PFC VS. CREDIT BASED FLOW CONTROL BUFFER REQUIREMENTS

Vahid Tabatabaee (Broadcom)
Chiara Piglione (Broadcom)
Pat Thaler (Broadcom)
Bruce Kwan (Broadcom)
5/14/2014

- **Objective**

- **Assumptions and Approach**

- **Analysis**

- **Conclusion**

OBJECTIVE

- **Compare buffer requirement for PFC and proposed Credit Based Flow Control (CBFC)**

  - Focus is on baseline buffer sizing for bare minimum requirements of a switch

  - This is not a detailed analysis for total switch buffer size requirements

OUTLINE

- **Objective**

- **Approach and Assumptions**

- **Analysis**

- **Conclusions**

- **Compare buffer size for a switch with PFC vs. CBFC for the following**

  - Baseline requirements

    - Lossless guarantees

    - High throughput

      - Line rate port pair streaming

    - Cut through switching

  - Switching buffer to control congestion spread

ASSUMPTIONS

- **Network Level**

  - MTU = 2KB

  - PFC Packet Size = 64B

  - CBFC Credit Packet Size = 64B

  - Cable Length = 100m

  - 4 Lossless class

- **Switch Level**

  - Buffer Unit = 100B

  - Credit Unit Size for CBFC = 64B

  - PFC response delay: IEEE Compliant

CBFC BASELINE OPERATION

- **At initialization**
  - Data Receiver advertises number of available credits to the Data Sender for each traffic class
  - Credit = 64-bytes

- **Data Sender**
  - Scheduler starts packet transmission if it has positive credits
  - Sender decrements credits when it sends packets

- **Data Receiver**
  - Upon reception of packet, increment buffer use count
  - Upon dequeue, decrement buffer use count and return credits to Data Sender

64B credits

Data sender   **Data Receiver**

Data packets
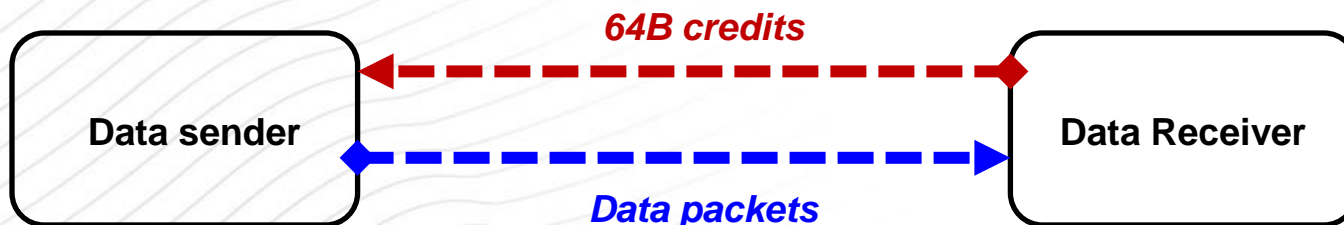
CBFC IMPACT ON SHARING BUFFER FOR LOSSLESS CLASSES

- **Basic CBFC does NOT allow buffer sharing for lossless traffic classes**

  - In order to guarantee lossless, the total buffer allocated for lossless classes must be consistent with the number of credits advertised for them at initialization

