# *DRAFT DISPOSITION OF BALLOT COMMENTS on P802.1d/D12: Draft Standard for Media Access Control (MAC) Bridges - Revisions*

Sponsor

**LAN MAN Standards Committee**
**of the**
**IEEE Computer Society**

Prepared by:

**Tony Jeffree, Project Editor**

11A Poplar Grove
Sale, Cheshire
M33 3AX
UK
Tel: +44-161-282-3824
Fax: +44-161-973-6534
Email: 100271.522@compuserve.com

*Commentary:*

This Draft Disposition of Ballot Comments has been prepared to document the ballot comments received in the ballot on P802.1d/D12, and to prepare the issues for discussion at the editing meeting. The document contains:

1) A table of responses received.
2) A listing of comments received, by author, each accompanied by a proposed disposition or a reference to the list of issues;

Note that for Editorial comments, the default disposition is that the Editor will take the comments into account in preparation of the next Draft. Explicit statements of how editorial comments have been resolved are only included for exceptions to this default.

3) A list of issues that were raised in ballot comments, and which require a solution to be agreed at the editing meeting.

This document, once agreed and completed, will constitute a record of the Instructions to the Editor for the preparation of P802.1d/D13.

<<Untill agreed by the Editing Meeting, any resolutions or proposed resolutions contained in this document are simply the opinion of the Editor; the contents of those resolutions is therefore subject to change.>>

**CONTENTS**

# 1. Table of Responses

The following table indicates the status of each ballot response received on the P802.1p/D3 ballot. Where comments have been received without an accompanying ballot, this is indicated by a reference only in the COMMENTS column.

**Table 1—P802.1d Ballot Analysis**

| NAME | STATUS | YES | NO | ABSTAIN | | | COMMENTS |
|---|---|---|---|---|---|---|---|
| | | | | (Time) | (Exp.) | (Other) | (Section, Page) |
| Floyd Backes | Voting | X | | | | | - |
| John Boal | Voting | X | | | | | - |
| Paul Brant | Nonvoting | | | X | | | - |
| Paul Carroll | Voting | X | | | | | - |
| Jeff Catlin | Voting | | | X | | | - |
| Steve Chan | Voting | | | X | | | - |
| Hon Wah Chin | Voting | X | | | | | - |
| Steve Cooper | Voting | X | | | | | - |
| David Delaney | Voting | X | | | | | - |
| Peter Ecclesine | Voting | X | | | | | - |
| JJ Ekstrom | Voting | X | | | | | - |
| Norm Finn | Voting | | | X | | | - |
| Paul Frantz | Voting | X | | | | | - |
| John Grinham | Voting | X | | | | | - |
| Steve Haddock | Voting | X | | | | | - |
| John Hart | Voting | X | | | | | - |
| Scott Harvell | Nonvoting | | X | | | | S12, P20 |
| Richard Hausman | Voting | X | | | | | - |
| Ariel Hendel | Voting | | | X | | | - |
| Steve Horowitz | Voting | | | X | | | - |
| Raj Jain | Nonvoting | | | X | | | - |
| Tony Jeffree | Voting | X | | | | | S3, P10 |

## Table 1—P802.1d Ballot Analysis

| NAME | STATUS | YES | NO | ABSTAIN | | | COMMENTS |
|---|---|---|---|---|---|---|---|
| | | | | (Time) | (Exp.) | (Other) | (Section, Page) |
| Hal Keen | Voting | | | X | | | |
| Keith Klamm | Voting | X | | | | | S14, P22 |
| Paul Langille | Voting | | | X | | | S15, P23 |
| Johann Lindmeyr | Voting | X | | | | | - |
| Peter Martini | Voting | | | X | | | - |
| Milan Merhar | Voting | X | | | | | - |
| Samba Murthy | Nonvoting | | | X | | | - |
| Yaron Nachman | Voting | X | | | | | - |
| Krishna Narayanaswami | Voting | | | X | | | - |
| Joerg Ottensmeyer | Voting | X | | | | | S6, P13 |
| Anand Padmanabhan | Nonvoting | | | X | | | - |
| Luc Pariseau | Voting | X | | | | | - |
| Yondav Perry | Voting | | | X | | | - |
| Gideon Prat | Voting | X | | | | | - |
| Anil Rijsinghani | Voting | X | | | | | S5, P12 |
| Doug Ruby | Voting | | | | X | | |
| Ayman Sayed | Voting | X | | | | | - |
| Mick Seaman | Voting | | X | | | | S16, P24 |
| Rich Seifert | Liaison | | X | | | | S9, P16 |
| Lee Sendelbach | Voting | X | | | | | - |
| Himanshu Shah | Voting | | X | | | | S10, P18 |
| Karl Shimada | Voting | X | | | | | |
| C. Fred Shu | Voting | | | X | | | - |
| PJ Singh | Voting | X | | | | | - |
| Rosemary Slager | Voting | X | | | | | - |
| Stuart Soloway | Voting | X | | | | | S4, P11 |

