

# P802.1Qdt Text Contribution

Lily Lv (Huawei)

March-2022

## Contents

1. Overview .....	2
1.3 Introduction.....	2
2. Normative references .....	3
5. Conformance .....	4
5.11 System requirements for Priority-based Flow Control (PFC).....	4
6. Support of the MAC Service .....	5
6.7.3 Support of MACsec protection on PFC frames .....	5
36. Priority-based Flow Control (PFC).....	7
36.1 PFC operation .....	7
36.1.1 Overview .....	7
36.3 Automatic PFC headroom calculation .....	7
36.3.1 Medium delay measurement.....	8
36.3.2. Internal processing delay calculation .....	9
36.3.3 PFC headroom calculation .....	9
38. Data Center Bridging eXchange protocol (DCBX) .....	9
38.2 Goals .....	9
D.2.10 Priority-based Flow Control Configuration TLV .....	9
D.5 IEEE 802.1/LLDP extension MIB .....	10
D.6 IEEE 802.1/LLDP extension YANG .....	10
Annex M - Support for PFC in link layers without MAC Control .....	10

<<This amendment is based on IEEE Std 802.1Q-Rev-d1-02>>

## **1. Overview**

### **1.3 Introduction**

*Insert the following items after item bh)*

bi) Defines a means for two participating systems to automatically calculate the minimum buffer requirements to assure lossless operation.

bk) Defines a means for MACsec protection of MAC control frames

## 2. Normative references

*Insert the following reference in the appropriate collating sequence:*

IEEE Std 1588: IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems

## **5. Conformance**

### **5.4.1.7 DCBX Bridge requirements**

*Insert the following items after item h)*

h) Support automatic PFC buffer requirement configuration (36.3)

### **5.11 System requirements for Priority-based Flow Control (PFC)**

*Insert the following items after item h)*

i) Support automatic configuration of PFC buffer requirements for lossless operation.

## 6. Support of the MAC Service

### Insert new subclause 6.7.3

#### 6.7.3 Support of MACsec protection on PFC frames

PFC functionality generates and processes MAC control primitives. MACsec functionality generates and processes MAC service primitives. In order to protect PFC frames with MACsec, it is necessary to provide a shim layer that converts the PFC MAC control primitives to MAC service primitives. As shown in figure aaa, after converting the control primitives from the upper layer PFC function to corresponding MAC service primitives, the MAC service primitives are sent to the MACsec function for encryption. Upon reception of an encrypted PFC frame, the reverse is performed. The decrypted MAC service primitives are converted to MAC control primitives and submit to upper layer PFC function.

