

Link Configuration of Pause Function

Howard M. Frazier, Jr.
Cisco Systems, Inc.
Workgroup Business Unit
28-January-1997

Howard M. Frazier, Jr.	IEEE P802.3z Gigabit Task Force	Page 1 of 15 28-January-1997
------------------------	------------------------------------	---------------------------------

Outline

- Background
- Issues
- Suggested Remedy
- Analysis
- Mismatch Handling
- Summary

Howard M. Frazier, Jr.	IEEE P802.3z Gigabit Task Force	Page 2 of 15 28-January-1997
------------------------	------------------------------------	---------------------------------

Background

- At the Vancouver meeting of IEEE P802.3z, the Task Force adopted a motion to include a minimal specification for asymmetric flow control to permit operation of Buffered Distributors
- The Task Force also received a presentation which described a proposal for configuring asymmetric flow control
- This proposal defined two bits in the link configuration word
 - One bit was named “Pause”, and defined to have the same meaning as the “Pause Operation for Full-Duplex Links” bit in the clause 28 Auto-negotiation base link code word (as modified by 802.3x)
 - The second bit was named “ASYM_DIR”, and if ASYM_DIR was set, the meaning of the “Pause” bit would be modified to indicate asymmetric configurations
- The proposal also contained a table reflecting the 16 possible combinations of “Pause” and ASYM_DIR that could be exchanged by the two ends of a link

Background

- I assume that a goal of the proposal was to maintain compatibility with the clause 28 definition (as modified by 802.3x) of the base link code word
- A virtue of this approach is that asymmetric flow control negotiation could conceivably be applied to 10/100BASE-T full duplex links
 - It was not stated, but possibly assumed, that the “Pause” and ASYM_DIR bits would be assigned to the same bit positions in the 1000BASE-X link configuration word that they would occupy in the clause 28 base link code word

Issues

- This approach has disadvantages
- The interpretation of the “Pause” and ASYM_DIR bits is non-intuitive, and requires extra logic to resolve, extra words to specify, and extra effort to configure
- The approach cannot meet the assumed goal of compatibility with clause 28 Auto-negotiation
 - “legacy” devices would not parse ASYM_DIR, and would not realize that, when set, this bit modifies the behavior of the “Pause” bit
- Rich Taborek’s proposal for the format of the Link Configuration code word assigned two bits (7:8) to PS1 and PS2, respectively
 - This would be in conflict with the assumed goal of the “Pause” + ASYM_DIR proposal, since the “Pause Operation for Full-Duplex Links” bit in clause 28 is assigned to bit 10 of the base link code word

Suggested Remedy

- Use two bits: Pause(t) and Pause(r)
 - Pause(t) = 1 indicates that the device might transmit Pause frames
 - Pause(r) = 1 indicates that the device can distinguish Pause frames, and implements the Pause function
 - Pause(t) = 0 indicates that the device will not transmit Pause frames
 - Pause(r) = 0 indicates that the device cannot distinguish Pause frames, and does not implement the Pause function
- Assign Pause(t), Pause(r) to bits 7,8 of the Link Configuration word

Suggested Remedy (cont)

- There are four possible combinations of the two bits, which can be discussed using the short hand:

Table 1—Shorthand Notation for Meaning of Pause(t) and Pause(r)

Pause(t)	Pause(r)	Shorthand	Description
0	0	N	Won't transmit them, can't receive them
0	1	R	Won't transmit them, can receive them
1	0	T	Might transmit them, can't receive them
1	1	B	Might transmit them, can receive them

- Each end of the link can resolve the configuration based on the simple formulae:
 - Allowed_to_transmit_Pause = local_Pause(t) & partner_Pause(r)
 - Expect_to_receive_Pause = local_Pause(r) & partner_Pause(t)

Suggested Remedy (cont)

- Consider the 16 possible cases that can occur between a local device and a link partner
- Table 2 shows what each end advertises
 - Local Advertise and Partner Advertise
- and the resulting Link Configuration
 - Local Config and Partner Config

Proposal

Table 2—Resolution of Link Configuration

Case	Local Advertise	Partner Advertise	Local Config	Partner Config
a	N	N	N	N
b	N	R	N	N
c	N	T	N	N
d	N	B	N	N
e	R	N	N	N
f	R	R	N	N
g	R	T	R ←	T
h	R	B	R ←	T
i	T	N	N	N
j	T	R	T →	R
k	T	T	N	N
l	T	B	T →	R
m	B	N	N	N
n	B	R	T →	R
o	B	T	R ←	T
p	B	B	B ↔	B

Analysis

- Cases a-f and i, k, m result in a link where Pause frames are not allowed to be sent in either direction
- Cases g, h and o result in a link where Pause frames can flow only from the partner (remote) end to the local end
- Cases j, l and n result in a link where Pause frames can flow only from the local end to the partner (remote) end
- Case p results in a link where Pause frames can flow in both directions

Analysis (cont)

- A device can implement a policy which restricts the configurations it will accept
- A Buffered Distributor, for instance, might always advertise “T”
 - Might transmit them, can’t receive them
- Therefore, the Buffered Distributor could wind up in case i, j, k or l
 - Depending on what the link partner advertises

Mismatch Handling

- If the partner is another Buffered Distributor advertising “T”:
 - The result would be case k
 - Either Buffered Distributor could reject the link, if they are not willing to accept a link without flow control
- The link is rejected by withholding C/ack transmission
 - Keep sending configuration word without the ACK bit set, thus preventing the link from coming up
- The link partner can be notified of the problem using the RF bits
 - Right now, we have two RF bits, based on Rich Taborek’s presentation from Vancouver

Mismatch Handling

- **Current Remote Fault Encoding:**

Table 3—Current Encoding of RF1 and RF2

RF1	RF2	Description
0	0	No error, link OK
0	1	Offline
1	0	Link Failure
1	1	Link Error

- **The meaning and purpose of the “Link Error” encoding is unclear**
 - **We aren’t going to recycle through the link configuration state machine every time there is a link error, so this encoding has no application**

Mismatch Handling

- **Suggest a new application for this encoding**
 - **Link Configuration Error**
- **Use this for mismatches like the case described above**
- **Make the following modification to the Remote Fault encoding:**

Table 4—Suggested Encoding of RF1 and RF2

RF1	RF2	Description
0	0	No error, link OK
0	1	Offline
1	0	Link Failure
1	1	Link Configuration Error

- **Withholding C/ack and signaling RF1,RF2 = Link Configuration Error provides a negative acknowledgement (C/nack) to the partner**

Summary

- The Pause(t) and Pause(r) bit semantics are much easier to understand and interpret than Pause + ASYM_DIR
- The Pause(t) and Pause(r) mechanism
 - Is general purpose
 - Provides a mechanism by which a device can reject an undesired link configuration, and notify the link partner
- The net effect is a much simpler scheme for negotiating flow control, including the asymmetric configurations