DCBX History Analysis

Motivation

- DCBX issues have been pointed out in January Interim Meeting
  - Presentation “DCBX PFC Configuration”
    https://www.ieee802.org/1/files/public/docs2024/dt-seaman-dcbx-pfc-config-1124-v02.pdf
  - Major issues
    - ETS tied to PFC?
    - Ambiguity of ‘PFC enable’
    - Conformance of DCBX
    - DCBX state machine
- Reviewing DCBX development history helps
  - Know the original goal of the designs
  - Understand how the issues appear
  - Evaluate the proposed solutions

DCBX Goals:
Configuration of Link Parameters for DCB Functions

- DCB Capability eXchange Protocol is responsible for configuration of link parameters for DCB functions
- It includes
  - A protocol to exchange (send and receive) DCB parameters between peers
  - Set local "operational" parameters based received DCB parameters
  - Resolve conflicting parameters

https://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf

Goal

- Discover DCB capability
- Detect misconfiguration
- Set operational configuration → **"Negotiation" of DCB features**

2.1 Goals

DCBX base spec-2008

The following lists the goals of DCBX.

Discovery of DCB capability in a peer: DCBX is used to know about the capabilities of the peer device. It is a means to know if the peer device supports a particular feature such as Priority Groups (PG) or Priority-based Flow Control (PFC). For example, it can be used to determine if two link peer devices support PFC.

DCB feature misconfiguration detection: DCBX can be used to detect misconfiguration of a feature between the peers on a link. Misconfiguration detection is feature-specific because some features may allow asymmetric configuration.

Peer configuration of DCB features: DCBX can be used by a device to perform configuration of DCB features in its link peer. The goal is to provide basic peer to peer configuration through DCBX in the initial version. Future versions of DCBX or another higher layer application can build on top of this to provide more complex configuration distribution mechanisms.

https://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbx-capability-exchange-discovery-protocol-1108-v1.01.pdf

802.1Q-2022

The goals of DCBX are as follows:

a) Discovery of DCB capability in a peer port; for example, it can be used to determine if two link peer ports support PFC.

b) DCB feature misconfiguration detection: DCBX can be used to detect misconfiguration of a feature between the peers on a link. Misconfiguration detection is feature-specific because some features allow asymmetric configuration.

c) Peer configuration of DCB features: DCBX can be used by a device to perform configuration of DCB features in its peer port if the peer port is willing to accept configuration.

DCBX Works for DCB Features ETS and PFC

Here's the initial version of ETS and PFC parameters

https://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf

Priority Groups Tables

Parameter	Syntax	Range	Default Value	Access (RO,RW, NA)	Scope	Description
Priority Group (PG) Allocation	Table					
PG ID (index)	Integer	0..7		RW	Exchanged	Queue bandwidth group
PG Percentage	Integer	0..100		RW	Exchanged	Percentage of link bandwidth
Strict Priority	Integer	0..2		RW	Exchanged	Strict priority settings: 0 – no strict priority 1 –Strict Priority
User Priority Allocation	Table					
Priority (index)	Integer	0..7		RW	Exchanged	
PG ID	Integer	0..7		RW	Exchanged	BWG to which the priority belongs

PFC Table

Parameter	Syntax	Range	Default Value	Access (RO,RW, NA)	Scope	Description
PFC Config	Table					
User Priority (index)	Integer	0..7		RW	Exchanged	
Admin mode	Integer	0..1	0	RW	Exchanged	Administrative PFC mode. 0: Disabled 1: Enabled PFC Enabled means that flow control in both directions (Rx and Tx) is enabled.

'PFC Enabled' means that flow control in both directions (Rx and Tx) is enabled, although TX and RX do not have to be tied togeterh.

There once was a proposal to converge ETS and PFC. But it was not pursued.

