Annex N Updates

Lily Lyu

March Plenary

Contents

N.1 Overview

- What is PFC headroom

  When activate PFC operation, “the PFC headroom (see 36.1.1) is the minimum buffer size that needs to remain available on receiver. It helps implementation to allocate buffer for PFC-enabled priorities. ”

- Introduce Annex N content

  “This annex explains delay model of PFC headroom, and provides an example of buffer allocation based on the PFC headroom calculation.”

N.2 Delay model of PFC headroom

PFC headroom calculation considers various delays accumulated

36.1.1

a) B’s reception processing to calculate the remaining buffering following frame receipt.
b) B’s PFC Initiator to initiate PFC following that buffering calculation.
c) Encoding of the PFC frame and any other transmission delays associated with B’s interface stack.
d) Any prior in-progress frame transmission by B (possibly of a maximum sized frame) to complete.
e) PFC frame transmission on the physical link.
f) The link delay for transmission from B to A.
g) PFC frame reception, including frame validation, by A’s interface stack.
h) A’s PFC Receiver to decode the PFC frame and halt transmission selection for specified priorities.
i) Any in-progress frame transmission by A (possibly of a maximum sized frame) to complete.
j) The link delay for transmission from A to B.
k) Reception delays associated with B’s interface stack, reception processing, and buffering.

Update of 36.1.1

PFC frame transmission in PFC initiator station B:
a) B’s reception processing to calculate the remaining buffering following frame receipt.
b) B’s PFC Initiator to initiate PFC following that buffering calculation and PFC frame encoded ready for transmission.
c) Any prior in-progress frame transmission by B (possibly of a maximum sized frame) to complete.
d) First bit of PFC frame sent to MAC service.
e) Last bit of PFC frame sent on the physical link.

PFC frame transmission across link from B to A:
f) The link delay for transmission from B to A.

PFC frame reception in PFC receiver station A (including PFC taking action):
g) PFC frame reception since the last bit of PFC frame received on link, including frame validation, by A’s interface stack.
h) A’s PFC Receiver to decode the PFC frame and halt transmission selection for specified priorities.

User data transmission in PFC receiver station A:
i) Any in-progress frame transmission by A (possibly of a maximum sized frame) to complete.
j) Last bit of last frame sent on the physical link.

User data transmission across link from A to B:
k) The link delay for transmission from A to B.

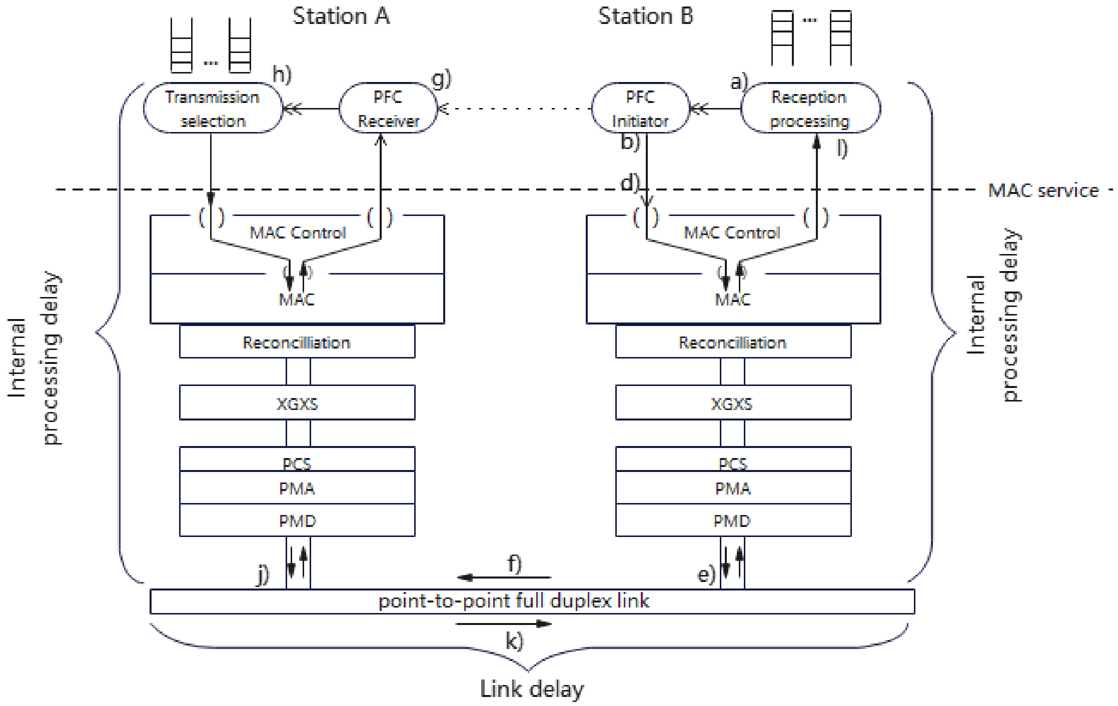
User data reception in PFC initiator station A:
l) Reception delays associated with B’s interface stack, reception processing, and buffering.

