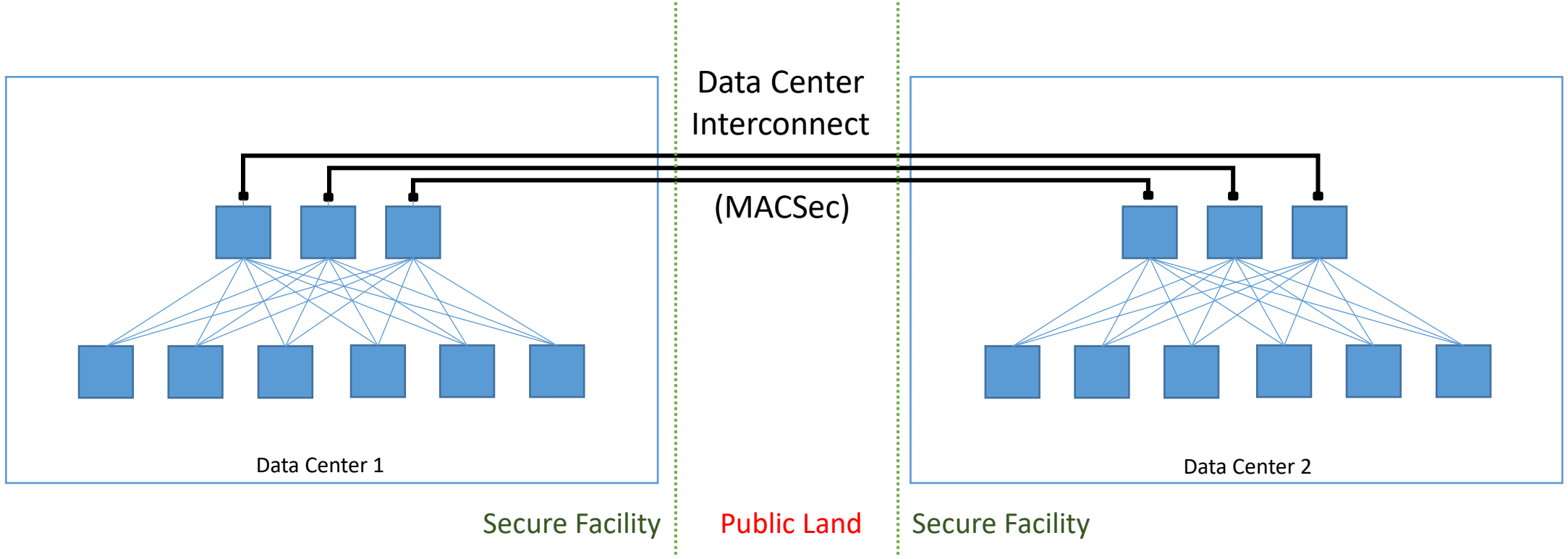
Incorporating MACSec into PFC Headroom Calculation

Paul Congdon, Lily Lv (Huawei)

Ruibo Han (CMCC)

