To: T10 Technical Committee
From: Tim Hoglund (Tim.Hoglund@lsil.com)
Date: 06 September 2003
Subject: T10/03-311r0 BackoffRetry priority for XL state machine

Revision History
Revision 0 (06 September 2003) first revision

Related Documents
sas1r00 - Serial Attached SCSI 1.1 revision 0

Overview
A potential livelock condition exists when OPEN address frames cross on a wire between two expanders. Because each source phy maintains (continues to increment) the AWT timer, the priority of a path request following the Backoff Retry condition may be elevated improperly. This can cause the ECM to select the wrong winner during arbitration, leading to a potential livelock condition.

A solution is provided which adds an additional argument to the Request Path request (expander phy to ECM connection request) in order to instruct the ECM to modify the arbitration priority for the path request for the phy which initiated the Backoff Retry condition.

Figure 1 illustrates a livelock case due to the re-arbitration that occurs after a Backoff and Retry condition.

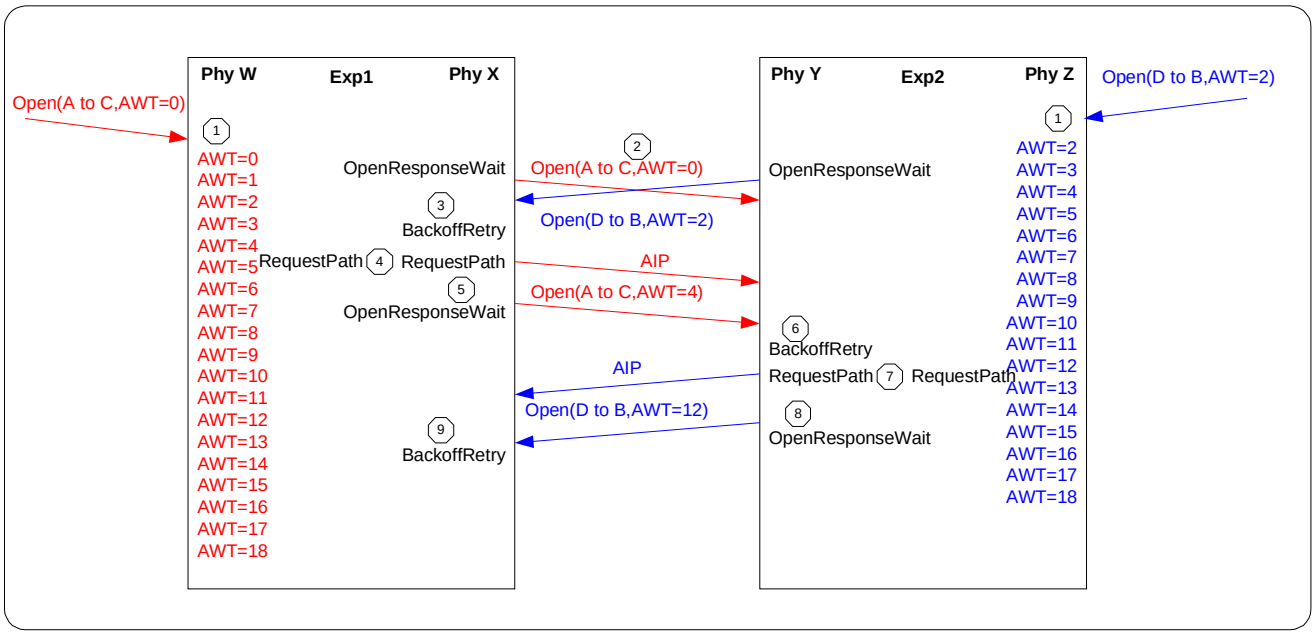

Figure 1: Livelock condition due to re-arbitration

1. Exp1[W] receives Open (A to C, AWT=0). Exp2[Z] receives Open(D to B, AWT=2).
2. Exp1[W] and Exp2[Z] both win arbitration and forward their Open address frame with the current AWT value to the other expander.
3. Exp1[X] receives a higher priority Open address frame and proceeds to XL1:Request_Path to rearbitrate on behalf of this new Open address frame.
4. Exp1[W] and Exp1[X] arbitrate and Exp1[W] wins because AWT has continued to increment at Exp1[W].
5. Exp1[W] forwards Open(A to C,AWT=4) through Exp1[X] to Exp2[Y].
6. Exp2[Y] receives a higher priority Open address frame and proceeds to XL1:Request_Path to rearbitrate.
7. Exp2[Y] and Exp2[Z] arbitrate and Exp2[Z] wins because AWT has continued to increment at Exp2[Z].
8. Exp2[Z] forwards Open(D to B,AWT=12) through Exp2[Y] to Exp1[X].
9. LIVELOCK...