A new term: Traffic Type

- **There are five different Traffic Types that must be considered:**
  - **AVB: see 802.1Qav**
  - **EP: ETS with PFC enabled**
  - **En: ETS without PFC enabled**
  - **nP: No ETS with PFC enabled**
  - **nn: everything else (i.e. non-ETS and non-PFC)**

A few proposed simplifications:

- **Disallow nP (i.e. PFC without ETS)**
  - **Allowing PFC on a non-bandwidth managed basis seems dangerous**
- **Require all Priorities within a Priority Group to all be either PFC or not PFC (no mixing of PFC and non-PFC within a Priority Group)**

https://www.ieee802.org/1/files/public/docs2008/az-pelissier-convergence-proposal-0708.pdf

The author later found the proposal was problematic.
https://www.ieee802.org/1/files/public/docs2009/az-pelissier-convergence-proposal-0309.pdf

Another contribution also disagreed the 'convergence', believing it would impose restrictions on product use cases.
https://www.ieee802.org/1/files/public/docs2008/az-carlson-ets-pfc-discussion-0908-v0.ppt

Stage1: DCBX is Mandatory for DCB Feature

DCBX is used to negotiate parameters for ETS, PFC and application priotity. Feature should be off on both sides if they do not have consistent configuration.

https://www.ieee802.org/1/files/public/docs2008/az-pelissier-dcbx-thoughts-1208.pdf

- **The current DCBX proposal provides for negotiation of the following parameters:**

Priority group parameters including:
Number of traffic classes supported
Priority group bandwidth percentage
Priority to priority group assignment
ETS

Priority Flow Control parameters including:
Number of traffic classes that support priority flow control
Priority flow control enabled per priority
PFC

Protocol parameters including:
Priority assignment for the protocol
Application Priority

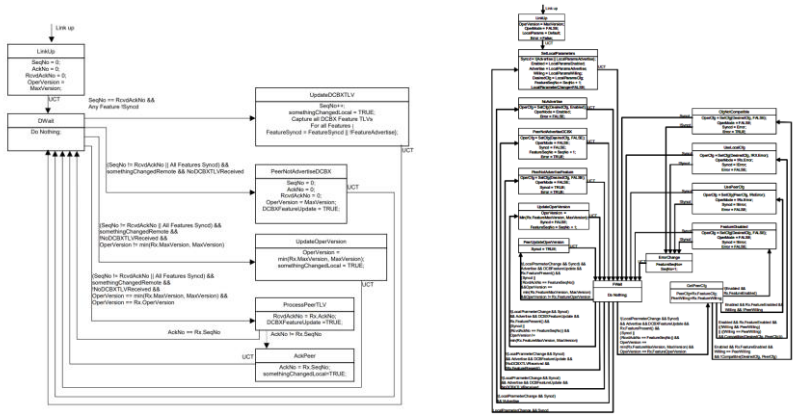
https://www.ieee802.org/1/files/public/docs2009/az-multanen-dcbx-req-0212-v01.pdf

Result: Baseline DCBX Proposal

The required goals and services of DCBX, plus the general principle that it would be "bad" to have a misconfigured DCB feature on a link:
- Feature is "on" only if both sides have a consistent configuration
- Otherwise feature should be off on both sides

DCBX is an acknowledged protocol using LLDP, including control state machine and feature state machine.

- Control state machine handles 2 peers in sync by exchanging LLDPDUs -- 'SeqNo''AckNo'
- Feature state machine runs on top of the control SM, handling local operational configuration by comparing and synchronizing with peer -- 'Enable''Willing''Error'…



"**A device capable of any DCB feature must have DCBX enabled by default with an option for DCBX to be administratively disabled.**" -- DCBX Base Spec 1.0

→ "**A port capable of any DCB feature shall have the capability for DCBX to be administratively disabled. The default state for DCBX is enabled.**" -- 802.1Q-2022

Stage 2: Simplify DCBX Protocol (1/2)

Not all DCB feature TLVs need negotiatiton

- **ETS bandwidth configuration is asymmetirc** https://www.ieee802.org/1/files/public/docs2009/az-pelissier-dcbx-bandwidth-0309v2.pdf
  - Configuration TLV
  - Recommendation TLV

⇨ Asymmetric attributes

- **'PFC enable' is negotiable parameter, while 'PFC cap' is not**
  https://www.ieee802.org/1/files/public/docs2009/az-pelissier-dcbxtlvs-0309.pdf

- Provides negotiation and information of PFC enabled / disabled per priority
- PFC Cap indicates the device's limitation of how many priorities may simultaneously support PFC (not negotiated).
- Utilizes Parameter Acceptance Framework
- PFC enable has 8 bits (one per priority)
  - A one indicates PFC is enabled on the priority
  - A zero indicates that PFC is disabled on the priority
  - Local policy in each end of the link decides whether to use the priority if the configuration does not match