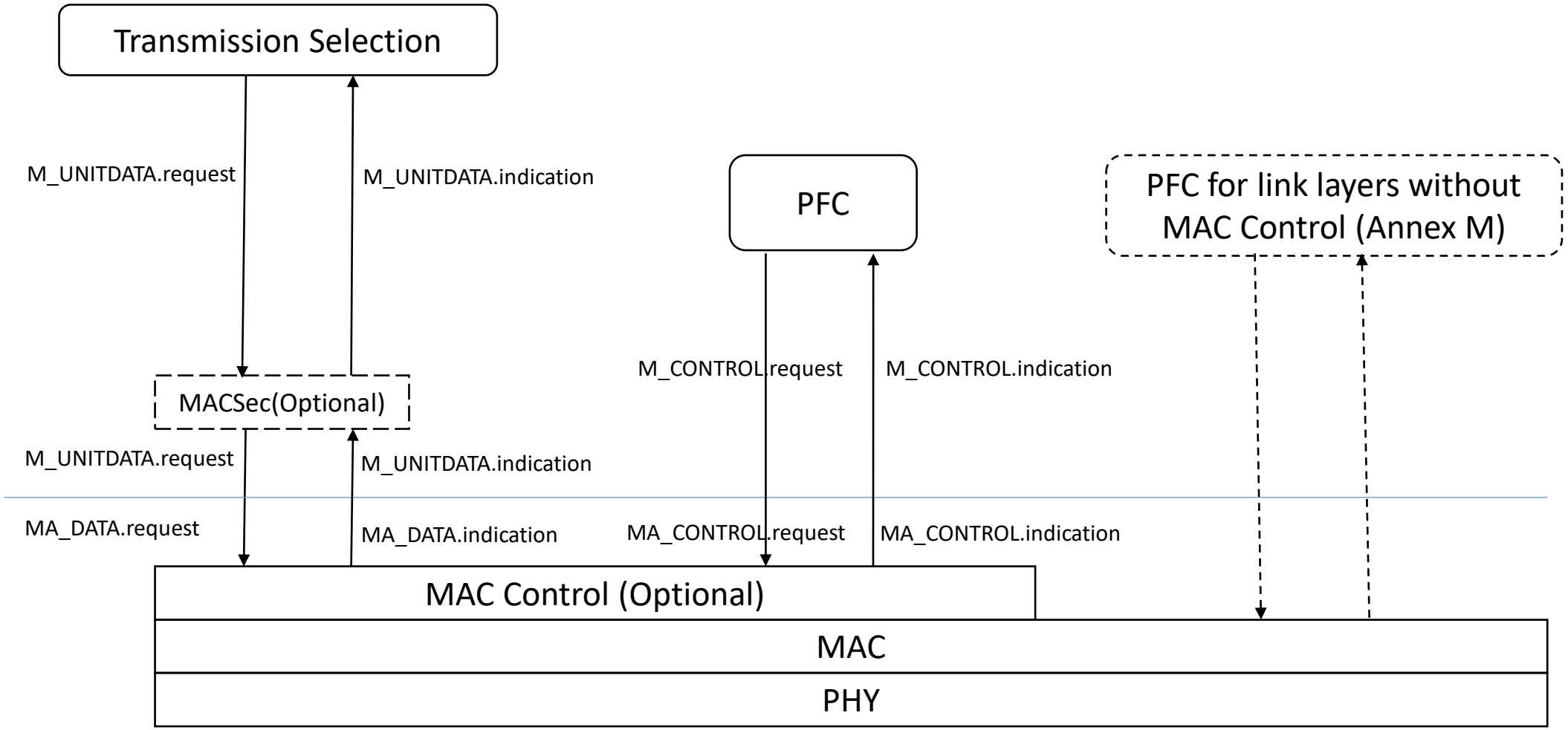
Mick Seaman (Independent)

A Use Case for MACSec To Consider



NOTE: The RDMA protocol over Ethernet (RoCEv2) necessitates the use PFC to avoid frame loss
It is desirable to protect PFC frames when they traverse data center interconnect links

Current Protocol Layers



NOTE: Figure indicates that PFC Frames are not encrypted

802.1Q MACSec Bypass Control is inconsistent

- While Figure 36-1 clearly indicates PFC frames are not encrypted
- 36.1.3.3 introduces MACSec Bypass Control (MBC)
- The DCBX Priority-based Flow Control Configuration TLV includes MBC as defined in D.2.10.4.
  - “The MACsec Bypass Capability Bit. If set to zero, the sending station is capable of bypassing MACsec processing when MACsec is disabled. If set to one, the sending station is not capable of bypassing MACsec processing when MACsec is disabled (see Clause 36).”