## Table 1—P802.1d Ballot Analysis

| NAME | STATUS | YES | NO | ABSTAIN | | | COMMENTS |
|---|---|---|---|---|---|---|---|
| | | | | (Time) | (Exp.) | (Other) | (Section, Page) |
| Richard Sweatt | Voting | X | | | | | - |
| Robin Tasker | Voting | X | | | | | S8, P15 |
| Dono van-Mierop | Voting | X | | | | | - |
| John Wakerly | Voting | | | X | | | - |
| Peter Wang | Voting | X | | | | | S13, P21 |
| Trevor Warwick | Liaison | | X | | | | S2, P7 |
| Alan Weissberger | Potential voting | | | X | | | - |
| Mike Witkowski | Voting | | X | | | | S7, P14 |
| Edward Wong | Potential voting | | | X | | | |
| Michael Wright | Voting | X | | | | | - |
| John Yang | Voting | | NO | | | | S11, P19 |
| Wayne Zakowski | Voting | X | | | | | - |
| ***TOTALS (Voters & Liaison only)*** | ***60 voting members of 802.1 (at start of ballot)*** | ***34*** | ***5*** | ***11*** | ***2*** | ***0*** | |
| ***PERCENTAGES (Voters & Liaisons only)*** | ***87% response (Yes, No or Abs)*** | ***87%*** | ***13%*** | | | | |

# 2. Trevor Warwick

## 2.1 Technical comments

### *Comment 1*

I am entering a Disapprove vote on this ballot because it does not address an issue that I had previously raised on D11, and I have not been notified of any subsequent discussion that took place on this topic. I include my original solution below, which in the absence of a more elegant method being found should be included in the next draft.

Bug in example spanning tree code in ISO 10038

There is one serious bug in the example "C" code in the standard. We can argue all day about exactly how closely implementors were intended to follow this code, but I think it is now a fact of life that many people used the code almost verbatim. Therefore, it is important that the bug is fixed.

The bug allows the propagation of stale information by a non-root bridge. It arises because the only mechanism in the code that checks for the root's message being too old is message_age_timer_expiry(), which is called from a single "for all ports" loop in the tick() function. This allows the following sequence of events to occur:

1) Receive BPDU from good root, on (root) port 2, with Message Age = 15, Max Age = 15. This BPDU is stored, and is valid, because it has not yet been timed out.

2) Hold timer prevents transmission of BPDU on (designated) port 1.

3) Tick() routine runs. The single loop in this routine will allow the transmission of a BPDU on port 1, using the stored value of Message Age, since the message_age_timer_expired() routine has not yet run for port 2.

4) The BPDU is transmitted on port 1, probably with Message Age = 16, Max Age = 15.

There is another sequence as follows:

1) Receive BPDU from good root, on (root) port 2, with Message Age = 15, Max Age = 15. This BPDU is stored, and is valid, because it has not yet been timed out.

2) Receive BPDU from a worse new root or worse new designated bridge on (designated) port 1.

3) Use reply() procedure to send back BPDU.

4) The BPDU is transmitted on port 1, probably with Message Age = 16, Max Age = 15.

There is nothing in the code that will stop this happening again, next time round. I have seen a situation with a circular configuration of four two-port bridges, where there was a valid spanning tree message from the true root circulating clockwise, and an invalid message from an old root circulating anti-clockwise with a continually incrementing Message Age field. All ports were forwarding...

