*Cl* **00** *SC* **0** *P* *L* # 4

Finn, Norman Cisco Systems

*Comment Type* **E** *Comment Status* **X**

All of my comments with regard to the use of the PD requested power value, PSE allocated power value, and reduced operation PD power value reduce to a lack of clarity of what this protocol can and cannot do, along with the assumption of request/ACK operation, which is not needed. Following are the fundamental facts that must be understood about *any* power negotiation protocol in this environment. These must be understood before looking at the protocol details, and very much need to be stated explicitly in the document, so that the reader understands the goals of the protocol.

1. The PSE has the final say-so about how much power the PD *SHOULD BE* using, because it (or the management protocol that drives it) has the overall view of the network and understands the operators' intentions.

2. The PSE has the final say-so about how much power it *IS* using.

3. If the PSE's final say-so on what the PD should be using disagrees with the PD's actual use, then:

a. If the PSE doesn't like how much power the PD is using, the PSE must choose whether to live with the situation or shut down the power to the PD entirely. (It is not at all clear that taking this drastic step is something that this protocol should define, e.g. by a time out. One can argue that it is sufficient to report the situation to the network administrator, and leave the shut-off to management action, whether programmatic or manual.)

b. If the PD doesn't like its allocation from the PSE, there is nothing it can do except complain to the network administrator (if its power allocation permits!).

4. The PSE's initial state must be that which was negotiated by the hardware.

5. The only reason for the PSE to initiate a change in the power level a PD is using is that it wants the PD to use *LESS* power. Unless the PD is asking for more power, there is no point in offering it.

6. The PD may ask for more power, to serve a user's desire, or for less power, to be a good citizen.

7. In order to protect against a hardware failure affecting multiple PDs, a PSE can cut power to any PD that either claims (threatens) to, or actually does, draw more than its allocated power.

*SuggestedRemedy*

Include the basic facts of negotiation, points 1-7, in the text, of course subject to adjustment by the editor.

*Proposed Response* *Response Status* **O**

*Cl* **00** *SC* **0** *P* *L* # 14

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

The method of interoperability between "new power TLV" implementations and "old power TLV" implementations is completely lacking, except for the "don't transmit both" injunction in 33.6. As mentioned in another comment, this is a serious flaw in the draft.

At present, the draft demands either a forklift upgrade of all systems, configuration in one system of the old/new capabilities of the neighboring system, or non-standard, unspecified, and therefore non-interoperable actions by the different implementations.

At a minimum, the interoperability scenarios between 802.3at-capable and 802.1AB-2004-capable systems must be defined, if 802.3at is to be successful. A non-normative appendix describing how 802.3at relates to the extremely limited capabilities of the widely-deployed TIA TR41 LLDP-MED standard would be very useful, and relatively easy to generate.

*SuggestedRemedy*

Given the suggestion for combining the 802.1AB power TLV and the 802.3at power TLV contained in my comment #1, .3at power can be combined fairly easily with .1AB power. When an 802.3at PSE implementation is receiving only the 802.1AB-2004 power TLV from the PD, it uses the power class field from the old TLV and Table 33-10 of 802.3af, instead of the (new) PD requested power value field, to determine the value for aMirroredDLLPDRequestedPowerValue, and otherwise uses the new state machines. Similarly, a PD uses the (old) power class field and 802.3af Table 33-10 to determine aMirroredDLLPSEAllocatedPowerValue. aMirroredLostCommunication is never set.

(There may be other ways to remedy this issue.)

*Proposed Response* *Response Status* **O**

*Cl* **00** *SC* **0** *P* *L* # 3

Finn, Norman Cisco Systems

*Comment Type* **E** *Comment Status* **X**

The PDF document's properties do not contain proper values for the document's title or author. (On the other hand, thanks to the editor for making the romand and arabic page numbers match correctly, and for the quantity of cross-references, often including variable names.)