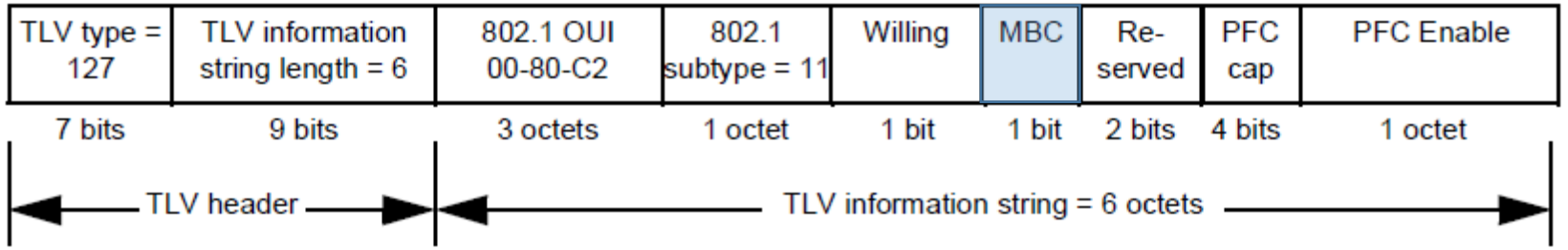
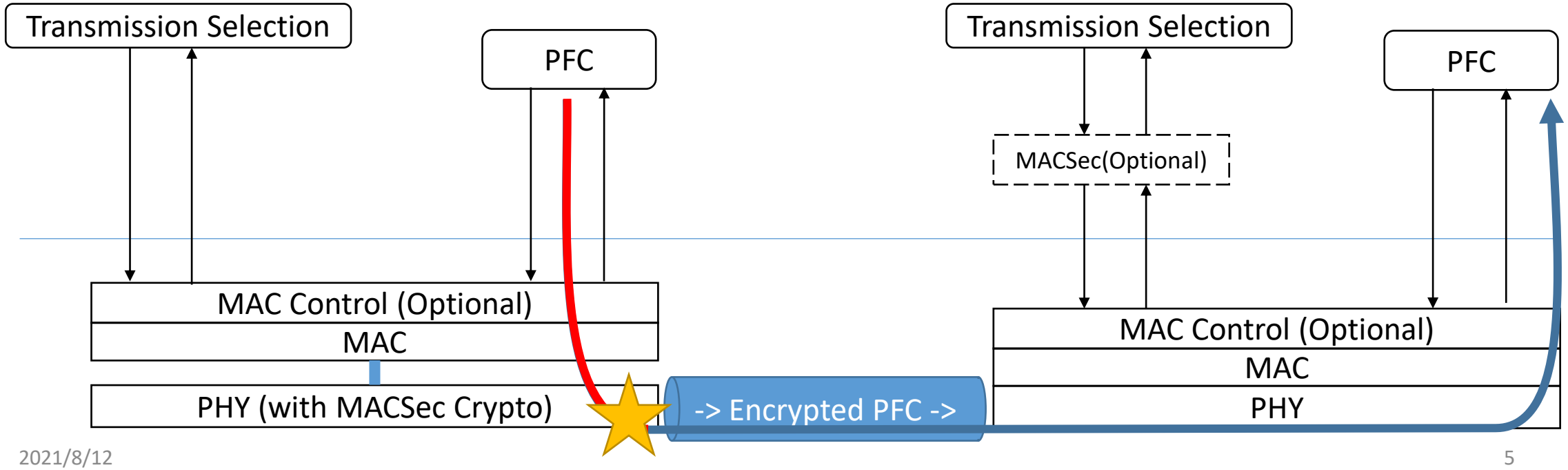


Figure D-10—Priority-based Flow Control Configuration TLV format

NOTE: IEEE Std 802.1AE has no concept of “MACSec Bypass”. Unprotected frames could use Uncontrolled Port

Interoperability issue in the field

- Early implementations of MACSec were implemented external to the MAC (i.e. within a PHY as a ‘bump in the wire’).
  - These early implementations encrypt everything coming out of the MAC
  - These early implementations were never compliant with 802.1AE
  - These early implementations do not run MKA and may suffer outages



Ethernet Data Encryption (EDE-M)

