Headroom Measurement Protocol Design

Lily Lv

To-Do List

- **Ethertype for Qdt**
  - **What EtherType should be used for the round trip delay?**
- **DCBX： PFC Configuration TLV format design**
  - **Important not to let packet formats and perceived encoding efficiencies to drive the protocol design.**
- Managed objects
  - The effort, implementation cost, and purpose of statistic gathering and retention requires careful consideration
- Timestamp point clarification
  - Will (t3-t2) be impacted (variably) by queue delay?
  - further specify t1, t4
- Timestamp accuracy
  - What is the accuracy of t1, t4?

Ethertype for Qdt

Reuse Qcz (CI) Ethertype 89-A2

Qcz definition

Table 47-1—Layer-2 CIM Encapsulation

	Octet	Length
PDU EtherType (89-A2)	1	2
Version	3	4 bits
Subtype	3	4 bits
CIM PDU	4	65-529

Table 47-4—CIM PDU

	Octet	Length
Version	1	4 bits
Reserved	1	3 bits
Add/Del	1	1 bit
destination_address	2	6
source_address	8	6
vlan_identifier	14	12 bits
Encapsulated MSDU length	16	2
Encapsulated MSDU	18	48-512

Subtype:

This field, 4 bits in length, shall be transmitted with the value 0 to indicate an encapsulated CIM PDU. The Subtype field occupies the least significant 4 bits of the first octet of the layer-2 CIM Encapsulation.

Qdt proposal

	Octet	Length
PDU Ethertype(89-A2)	1	2
Version	3	4 bits
Subtype	3	4 bits
Headroom Measurement PDU	4	65-529

Subtype 0, CIM

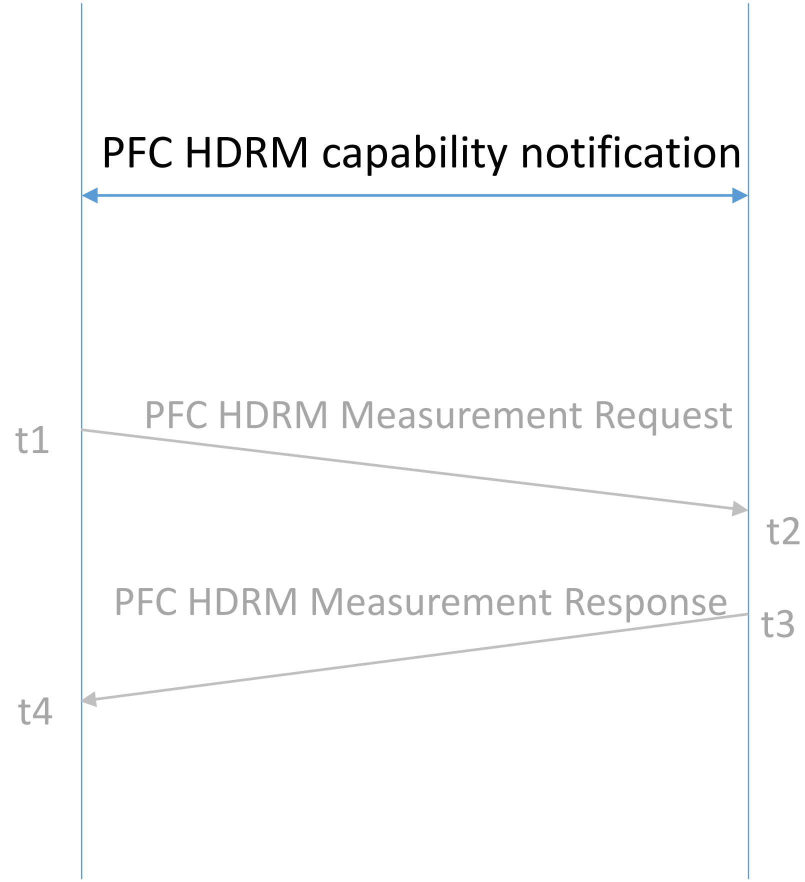
Subtype 1, Headroom Measurement Message

Question:

Is "65-529" too big for headroom measurement PDU?

PFC Configuration TLV format design (1/7)

- Protocol design for capability notification.
  - Augment DCBX by extending PFC configuration TLV



Phase 1: Capability notification

-- If both sides support PFC HDRM, initiate PFC HDRM Measurement Request, otherwise, stop the procedure.

Phase 1 answers:

- ➢ Which option(s) of measurement is supported?
- ➢ Which option of measurement will be used?

