Impact of expanding maximum frame size to PAUSE Flow Control operation

Kevin Daines
World Wide Packets

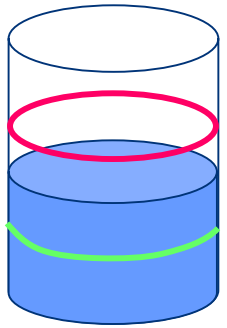
30 September 2004

Outline

- Review PAUSE operation
- Sample buffering requirements
- Analyze impact of increasing frame length to buffering requirements
- Workarounds
- Conclusion

PAUSE XON/XOFF operation

- PAUSE Flow Control may be used to prevent frame loss due to ingress buffer overflow
- Operation
  - Device A monitors ingress buffer depth
  - Before ingress buffer fills (high watermark), causing frame loss, device A sends XOFF PAUSE Flow Control message to device B
  - Device B receives XOFF and ceases sending traffic
  - When ingress buffer drains (low watermark), device A sends XON PAUSE Flow Control to device B
  - Device B receives XON and resumes sending traffic



Ingress buffer requirements

- To prevent frame loss due to lack of available ingress buffer, device A needs to send an XOFF PAUSE Flow Control frame and still receive, worst case:

Link speed	Port Type	Link length (m)	Octets
100 Mb/s	100BASE-FX	2,000	3,426
1 Gb/s	1000BASE-LX	5,000	9,486
1 Gb/s	1000BASE-LX10	10,000	15,800
10 Gb/s	10GBASE-ER	40,000	512,000

High watermark math

- **Calculation of watermark placement from top of FIFO:**

	1522	*Maximum frame (sent just before watermark)*
	64	*XOFF frame*
	varies	*Trip delay*
	varies	*Response time to react to XOFF*
	1522	*Maximum frame (sent just before XOFF Response)*
+	*varies*	*Trip delay*
	Total	*Octets to receive after crossing upper watermark*

Detail behind numbers

In octets (except where noted)	100BASE-FX	1000BASE-LX	1000BASE-LX10	10GBASE-ER
Maximum frame	1,522			
XOFF frame	64			
c (m/s)	*300,000,000*			
Fiber delay estimate	*0.66 of c*			
Fiber delay (BT/m)	*0.505*	*5.051*		*50.505*
Round-trip delay (BT/m)	*1.01*	*10.101*		*101.010*
Trip delay	140	3,157	6,314	252,526
Response time	64			3,840
Maximum frame	1,522			
Trip delay	140	3,157	6,314	252,526
Total octets to receive	3,426	9,486	15,800	512,000

Impact of larger frames

- **Additional space needed between upper watermark and top of ingress buffer:**

In octets (except where noted)	100BASE-FX	1000BASE-LX	1000BASE-LX10	10GBASE-ER
Maximum frame = 1522	3,426	9,486	15,800	512,000
Maximum frame = 2048[1]	4,478	10,538	16,852	513,052
Increase	*1,052*	*1,052*	*1,052*	*1,052*
Increase (%)	*30.7%*	*11.1%*	*6.6%*	*0.2%*

[1] Assumes 2,048 as new maximum frame length. This is TBD.

Possible workarounds

- Upper watermark could be moved lower in FIFO to account for additional 1,052[1] octets
- Depending on size of ingress buffer, two outcomes possible:
  - No change in behavior, it works!
  - Slightly lower efficiency when ingress buffer falls below existing lower watermark triggering XON frame
    - If space in ingress buffer allows, lower watermark could be moved higher

[1] Assumes 2,048 as new maximum frame length. This is TBD.

Conclusion

- **Impact of adding up to 526 octets has nominal impact on newer (*faster/longer*) links**
  - **Buffering requirements dominated by round trip delay of medium**
- **With programmable ingress buffer watermarks, workaround exists**
- **Limited use of PAUSE Flow Control does not warrant further analysis or concern**