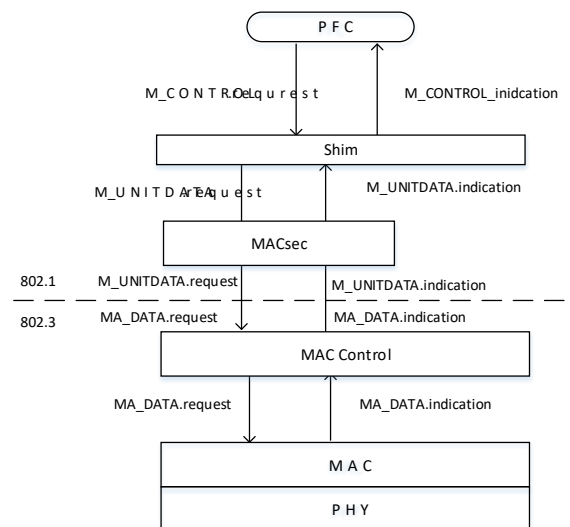


Figure aaa-- MACsec enable PFC frame transmission

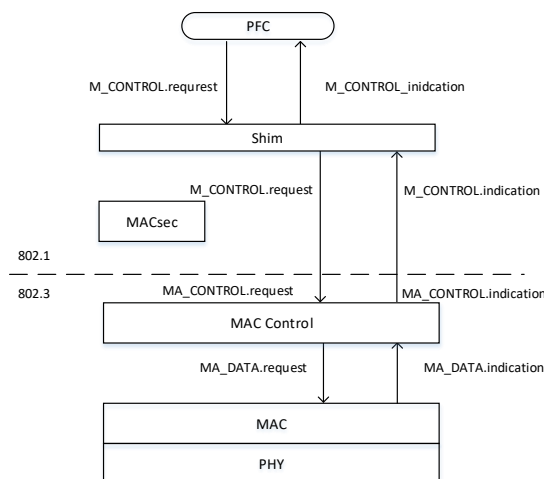


Figure bbb --- MACsec disabled PFC frame transmission

The shim function is transparent when MACsec is not enabled, as shown in figure bbb. When the shim function receives an M\_CONTROL.request primitive and MACsec is not enabled, it transparently passes the primitive to the underlying MAC control interface. When MACsec is enabled, the shim function generates a corresponding M\_UNITDATA.request to the underlying MAC service. The parameters of M\_CONTROL.request primitive are destination\_address, opcode,

and request\_operand\_list. The parameters of M\_UNITDATA.request primitive are destination\_address, source\_address, mac\_service\_data\_unit, priority, drop\_eligible, frame\_check\_sequence, service\_access\_point\_identifier, connection\_identifier.

— The destination\_address parameter is passed unaltered

— The opcode (2 octets) and request\_operand\_list (18 octets) are combined as mac\_service\_data\_unit

—Generate parameters of source\_address, frame\_check\_sequence priority, drop\_eligible, service\_access\_point\_identifier, and connection\_identifier.

- Priority parameter is set to 7, to give the MAC control frame highest priority for processing.
- Drop\_eligible parameter is set to FALSE.
- Service\_access\_point\_identifier and connection\_identifier parameters are set to NULL.

When the shim function receives an M\_UNITDATA.indication primitive from the underlying MAC service, it reads the destination\_address parameter. If destination\_address parameter is not 01-80-C2-00-00-01, the shim transparently passes the primitive to the upper layer MAC service; otherwise it is a MAC control frame, the shim converts M\_UNITDATA.indication to a corresponding M\_CONTROL.indication. The parameters of M\_UNITDATA.indication are destination\_address, source\_address, mac\_service\_data\_unit, priority, drop\_eligible, frame\_check\_sequence, service\_access\_point\_identifier, connection\_identifier. The parameters of M\_CONTROL.indication primitive are opcode, and indication\_operand\_list.

— The parameters of destination\_address, source\_address, priority, drop\_eligible, frame\_check\_sequence, service\_access\_point\_identifier, connection\_identifier are dropped.

— The mac\_service\_data\_unit parameter is parsed to extract opcode parameter and indication\_operand\_list parameter.

- The first 2 octets is opcode parameter.
- The following 18 octets is operand\_list parameter.

## 36. Priority-based Flow Control (PFC)

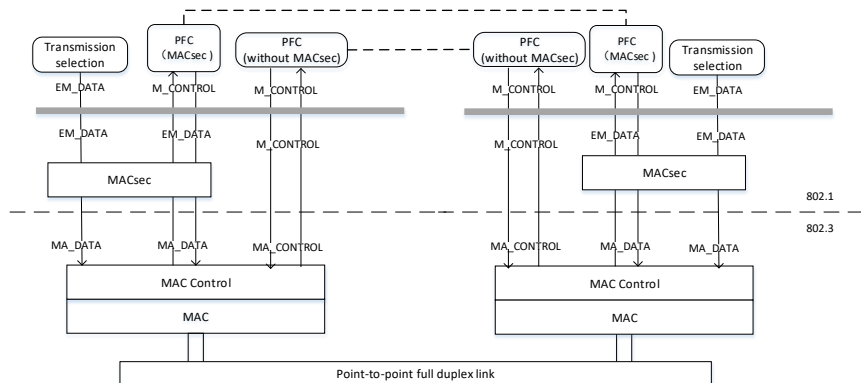
*Modify the description as following.*

This clause specifies the operation of PFC (see 36.1), ~~and~~ the architecture of Priority-based Flow Control in a PFC-aware system (see 36.2) and the automatic PFC headroom calculation (see 36.3).