N.2 Delay model of PFC headroom

Internal processing delay(ID): the time spent on frame processing within PFC initiator station and PFC receiver station, including interface stack delay, queue status change delay, and buffering delay etc., assuming no prior in-progress frame transmission

Link delay(LD): the time spent on physical link between PFC initiator station and PFC receiver station

Worst-case delay(WD): the additional time needed for a maximum sized frame transmission before the PFC frame transmission at PFC initiator station, and the additional time needed for a maximum sized frame transmission after the PFC frame taking effect at PFC receiver station



PFC frame transmission in PFC initiator station B:

a) B's reception processing to calculate the remaining buffering following frame receipt.

b) B's PFC Initiator to initiate PFC following that buffering calculation and PFC frame encoded ready for transmission.

c) Any prior in-progress frame transmission by B (possibly of a maximum sized frame) to complete.

d) First bit of PFC frame sent to MAC service.

e) Last bit of PFC frame sent on the physical link.

PFC frame transmission across link from B to A:

f) The link delay for transmission from B to A.

PFC frame reception in PFC receiver station A (including PFC taking action):

g) PFC frame reception since the last bit of PFC frame received on link, including frame validation, by A's interface stack.

h) A's PFC Receiver to decode the PFC frame and halt transmission selection for specified priorities.

User data transmission in PFC receiver station A:

i) Any in-progress frame transmission by A (possibly of a maximum sized frame) to complete.

j) Last bit of last frame sent on the physical link.

User data transmission across link from A to B:

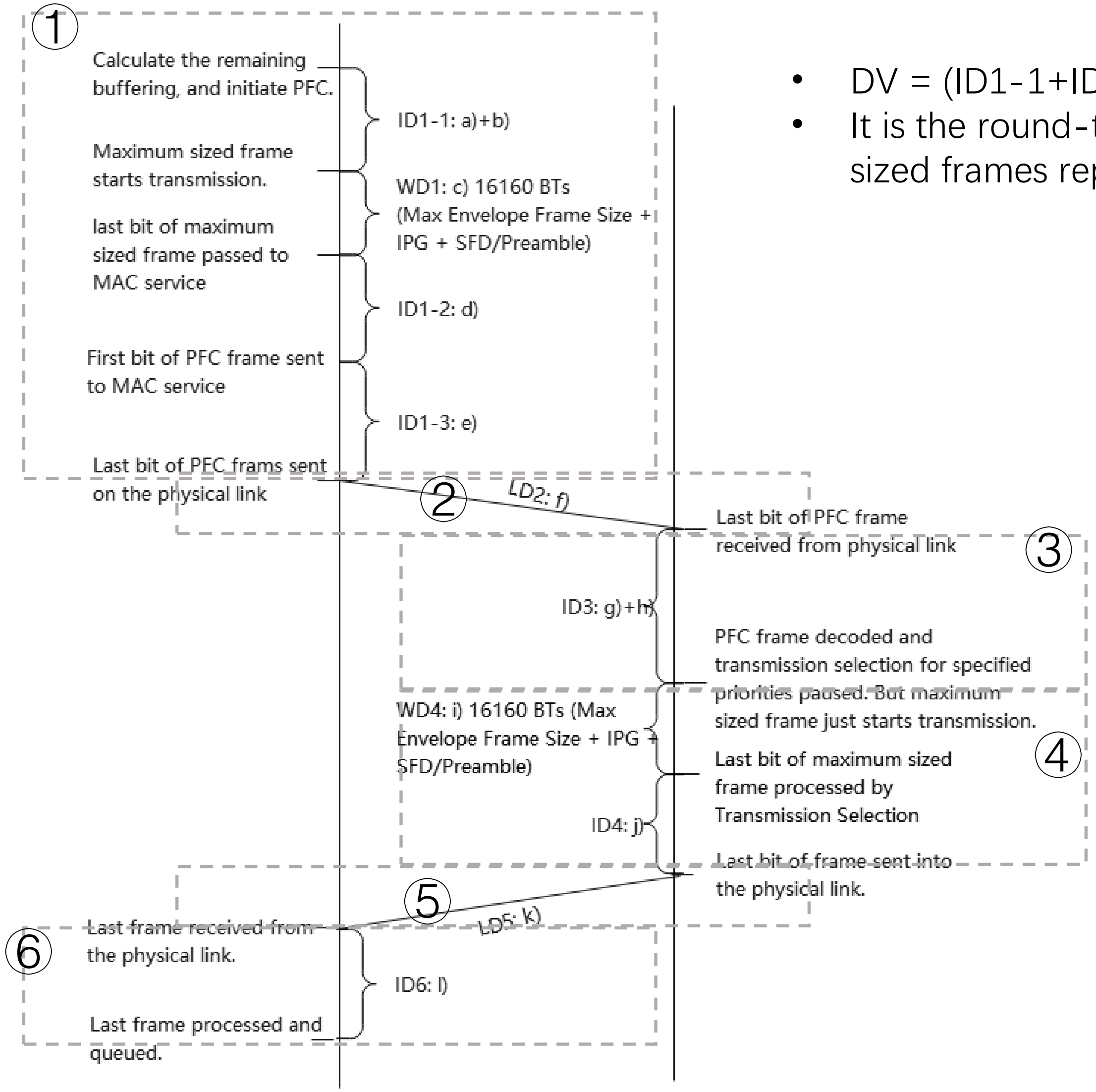
k) The link delay for transmission from A to B.

User data reception in PFC initiator station A:

l) Reception delays associated with B's interface stack, reception processing, and buffering.

N.2 Delay model of PFC headroom

Figure shows a possible worst-case delay example where MACsec is disabled for PFC frame and user data.



- DV = (ID1-1+ID1-2+ID1-3) + WD1 + LD2 + ID3 + WD4 + ID4 + LD5 + ID6
- It is the round-trip delay from PFC initiation to last frame reception and buffered, plus 2 maximum sized frames represented by bit times.

①	**PFC frame transmission in PFC initiator station B:** a) B's reception processing to calculate the remaining buffering following frame receipt. b) B's PFC Initiator to initiate PFC following that buffering calculation and PFC frame encoded ready for transmission. c) Any prior in-progress frame transmission by B (possibly of a maximum sized frame) to complete. d) First bit of PFC frame sent to MAC service. e) Last bit of PFC frame sent on the physical link.
②	**PFC frame transmission across link from B to A:** f) The link delay for transmission from B to A.
③	**PFC frame reception in PFC receiver station A (including PFC taking action):** g) PFC frame reception since the last bit of PFC frame received on link, including frame validation, by A's interface stack. h) A's PFC Receiver to decode the PFC frame and halt transmission selection for specified priorities.
④	**User data transmission in PFC receiver station A:** i) Any in-progress frame transmission by A (possibly of a maximum sized frame) to complete. j) Last bit of last frame sent on the physical link.
⑤	**User data transmission across link from A to B:** k) The link delay for transmission from A to B.
⑥	**User data reception in PFC initiator station A:** l) Reception delays associated with B's interface stack, reception processing, and buffering.