Architecturally, I am unsure of the correct way to go about fixing this. The same problem occurs in the textual description in 4.6 onwards, because there is no formal statement about exactly how the timers are supposed to operate.

I would actually prefer to see 4.6 to 4.8 replaced with a formal state table, but I accept that there may not be much interest in doing this after five years.

Pragmatically, there are two "C" code changes that can be made to achieve a simple, though rather inelegant, fix. Firstly, the loop in tick() is split in two, so that all message_age_timer_expired() routines are called first.

Secondly, a check is added to transmit_config() to ensure that a message is never transmitted with a Message Age greater than Max Age. These changes can be used to fix broken implementations today while a better solution is found.

```
/*=====================================================================*/
/* section of transmit_config() */
      config_bpdu[port_no].max_age = bridge_info.max_age; /* (4.6.1.3.2(7))
*/
      config_bpdu[port_no].hello_time = bridge_info.hello_time;
      config_bpdu[port_no].forward_delay = bridge_info.forward_delay;
      config_bpdu[port_no].topology_change_acknowledgment
        = port_info[port_no].topology_change_acknowledge;
                                        /* (4.6.1.3.2(8))
*/
/*
** This line removed
*/
/*      port_info[port_no].topology_change_acknowledge = FALSE; */ /* TW Fix
*/
                                        /* (4.6.1.3.2(8))
*/
      config_bpdu[port_no].topology_change
        = bridge_info.topology_change;          /* (4.6.1.3.2(9))
*/
/*
** This clause added
*/
      /*
      ** TW Fix
      ** We have to be sure that we're not about to propagate a stale
message
      */
      if (config_bpdu[port_no].message_age < bridge_info.max_age)
      {
            send_config_bpdu(port_no, &config_bpdu[port_no]);
            port_info[port_no].topology_change_acknowledge = FALSE;
                                        /* (4.6.1.3.2(8))
*/
      }
/*=====================================================================*/
/* section of tick() */
  /*
  ** TW Fix
  ** Processing of message_age_timer moved out of loop, as
  ** this is part of the fix to avoid propagating stale
  ** information. It ensures that all received messages have been
  ** aged out before we do any further timer-based processing.
  */
```

```
    for (port_no = 1; port_no <= No_of_ports; port_no++)
    {
      if (message_age_timer_expired(port_no))
      {
        message_age_timer_expiry(port_no);
      }
    }
    for (port_no = 1; port_no <= No_of_ports; port_no++)
    {
      if (forward_delay_timer_expired(port_no))
      {
        forward_delay_timer_expiry(port_no);
      }
/*    if (message_age_timer_expired(port_no))
**    {
**      message_age_timer_expiry(port_no);
**    }
*/
      if (hold_timer_expired(port_no))
      {
        hold_timer_expiry(port_no);
      }
    }
/*=====================================================================*/
```

***Disposition of Comment 1***

See Issue 1.

# 3. Tony Jeffree

## 3.1 Editorial comments

### *Comment 2*

There are a number of Editor's Notes in the text (surrounded by <<>>), some of which should simply get stripped out in the next revision or prior to publication, others which need fixing. The following are in the latter category:

Introduction, Page 8 Line 15. Insert the correct reference to EIA xxxx.

Page 20, lines 42, 43. Inconsistency needs to be resolved between Table 4.2 and B.4.1.

Page 21, Line 34. If no improvements are forthcoming, this instruction should be removed.

Page 26, Lines 4, 5. Need to resolve whether or not the Annex A PICS Proforma needs a link to the Annex D SRT PICS Proforma.

Page 28, Line 12. The material in C.1 could reasonably be incorporated into clause 1.

Page 200, Line 1, 52 (C.5). If no managed object definitions are forthcoming, I suggest that the <<>> note be removed.

### *Disposition of Comment 2*

See Issue 2.

# 4. Stuart Soloway

## 4.1 Technical comments

### *Comment 3*

I have voted to approve, but I find it very strange that Trevor Warwick's apparent discovery of a bug in the algorithm does is not being addressed. Perhaps this pro-forma approval process is not the time to address it, but I hope this issue will be resolved some time in the near future.

### *Disposition of Comment 3*

See Issue 1.