### 36.1 PFC operation

#### 36.1.1 Overview

*Replace figure 36-1 with the following figure*



**Figure 36-1—PFC peering**

*Insert new paragraph at the end of this subclause:*

PFC is intended to be used on full-duplex links. When PFC is invoked, there is a time delay between the PFC invocation on the PFC initiator and the pause action on the PFC receiver. In order to guarantee no data frames are dropped by the PFC initiator, a certain amount of buffer needs to be available at the PFC initiator to absorb the data in flight after the PFC frame has been transmitted. The reserved buffer space is also known as PFC headroom. A method to automatically calculate the headroom is specified in subclause 36.3.

*Insert new subclause 36.3 and its subclauses and tables, as shown, re-numbering as necessary.*

### 36.3 Automatic PFC headroom calculation

Automatic PFC headroom calculation provides a method to configure the minimum amount of buffer space required on the PFC initiator to guarantee no packet loss when using PFC.

Automatic PFC headroom calculation follows a worst-case delay model to determine the headroom requirement (see figure N-3 of Annex N). The calculation considers the time between the PFC frame invocations by the PFC initiator, until the last bit of the PFC frame is received by the PFC receiver. The total delay value formula is specified in Annex as below.

$$DV = 2 \times (\text{Max Frame}) + (\text{PFC Frame}) + 2 \times (\text{Cable Delay}) + \text{TXd}_{s1} + \text{RXd}_{s2} + \text{HD}_{s2} + \text{TXd}_{s2} + \text{RXd}_{s1}$$

Cable delay is the propagation delay over the transmission medium.

TXds1 and RXds1 are the interface delay of PFC initiator. RXds2 and TXds2 are the interface delay of PFC receiver. Interface delay is specified N.3.

HDs2 is higher layer delay of PFC receiver. Higher layer delay is specified in N.4.

The total delay value can be divided into medium delay, internal processing delay and fixed delay, shown in figure xxx.

$$\text{Delay Value} = \underbrace{2 \times (\text{Cable Delay})}_{\text{Medium delay}} + \underbrace{\text{TXds1} + \text{RXds2} + \text{HDs2} + \text{TXds2} + \text{RXds1}}_{\text{Internal processing delay}} + \underbrace{2 \times (\text{Max Frame}) + (\text{PFC Frame})}_{\text{Fixed delay}}$$

Figure xxx --- Delay value formula

In the figure xxx, medium delay depends on the deployment environment. A measurement mechanism is described in clause 36.3.1. Internal processing delay is vendor specific, comprises interface delay and higher layer delay. The value of internal processing delay is calculated using the mechanism described in clause 36.3.2. Fixed delay equals to length of time to transmit 2 times maximum frame and PFC frame. With medium delay, internal processing delay and fixed delay, clause 36.3.3 describes the calculation of PFC headroom.

### 36.3.1 Medium delay measurement

Medium delay is the time of a full-duplex point-to-point round trip transmission. If it is symmetric point-to-point link, medium delay is 2 times cable delay. The measurement uses the peer-to-peer delay mechanism shown in figure yyy. It is the same as the mechanism specified in IEEE Std 1588-2019, supporting both one-step procedure and two-step procedure.

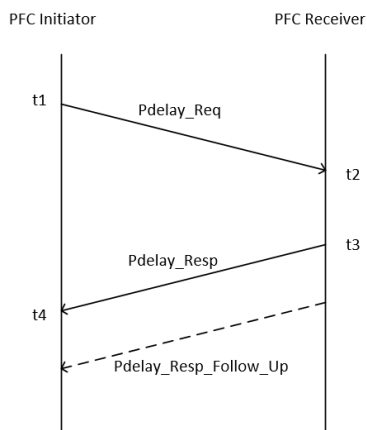


Figure yyy --- Medium delay measurement

For one-step procedure,

- PFC initiator issues a Pdelay\_Req message and generates a timestamp, t1.
- PFC receiver generates a timestamp, t2, upon receipt of the Pdelay\_Req message.
- Upon receipt of Pdelay\_Req message, PFC receiver issues a Pdelay\_Resp message and generate a timestamp, t3.  
Pdelay\_Resp message conveys the difference between the timestamp t2 and t3.
- PFC initiator generates a timestamp, t4, upon receipt of the Pdelay\_Resp

For two-step procedure,

- PFC initiator issues a Pdelay\_Req message and generates a timestamp, t1.