Specifies ‘Bump in the Wire’ Behavior

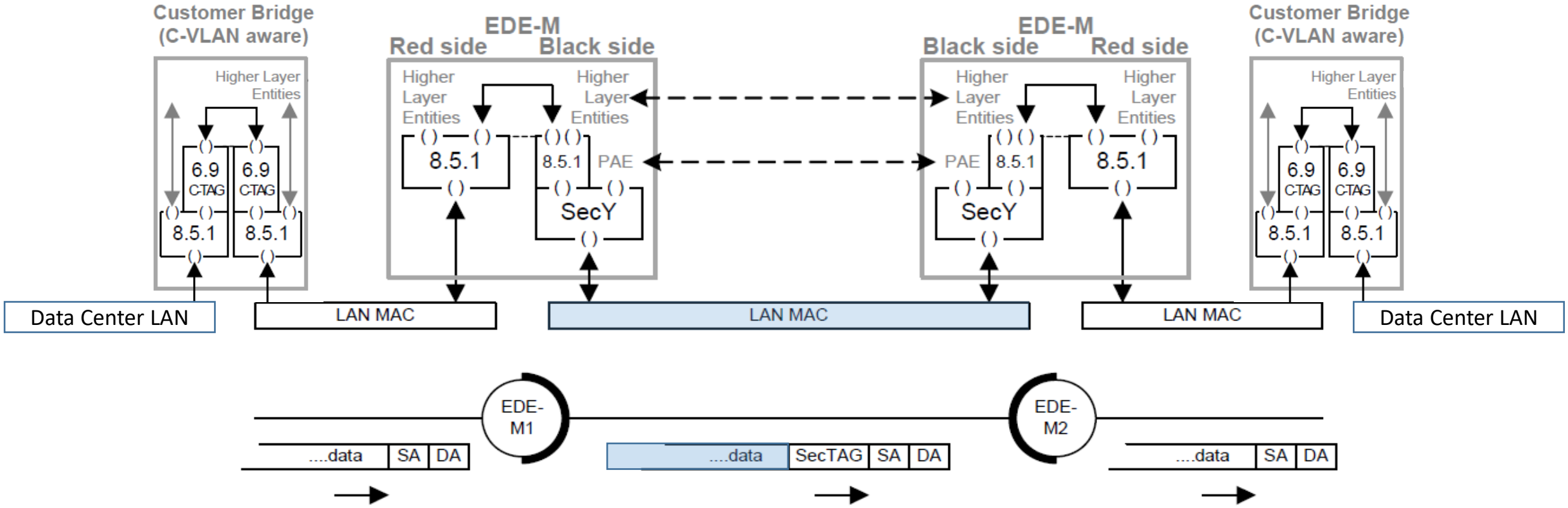
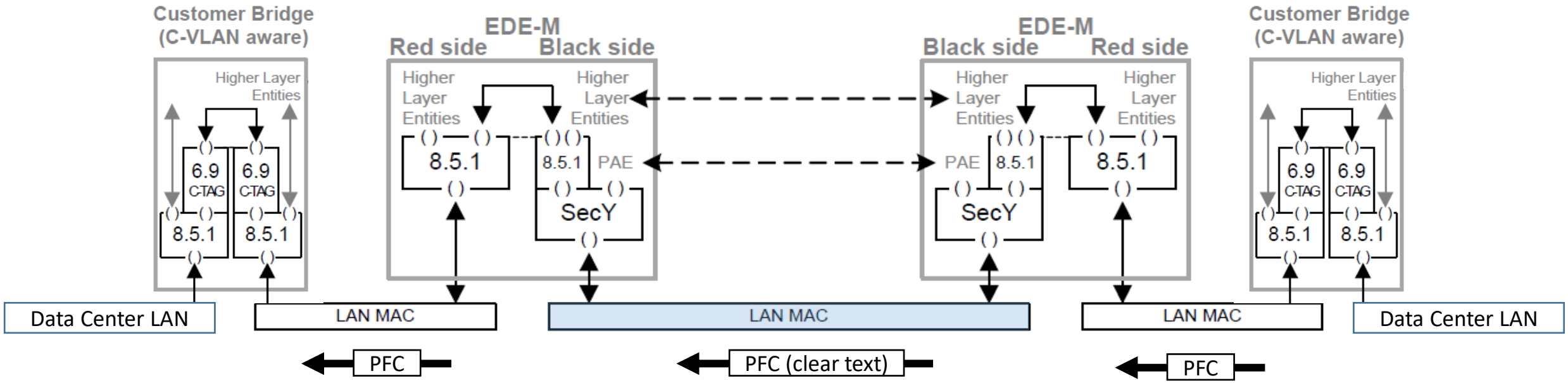


Figure 15-1—EDE-Ms connected by a point-to-point LAN

EDE-M Conformance:

a) Comprise a VLAN-unaware MAC Bridge as specified by IEEE Std 802.1Q (5.14 of IEEE Std 802.1Q-2018) with the constraints and exceptions specified in this standard.