# 5. Anil Rijsinghani

## 5.1 Technical comments

### *Comment 4*

Further discussion and understanding of an issue raised in mail to the exploder by Trevor regarding propagation of old spanning tree information, resulting in loops in certain conditions, would be very useful indeed. (perhaps in the form of a presentation to the WG??)

#### *Disposition of Comment 4*

See Issue 1.

# 6. Joerg Ottensmeyer

## 6.1 Technical comments

### *Comment 5*

I approve although I am not sure if the PICS conform to the latest format required for ISO standards. Since this is a revision we should check it before submitting it to ISO/IEC.

#### *Disposition of Comment 5*

My understanding is that the PICS does indeed conform to current ISO format, as it is based on the PICS in 802.1j, which Alan Chambers has developed in accordance with current ISO rules; however, I will take an action to check this out with Alan.

# 7. Mike Witkowski

## 7.1 Technical comments

### *Comment 6*

The first sentence in section 3.12.2 (line 12) reads as "The individual MAC Entity associated with each Bridge Port shall have a separate specific MAC Address." It is my understanding that many implementations of bridges and switches actually support a single MAC address for the entire device. Each port recognizes this address and forwards frames that match this address to the local LLC layer entity (e.g. the local CPU). This is done in many cases to ease the consumption of MAC addresses, especially for large port count switch devices. To my knowledge, it seems that having a single MAC address for an entire bridge/switch does not break the STP. Shouldn't this sentence be changed to something like: "The individual MAC Entity associated with each Bridge Port may have a separate specific MAC address, or it may share a MAC address with any subset of the other MAC Entities associated with the other Bridge Ports on the Bridge."

#### *Disposition of Comment 6*

See Issue 3.

### *Comment 7*

The first paragraph in section 3.12.5 (lines 49-52) references section 4.5.3 and section 5.2.5 in regards to a definition of Bridge Identifier. In reality, Bridge Identifiers are defined in section 4.5.1.3, section 4.5.3.7, and section 5.2.5.

#### *Disposition of Comment 7*

Editor to correct the references.

### *Comment 8*

Along with my comment above, the definitions of the Bridge Identifier in sections 4.5.1.3 and section 4.5.3.7 are similar but stress different aspects of the Bridge Identifier. Neither of these definitions are specific to where they sit (Configuration BPDU Parameters and Topology Change Notification BPDU, respectively). I propose that we merge the definitions so they say the same thing, then make section 4.5.3.7 just reference section 4.5.1.3 so that there is only one definition of the purpose of Bridge Identifiers.

#### *Disposition of Comment 8*

As these are distinct parameters with (currently) quite distinct descriptions, it is not clear to the Editor how the proposed merge would be achieved, or what value would be added by doing so.

# 8. Robin Tasker

## 8.1 Technical comments

There are five outstanding issues in the current ballot which need to be addressed. Assuming the default is "do nothing", I've the following comments on some of the issues.

### *Comment 9*

Page 11 (r) and (s). Not sure (r) needs to change but maybe (s) should just be silently discarded. Does anyone use 802.1 remote management (802.1B)?

#### *Disposition of Comment 9*

See Issue 4.

### *Comment 10*

Page 20 Table 4.2 and B.4.1. The values here differ. B.4.1 using 2.0s for absolute values and discusses interoperability with bridges employing recommended values with a bridge diameter of at least 3. Table 4.2 uses values of 4.0s with a maximum bridge diameter of 7. Are 4.2 and B.4.1 actually saying the same thing?

#### *Disposition of Comment 10*

See Issue 2.

## 8.2 Editorial comments

### *Comment 11*

Page 8, line 15 Building wiring infrastucture is specified in ISO/IEC 11801

Page 17, line 21 In the second sentence on the line, add the word "be" so the sentence begins "It may be the specific.....".

# 9. Rich Seifert

## 9.1 Technical comments

### *Comment 12*

In clause 2.3.3: Frame Misordering, it states:

"MA_UNITDATA.indication service primitives corresponding to MA_UNITDATA.request primitives with the same requested priority are received in the same order as the request primitives were processed. The operation of the Bridges does not misorder frames transmitted with the same user priority."