*SuggestedRemedy*

This can be remedied using FrameMaker's File, Properties pull-down menu item in the .book file, after selecting top-level book, itself, in the window.

*Proposed Response* *Response Status* **O**

*Cl* **30** *SC* **30.9.1.1.22** *P* **29** *L* **38** # 10

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

(Also 30.9.2.1.12, p33 line 2) In both cases, aMirroredLostCommunication returns the number of times the remote system has lost communications. Since this is not conveyed in the LLDP TLV (the TLV conveys only a single bit in the loss of communications field) there is no way to obtain this information.

*SuggestedRemedy*

Either of the following two remedies will satisfy this comment:

1. Change the definition of the loss of communications field to match 30.9.2.1.12.

2. Delete aMirroredLostCommunication and loss_of_comms. (See my Comment #15.)

*Proposed Response* *Response Status* **O**

*Cl* **30** *SC* **30.9.2.1.13** *P* **33** *L* **12** # 17

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

The definition of aDLLPDResponseTime states that aDLLPDRequestedPowerValue (the transmnitted PD requested power value field) is updated from the received aReceivedDLLPSEAllocatedPowerValue (the received PSE allocated power value). This is unnecessary, it denies a useful feature, and can lead to an infinite loop.

It is unnecessary, because the number transmitted by the PD in the PSE allocated power value properly reflects the PD's understanding of what the PSE wants it to do.

It denies a useful feature, and complicates the protocol, as follows. The fact that the PSE cannot or will not allocate what the PD wants does *not* change what the PD wants. It changes what the PD *gets*. If the PD changes its "want" to match the "allocated", then it raises the question of when to ask again for more power, how often it can ask, how many times it should ask to make sure the PSE knows it has asked, etc., etc. The protocol is much simpler, more useful, and the timer aDLLPDResponseTime can be eliminated, if the PD's wants do not reflect the allocated power.

It can lead to an infinite loop, because the protocol, as defined, has a circular chain of causality, which is a very fundamental flaw in any protocol. For example, if the PD requests a higher value for power at the same time the PSE informs it that it should change to a lower value. If the PSE and PD both respond (as the state machines say they can), then they flip-flop back and forth, wasting time and resources. This requires Yet Another Timer and/or random delay to resolve. Again, you are setting yourself a problem and having to solve it.

*SuggestedRemedy*

See the slide presentation from Anoop Vitteth.

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **33.6** *P* **100** *L* **5** # 15

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

As mentioned in my comment #1 regarding interoperability between 802.1AB-2004 and 802.3at implementations of the Power TLVs, 802.1AB unfortunately failed to specify that all reserved fields in transmitted TLVs shall contain 0, and all reserved fields in received TLVs shall be ignored. This has the consequence of limiting the options for .1AB/.3at interoperability, now. This mistake should not be repeated.

*SuggestedRemedy*

State somewhere, either in 33.6 or in a subclause thereof, that all reserved fields in transmitted TLVs shall contain 0, and all reserved fields in received TLVs shall be ignored.

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **33.6.2** *P* **100** *L* **48** # 6

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

The goals of protocol revision control are:

1. To allow new versions of the protocol to be introduced without requiring all communicating systems to be upgraded simultaneously.

2. To leave no ambiguities in the proper behavior of systems when implementations supporting different versions communicate.

3. To never require an implementation to transmit multiple versions of the same PDU.

(See IEEE 802.1ag-2007 subclause 20.46 for a full explanation of a set of techniques that meet these goals.)

The cited paragraph satisfies 3, at the (unacceptable) cost of violating one or both of the first two.

Unless the TG is very confident that the IEEE 802.1AB-2005 power TLV was not implemented, interoperability with systems that only know the old TLV is important.

The new power TLV seems to supersede the old power TLV in Draft 3.1. The paragraph at line 48 states that, "when the DTE Power via MDI classification TLV is being transmitted, the Power via MDI TLV shall not be transmitted." This statement makes the protocol unusable, because there is no means specified for a system to decide which TLV to transmit. The choice cannot be left as an exercise by the implementor, or interoperability will suffer. So, what obvious choices are possible?