No PFC Forwarding, but backward propagation



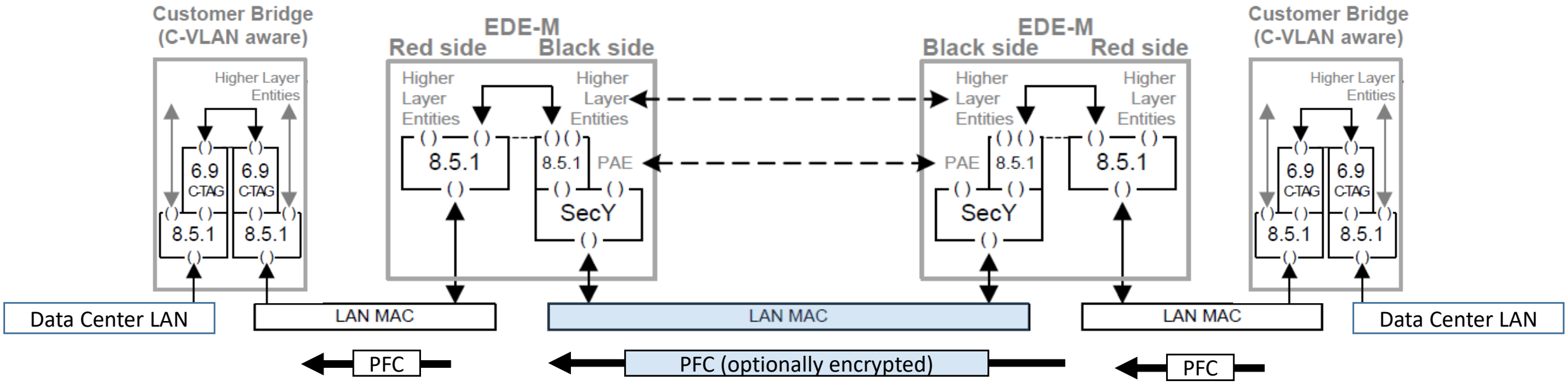
802.1 Reserved Addresses are not forwarded:

a) PFC uses reserved address 01-80-C2-00-00-01

b) C-VLAN Bridges, MAC Bridges, TPMRs and EDEs do not forward 01-80-C2-00-00-01

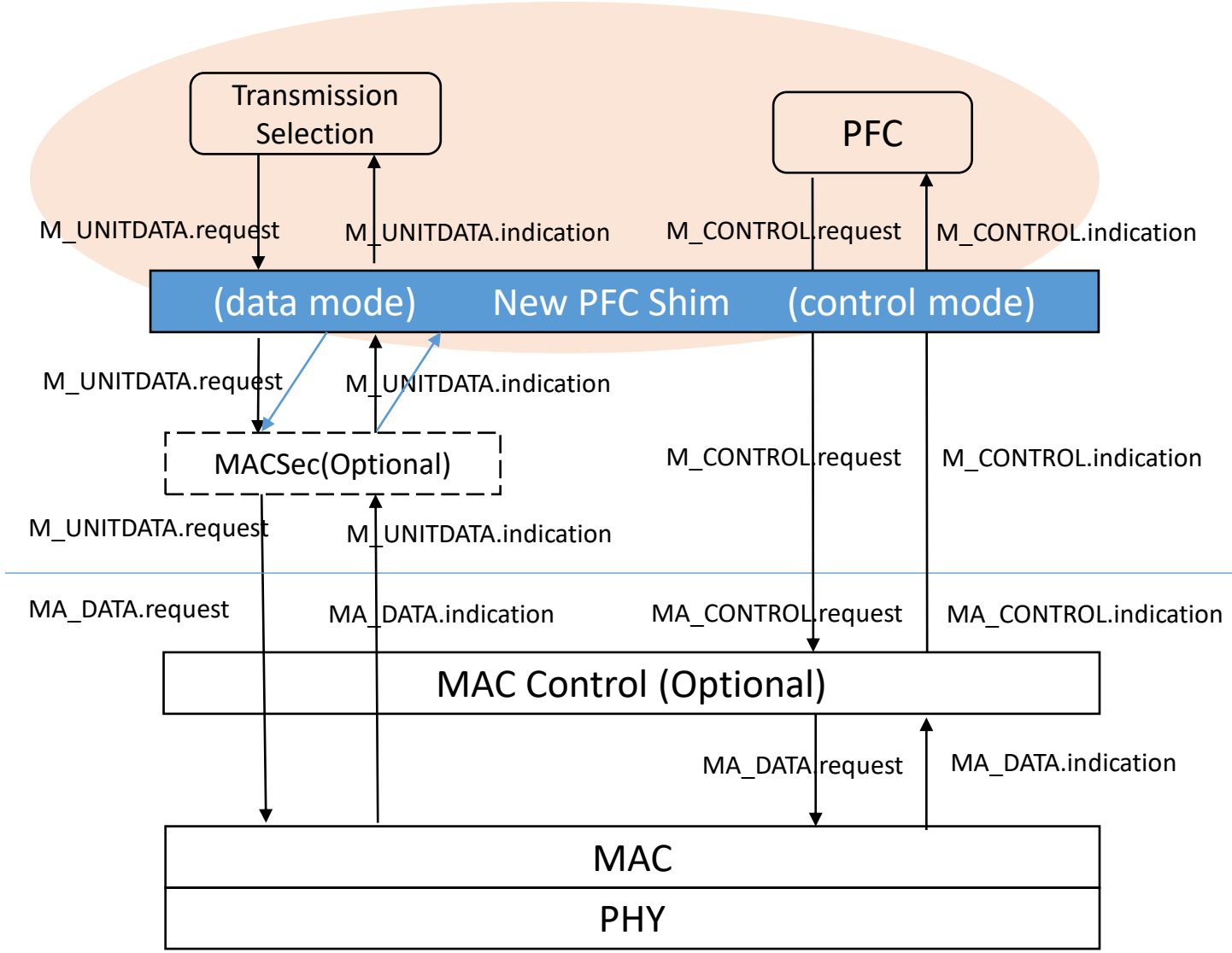
Desired Future State

- Allow an option for protecting/securing PFC frames
- Do not change the backward propagation model of PFC
- Include MACSec delay into the PFC Headroom calculation