This is not strictly true. It should read:

"MA_UNITDATA.indication service primitives corresponding to MA_UNITDATA.request primitives with the same source address, destination address and requested priority are received in the same order as the request primitives were processed. The operation of the Bridges does not misorder frames between a given pair of addresses transmitted with the same user priority."

That is, frames are not necessarily received in the same order transmitted or even received by the bridge. The bridge is allowed to reorder as long as it does not reorder between a given SA-DA pairing. This allows multicasts to be reordered relative to unicasts.

#### *Disposition of Comment 12*

See Issue 5.

### *Comment 13*

In clause 2.5.1: 802.3 now supports both the use of a Length field with pad, and a Type field. The discussion in this section should be updated to reflect this. Currently, it assumes that all 802.3 frames have a length field and optional pad

#### *Disposition of Comment 13*

See Issue 2.

### *Comment 14*

In 3.5, it states: "Frames with a frame_type of user_data_frame and addressed to the Bridge Port as an end station shall be submitted to LLC. Frames submitted to LLC can also be submitted to the Learning and Forwarding Processes, as specified above. Frames addressed to a Bridge Port as an end station, and relayed to that Bridge Port from other Bridge Ports in the same Bridge by the Forwarding Process, shall also be submitted to LLC."

802.3 now specifies a MAC client OTHER THAN LLC (i.e., MAC Control), which would typically be used for flow control. Frames, in general, should be passed to the MAC Client, which may be LLC or some other entity. Since flow control is specifically targeted for use within bridges/switches, and is performed below LLC, this section (as well as 2.5.1 and 3.12.2) should reflect this architectural change.

### *Disposition of Comment 14*

See Issue 2.

## 9.2 Editorial comments

### *Comment 15*

In 3.12.5: Change: “An unique MAC address, known as the Bridge Address,” to: “A unique MAC address, known as the Bridge Address,”

3.7.1: "d) The Maximum Service Data Unit Size supported by the LAN to which the transmission Port is attached and would not be exceeded." should be: "d) The Maximum Service Data Unit Size supported by the LAN to which the transmission Port is attached would not be exceeded." (delete word "and")

In Table 3-5: The address 01-80-C2-00-00-01 has been reserved for: "IEEE 802.3x MAC Control PAUSE Operation", not "IEEE 802.3 Flow Control"

# 10. Himanshu Shah

## 10.1 Technical comments

### *Comment 16*

I agree with Rich Seifert concerns. Frame ordering should be reworded in order to align with the new limitations that we can comply to.

#### *Disposition of Comment 16*

See Issue 5.

# 11. John Yang

## 11.1 Editorial comments

### *Comment 17*

All Port States of Figure 4.3 are in Disabled state. They should be Disabled, Blocking, Listening, Learning, and Forwarding, respectively.

### *Disposition of Comment 17*

Cut & paste bug. The Editor will correct the diagram.

# 12. Scott Harvell

## 12.1 Technical comments

### *Comment 18*

While I am not yet a voting member I would like to comment on something that I have run into recently. It has to do with TCN messages. Working in the desktop switching space I have discovered something that, I believe, was not really the intended function. Anytime that a PC, directly connected to a 10baseT switch port, is powercycled it causes the switch port to go into the disabled state (physical port loses link) which means when the PC is turned back on it will transition back into the forwarding state, and send a TCN massage to the root.

This series of events causes the network to transition into topology change mode and all switches change over to short entry age, which increases the flooding in the network. All of this because someone power cycled a PC in their office. Now I work for a small company but we a seeing an average of 2.6 topology changes an hour due to this, so I can imagine what a affect this could have on a large network.

I believe that this became this way because when 802.1d was established there really was no desktop switches and end users could not really affect the port state of a switch. That is no longer the case. Currently the 802.1d spec calls for the sending of a TCN message when a port transitions into the forwarding state or into the blocking state. I would say that it would be fine to just send TCN when a port went into blocking except for one case that breaks that setup. That case is where a LAN has a port from each of two different switches connected to it. If the port that was forwarding goes into broken state then the remaining port needs to send TCN when it goes into the forwarding state. Therefore I would like to propose that the following option be added to the 802.1d spec. Please not the word OPTION here as I believe that it should not be required that this feature be supported on a backbone switch.