N.3 Internal processing delay

The Internal Processing Delay is implementation dependent.
It comprises

- frame processing delays above MAC service which is between MAC control client and transmission selection
- MAC and PHY layer interface delays.

Sublayer	Maximum RTT (bit times)	Maximum RTT (pause quanta)	Reference (subclause of IEEE Std 802.3-2018 [B14])
10G MAC Control, MAC, and RS	8192	16	46.1.4
XGXS and XAUI	2048	4	48.5
10GBASE-X PCS	2048	4	49.2.15
10GBASE-R PCS	3584	7	50.3.7
LX4 PMD	512	1	53.2
CX4 PMD	512	1	54.3
Serial PMA and PMD	512	1	52.2
10GBASE-T	25 600	50	55.11

IEEE 802.3 Interface Delays

Example of processing delays above MAC service are MACsec and entering pause state delays.

- MACsec delay: For link speeds of up to 10Gb/s, MACsec constrains each of the transmit delay and the receive delay to a maximum of 19 360 bit times (see 36.1.3.3).
- Entering pause state delay: This standard defines a queue shall go into paused state in no more than 614.4 ns (see 36.1.3.3). This delay is equivalent to 6144 bit times at the speed of 10Gb/s.

N.4 Link delay

The Link Delay is the propagation delay over the transmission medium and can be approximated by the following equation:

$$\text{Link Delay} = \text{Medium Length} \times \frac{1}{BT \times \upsilon}$$

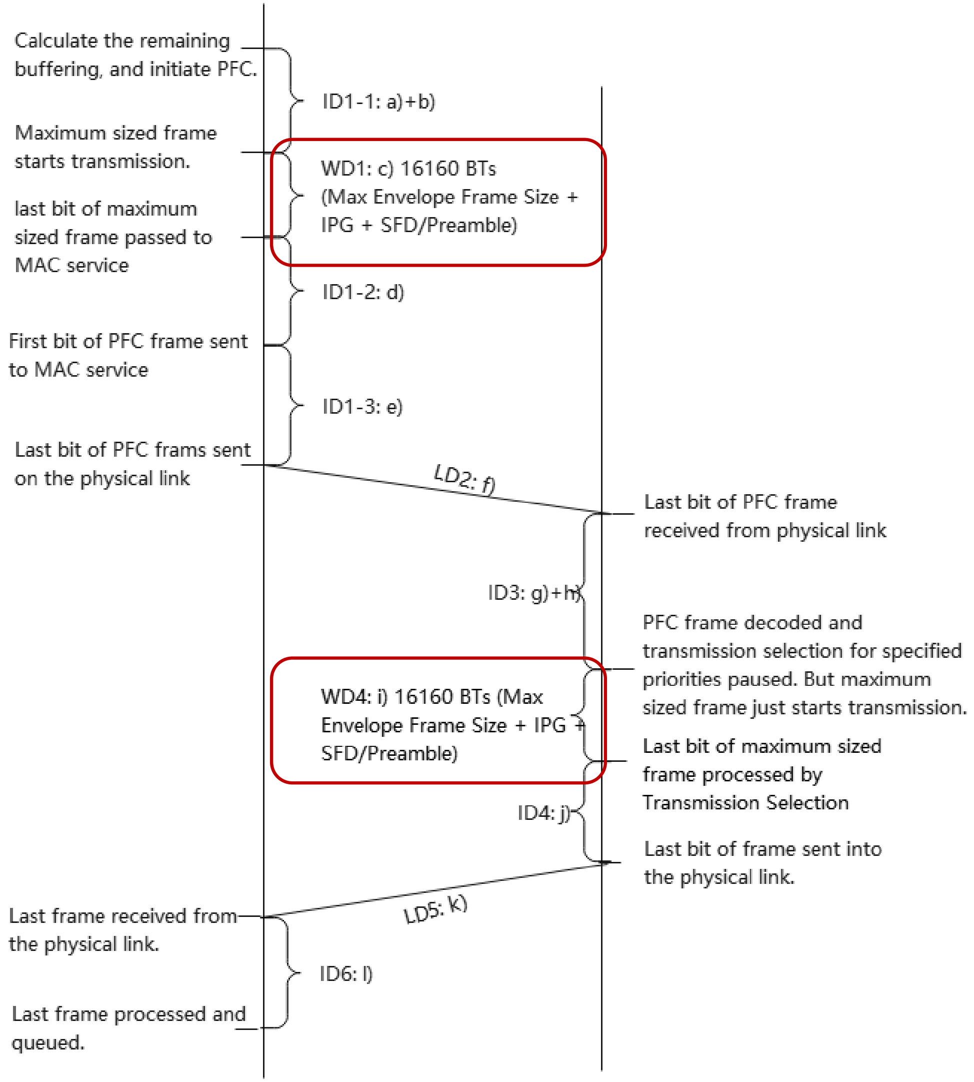
where u is the signal propagation speed in the medium and BT is the bit time of the medium.

