To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 18 October 2004
Subject: 04-341r0 SAS-1.1 Do not reset Arbitration Wait Time timer on incoming OPEN

Revision history
Revision 0 (18 October 2004) First revision

Related documents
sas1r06 - Serial Attached SCSI 1.1 revision 6

Overview
A livelock exists if:

a) a target is trying to open a connection to send a frame to an initiator the same time the initiator tries to open a connection to the target;
b) the OPENs cross on the wire;
c) the initiator's OPEN wins arbitration;
d) the target replies with OPEN_REJECT (RETRY); and
e) they try again.

SAS-1.1 currently requires the initiator (the winner) to reset its Arbitration Wait Time timer, because it won the crossing-on-the-wire comparison and got a response (either OPEN_ACCEPT or OPEN_REJECT) from the target (the loser).

The main text could be interpreted as also requiring the target (the loser) to reset its Arbitration Wait Time timer; it says it must do so "if the incoming connection request satisfies [the target's] arbitration request" but doesn't differentiate between the target rejecting rather than accepting the request. If the target does reset its Arbitration Wait Time timer, then a livelock occurs, as the same scenario occurs over and over.

Suggested fix (part 1): The loser should not reset its AWT just because the incoming request (accepted or not) is from the same port the loser was trying to open. If the loser rejects the requst, it must not reset its AWT.

If the loser accepts the request, the winner is not required to grant RRDY credit - only the loser must grant RRDY credit. The loser could be starved from ever sending a frame if each time it sends its OPEN, it sets the ARBITRATION WAIT TIME field set to 0, loses, and the winner doesn't grant it credit. If the winner is an initiator, it is eventually required to grant credit (a target is not), but it doesn't necessarily have to do so in an outbound connection request.

The port layer state machine does includes a note mentioning that a Tx Open may be created from a Retry Open if credit is not granted (i.e. the Arbitration Wait Time timer can keep running if no credit is received).

Suggested fix (part 2): promote that rule from a note into the state machine text, and change it to a shall. The loser should only reset its Arbitration Wait Time timer if it is not able to send a frame on the connection that is established.

Suggested changes

7.12.3 Arbitration fairness

SAS supports least-recently used arbitration fairness.

Each SAS port and expander port shall include an Arbitration Wait Time timer which counts the time from when the port makes a connection request until its request is granted. 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.

SAS ports (i.e., SAS initiator ports and SAS target ports) shall start the Arbitration Wait Time timer (see 8.2.2) when they transmit 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. SAS ports should set the Arbitration Wait Time timer to zero when they transmit 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, unfair SAS ports 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 port shall stop the Arbitration Wait Time timer and set it to zero when it wins arbitration (i.e., it receives either OPEN_ACCEPT or OPEN_REJECT from the destination SAS port rather than from an intermediate expander device). ~~A port shall stop the Arbitration Wait Time timer when it loses arbitration to a connection request that satisfies its arbitration request (i.e., it receives an OPEN address frame from the destination SAS port with the INITIATOR PORT bit set to the opposite value and a matching PROTOCOL field).~~If a port receives a connection request that satisfies its arbitration request (i.e., it receives an OPEN address frame from the destination SAS port with the INITIATOR PORT bit set to the opposite value and a matching PROTOCOL field), it shall not stop the Arbitration Wait Time timer unless it accepts the request and it is able to make forward progess for the arbitration request in the connection (e.g., the source phy provides RRDY credit so the port is able to transmit a frame).

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.

8.2.2 PL_OC (port layer overall control) state machine

8.2.2.3.2 PL_OC2:Overall_Control state establishing connections

...

If this state receives a Retry Open message and there are pending Tx Frame messages for which pending Tx Open messages have not been created, then this state should create a pending Tx Open message from the Retry Open message.

If this state does not create a pending Tx Open message from a Retry Open message (e.g., the current number of pending Tx Open messages equals the number of phys), then this state shall discard the Retry Open message. This state may create a new pending Tx Open message at a later time for the pending Tx Frame message that resulted in the Retry Open message.

If this state receives a Retry Open (Opened By Destination) message ~~and~~, the initiator and protocol arguments match those in the Tx Open messages that resulted in the Retry Open message, then:

a) if credit is received in the connection for transmission of at least one frame, then this state ~~may discard the Retry Open message and~~ shall use the established connection to send at least one pending Tx Frame message~~s~~ as a Tx Frame message~~s~~ to the destination SAS address. If this state has a pending Tx Open slot available (e.g., in a wide port) and still has pending Tx Frame message(s) in addition to the one destined for that connection, this state may either discard the Retry Open message or create a pending Tx Open message from the Retry Open message; and

b) If credit is not received in the connection, this state shall create a pending Tx Open message from the Retry Open message in order to establish a connection where credit is granted (i.e., it shall keep running its Arbitration Wait Time timer and reissue the connection request).

~~If this state receives a Retry Open (Opened By Destination) message, then, if this state has a pending Tx Open slot available, this state may create a pending Tx Open message from the Retry Open message.~~

~~NOTE 40 If a connection is established by another port as indicated by a Retry Open (Opened By Destination) message, credit may not be granted for frame transmission. In this case this state may create a pending Tx Open message from a Retry Open message in order to establish a connection where credit is granted.~~

If a connection is established by another port as indicated by a Retry Open (Opened By Destination) message, credit may not be granted for frame transmission. In this case this state may create a pending Tx Open message from a Retry Open message in order to establish a connection where credit is granted.