OUTLINE

- **Objective**

- **Approach and Assumptions**

- **Analysis**

- **Conclusions**
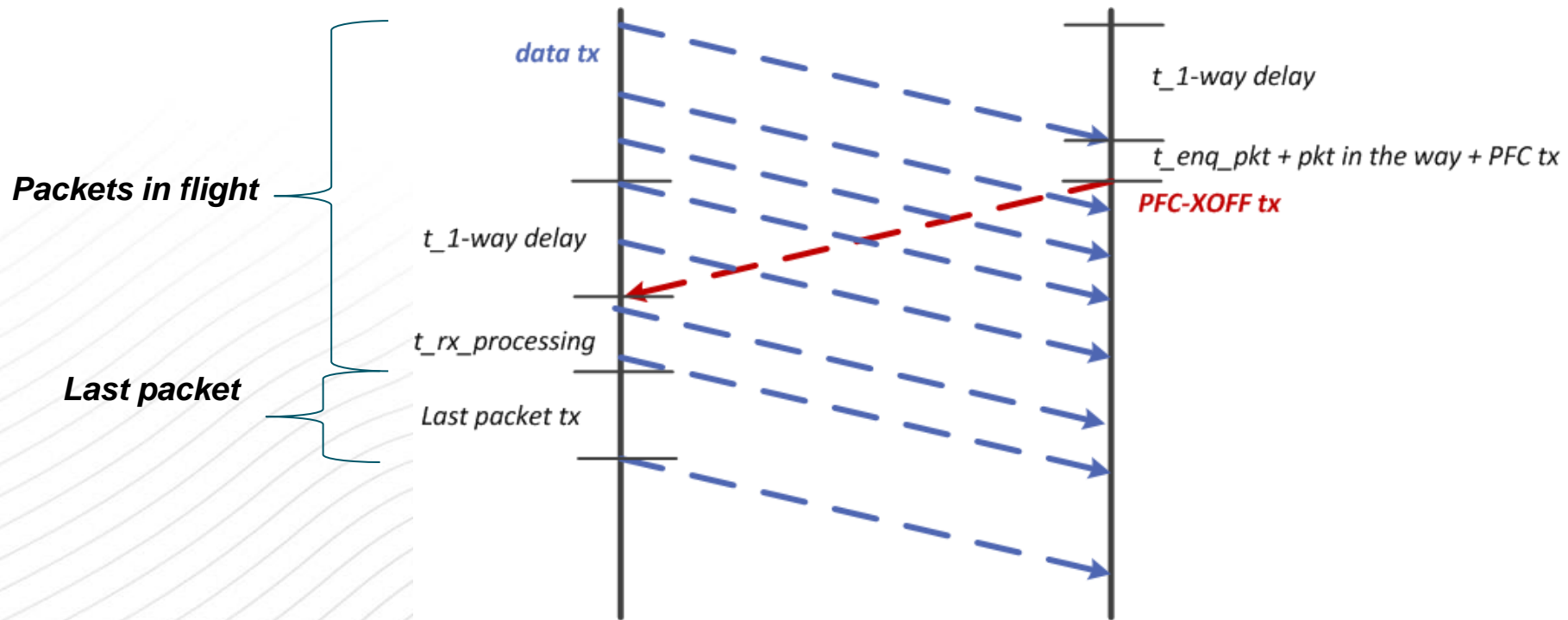
PFC SETTING FOR BASELINE BUFFER SIZING

- **Headroom Buffer to Support Lossless**
  - Absorbs packets in flight + last packet sent

- **Switching Buffer for Line-rate Port-pair Streaming**
  - 1 MTU per port (assuming cut-through is used)

- **Additional Switching Buffer for Cut-through**
  - None: 1 MTU switching buffer per port is sufficient to support port pair streaming and cut-through

- **Headroom is the main buffer size driver for PFC setting**
  - Must absorb packets in flight + last packet sent

BASELINE BUFFER SIZING FOR PFC

- **Switch**

  - 8 x 100G switch

- **Requirement assumptions**

  - Support 4 lossless classes

    - Support lossless for stream of 64B packets

  - Support line rate port pair streaming for 64B packets

  - Support cut-thru

- **MTU**

  - 2KB

	Headroom to support lossless for 64B packets	Switching Buffer for line port pair streaming	Switching buffer to support cut-through	Total
PFC	1.196MB	16.4 KB	0	1.21 MB

CBFC SETTING FOR BASELINE BUFFER SIZING

- **Headroom Buffer to Support Lossless**

  - None

- **Switching Buffer for Line-rate Port-pair Streaming**

  - Absorbs packets in flight

- **Additional Switching Buffer for Cut-through**

  - 1 MTU per (port, lossless class)

    - Packet size is not known upon arrival of packet to sender

    - Sender starts packet cut-through if it has positive credit before full size of packet is known

    - To insure sufficient buffer at receiver, additional buffer for MTU size packet is needed per (port, lossless class)