N.5 Worst-case delay

The Worst-case Delay comprises 2 parts.

- At PFC initiator station, it is assumed a maximum sized frame just start transmission from Transmission Selection when PFC is invoked. PFC frame has to wait until this in-progress frame complete transmission.
- At PFC receiver station, it is assumed queue is paused but a maximum sized frame just starts transmission. Thus, bit times of the maximum sized frame is added into the total delay.

Note:
Enabling MACsec impacts worst-case delay model. That is implementation dependent.



N.6 Buffer allocation example

10GBASE-T PHY example:

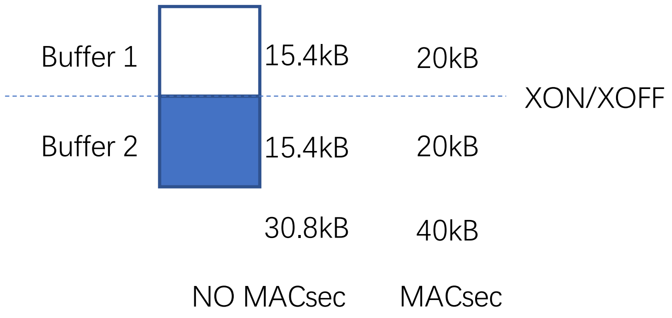
- PFC frame generation: 200 bit times;
- Maximum envelope frame size: 2000 octets, 16 160 bit times;
- PFC frame size: 64 octets, 672 bit times;
- XGMII MAC/RS and XAUI interface: 8192 + 2 × 2048 = 12 288 bit times;
- 10GBASE-T Delay: 25 600 bit times;
- 100 meters Cat6 cable: 5556 bit times (computed assuming u = 0.6 × c, where c is the speed of the light in meters per second);
- Entering paused state = 6144 bit times

MACsec not supported:
DV = (ID1-1+ID1-2+ID1-3) + WD1 + LD2 + ID3 + WD4 + ID4 + LD5 + ID6

DV =((200) + (672 + (12 288 + 25 600) /2))+ (16 160) + (5556) + ((12 288 + 25 600) /2 + 6144) + (16 160) + ((12 288 + 25 600) /2) + (5556) + ((12 288 + 25 600) /2) = 126 224 bit times
---> PFC headroom is 15.4kB

MACsec supported (data frame):
WD1 and ID4, each is incremented by 19 360 bit times
DV = 126 224 + 19 360 + 19 360 = 164 944 bit times
---> PFC headroom is 20kB

Buffer allocation strategy:
2 times PFC headroom is allocated to PFC enabled priority queue. XON/XOFF threshold is set to the size of PFC headroom.



PFC guarantees no frame loss and no throughput loss.

When priority queue is increased to XOFF, PFC pause is invoked. Buffer 1 is used to absorb inflight frames.

When priority queue is decreased to XOFF, PFC resume is invoked. Buffer 2 is used to prevent traffic stop.