Something fairly simple, like "Start sending the new, switch to the old if you receive the old" does not work. To see why, consider the case of a PD with software in ROM that knows the old TLV. Suppose that after booting, it downloads software that knows the new TLV. Since the PSE doesn't know about the reboot, it is very easy to get into a mode where the two devices exchange LLDPDUs more or less simultaneously, forever out of sync as to which TLV to use.

As pointed out in the text, sending both TLVs is not a good option, because it is wasteful of a very scarce resource (LLDPDU TLV space), especially for IP telephones.

The trivial choice of configuring which TLV to send is unacceptable. LLDP is a discovery protocol. Requiring proper configuration at both ends in order for LLDP to perform correctly is a fundamental violation of its reason for existing.

*SuggestedRemedy*

The usual 802.1 plan, which would simply extend the existing TLV, is one option. This solution places all of the new information immediately following the old information, using the old TLV's subtype. The total length of the Value part of the TLV is then the sum of the old and new Value lengths. A new implementation sends both kinds of information, but listens to only the new information. An old implementation, of course, pays attention to only the old information.

This solution will work, because 802.1AB-2005 subclause 10.3.2.1 point b requires old implementations to ingore the extra bytes in the TLV that carry the new information. This solution would have extra bytes in the TLV, but it interoperates correctly, and requires no extra state machines.

*Proposed Response* *Response Status* **O**

---

*Cl* **33** *SC* **33.6.2.1** *P* **101** *L* **36** # 5

Finn, Norman Cisco Systems

*Comment Type* **T** *Comment Status* **X**

This field and the Loss of communication field (33.6.2.4, p103, line 10) should be combined. There is no need for wasting bits, because the TLV size can be increased in future revisions of the standard. (Old implementations are required to not care if extra bytes are added to a TLV by a new rev of the standard.)

*SuggestedRemedy*

Delete the Loss of communication field. Place the loss of communication bit in bit 3 (or bit 2) of the Power type/source/priority field. (This comment is simplified if either the loss of communication field is deleted, or is irrelevant, if the loss of communication field is changed from a bit to a counter.)

*Proposed Response* *Response Status* **O**

---

*Cl* **33** *SC* **33.6.2.4** *P* **103** *L* **12** # 16

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

The loss of communication bit seems unnecessary, because the PSE or PD should not need to know whether the other side sees their LLDPDUs and/or power TLVs.

If the PD's LLDPDUs are not being received by the PSE, then the PSE's transmitted allocated power value field will not change from its last value, whether that came from a received LLDPDU or from the hardware negotiation.

If the PSE's LLDPDUs are not being received by the PD, then the allocated power value field transmitted by the PD will not change from its last value, whether that came from a received LLDPDU or from the PD's knowledge of its hardware-requested power level.

Defining the use of the fields in this way, and particularly their initial values (obtained from the hardware negotiation), eliminates much of the complexity of the state machines in Figure 33-30 and 33-31, and elminiates the need either for a loss of communication bit, loss of communication state variables.

Note that, as mentioned in my Comment #6, resetting a brain dead PD can be done by detecting the reception, followed by the loss of reception, of the PD's LLDP PDUs (not the power negotiation TLV). That still does not require the loss of communication field in the TLV, nor for that matter, does it need to be a feature of 803.3at.

*SuggestedRemedy*

Make the suggested changes.

*Proposed Response* *Response Status* **O**

---

*Cl* **33** *SC* **33.6.2.4** *P* **103** *L* **3** # 12

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

The phrase, "the device believes it has lost communication with the far end" lacks sufficient precision to implement interoperably. Perhaps the correct phrase is, "loss_of_comms = FALSE".

*SuggestedRemedy*

Provide a precise definition in terms of state machine variables and/or attributes. (Better yet, delete the notion of loss of communication. See my Comment #15.)