- **Switching buffer for line-rate port pair streaming is the main buffer size driver for CBFC setting**
  - Must absorb packets in flight
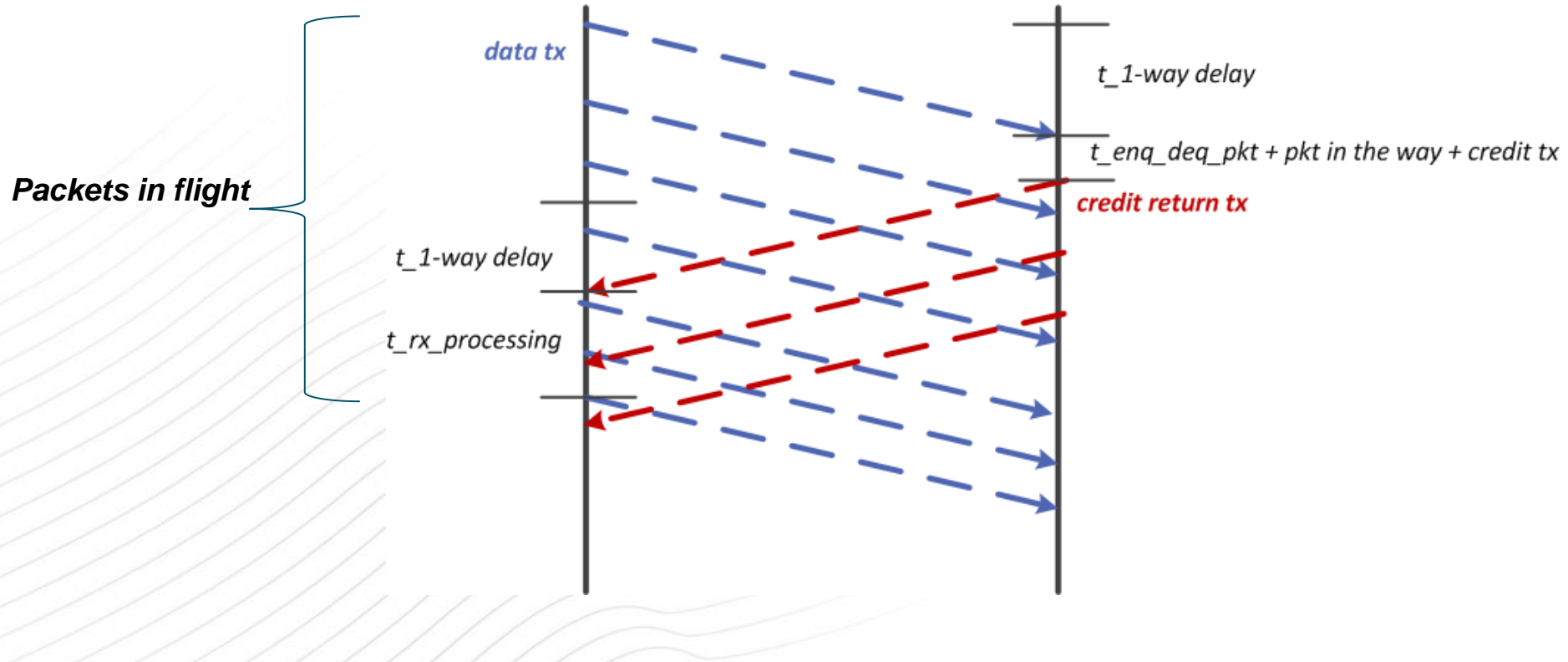
BASELINE BUFFER SIZING FOR CBFC

- **Switch**
  - 8 x 100G switch

- **Requirement assumptions**
  - Support 4 lossless classes
    - Support lossless for stream of 64B packets
  - Support line rate port pair streaming for 64B packets
  - Support cut-thru

- **MTU**
  - 2KB

	Headroom to support lossless for 64B packets	Switching Buffer for line port pair streaming	Switching buffer to support cut-through	Total
CBFC	0	1.13 MB	65.60 KB	1.196 MB

BASELINE BUFFER REQUIREMENT COMPARISON

- **CBFC provides little to no benefit over PFC (1.3% in this example) for baseline buffer requirements**

	Headroom to support lossless for 64B packets	Switching Buffer for line port pair streaming	Switching buffer to support cut-through	Total
PFC	1.196MB	16.4 KB	0	1.212 MB
CBFC	0	1.13 MB	65.60 KB	1.196 MB

Savings does not warrant adding CBFC and compromising Ethernet simplicity
by imposing need to support 2 different link level flow control protocols

SWITCHING BUFFER TO CONTROL CONGESTION SPREAD

- **Additional switching buffer is required to control congestion spread in networks**

  - Additional switching buffer, allows for burst absorption when there is congestion in switch without stopping traffic flow from sender and spreading the congestion

- **Comparison of PFC and CBFC**

  - **PFC**: Inherently enables buffer sharing between (port, traffic class)
    - Highly efficient in utilization of switch buffer for burst absorption
  - **CBFC**: Inhibits buffer sharing between (port, traffic class)
    - Inefficient utilization of congestion buffer for burst absorption

> *PFC allows for sharing and higher utilization of switching buffer and controls congestion spread more efficiently*

OUTLINE

- **Objective**

- **Approach and Assumptions**

- **Analysis**

- **Conclusions**

CONCLUSION

- **Do not pursue CBFC as it compromises simplicity of Ethernet for little to no benefit**

  - No real benefit for baseline buffer requirements

  - CBFC is less efficient than PFC in controlling congestion spread

    - PFC takes advantage of sharing buffer where as baseline CBFC allocates buffer statically per (port, lossless class)