Should phase 1 answer

- ➢ What the compensation value is for each option?

PFC Configuration TLV format design (2/7)

- DCBX mechanism
  - DCBX has 3 types of attributes:
    - Informational attributes
    - Asymmetric attributes
    - Symmetric attributes
  - PFC configuration TLV is sent using symmetric attributes passing.
    - Symmetric attributes: “the passing of a attribute from one port to its peer port with objective of both ports utilizing the same attribute value.”
  - ‘Willing’ is important in symmetric attribute passing.
    - “A Willing port shall set its operational attribute to that indicated in the received TLV if the received TLV has the W bit set to zero. If both the local port and remote port are willing, then the attribute values of the port with the lower numerical MAC address shall take precedence.”

PFC configuration TLV

PFC configuration TLV “**Shall be sent using Symmetric attribute passing**”

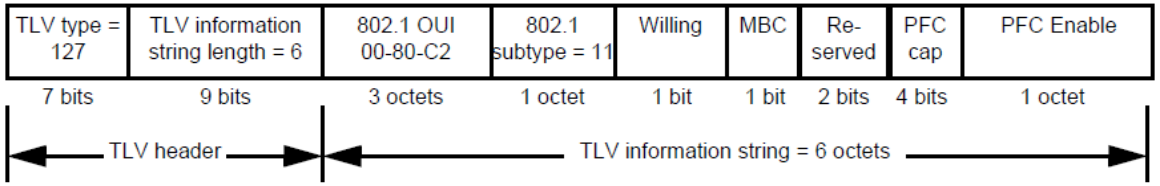


Figure D-10—Priority-based Flow Control Configuration TLV format

DCBX symmetric attributes

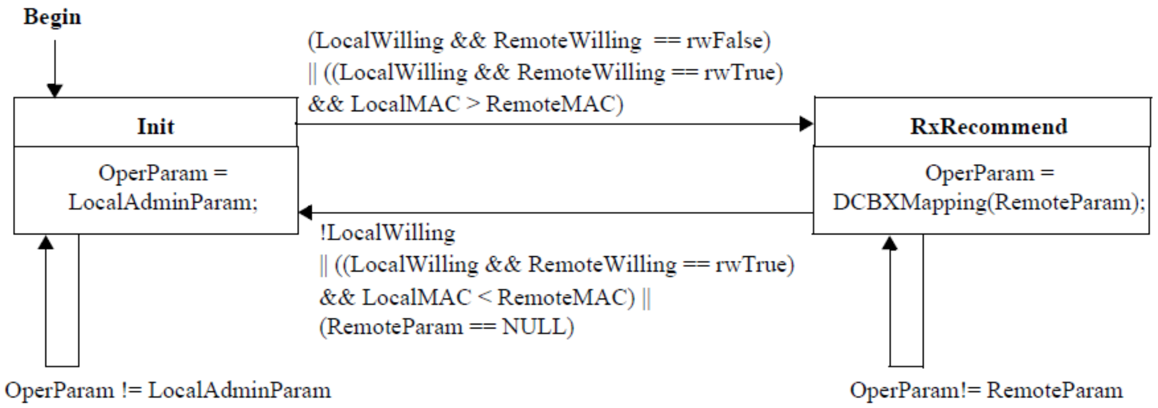


Figure 38-2—Symmetric state machine

PFC Configuration TLV format design (3/7)

- Non-PTP measurement required information from PFC configuration TLV
  - Non-PTP capability of remote port
  - (Compensation value is transmitted in response message)

TLV type = 127	TLV information string length = 6	802.1 OUI 00-80-C2	802.1 subtype = 11	Willing	MBC	Re-served	PFC cap	PFC Enable
7 bits	9 bits	3 octets	1 octet	1 bit	1 bit	2 bits	4 bits	1 octet
TLV header		TLV information string = 6 octets						

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

TLV type = 127	TLV information string length = 6	802.1 OUI 00-80-C2	802.1 subtype =11	Willing	MBC	MACsec cap	HDR cap	PFC cap	PFC Enable
7 bits	9 bits	3 octets	1 octets	1 bit	1 bit	1 bit	1 bit	4 bits	1 octets
TLV header		TLV information string = 6 octets							

Only capability is included in PFC configuration TLV.
Compensation value will be included in request-response procedure.

