T10/07-392 revision 0

Date: September 13, 2007

To: T10 Committee (SCSI)

From: George Penokie (IBM)

Subject: SAS-2: Remove AWT reset on receipt of OPEN_REJECT (RETRY)

1 Overview

The fairness algorithms in SAS-2 have a flaw in that when a SAS device receives receipt an OPEN_REJECT (RETRY) it is required to reset the AWT to zero before sending another OPEN address frame to the SAS device that rejected the open. In a large SAS topology running a heavy workload this behavior can result in SAS devices not being able to get an open through in a reasonable length of time.

This proposal adds a NO_AWT_REZERO bit to Protocol-Specific Port mode page SAS-2.

2 Proposed SAS-2 changes

7.12.3 Arbitration fairness

SAS supports least-recently used arbitration fairness for connection requests.

Each SAS port and expander port shall include an Arbitration Wait Time timer which counts the time from the moment when the port makes a connection request until the request is accepted or rejected. The Arbitration Wait Time timer is in the port layer state machine (see 8.2.2). The Arbitration Wait Time timer shall count in microseconds from 0 µs to 32 767 µs and in milliseconds from 32 768 µs to 32 767 ms + 32 768 µs. The Arbitration Wait Time timer shall stop incrementing when its value reaches 32 767 ms + 32 768 µs.

A SAS port (i.e., SAS initiator ports and SAS target ports) shall start the Arbitration Wait Time timer when it transmits the first OPEN address frame (see 7.8.3) for the connection request. When the SAS port retransmits the OPEN address frame (e.g., after losing arbitration and handling an inbound OPEN address frame), it shall set the ARBITRATION WAIT TIME field to the current value of the Arbitration Wait Time timer.

A SAS port should set the Arbitration Wait Time timer to zero when it transmits the first OPEN address frame for the connection request. A SAS initiator port or SAS target port may be unfair by setting the ARBITRATION WAIT TIME field in the OPEN address frame (see 7.8.3) to a higher value than its Arbitration Wait Time timer indicates. However, an unfair SAS port shall not set the ARBITRATION WAIT TIME field to a value greater than or equal to 8000h; this limits the amount of unfairness and helps prevent livelocks.

The expander port that receives an OPEN address frame shall set the Arbitration Wait Time timer to the value of the incoming ARBITRATION WAIT TIME field and start the Arbitration Wait Time timer as it arbitrates for internal access to the outgoing expander port. When the expander device transmits the OPEN address frame out another expander port, it shall set the outgoing ARBITRATION WAIT TIME field to the current value of the Arbitration Wait Time timer maintained by the incoming expander port.

A SAS port shall stop the Arbitration Wait Time timer and set it to zero when it has no more frames to send.

A SAS port shall stop the Arbitration Wait Time timer and set it to zero when it receives one of the following connection responses:

a) OPEN_ACCEPT;
b) OPEN_REJECT (PROTOCOL NOT SUPPORTED);
c) OPEN_REJECT (RESERVED ABANDON 1);
d) OPEN_REJECT (RESERVED ABANDON 2);
e) OPEN_REJECT (RESERVED ABANDON 3);
f) OPEN_REJECT (STP RESOURCES BUSY);
g) OPEN_REJECT (WRONG DESTINATION);
h) OPEN_REJECT (RESERVED CONTINUE 0); or
i) OPEN_REJECT (RESERVED CONTINUE 1).~~; or~~
~~j) OPEN_REJECT (RETRY).~~

NOTE 1 - Connection responses that are conclusively from the destination phy (see table 112 and table 113 in 7.2.5.13) are included in the list. Except for OPEN_REJECT (RETRY), connection responses that are only from or may be from expander phys are not included.

A connection response of OPEN_REJECT (RETRY) shall not stop the Arbitration Wait Time timer and set it to zero if the NO AWT REZERO bit in the Protocol-Specific Port mode page (see 10.2.7.4) set to one. If the NO AWT REZERO bit set to zero a SAS port shall stop the Arbitration Wait Time timer and set it to zero.

A SAS port should not stop the Arbitration Wait Time timer and set it to zero when it receives an incoming OPEN address frame that has priority over the outgoing OPEN address frame according to table 1, regardless of whether it replies with an OPEN_ACCEPT or an OPEN_REJECT.

A SAS port should not stop the Arbitration Wait Time timer and set it to zero when it receives an incoming OPEN address frame that has priority over the outgoing OPEN address frame according to table 1, regardless of whether it replies with an OPEN_ACCEPT or an OPEN_REJECT.

When arbitrating for access to an outgoing expander port, the expander device shall select the connection request based on the rules described in 7.12.4.