If a port is transitioning to the forwarding state from the disabled state that no TCN message need be sent. If a port is transitioning from the blocking state to the forwarding state then a TCN message must be sent. Of course this does mean that one must keep track of the previous state before one got into the listen-learn-forward states. This should allow the desired behavior on the extended LAN without having an unduly large number of topology changes in the network.

Comments?

### *Disposition of Comment 18*

See Issue 6.

# 13. Peter Wang

## 13.1 Technical comments

### *Comment 19*

Sec. 2.3.5, last paragraph: 'Note that the frame is completely received before it is relayed as the Frame Check Sequence (FCS) is to be calculated and the frame discarded if in error.'

- Cut-through switches today don't wait for the frame to be completely received.

#### *Disposition of Comment 19*

See Issue 2.

### *Comment 20*

Sec. 4.3.3, 3rd paragraph: '... become the Designated Port for the LAN....' - Shouldn't it be 'Designated Bridge'?

#### *Disposition of Comment 20*

See Issue 2.

## 13.2 Editorial comments

### *Comment 21*

Sec. 2.4 references ISO/IEC 10039. - I thought this document was renumbered to 15802-1?

# 14. Keith Klamm

## 14.1 Technical comments

### *Comment 22*

Minor Technical- Page 34, line 12: Item c) in this list of conformance statements appears to relate to the former section 3.7.2 "LLC Duplicate Address Check". Should item c) be removed since section 3.7.2 was removed?

#### *Disposition of Comment 22*

True. Editor to remove this item from the list.

## 14.2 Editorial comments

### *Comment 23*

Minor Editorial- Page 63, line 44: I think it would be clearer for the reference to "table 5" be changed to "table 3.5".

# 15. Paul Langille

## 15.1 Technical comments

### *Comment 24*

I think we should consider adding the changes proposed by Trevor Warwick to address the Spanning Tree issues he raised.

#### *Disposition of Comment 24*

See Issue 1.

### *Comment 25*

I suggest we investigate the Spanning Tree issue raised by Scott Harvell. If there is a `simple' solution then we should add it into this version.

#### *Disposition of Comment 25*

See Issue 6.

# 16. Mick Seaman

## 16.1 Technical comments

### *Comment 26*

We need to address the points raised by Trevor Warwick re: spanning tree messages which should be aged out on receipt or transmit.

#### *Disposition of Comment 26*

See Issue 1.

# 17. Summary of Issues

The following list summarizes the major issues raised as ballot comments. Each issue is cross-referenced to the ballot comment(s) that raised it.

## *Issue 1. Bug in Spanning Tree code*

Summary of points made:

a) There is a bug that allows the propagation of stale information by a non-root bridge (Comment 1);
b) Need to investigate/address the bug reported by Trevor Warwick (Comment 3, Comment 4, Comment 24, Comment 26)

### *Resolution of Issue 1*

### *Issue 2. Editorial corrections*

Summary of points made:

a) Various editor's notes need to be resolved (Comment 2);
b) 2.5.1 needs to reflect the fact that 802.3 supports Ethernet-style frames (Comment 13);
c) 3.5 needs to reflect the fact that not all users of MAC services are LLC entities (Comment 14);
d) 2.3.5 needs to reflect reality of cut-through operation (Comment 19);
e) 4.3.3, 3rd paragraph - Shouldn't it be 'Designated Bridge'? (Comment 20)

### *Resolution of Issue 2*

### *Issue 3. Addressing Bridge Ports*

Summary of points made:

a) The standard should not mandate a separate MAC address for each Port (Comment 6)

### *Resolution of Issue 3*

### *Issue 4. Use of 802.1 Management*

Summary of points made:

a) Remove conformance reference to 802.1B (Comment 9);
b) Need to resolve apparent discrepancy between Table 4.2 and B.4.1 (Comment 10)

### *Resolution of Issue 4*

### *Issue 5. Rules for frame (mis)ordering*

Summary of points made:

a) Current rule is incorrect - should only guarantee to maintain ordering for a given priority and SA/DA pairing (Comment 12, Comment 16)

### *Resolution of Issue 5*

### *Issue 6. Generation of TCN*

Summary of points made:

a) Generation of TCN should be optional on transition from Disabled -> Forwarding (Comment 18, Comment 25)

### *Resolution of Issue 6*