Suggested Changes
7.12 Connections
7.12.3 Arbitration fairness

Delete the following text – it is redundant and inconsistent with 7.12.4.1

~~When arbitrating for access to an outgoing expander port, the expander device shall select the connection request from the expander port with the largest Arbitration Wait Time timer value. If the largest Arbitration Wait timer values are identical, then the connection request with the largest SOURCE SAS ADDRESS shall win arbitration.~~

7.12.4 Arbitration and resource management in an expander device
7.12.4.1 Arbitration overview

~~Each path request contains the Arbitration Wait Time and the Source SAS Address arguments from the received OPEN address frame.~~

~~If two or more path requests contend, the winner shall be selected by comparing the OPEN address frame contents using the arbitration priority described in table 79.~~

~~Table 79 — Arbitration priority for contending path requests in the ECM~~

~~Bits 83-68 (83 is MSB)~~	~~Bits 67-5~~	~~Bits 3-0 (0 is LSB)~~
~~ARBITRATION WAIT TIME field value~~	~~SOURCE SAS ADDRESS field value~~	~~CONNECTION RATE field value~~

Each path request contains the Arbitration Wait Time, Source SAS Address, and Connection Rate arguments from the received OPEN address frame.

Each path request also contains a Retry Priority status argument. The Retry Priority status argument is used to prevent Arbitration Wait Time from being considered during the arbitration which occurs after a BackoffRetry response is sent by an expander phy (see 7.15.4).

An expander phy shall set the path request Retry Priority status argument to RETRY PRIORITY IGNORE AWT when it requests a path after it has forwarded an OPEN address frame to a physical link and an OPEN address frame is received with higher arbitration priority and the destination SAS address and connection rate of the received OPEN address frame are not equal to the source SAS address and connection rate of the transmitted OPEN address frame (see 7.15.4 and 7.15.9); otherwise the Retry Priority status argument shall be set to RETRY PRIORITY NORMAL.

If two or more path requests contend and all of the path requests include a Retry Priority status equal to RETRY PRIORITY NORMAL, the winner shall be selected by comparing OPEN address frame contents using the arbitration priority described in table 79.

Table 79 — Arbitration priority for contending path requests in the ECM (all path requests include Retry Priority status equal to RETRY PRIORITY NORMAL)

Bits 83-68 (83 is MSB)	Bits 67-5	Bits 3-0 (0 is LSB)
ARBITRATION WAIT TIME field value	SOURCE SAS ADDRESS field value	CONNECTION RATE field value

If two or more path requests contend and any of the path requests include a Retry Priority status equal to RETRY PRIORITY IGNORE AWT, the winner shall be selected among the set of path requests which include a

Retry Priority status equal to RETRY PRIORITY IGNORE AWT, using the arbitration priority described in table 80.

Table 80 — Arbitration priority for contending path requests in the ECM (at least one path request includes Retry Priority status equal to RETRY PRIORITY IGNORE AWT)

Bits 67-5	Bits 3-0 (0 is LSB)
SOURCE SAS ADDRESS field value	CONNECTION RATE field value

7.15 XL (link layer for expander phys) state machine
7.15.4 XL1:Request_Path state
7.15.4.1 State description
This state is used to arbitrate for connection resources and to specify the destination of the connection.

Upon entry into this state, this state shall repeatedly send a Transmit AIP (Normal) parameter to the XL transmitter.

If an Arbitrating (Waiting On Partial) or Arbitrating (Blocked On Partial) confirmation is received, this state shall repeatedly send a Transmit AIP (Waiting On Partial) parameter to the XL transmitter.

If an Arbitrating (Waiting On Connection) confirmation is received, this state shall repeatedly send a Transmit AIP (Waiting On Connection) parameter to the XL transmitter.

If this state is entered from the XL6:Open_Response_Wait state Retry Priority status shall be set to RETRY PRIORITY IGNORE AWT otherwise Retry Priority status shall be set to RETRY PRIORITY NORMAL.

Upon entry into this state, this state shall send a Request Path request to the ECM with the following arguments:

a) destination SAS address;
b) source SAS address;
c) protocol;
d) connection rate;
e) arbitration wait time;
f) initiator port bit;
g) initiator connection tag;
h) pathway blocked count; ~~and~~
i) partial pathway timeout status~~.~~; and
j) retry priority status.