If two connection requests pass on a physical link, the phy shall determine the winner by comparing OPEN address frame field contents using the arbitration priority described in table 1.

Table 1 — Arbitration priority for OPEN address frames passing on a physical link

Bits 79-64 (79 is MSB)	Bits 63-0 (0 is LSB)
ARBITRATION WAIT TIME field value	SOURCE SAS ADDRESS field value

See 7.8.3 for details on the OPEN address frame and the ARBITRATION WAIT TIME field.

10.2.7.4 Protocol-Specific Port mode page

The Protocol-Specific Port mode page (see SPC-4) contains parameters that affect SSP target port operation. If the mode page is implemented by one logical unit in a SCSI target device, it shall be implemented by all logical units in the SCSI target device that support the MODE SELECT or MODE SENSE commands.

The mode page policy (see SPC-4) for this mode page shall be either shared or per target port. If a SAS target device has multiple SSP target ports, the mode page policy should be per target port.

Parameters in this mode page:

a) shall affect all phys in the SSP target port if the mode page policy is per target port; and
b) shall affect all SSP target ports in the SAS target device if the mode page policy is shared.

Table 2 defines the format of the page for SAS SSP.

Table 2 — Protocol-Specific Port mode page for SAS SSP

Byte\Bit	7	6	5	4	3	2	1	0
0	PS	SPF (0b)	PAGE CODE (19h)					
1	PAGE LENGTH (06h)							
2	Reserved	NO AWT REZERO	BROADCAST ASYNCHRONOUS EVENT	READY LED MEANING	PROTOCOL IDENTIFIER (6h)			
3	Reserved							
4	(MSB)	I_T NEXUS LOSS TIME						
5								(LSB)
6	(MSB)	INITIATOR RESPONSE TIMEOUT						
7								(LSB)

The PARAMETERS SAVEABLE (PS) bit is defined in SPC-4.

The SUBPAGE FORMAT (SPF) bit shall be set to zero for access to this mode page.

The PAGE CODE field shall be set to 19h.

The PAGE LENGTH field shall be set to the number of bytes in the page after the PAGE LENGTH field (i.e., 06h).

A NO AWT REZERO bit set to one specifies that the a SAS port shall not stop the Arbitration Wait Time timer and set it to zero when it receives an OPEN_REJECT (RETRY). A NO AWT REZERO bit set to zero specifies that the a SAS port shall stop the Arbitration Wait Time timer and set it to zero when it receives an OPEN_REJECT (RETRY).

A BROADCAST ASYNCHRONOUS EVENT bit set to one specifies that the device server shall enable origination of Broadcast (Asynchronous Event) (see 4.1.13). A BROADCAST ASYNCHRONOUS EVENT bit set to zero specifies that the device server shall disable origination of Broadcast (Asynchronous Event).

The READY LED MEANING bit specifies the READY LED signal behavior (see 10.4.1). Regardless of the mode page policy (see SPC-4) for this mode page, the shared mode page policy shall be applied to the READY LED MEANING bit.

The PROTOCOL IDENTIFIER field shall be set to 6h indicating this is a SAS SSP specific mode page.

The I_T NEXUS LOSS TIME field contains the time that the SSP target port shall retry connection requests to an SSP initiator port that are rejected with certain responses indicating that the SSP initiator port may no longer be present (see 8.2.2) before recognizing an I_T nexus loss (see 4.5). Table 3 defines the values of the I_T NEXUS LOSS TIME field. If this mode page is not implemented, the I_T nexus loss time is vendor specific. This value is enforced by the port layer (see 8.2.2).

Table 3 — I_T NEXUS LOSS TIME field

Code	Description
0000h	Vendor-specific amount of time.
0001h to FFFEh	Time in milliseconds.
FFFFh	The SSP target port shall never recognize an I_T nexus loss (i.e., it shall retry the connection requests forever).

NOTE 2 - If this mode page is implemented, the default value of the I_T NEXUS LOSS TIME field should be non-zero. It is recommended that this value be 2 000 ms.

NOTE 3 - An SSP initiator port should retry connection requests for the time indicated by the I_T NEXUS LOSS TIME field in the Protocol-Specific Port mode page for the SSP target port to which it is trying to establish a connection (see 4.5).

The INITIATOR RESPONSE TIMEOUT field contains the time in milliseconds that the SSP target port shall wait for the receipt of a frame (e.g., a write DATA frame) before aborting the command associated with that frame. An INITIATOR RESPONSE TIMEOUT field value of zero indicates that the SSP target port shall disable the initiator response timeout timer. If this mode page is not implemented, the logical unit shall not implement an initiator response timeout timer. This value is enforced by the transport layer (see 9.2.6.3).