- b) PFC receiver generates a timestamp, t2, upon receipt of the Pdelay\_Req message.
- c) Upon receipt of Pdelay\_Req message, PFC receiver issues a Pdelay\_Resp message and generate a timestamp, t3.  
Pdelay\_Resp message conveys the timestamp t2.
- d) PFC receiver issues a Pdelay\_Resp\_Follow\_Up message  
Pdelay\_Resp\_Follow\_Up message conveys the timestamp t3.
- e) PFC initiator generates a timestamp, t4, upon receipt of the Pdelay\_Resp

PFC initiator uses these 4 timestamps to compute medium delay.

Medium delay (MD) =  $t4 - t1 - (t3 - t2)$

### 36.3.2. Internal processing delay calculation

Shown in figure xxx, the total value of internal processing delay equals to the sum of PFC initiator interface delay including TXds1, RXds1, and PFC receiver interface delay including TXds2, RXds2 and HDs2. The values are implementation specific. Although some MAC interfaces, such as IEEE Std 802.3 specify the maximum value of the interface delay, implementations always are much smaller. Both PFC initiator and PFC receiver obtain their own internal processing delays. PFC receiver conveys its internal processing delay to PFC initiator using DCBX (see clause 38). The calculation of total internal processing delay is done at PFC initiator. After receiving PFC receiver's internal processing delay, PFC initiator adds the received value and its internal processing delay, to get the total value of internal processing delay.

<<Note -- Consider to add 'reference plane' description.>>

### 36.3.3 PFC headroom calculation

The calculation of PFC headroom takes place at the PFC initiator. PFC initiator gets the delay value by adding medium delay (see subclause 36.3.1), internal processing delay (see subclause 36.3.2) and fixed delay.

Besides delay value, PFC headroom calculation needs a correction coefficient. That is to adjust the accuracy considering implementation specific impact, such as internal buffer fragmentation. So the headroom calculation formula is illustrated as below.

PFC headroom = Delay value \* alpha

alpha is implementation specific coefficient.

## 38. Data Center Bridging eXchange protocol (DCBX)

### 38.2 Goals

#### *Modify item a) as following*

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, and if automatic PFC headroom configuration is supported.

### D.2.10 Priority-based Flow Control Configuration TLV

*Replace figure D-10 with the following figure.*

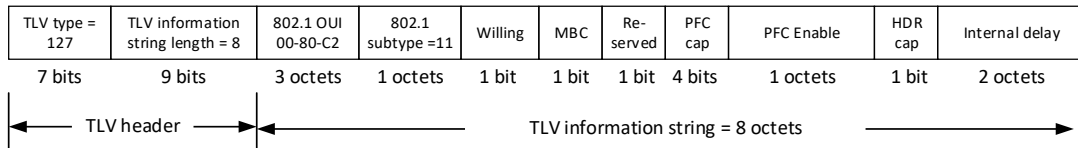


Figure D-10 Priority-based Flow Control Configuration TLV format

### D.2.10.2 TLV information string length

A 9-bit unsigned integer, occupying the LSB of the first octet of the TLV (the MSB of the TLV information string length) and the entire second octet of the TLV, containing the total number of octets in the TLV information string of the Priority-based Flow Control Configuration TLV. This does not count the TLV type and TLV information string length fields. It is equal to 68.

### D.2.10.x HDR cap

A 1-bit unsigned integer that indicates the device support of automatic PFC headroom calculation. If the HDR cap bit is 1, and PFC is enabled on at least one traffic class, the automatic headroom calculation is enabled.

### D.2.10.y Internal delay

A 2-octet unsigned integer contains the length of time for which the device process received PFC pause frame. It includes TX interface delay, RX interface delay and higher layer delay. The value is measured in units of pause\_quanta, equal to the time required to transmit 512 bits of a frame at the data rate of the MAC.

## D.5 IEEE 802.1/LLDP extension MIB

TBD content

## D.6 IEEE 802.1/LLDP extension YANG

TBD content

## Annex M - Support for PFC in link layers without MAC Control

TBD content