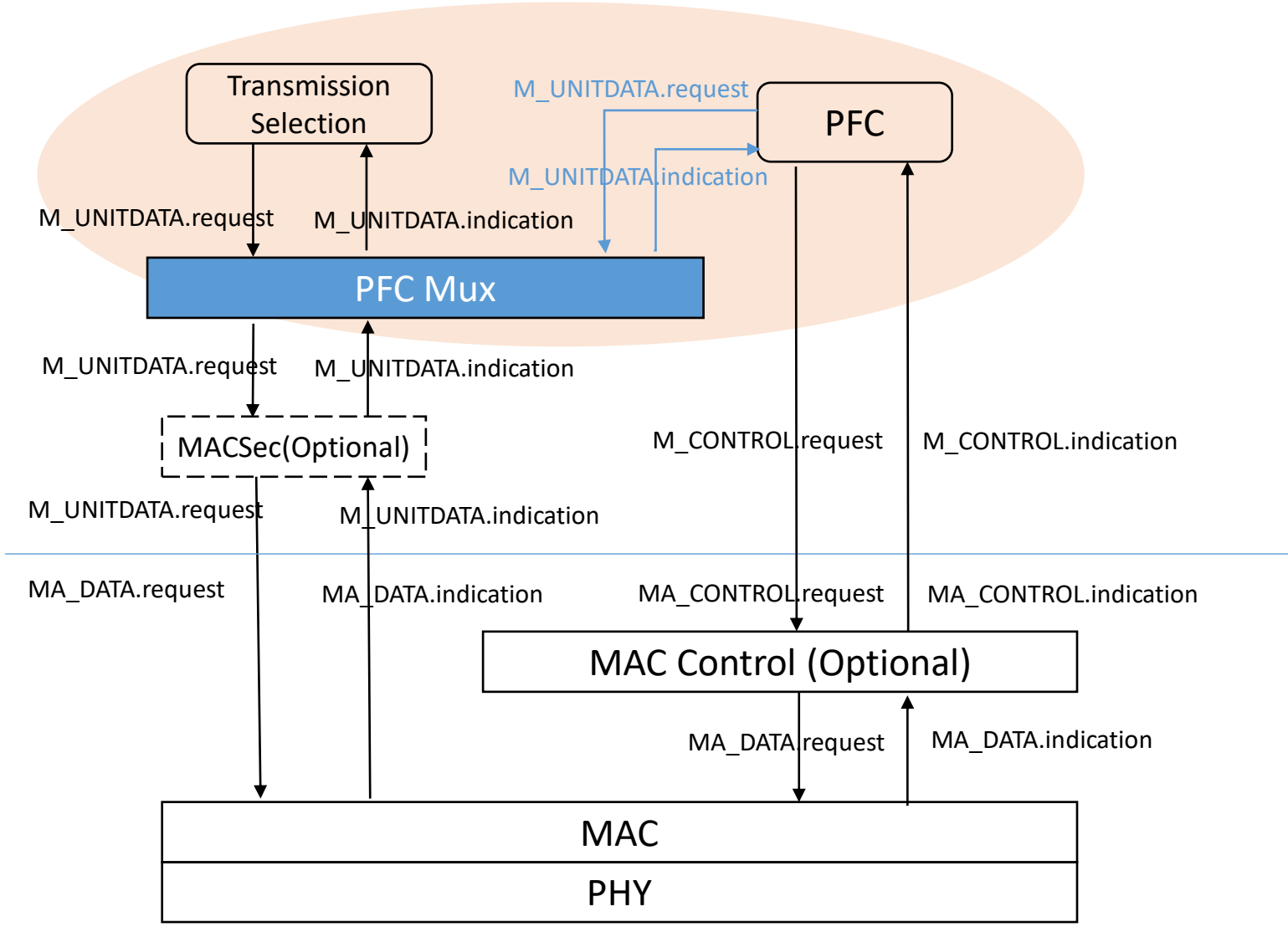
Option A: New PFC shim/layer

- Minimal impact to current PFC
- Shim passes through existing MAC Control interface in ‘control mode’ (with no delay)
- Shim configured to generate and intercept PFC frames if ‘data mode' is desired
- Internal delay calculation depends on Shim configuration



