1. Introduction

This document proposes editing instructions to add the DCBX related TLVs to P802.1AB-REV/D4.0. These additions affect Annex E, Organizationally Specific TLVs. The clause numbers in this document correspond to the clauses to be edited in P8021AB-REV/D4.0.

This proposal covered in this document is intended to provide identical functionality to that proposed in the CEE Author's DCBX proposal. As such, it contains certain functionality (or lack thereof) which the author finds inadvisable. As such, this proposal does not reflect the author's position on many of these items. Several of these have been highlighted in Editor's Notes.

More importantly, this document does not in anyway reflect the consensus position of the DCB group.

The intent of this document is to establish that the functionality provided in the CEE Author's DCBX proposal may be achieved directly over LLDP without the protocol and feature state machines proposed in CEE Author's DCBX, resulting in a much simpler, testable, and interoperable protocol.

The CEE Author's DCBX proposal essentially creates a new protocol that operates over LLDP. The author believes that we have a responsibility to attempt to use existing IEEE protocols to meet the needs of CN, ETS, and PFC. If and only if it is not practical or reasonable to do so are we justified in inventing new protocols.

It is hoped that this will drive consensus to operate DCBX directly over LLDP at which point the details of the operation of the individual TLVs may be worked out.

Comments and suggestions are always welcomed!

Joe Pelissier
jopeliss@cisco.com

Joe Pelissier, Cisco
February 9, 2009

E.1 Requirements of the IEEE 802.1 Organizationally Specific TLV set

Change the Note to read:

NOTE- These TLVs are intended to be used by IEEE 802.1 Bridges, end stations, or both, as indicated in Table E-1.

Change 3rd paragraph to read:

The IEEE 802.1 Organizationally Specific TLVs are divided into groups as indicated in Table E-1. If any 802.1 Organizationally Specific TLV within a group is supported, all 802.1 Organizationally Specific TLVs within that group shall be supported. All IEEE 802.1 Organizationally Specific TLVs shall conform to the LLDPDU bit and octet ordering conventions of 9.1.

Update Table E-1 As indicated below:

Table E-1— IEEE 802.1 Organizationally Specific TLVs

IEEE 802.1 subtype	TLV name	Reference	Group	Used by
01	Port VLAN ID	E.2	Base	Bridges
02	Port And Protocol VLAN ID	E.3	Base	Bridges
03	VLAN Name	E.4	Base	Bridges
04	Protocol Identity	E.5	Base	Bridges
05	VID Usage Digest	E.6	Base	Bridges
06	Management VID	E.7	Base	Bridges
07	Link Aggregation	E.8	Base	Bridges
08	Traffic Classes Supported	E.9	DCBX	Bridges & end stations
09	Priority Group	E.10	DCBX	Bridges & end stations
10	Priority-based Flow Control	E.11	DCBX	Bridges & end stations
11	Application Priority	E.12	DCBX	Bridges & end stations
12	Congestion Notification	E.13	DCBX	Bridges & end stations

E.9 Traffic Classes Supported

Insert this clause after the existing clause E.8 and renumber existing clauses E.9 onward appropriately.

The Traffic Classes TLV is an optional fixed length TLV that allows a bridge port or end station port to advertise the number of traffic classes it supports for transmission.

Figure E-8 shows the Traffic Classes Supported TLV:

TLV type = 127	TLV Information string length = 5	802.1 OUI 00-80-C2	802.1 subtype = 8	Reserved	Traffic Classes Supported
7 bits	9 bits	3 octets	1 octet	5 bits	3 bits
← TLV Header →		← TLV information string →			

Figure E-8—Traffic Classes Supported TLV format

The value in the Traffic Classes Supported field indicates the number of traffic classes supported by the originating port. A zero in this field indicates the originating port supports eight traffic classes.

Editors Note - It is exceedingly unclear that there is value in sending this parameter from the end station to the bridge. A bridge has no specific knowledge of the applications running on the end station and therefore no rational way to make a judgement as to whether the number of traffic classes supported is sufficient. Furthermore, even if it could make such a judgement, what would it do with the information? It seems unreasonable that it would prevent the end station from connecting to the fabric.

Editors note - It is exceedingly unclear that there is any value in a bridge sending this information to an end station. The end station would have no basis on which to judge whether the number of supported traffic classes is sufficient. For example, if the end station is connected with a 10G link to a bridge with a 10G uplink, then the link to the end station would never saturate and the number of traffic classes supported is irrelevant.

Editors note - unless the usefulness of this TLV can be established, it should be removed.

E.10 Priority Group TLV

Insert this clause after the new clause E.9 above and renumber the existing clauses appropriately.

The Priority Group TLV is an optional fixed length TLV that allows a bridge port or end station port to advertise a priority group configuration and bandwidth allocation configuration. See {add reference to ETS} for specifics on how these parameters are used.