'PFC enable' needs negotiation

'PFC cap' does not need negotiation

Local policy decides how to handle mismatch

⇨ **PFC Configuration TLV follows Symmetric attributes passing. (Actually 'PFC enable' is, but 'PFC cap' is not negotiable parameters)**

- **Application Priority TLV does not need negotiation** https://www.ieee802.org/1/files/public/docs2009/az-pelissier-dcbx-tlv-thoughts-0209.pdf
  - **Application Priority TLV:**

    It is not clear that there is a valid use case to negotiate this from end station to bridge or bridge to bridge

⇨ Informational attributes

Stage 2: Simplify DCBX Protocol (2/2)

Simplified state machine

- **Rule of 'willing'** : W bit is used in Asymmetric attributes passing and Symmetric attibutes passing.

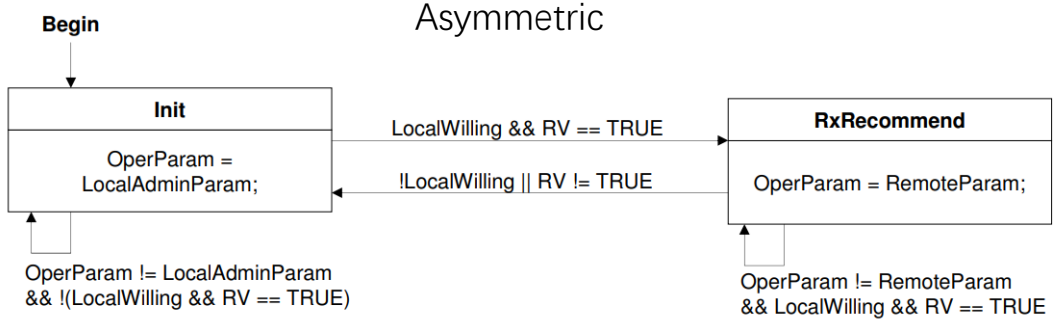
1) If peer willing=0 and local willing=1, configure local operational state to match peer
2) If peer willing=1 or local willing=0, local operational state is not changed

https://www.ieee802.org/1/files/public/docs2009/az-pelissier-dcbx-simplified-0209v2.pdf

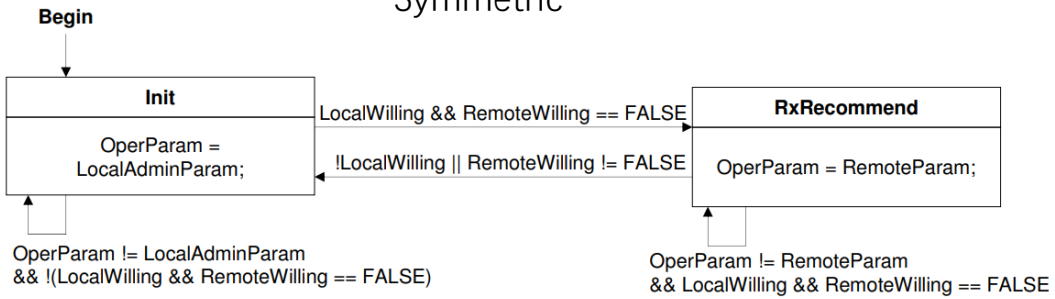
- **Rudimentary form of today's DCBX state machines**

https://www.ieee802.org/1/files/public/docs2009/az-pelissier-dcbx-framework-0509.pdf

Asymmetric



Symmetric



Conside both sides are 'willing' in later comment resolution

The priciple to design SMs (negotiation, no acknowledgement)

Note: RV can take on three values: TRUE, FALSE, and NULL. NULL indicates that the Recommendation TLV was not part of the last LLDP PDU received, or that no LLDP PDUs have been received.