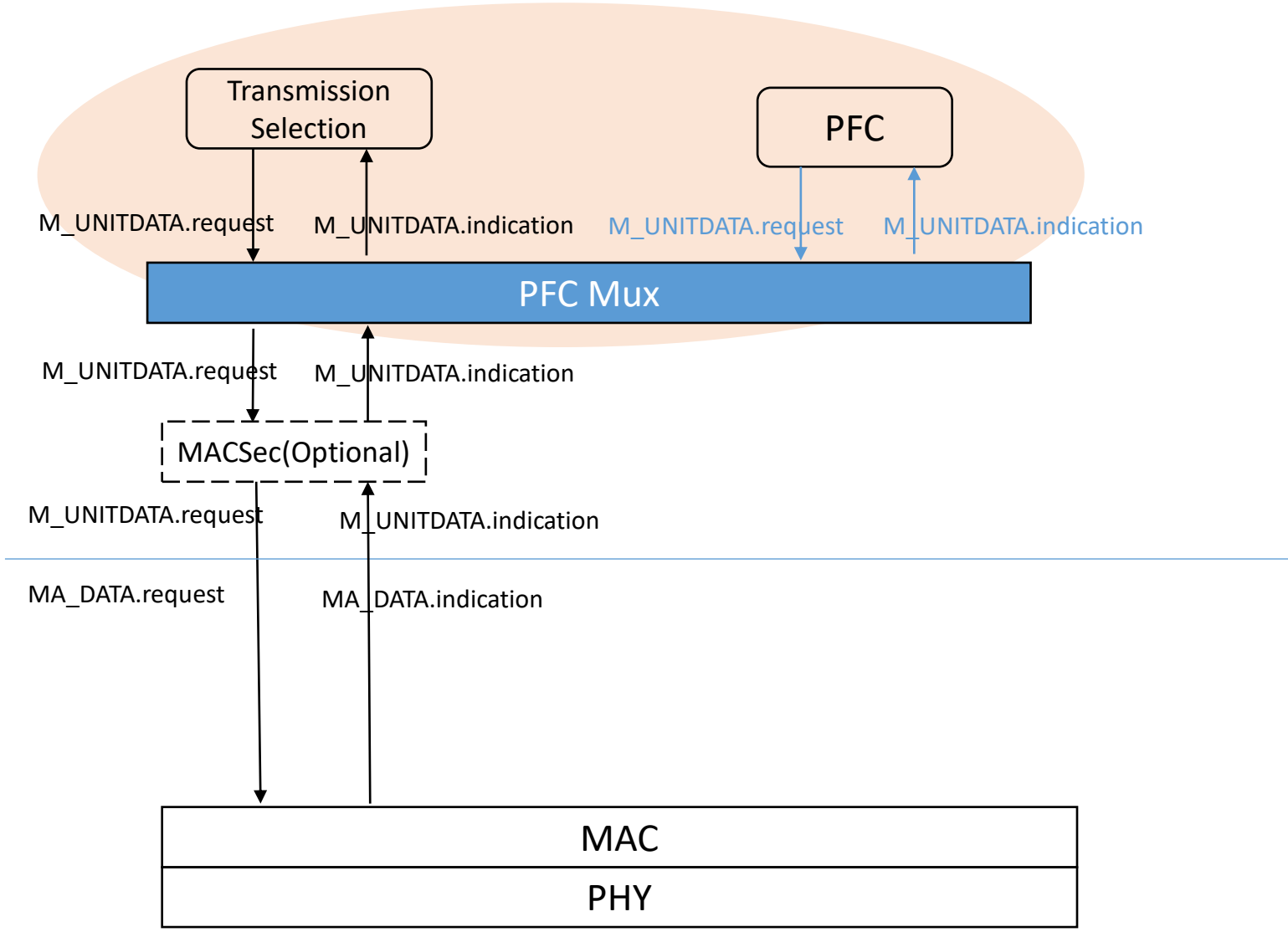
Option B: PFC Mux in Data Path

- PFC configured to generate frames or existing MAC control
- PFC Mux inserts and extracts PFC frames
- Internal delay calculation includes MACSec delay and PFC frame processing delay



Option C: Eliminate MAC Control Interface

- Redefinition of PFC interface
- PFC Mux inserts and extracts PFC frames
- PFC configured to use controlled or uncontrolled port
- Supports all point-to-point MACs
- Internal delay calculation includes MACSec delay and PFC frame processing delay



MACSec, PFC and headroom calculation

- Encryption of PFC increases delay and buffer requirements, making headroom calculation more important.
- Relationship to previous headroom calculation options (see: https://www.ieee802.org/1/files/public/docs2021/new-lv-adaptive-pfc-headroom-and-PTP-0602-v03.pdf)
  - Option 1: new timestamp reference + PTP Pdelay messages
    - May need multiple reference points
  - Option 2: new timestamp reference + new MAC control frames
    - New reference points, limits MACSec/PFC options A, B, C, A new MAC control frame has same issue as PFC
  - Option 3: new procedure for internal delay + PTP Pdelay messages
    - Different internal delays based upon configuration
- Option 3 still works for such changes. Option 1&2 need to redefine timestamp points. Option 3 is more flexible.
- LLDP carries internal processing delay including shim layer interface delay and MACSec delay.

Summary

- Project Need has expanded (PFC enhancements)
  - PFC usage in DCN and DCI
  - Simplify the deployment of PFC, especially in DCI
    - No MACSec to protect against threats such as PFC frame insertion or privacy protection
    - Hard to allocate proper buffer to save buffer resource
  - Need to specify internal interface to enable both data and control channel for PFC and a headroom measurement mechanism.
- Proposed scope of work
  - Update DCBX to discover the capability and auto-enable the feature. Address MBC inconsistency
  - LLDP + Pdelay procedure to measure delay
  - State machines and protocol description
  - Updates to DCBX MIBs and YANG
  - Enhanced descriptions in Annex M & N
  - Define shim/mux layer in 802.1 Q instead of 802.1AE, in order to unbind shim layer from MACSec.
  - Document the PFC propagation model as opposed to allowing PFC frames to be forwarded transparently.
- Market potential
- Technical feasibility
  - New shim/mux layer mapping M_CONTROL primitives and/or M_UNITDATA primitives
  - Explain Option 3 for headroom measurement