Figure E-9 shows the Priority Group TLV:

TLV type = 127	TLV Information string length = 17	802.1 OUI 00-80-C2	802.1 subtype = 9	Reserved	Willing	Priority Group BW Table	Priority Group Assignment Table
7 bits	9 bits	3 octets	1 octet	7 bits	1 bit	8 octets	4 octets
← TLV Header →		← TLV information string →					

Figure E-9—Priority Group TLV format

The Priority Group BW Table provides a bandwidth percentage to be specified for each priority group.

Figure E-10 shows the format of the Priority Group BW Table:

Priority Group 0 Weight	Priority Group 1 Weight	Priority Group 2 Weight	Priority Group 3 Weight	Priority Group 4 Weight	Priority Group 5 Weight	Priority Group 6 Weight	Priority Group 7 Weight
4 bits	4 bits	4 bits	4 bits	4 bits	4 bits	4 bits	4-bits

Figure E-10—Priority Group BW Table format

The Priority Group Assignment Table provides the assignment of priorities to Priority Groups. A value between 0 and 7 indicated the Priority group to which the priority is assigned. A value of 15 indicates the priority is not part of a priority group. Values between eight and 14 are not permitted.

Figure E-11 shows the format of the Priority Group Assignment Table:

Pri. 0 Group Assignment	Pri.1 Group Assignment	Pri. 2 Group Assignment	Pri. 3 Group Assignment	Pri. 4 Group Assignment	Pri. 5 Group Assignment	Pri. 6 Group Assignment	Pri. 7 Group Assignment
4 bits	4 bits	4 bits	4 bits	4 bits	4 bits	4 bits	4-bits

Figure E-11—Priority Group Assignment Table

The Willing bit provides a mechanism for a port to obtain these configuration parameters from its peer port. If a port wishes to obtain these parameters from its peer port, it shall set to the Willing bit to one; otherwise, it shall set to willing bit to zero. If a port receives a Priority Group TLV from its peer with the Willing bit set to zero and the receiving port has set the Willing bit to one in the Priority Group TLV it transmits, then it shall configure its priority to Priority Group assignments and its Priority Group Bandwidth percentages to match that received in the Priority Group TLV. If the Willing bit is set to one in the received Priority Group TLV or the Willing bit is set to zero in the Priority Group TLV being transmitted, then the port shall not update its priority to Priority Group assignments and its Priority Group Bandwidth percentages to match that received in the Priority Group TLV (this does not imply that either end is required to change its configuration if the configurations happened to match prior to reception of the TLV).

Note - the above restriction is required to prevent a endless cycle of each port updating its Priority-based Flow Control based on that of its peer.

Editor's Note - This carries over two restrictions form the CEE author's version of DCBX that are probably unnecessary. Specifically, if the mechanism is used to configure the these parameters, both ends of the link are forced to have the same configuration. Given that the bandwidth consumed by most traffic types is not symmetric, it may well be desirable that the bandwidth percentages be programmed differently on each end of the link. This limitation can be overcome by sending two sets of the two tables. The first set contains the local settings. The second set contains the desired setting for the far end.

Editor's Note - The above mechanism also would enable removal of the restriction that the passing of this configuration can only occur in one direction. It is reasonable to assume in many cases that each end of the link would know what would be the best configuration of these parameters for the data sent to it. The above mechanism would allow the Willing bit to be dispensed with and the assignments could operate in both directions.

Editor's Note - If the above mechanism is implemented, it is not completely clear that there is value in passing the local configuration. Further investigation is necessary.

E.11 Priority-based Flow Control TLV

Insert this clause after the new clause E.10 above and renumber the existing clauses appropriately.

The Priority-based Flow Control TLV is an optional fixed length TLV that allows a bridge port or end station port to advertise for each priority whether or not Priority-based Flow Control is enabled.

Figure E-12 shows the format of the Priority-based Flow Control TLV:

TLV type = 127	TLV Information string length = 6	802.1 OUI 00-80-C2	802.1 subtype = 10	Willing	Reserved	Number of PFC TCs	PFC Enabled Mask
7 bits	9 bits	3 octets	1 octet	1 bit	4 bits	3 bits	1 octet
◄── TLV Header ──►		◄── TLV information string ──►					

Figure E-12—Priority-based Flow Control TLV format

Each bit in the PFC Enabled Mask corresponds to one priority. The least significant bit refers to priority 0, and each increasing order bit corresponds to the next greater priority. A one in the bit position indicates that Priority-based Flow Control is enabled for the corresponding priority, a zero indicates that it is not.

The Number of PFC TCs field indicates the number of Traffic Classes that may simultaneously support Priority Flow Control. Setting this field to zero indicates that all available traffic classes may simultaneously support Priority Flow Control.

The Willing bit provides a mechanism for a port to obtain the configuration of Priority-based Flow Control from its peer port. If a port wishes to obtain this configuration from its peer port, it shall set to the Willing bit to one; otherwise, it shall set to willing bit to zero. If a port receives a Priority-based Flow Control TLV from its peer with the Willing bit set to zero and the receiving port has set the Willing bit to one in the Priority Group TLV it transmits, then it shall configure its Priority-based Flow Control to match that received in the Priority-based Flow Control TLV. If the Willing bit is set to one in the received Priority-based Flow Control TLV or the Willing bit is set to zero in the Priority-based Flow Control TLV being transmitted, then the port shall not update its priority Priority-based Flow Control configuration to match that received in the Priority-based Flow Control TLV (this does not imply that either end is required to change its configuration if the configurations happened to match prior to reception of the TLV).