Pending status only for symmetric attributes passing

The DCBX Exchange Status has 2 basic values:

1. Pending – a Feature TLV with different information than currently known is expected from the peer
2. Done – the Feature TLV information received from the peer (including no TLV) is final. (Won't change except for administrative action on the peer.)

https://www.ieee802.org/1/files/public/docs2009/az-multanen-dcbx-xchg-status-v01.pdf

```
Pending = RemoteParam == NULL
          || !LocalWilling && RemoteWilling == TRUE && OperParam != RemoteParam;
```

Stage 3: Optimize DCBX by Comment Resolutions

DCBX is Mandatory for PFC and ETS

- **DCBX is mandatory for PFC and ETS**
  - "Configuration of ETS and PFC is done using DCBX which is based on LLDP"

Cl **17** *SC* **17.2.14** *P* **10** *L* **19** # 2

Dan Romascanu None entered

Comment Type **ER** *Comment Status* **D** *Done*

There is a fondamental design decision taken by the authors to use an extension of the LLDP MIB as the principal mean to configure BW and priority on a DCBX bridge. This needs some more justification and description - maybe in this section, maybe in some other part of the document.

SuggestedRemedy

add description of operational flow and remote configuration using LLDP MIB extension

Proposed Response *Response Status* **W**

PROPOSED ACCEPT IN PRINCIPLE. State something along the lines of: "Configuration of ETS and PFC is done using DCBX which is based on LLDP. For this reason the MIB is an LLDP extension."

- **DCBX is mandatory for ETS**
  - "DCBX is required for ports using ETS"
  - "To A.31 add DCBX as mandatory for ETS."

Cl **12** *SC* **12.22** *P* **9** *L* **19** # 48

Ben Mack-Crane None entered

Comment Type **TR** *Comment Status* **A** *Done*

Where is the MIB for the ETS objects described in clause 12.22? The comment resolution from the last ballot indicates that the DCBX MIB is required to configure ETS. Does this imply that no configuration for ETS is provided unless DCBX is supported? Since DCBX is optional this would seem to be an undesirable dependence. Furthermore, it is not clear that the DCBX MIB allows all the necessary parameters to be configured (e.g., ETSEnable).

SuggestedRemedy

Add MIB for ETS configuration. For the DCBX MIB follow the conventions of previous LLDP MIBs (i.e., the LLDP MIB allows values exchanged by LLDP to be read but is not the vehicle for configuration of bridge port parameters except those directly related to LLDP itself).

Response *Response Status* **C**

ACCEPT IN PRINCIPLE. Add a requrement stating that DCBX is required for ports using ETS.

To A.31 add DCBX as mandatory for ETS. Also, need to add text to body of document that states this.

Remove "ETSEnable" from 12.22 (and Table 12-6). Also, remove ETSEnable from 37.3.

Stage 3: Optimize DCBX by Comment Resolutions

DCBX Conformance

5.4.1.6 Enhanced Transmission Selection bridge requirements

A device supporting ETS shall:

a) Support at least 3 traffic classes (see 37.3);

NOTE —A minimum of 3 traffic classes allows a minimum configuration such that one traffic class contains priorities with PFC enabled, one traffic class contains priorities with PFC disabled, and one traffic class using strict priority.

b) Support bandwidth configuration with a granularity of 1% or finer (see 37.3);
c) Support bandwidth allocation with a precision of 10% (see 37.3);
d) Support a transmission selection policy such that if one of the traffic classes does not consume its allocated bandwidth, then any unused bandwidth is available to other traffic classes (see 37.3); and
e) Support DCBX (see Clause 38).

Insert the following new subclause 5.4.1.7.

5.4.1.7 DCBX bridge requirements

A device supporting DCBX shall:

a) Support LLDP transmit and receive mode (see IEEE Std. 802.1AB).
a) Support the DCBX ETS Configuration TLV (see D.2.9).
b) Support the ETS Recommendation TLV (see D.2.10).
c) Support the Priority-based Flow Control Configuration TLV (see D.2.11).
d) Support the Application Priority TLV (see D.2.12).
e) Support the asymmetric and symmetric DCBX state machines (see 38.4).

