To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 3 November 2004
Subject: 04-363r0 SAS-1.1 DONE Timeout vs ACK NAK Timeout race condition note

Revision history
Revision 0 (3 November 2004) First revision

Related documents
sas1r06 - Serial Attached SCSI 1.1 revision 6

Overview
As has been mentioned in an editor's note in the last few revisions of SAS-1.1, a note needs to be added explaining a case where DONE Timeout can race ACK/NAK Timeout after an error occurs (a lost SOF).

The scenario is:

1) Connection is open.
2) Target sends DONE indicating it has no more frames to send.
3) Initiator continues sending write DATA frames (non-interlocked).
4) One frame's SOF is lost.
5) After sending the last non-interlocked frame, the initiator waits for an ACK that is never coming, so in 1 ms experiences an ACK/NAK Timeout.
6) Meanwhile, after receiving the last non-interlocked frame, the target's DONE Timeout timer runs and expires in 1 ms because no DONE or new SOF shows up.

The target ends up sending BREAK concurrently with the intiator sending DONE (ACK/NAK TIMEOUT) - a race condition. Regardless of which one happens first, though, the result is the same, since "connection lost without ACK/NAK" is handled the same as an ACK/NAK timeout.

If transport layer retries are not being used, the initiator aborts the command. If it sent DONE (ACK/NAK TIMEOUT) first, it does so because of the ACK/NAK timeout; if it receives the BREAK first, it does so because the transport layer receives Transmission Status (Connection Lost Without ACK/NAK) for the last frame.

If transport layer retries are being used, the write data frames since the previous XFER_RDY are retransmitted.

The same is true if this happens in the other direction, and the target was sending non-interlocked read DATA frames after the initiator sent DONE.

Suggested changes

7.16.7.5 SSP_D (DONE control) state machine

The SSP_D state machine's function is to ensure a DONE has been received and transmitted before the SL_CC state machine disables the SSP state machines. This state machine consists of one state.

This state machine ensures that a DONE is received and transmitted before the connection is closed. The DONE may be transmitted and received in any order.

If the DONE Received message has been received before the Transmitted DONE message is received, this state machine shall send the Request Close message to the SL_CC state machine (see 7.14) and all the SSP state machines after receiving the Transmitted DONE message.

If a DONE Received message, the Transmitted DONE (Normal) message, or the Transmitted DONE (Credit Timeout) message has not been received and the Rx Credit Status (Extended) message or the Rx Credit Control (Blocked) message has been received, then this state shall initialize and start the DONE Timeout timer after receiving the Transmitted DONE (Normal) message or the Transmitted DONE (Credit Timeout) message.

If the DONE Received message has not been received and the Transmitted DONE (Normal) message or the Transmitted DONE (Credit Timeout) message has been received, this state machine shall initialize and start the DONE Timeout timer each time:

a) the Rx Credit Status (Extended) message is received; or
b) the Rx Credit Control (Blocked) message is received.

If the Transmitted DONE (Normal) message or the Transmitted DONE (Credit Timeout) message has been received, the DONE Timeout timer shall be reinitialized each time the EOF Received message is received.

If the Transmitted DONE (Normal) message or the Transmitted DONE (Credit Timeout) message has been received, the DONE Timeout timer shall be stopped after:

a) the Rx Credit Status (Exhausted) message is received; and
b) the Rx Credit Control (Blocked) message has not been received.

NOTE 1 Stopping the timer ensures that, if credit remains exhausted long enough that the Credit Timeout timer of the other phy in the connection expires, the other phy is able to transmit a DONE (CREDIT TIMEOUT).

If the DONE Received message has not been received and the Transmitted DONE (ACK/NAK Timeout) message has been received:

a) this state machine shall initialize and start the DONE Timeout timer; and
b) this state shall not reinitialize the DONE Timeout timer if an EOF Received message is received.

If the DONE Received message is received before the DONE Timeout timer expires, this state machine shall send the Request Close message to the SL_CC state machine and all the SSP state machines.

If the DONE Received message is not received before the DONE Timeout timer expires, this state machine shall:

a) send a DONE Timeout confirmation to the port layer; and
b) send a Request Break message to the SL_CC state machine and all the SSP state machines.

Any time a DONE Received message is received this state machine shall send a DONE Received confirmation to the port layer. A DONE Received (ACK/NAK Timeout) confirmation informs the port layer that the SSP transmitter is going to close the connection within 1 ms; other DONE Received confirmations (e.g., DONE Received (Close Connection) and DONE Received (Credit Timeout)) may be used by the application layer to decide when to reuse tags.

NOTE 2 The DONE Timeout timer in one phy (e.g., phy A) may expire concurrently with the ACK/NAK Timeout timer in the other phy (e.g., phy B) in a connection.

For example, if phy A receives DONE (NORMAL) indicating phy B has no more frames to transmit, and phy A then transmits a series of non-interlocked frames where one or more of the SOFs is corrupted, then phy A waits to receive all the ACKs and/or NAKs after transmitting the series of non-interlocked frames. However, since phy B did not receive the full number of SOFs, it does not transmit as many ACKs and/or NAKs as phy A is expecting. The ACK/NAK Timeout timer in phy A expires and phy A transmits DONE (ACK/NAK TIMEOUT). Meanwhile, despite having transmitted DONE, phy B stops receiving frames while phy A is waiting for the final ACKs and/or NAKs. Since phy B does not receive DONE or any more frames, its DONE Timeout timer expires and phy B transmits BREAK.

Since the timers may expire at slightly different times (e.g., due to timer resolution differences), the DONE (ACK/NAK TIMEOUT) may be transmitted before, concurrently with, or after the BREAK. Nevertheless, the phys handle the link layer error (i.e., the ACK/NAK timeout or the DONE timeout) the same way (see 9.2.4.5).