Note - the above restriction is required to prevent a endless cycle of each port updating its Priority-based Flow Control based on that of its peer.

E.12 Application Priority TLV

Insert this clause after the new clause E.11 above and renumber the existing clauses appropriately.

The Application Priority TLV is an optional variable length TLV that allows a port to advertise a priority over which a particular protocol should operate.

Figure E-13 shows the format of the Application Priority TLV:

TLV type = 127	TLV Information string length	802.1 OUI 00-80-C2	802.1 subtype = 11	Reserved	Willing	Protocol Table
7 bits	9 bits	3 octets	1 octet	7 bits	1 bit	variable
◄── TLV Header ──►		◄── TLV information string ──►				

Figure E-13—Application Priority TLV format

The Protocol Table contains one or more Protocol Entries. Each Protocol Entry identifies a protocol and the priority over which the protocol should operate.

Figure E-14 shows the format of the Precool Entry:

Sel	Reserved	Priority	Protocol ID
3 bits	4 bits	3 bits	2 octets

Figure E-14—Priority Entry format

The protocol to which this TLV refers is indicated either by a Ethertype Value or a Port number. A list of Ethertypes may be obtained from the IEEE at http://standards.ieee.org/regauth/ethertype/eth.txt. A list of assigned port numbers may be obtained from IANA at www.iana.org/assignments/port-number. The Sel Field identifies to which protocol the Protocol ID field applies as indicated in Table E-2.

Table E-2—Sel Field Interpretation

SEL	Protocol ID Interpretation
0	Protocol Entry applies to the protocol identified by the Ethertype in Protocol ID.
1	Protocol Entry to the Protocol identified by the Port Number in Protocol ID when operating over UDP.
2	Protocol Entry to the Protocol identified by the Port Number in Protocol ID when operating over TCP.
3	Protocol Entry to the Protocol identified by the Port Number in Protocol ID when operating over TCP or UDP.
4	Protocol Entry to the Protocol identified by the Port Number in Protocol ID when operating over neither UDP nor TCP.
5-7	Reserved for future use - Entries containing these values shall be ignored.

The Willing bit provides a mechanism for a port to obtain the protocol priority assignments from its peer port. If a port wishes to obtain this configuration from its peer port, it shall set to the Willing bit to one; otherwise, it shall set to willing bit to zero. If a port receives a Application Priority TLV from its peer with the Willing bit set to zero and the receiving port has set the Willing bit to one in the Application Priority TLV it transmits, then it shall configure its Application Priorities to match that received in the Priority-based Flow Control TLV. If the Willing bit is set to one in the received Application Priority TLV or the Willing bit is set to zero in the Application Priority TLV being transmitted, then the port shall not update its application priorities configuration to match that received in the Application Priority TLV (this does not imply that either end is required to change its configuration if the configurations happened to match prior to reception of the TLV).

Note - the above restriction is required to prevent a endless cycle of each port updating its Priority-based Flow Control based on that of its peer.

Editor’s note - It is not clear that it ever makes sense for an end station to attempt to set these priority assignments for a bridge (i.e. end station not willing, bridge willing). This TLV should probably be reworked such that only an end station may set its priorities based on receiving this TLV from a bridge.

E.13 Congestion Notification TLV

Insert this clause after the new clause E.12 above and renumber the existing clauses appropriately.

The Congestion Notification TLV is an optional fixed length TLV that allows a bridge port or end station port to advertise which priorities are operating congestion notification and to control the defense of those priorities.

Figure E-15 shows the format of the Congestion Notification TLV:

TLV type = 127	TLV Information string length = 6	802.1 OUI 00-80-C2	802.1 subtype = 12	CN Mask	Ready Mask
7 bits	9 bits	3 octets	1 octet	1 octet	1 octet
TLV Header		TLV information string			

Figure E-15—Congestion Notification TLV format

The CN Mask field indicates on which priorities Congestion Notification is enabled. The least significant bit refers to priority 0, and each increasing order bit corresponds to the next greater priority. A one in the bit position indicates that Congestion Notification is enabled for the corresponding priority; a zero indicates that it is not.

Devices that support Congestion Notification may implement defenses that block traffic on priorities that have congestion notification enabled or remap that traffic to other priorities until it is confirmed that the peer port also has Congestion Notification enabled for that priority. The Ready Mask indicates which priorities are ready to receive traffic with Congestion Notification disabled (i.e. the receive defenses are disabled). The least significant bit refers to priority 0, and each increasing order bit corresponds to the next greater priority. Bit positions in Ready Mask are reserved for which the corresponding bit in the CN Mask field are set to zero. Otherwise, a one in the bit position indicates that the port is ready to receive Congestion Notification enabled traffic; a zero indicates that it is not. See {802.1au} for a complete description of the use of these TLVs their normative controlling state machine.