

First Burst Simplification

Bob Nixon

bob.nixon@emulex.com

1 Overview

It is difficult for initiators in multi-initiator SCSI domains to properly support first burst transfers because the transfer size may change unpredictably; however, a SCSI initiator will fail if it uses an incorrect first burst size. This proposal presents the issue and several alternate approaches that simplify or eliminate support for first burst.

2 Issue

SAS 1.0 specifies support for the FIRST BURST SIZE field in the Disconnect - Reconnect mode page. Parameters in the Disconnect - Reconnect mode page pertain to a specific target port regardless of the initiator port or target logical unit. As these parameters are not specifically read-only in either SPC-3 or SAS, an initiator may alter them by a MODE SELECT command. Such a change immediately affects all other initiators using any logical unit via the same port. In a multi-initiator SCSI domain such as SAS, there is no reason to expect various initiators to coordinate setting of first burst. In fact, as this value is provided explicitly for transport tuning, each initiator may see some advantage (a hair better performance, an increase in stability) to forcing its own optimal value (or zero) for first burst.

The result of an initiator not handling a change of the first burst size is that the initiator will experience write timeouts. It doesn't matter who changed the value: in the eyes of the customer, the fault is with the initiator that times out. So a viable initiator has to design to accomodate unpredictable changes of the first burst size.

Designing an initiator for unpredictable changes of the first burst size requires active assurance that the setting of first burst is at least known by the initiator (if not actually set to a preferred value) before every comand is sent. Since this involves awareness of unit attentions and sending of MODE SENSE/SELECT commands, the initiator transport layer becomes functionally entangled with the initiator application(s). This is bad, as evidenced by the acceptance of 02-403 (Elliot), which attempted to untangle exactly this issue.

In SAS, setting first burst to zero disables the feature, but this does not resolve the problem: Another initiator may at any time set it nonzero. In fact, it was pointed out to me that a transport has to support any possible setting of first burst size in order to deliver a MODE SELECT to disable it.

SPC-3 (thanks to 02-403) hints that using a setting of FIRST BURST SIZE (e.g., zero) to disable first burst is an incomplete implementation...in the specification of the FIRST BURST SIZE field it says "SCSI transport protocols supporting this field shall provide an additional mechanism to enable and disable the first burst function." Additional to what? The only reasonable interpretation in this context is "other than the mode page setting itself". SAS does not provide such an additional mechanism.

As an (a?) historical precedent, FCP-2 provides an additional mechanism by means of a flag during login that enables or disables first burst on a per-initiator basis. Subsequently, the FC-PLDA interoperability profile required that the flag is to be set to one, unconditionally disabling first burst by specification. In T10/02-324 (Gardner) and in T10 reflector conversations leading to this proposal, serious questions have been raised about the value of first burst behavior to SAS versus the effort to implement it. SAS should carefully consider the relevance of the FCP-2 precedent to SAS.

3 Proposed resolutions

Here are four approaches that could be taken in SAS 1.1 to ameliorate the issues raised by first burst size. They are presented in the order of preference of my company:

- 1) Obsolete the first burst feature.
- 2) Provide a true "additional mechanism" to disable first burst on a per initiator basis. This proposal suggests doing it by adding a bit to the SSP COMMAND IU indicating whether the target is to honor (bit=0) or ignore (bit=1) first burst size for the command.
- 3) Make first burst a fixed value by SAS specification (e.g., all commands causing data transfer from an SSP initiator to an SSP target shall have a first burst transfer size of 10000h bytes).
- 4) As a marginally acceptable resolution, make first burst a fixed property of the target port (i.e., the value of first burst shall not be changable by MODE SELECT).

4 Instructions to editor

4.1 Overview

Only one of the following subclauses should be implemented. The reference document is SAS-r04a.

4.2 Obsolete first burst

In subclause 9.2.1, eliminate list item b.

In subclause 9.2.2.3 second paragraph, eliminate the text “unless the FIRST BURST SIZE field in the Disconnect-Reconnect mode page (see 10.2.6.1.5) is not set to zero”.

In subclause 9.2.2.3 third paragraph, eliminate the first sentence and join the remainder to the second paragraph.

In subclause 9.2.2.4 unordered list, merge item a into the prior paragraph and eliminate item b.

in subclause 9.2.5.1, change the paragraph “If an SSP target port receives a DATA frame that does not contain first burst data and for which there is no XFER_RDY frame outstanding, it shall discard the frame (see 9.2.6.3.2)” to “If an SSP target port receives a DATA frame for which there is no XFER_RDY frame outstanding, it shall discard the frame (see 9.2.6.3.2)”.

In subclause 9.2.6.2.2.1, eliminate the paragraph “If the request is for a data-out command, then the request also includes the number of bytes for the first burst size for the logical unit”.

In subclause 9.2.6.2.2.3 list, eliminate list item a.

In subclause 9.2.6.2.2.3 list item c, eliminate the text “the first burst size or”.

