected together into one state machine.

Figure 5—MSTP state machines overview (updated to show variables L2GPRX uses to signal)