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **33.6.2.6** *P* **103** *L* **25** # 7

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

The PD model number field as defined in 33.6.2.6 is neither necessary, sufficient, safe, nor in practice, useful, to accomplish any purpose suggested by the text or by the name of the field.

The field is not necessary, because TIA T.R. 41 LLDP-MED standard defines a globally unique vendor / model number combination. The LLDP-MED has the same uniqueness properties as the one defined by subclause 30.9.2.1.14. Furthermore, the uses of a system's model number are not correlated with PSE/PD power. The model number may or may not be of utility to power negotiation (see below, "useful"). The model number may well be of utility beyond power negotiation, e.g. for selecting the right icon in a management display. In addition, the PSE's model number can be equally informative to the PD.

The two-byte field is not sufficient, because there is no means specified for determining the "implementor" that defines the meaning of the PD model number field. As mentioned in the note in 33.6.2.6, two different implementors can use the same PD model number, with totally different meanings behind those numbers. This makes interoperable use of this field, based on this standard alone, impossible.

The 2-byte field is not safe, in that one company could deliberately choose to use a model number that conflicts with another company's number, in order to inhibit interoperability and/or initiate legal battles. The large, globally unique field is not safe because the standard does not define how the receiving side is to use the field. In the absence of that definition, a vendor could define its use, protect that use via patents, and claim that use is both conformant to the standard, and not covered by the fair and non-descriminatory rules of the IEEE 802 patent policy.

The field is not practically useful, in that the introduction of any new model powered device to a network requires the updating of the PSEs' PD model number tables. While the updating of the PSEs is typically managed by the network administrators, the addition of PDs can be almost entirely out of control. Many of the members of 802.3 are familiar, as consumers, with the problem of home electronics devices purchased after the purchase of a "universal remote controller" containing an out-of-date list of other vendors' model numbers.

To sum up, the 2-byte field defined in 33.6.2.6 is clearly broken, and must be removed. A large field containing the model number defined in 30.9.2.1.14 is not related solely to power negotiation, is redundant to that specified by TR41, has insufficient semantics to supply interoperability, amd so should be removed.

*SuggestedRemedy*

Two possible remedies:

1. Delete the PD model number field from the TLV.

2. Update Figure 33-29 and 33.6.2.2 to agree with the text of 30.9.2.1.14, which defines a globally unique model number, send the system's model number, whether a PSE or a PD, and define *exactly* how it is used on the receiving end.

Either remedy will satisfy this comment, but I much prefer #1. The LLDP-MED model number is still available for those who want to use it for proprietary purposes.

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **33.7** *P* **111** *L* **1** # 2

Finn, Norman Cisco Systems

*Comment Type* **E** *Comment Status* **X**

"Loss of management frame communication" is an unfortunate choice of words. The term, "management frame" could cover a very large territory, including:

* SNMP over UDP over IP management queries and responses.

* Bridge Protocol Data Units (BPDUs)

The proper term is either, "LLDPDUs" as defined in 802.1AB, or "DTE Power via MDI classification TLVs".

*SuggestedRemedy*

Replace "management frame" with "LLDPDU". See also my Comment 15. Changing it to "DTE Power via MDI classification TLVs" would be done only if my Comment 6 is rejected.

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **33.7** *P* **111** *L* **16** # 1

Finn, Norman Cisco Systems

*Comment Type* **E** *Comment Status* **X**

The statement, "If a loss of management frame communication is asserted and persists for a time duration ..., a PSE may remove power." is semantically equivalent to, "A PD shall transmit LLDPDUs containing the DTE Power via MDI classification TLV forever." This appears at first glance to be in direct conflict with subclause 33.6, which states that, "Type 2 PDs that require more than 12.95 W must support Data Link Layer classification (see 33.3.5). Data Link Layer classification is optional for all other devices." If a PSE implementation takes advantage of the "may" and requires LLDP, and a PD implementation takes advantage of the "optional" and is unable to send them, then those two standard-conformant devices are non-interoperable.