In subclause 9.2.6.3.2, change the paragraph “If the frame type is DATA, and the tag matches a tag for an outstanding data-out command without first burst data for which no XFER_RDY frame is outstanding, then this state machine shall discard the frame and terminate” to “If the frame type is DATA, and the tag matches a tag for an outstanding data-out command for which no XFER_RDY frame is outstanding, then this state machine shall discard the frame and terminate”.

In subclause 9.2.6.3.3.1 last list, merge item a into the prior paragraph and eliminate item b.

In subclause 9.2.6.3.3.5.1, eliminate the first paragraph after the first unordered list.

In subclause 9.2.6.3.3.5.1 first ordered list item 3, change “If an XFER_RDY frame was sent for the data (i.e., it is not first burst data) and the length of the data exceeds that specified by the XFER_RDY frame that requested the data” to “If the length of the data exceeds that specified by the XFER_RDY frame that requested the data”

In subclause 9.2.6.3.3.5.1 last paragraph, eliminate the text “If this state has no more bytes in its first burst buffer, then this state machine shall terminate after sending the confirmation. If this state has more bytes to move in its first burst buffer, then this state machine shall wait for a Receive Data-Out transport protocol service request”.

In subclause 9.2.6.3.3.5.2 unordered list, eliminate the text “and first burst is not enabled; or” from item a, merge item a into the prior paragraph and eliminate item b.

In subclause 9.2.6.3.3.6.1, eliminate the first paragraph after the first unordered list.

In subclause 10.2.1.2 Send SCSI Command service request prototype, eliminate the First Burst Enabled parameter.

In subclause 10.2.1.2 table 113, eliminate the row defining the First Burst Enabled parameter.

In subclause 10.2.1.3 table 114, eliminate the row defining the First Burst Enabled parameter.

In subclause 10.2.6.1.1 table 127, change "FIRST BURST SIZE" to "Obsolete".

In subclause 10.2.6.1.1, eliminate the paragraph "The FIRST BURST SIZE field is defined in 10.2.6.1.5".

Eliminate subclause 10.2.6.1.5.

4.3 Provide an additional mechanism for first burst enable/disable

In subclause 9.2.2.1, table 24, add a flag field in byte 9 bit 7 labeled "FB".

In subclause 9.2.2.1, add a paragraph "The FB bit directs the target SSP transport state machine to attend or ignore the FIRST BURST SIZE field in the Disconnect - Reconnect mode page for the target port (see 10.2.6.1.5). If the value in the FB bit is zero, the initiator and target SSP transport state machines shall transfer any data in accord with the value of the FIRST BURST SIZE field in the Disconnect - Reconnect mode page for the target port. If the value in the FB bit is one, the initiator and target SSP transport state machines shall transfer any data as though the value of the FIRST BURST SIZE field in the Disconnect - Reconnect mode page for the target port were zero (i.e., first burst data transfer shall be disabled for the command)."

In subclause 9.2.2.3 second paragraph, eliminate the last sentence.

In subclause 9.2.2.3 third paragraph, change the first sentence to "In the initial XFER_RDY frame for a given command, the SSP target port shall set the requested offset to the number of bytes of first burst data enabled for the command (see 9.2.2.1)". Join the third paragraph to the second paragraph.

In subclause 9.2.2.4 list item b, change "if the FIRST BURST SIZE field in the Disconnect-Reconnect mode page is not zero (see 10.2.6.1.5)." to "if first burst data is enabled for the command (see 9.2.2.1)".

In subclause 9.2.6.2.2.1, change the paragraph "If the request is for a data-out command, then the request also includes the number of bytes for the first burst size for the logical unit" to "If the request is for a data-out command, then the request also includes the number of bytes of first burst data enabled for the command (see 9.2.2.1)".

In subclause 10.2.6.1.5 first paragraph, prefix the first sentence with "For commands sent with first burst data disabled (see 9.2.2.1), the First Burst Data Size shall be ignored. For commands sent with first burst data enabled,".

4.4 Make first burst a fixed value

In subclause 10.2.6.1.5 first paragraph, change "contains the maximum amount" to "contains 128, which is the maximum amount".

In subclause 10.2.6.1.5 first paragraph, change "(e.g., a value of one in this field means that the number of bytes transferred by the SSP initiator port is less than or equal to 512 and a value of two in this field means that the number of bytes transferred by the SSP initiator port is less than or equal to 1 024)" to "(i.e., the number of bytes transferred by the SSP initiator port before the initial XFER_RDY shall be less than or equal to 512 times 128, which is 65 536)".

In subclause 10.2.6.1.5, add a new last paragraph "SAS target devices shall not allow the value of the FIRST BURST SIZE field in the Disconnect-Reconnect mode page to be changed".

4.5 Make first burst a fixed characteristic of the target port

In subclause 10.2.6.1.5, add a new last paragraph "SAS target devices shall not allow the value of the FIRST BURST SIZE field in the Disconnect-Reconnect mode page to be changed".