PFC Configuration TLV format design (4/7)

- PTP-based measurement required information from PFC configuration TLV
  - PTP-based capability of remote
  - Compensation value (internal processing delay) of remote

TLV type = 127	TLV information string length = 6	802.1 OUI 00-80-C2	802.1 subtype = 11	Willing	MBC	Re-served	PFC cap	PFC Enable
7 bits	9 bits	3 octets	1 octet	1 bit	1 bit	2 bits	4 bits	1 octet
TLV header		TLV information string = 6 octets						

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

TLV type = 127	TLV information string length = 8	802.1 OUI 00-80-C2	802.1 subtype =11	Willing	MBC	MACsec cap	Rese rved	PFC cap	PFC Enable	PTP comp
7 bits	9 bits	3 octets	1 octets	1 bit	1 bit	1 bit	1 bit	4 bits	1 octets	16 bit
TLV header		TLV information string = 8 octets								

PTP comp > 0 , PTP-based measurement is capable, compensation value equals to 'PTP comp'
PTP comp = 0 , PTP-based measurement is incapable.

PFC Configuration TLV format design (5/7)

- PTP-based measurement required information from PFC configuration TLV

TLV type = 127	TLV information string length = 8	802.1 OUI 00-80-C2	802.1 subtype =11	Willing	MBC	MACsec cap	Rese rved	PFC cap	PFC Enable	PTP comp
7 bits	9 bits	3 octets	1 octets	1 bit	1 bit	1 bit	1 bit	4 bits	1 octets	16 bit
TLV header		TLV information string = 8 octets								

Does it contradict the symmetric attribute passing rule?

Assuming below case.
Local para: W=0, PTP comp = ‘a’ (>0)
Remote para: W=1, PTP comp = ‘b’ (>0)

According to symmetric attribute passing rule, local operation parameter of PTP comp will still be ‘a’.
However, what local system really needs to calculate headroom is ‘b’.

PTP-based capability and PTP comp cannot be combined together in PFC configuration TLV.

- PTP-based capability is symmetric attribute
- PTP comp is informational attribute

PFC Configuration TLV format design (6/7)

- Proposal :
  - ➢ PFC configuration TLV only includes 'capability'

Question： if both non-PTP and PTP-based are capable, which will be used ?

Each bit indicates one capability.

Option 1: Define priority of the 2 methods.

TLV type = 127	TLV information string length = 7	802.1 OUI 00-80-C2	802.1 subtype =11	Willing	MBC	MACsec cap	Rese rved	PFC cap	PFC Enable	non-PTP	PTP-based	Reser ved
7 bits	9 bits	3 octets	1 octets	1 bit	1 bit	1 bit	1 bit	4 bits	1 octets	1 bit	1 bit	6 bits
TLV header		TLV information string = 7 octets										

Option 2: Only one capability can be set in TLV.

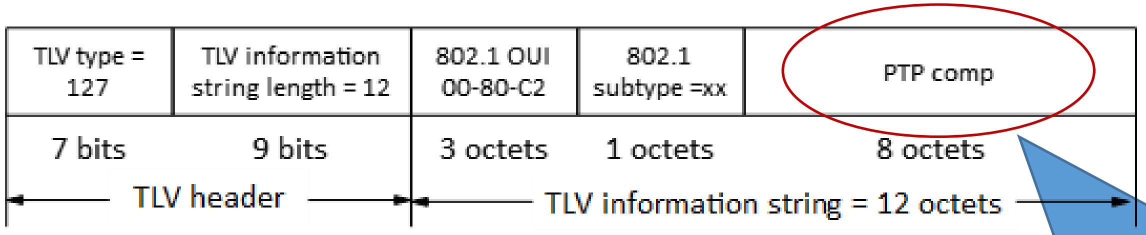
TLV type = 127	TLV information string length = 7	802.1 OUI 00-80-C2	802.1 subtype =11	Willing	MBC	MACsec cap	Rese rved	PFC cap	PFC Enable	Measurement Mode	Reser ved
7 bits	9 bits	3 octets	1 octets	1 bit	1 bit	1 bit	1 bit	4 bits	1 octets	2 bits	6 bits
TLV header		TLV information string = 7 octets									

Mode：
00 non-capable
01 non-PTP
10 PTP based
11 reserved

PFC Configuration TLV format design (7/7)

- Proposal :
  - ‘PTP comp’ for PTP-based measurement passes to peer separately.

Option 1: Define a new informational TLV - **PFC informational TLV**



Compensation value for PTP-based measurement

DCBX informational attributes: “Informational attributes are exchanged via LLDP without any participation in a DCBX state machine.”

Option 2: Do not specify it in Qdt.

Mention the internal processing delay impact, but allow vendor specific way to implement.

Thanks