It is possible (I have not participated in the debates in the TG) that the intention of the "may" in 33.7 and the variable pse_power_cycles that controls it is to reset a PD that has gone "brain dead", and that the even occurs only if the PD a) transmits LLDP + Power TLV, and then b) stops. In that case, the "may" in 33.7 still seems inappropriate; the operator "can" set pse_power_cycles either to true or to false, in which case the implementation "shall" do whatever the state machines say to do, given the state of pse_power_cycles. At least, in 802.1 parlance, "may" is reserved for an implementation decision, made perhaps via outside-the-standard controls.

In this latter case, the detection of loss of connection (but not the loss of connection field in the TLV) is useful, and should be retained, in spite of my Comment #15.

*SuggestedRemedy*

Pick one:

1. Make it clear that pse_power_cycles is intended to turn on "reset on brain death" mode in the PSE, and preferably, point out that this reset is not triggered if the PD never sends LLDP. Definitely point out that a management action on the PD to turn off LLDP can result in the PSE removing power and thus resetting the device. (In which case, this is largely an Editorial, instead of Technical, comment.)

2. Remove permission for the PSE to remove power if a loss of management frame communication is asserted from 33.7.

See also my Comment 15.

*Proposed Response* *Response Status* **W**

No Comment Type, set to 'E' as a default

*Cl* **33** *SC* **33.7** *P* **111** *L* **3** # 9

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

No initial value for the loss of communications field is defined. No means of specifying when or how it is reset is defined.

*SuggestedRemedy*

Either:

1. Define the bit's initial value, specify when to reset it, and specify how it is used in the receivers' state machines. (I suspect this is a matter of specifying the relationship between the variable "loss_of_comms" and the transmitted field value.)

2. Delete the loss of communication bit from the TLV.

I prefer solution 2. Note that deleting the bit from the TLV does not in iteself require deleting the notion of loss of communication from the state machines. (But see also my Comment #15.)

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **33.7** *P* **111** *L* **3** # 13

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

No distinction is made between loss of LLDPDUs and loss of the DTE Power via MDI classification TLV in those LLDPDUs. The assumption seems to be made that, if loss_of_comms is true (meaning that the LLDPDUs are being received) that the DTE Power via MDI classification TLV is being received. That is not a valid assumption.

If my other comments are accepted, only the loss of LLDPDUs is relevant, and only for resetting a brain-dead PD. See my Comment #6.

*SuggestedRemedy*

Describe what happens when the DTE Power via MDI classification TLV is gained or lost, perhaps by including lack of the DTE Power via MDI classification TLV in "loss of management frames", or perhaps by distinguishing the two events. See also my comments 6 and 15.

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **Figure 33-30** *P* **107** *L* **11** # 8

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

Neither "loss_bit" nor "LOSS" is defined in this document.  Same problem in Figure 33-31, p108, line 9.  Does "TRUE" in Figure 33-31 mean the same as "LOSS" in Figure 33-30?

*SuggestedRemedy*

Either change the state machine diagram to reflect the proper variable and value, or define "loss_bit" and "LOSS".  (Better yet, follow my comment #15 and delete loss of communication detection.)

*Proposed Response* *Response Status* **O**

*Cl* **33** *SC* **Table 33-29** *P* **106** *L* **27** # 11

Finn, Norman Cisco Systems

*Comment Type* **TR** *Comment Status* **X**

Table 33-29 is nowhere referenced in the text.  More specifically, the mapping from the attributes aMirroredLostCommunication and aLostCommunication, both of which are counters, to the variable loss_of_comms, which is a Boolean, is not defined.  Given that loss_of_comms is reset by the state machines, it is not clear how this mapping would work.

*SuggestedRemedy*

Define the mapping of aMirroredLostCommunication and aLostCommunication to loss_of_comms, including additional state machines and/or variables, if required.  See also my Comment #15.

*Proposed Response* *Response Status* **O**