A.5 Major capabilities

Insert the following at the end of Table A.5

ETS	Does the implementation support bandwidth management using ETS?	O	37	Yes [ ] No [ ]
DCBX	Does the implementation support configuration management via DCBX?	O	38	Yes [ ] No [ ]

A.32 DCBX

DCBX-1	Support LLDP	DCBX:M	IEEE Std. 802.1AB	Yes [ ]
DCBX-2	Support the DCBX ETS Configuration TLV	DCBX:M	D.2.9	Yes [ ]
DCBX-3	Support the ETS Recommendation TLV	DCBX:M	D.2.10	Yes [ ]
DCBX-4	Support the Priority-based Flow Control Configuration TLV	DCBX:M	D.2.11	Yes [ ]
DCBX-5	Support the Application Priority TLV	DCBX:M	D.2.12	Yes [ ]
DCBX-6	Support the DCBX asymmetric state machine	DCBX:M	38.4.1	Yes [ ]
DCBX-7	Support the DCBX symmetric state machine	DCBX:M	38.4.2	Yes [ ]

Stage 3: Optimize DCBX by Comment Resolutions

State Machine is Based on LLDP

- **DCBX is based on LLDP, so LLDP takes care of parameters' update and expiration.**

"What we need to say is that the feature state machine is invoked if that feature's TLV is present."

"DCBX state machines are invoked when a remote MIB DCBX TLV changes or ages out"

"DCBX state machine transitions are based on the DCBX objects in the LLDP MIB module. Operation of the DCBX state machine may affect the values of the DCBX objects in the LLDP MIB module"

- **Consider both sides are willing**
  - For asymmetric attribute passing (recommendation value), each side adopts the other's recommendation.
  - For symmetric attribute passing, the side with lower numerical MAC address is adopted.
- **Consider remote MIB DCBX TLV ages out**
  - Adding condition 'RemoteParam == NULL' when transition from RxRecommend to Init

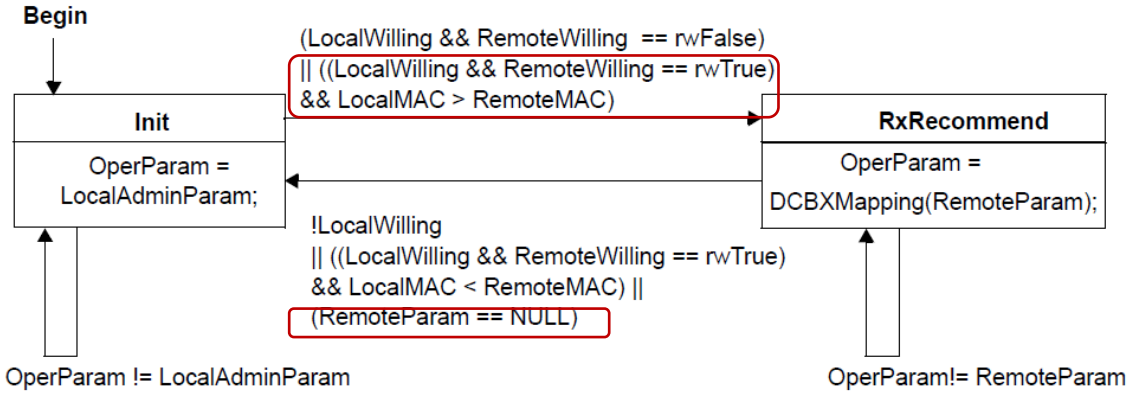


Figure 38-2—Symmetric state machine

Conclusions of the Historical Findings

- DCBX was designed to be a protocol covering all DCB features configuration.
  - DCB features include ETS, PFC, Application priority, CN(was removed)

→ There is confusion regarding the relationship between PFC and ETS.

- DCBX 'negotiates' peer sides operational configurations
  - Application priority: informational --- same as LLDP
  - ETS bandwidth: asymmetric --- provide recommendation
  - PFC enable: symmetric --- target for same configuration

→ Mandatory for PFC/ETS, but the simplified state machines are confusing (pending state, oper/admin config…)

- 'PFC enable' means both Tx and Rx are enabled.

→ Not necessary to be tied, but implementation already follows it.

Expected DCBX

Based on current spec and today’ s implementation, the expected DCBX behavior is as follows.

- DCBX is an optional feature.
- DCBX is mandatory for PFC.
  - Not only mandatory to be implemented, but also mandatory to be used
- DCBX is used to exchange **admin configuration (TBD)** of PFC peer sides
  - Each peer deduces the other peer’s operational configuration
- It might be a short period during which peer sides do not have a synchronized states, that may cause packet loss of PFC storm
  - DCBX does not have a handshake procedure.
  - PFC Tx and Rx are tied together.

Other considerations:

- PFC is not tied to ETS
- ‘PFC Enable’ means both Tx and Rx are enabled