

Date: 14 January 2003
To: Technical Committee T10
From: Cris Simpson *cris.simpson@intel.com*
Subject: **SRP-2 Immediate Data**

ABSTRACT

T10/02-096r0 provided the rationale for including data with SRP write commands: the round trips need to transfer the write data with the existing SRP model reduce performance as latencies increase.

SOLUTION

Include up to a pre-defined number (`MAX_IMMED_LEN`) of bytes of write data within the `SRP_COMMAND` information unit. This document uses the term 'immediate data' to refer to data transferred within an `SRP_COMMAND` IU.

DEFINITIONS

`MAX_IMMED_LEN` - The maximum number of bytes of data-out that may be included within an `SRP_COMMAND` information unit.

small write - A write command where the number of data-out bytes does not exceed `MAX_IMMED_LEN`. All data is transferred within the `SRP_COMMAND` information unit.

large write - A write command where the number of data-out bytes exceeds `MAX_IMMED_LEN`, requiring the target port to issue one or more RDMA reads to fetch the additional data.

ISSUES

For small writes, all the data is transferred with the command. For larger writes, where the data length exceeds `MAX_IMMED_LEN`, there are two alternatives: include some data with the command (target issues RDMA to get remaining data) or include no data with the command. Since the target will need to allocate receive buffers of a size equal to `MAX_IMMED_LEN`, there appears to be no disadvantage to allowing the inclusion of some data with the command.

Application model

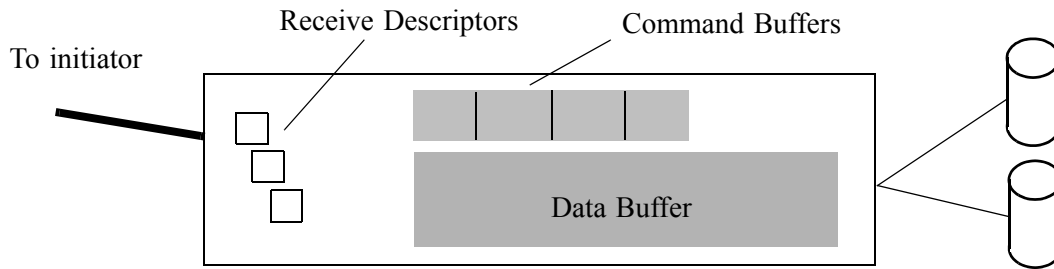


Figure 1 Application model

Figure 1 shows a model of an SRP target, with receive descriptors and command and data buffers. In this proposal, the data within the information unit begins at a known offset, allowing receive descriptors to split an incoming information unit into a ‘command part’ placed into a command buffer and a ‘data part’ placed into a data buffer.

There is a cost to the target for supporting immediate data: data buffers must be allocated at the time the receive descriptors are posted. When immediate data is not enabled, data buffers need not be allocated until the target issues the RDMA read operation to fetch the write data. Early allocation increases the time that buffers are allocated, resulting in higher buffer utilization and decreasing the number of commands the target may support at one time.

This proposal considers the immediate data sent in a large write as a cached copy of the data in initiator memory. This allows the target to discard the data and re-fetch it later. This is not the case with short writes, which do not specify a data-out buffer.

Table 1: SRP_CMD request

Byte	Bit	7	6	5	4	3	2	1	0
0		TYPE (02h)							
1		Reserved					UCSOLNT	SCSOLNT	Reserved
2		Reserved							
3									
4									
5		DATA-OUT BUFFER DESCRIPTOR FORMAT				DATA-IN BUFFER DESCRIPTOR FORMAT			
6		DATA-OUT BUFFER DESCRIPTOR COUNT							
7		DATA-IN BUFFER DESCRIPTOR COUNT							
8		(MSB)	TAG						
...									
15		(LSB)							
16		bReserved							
...									
19									
20		(MSB)	LOGICAL UNIT NUMBER						
...									
27		(LSB)							
28		Reserved							
29		Reserved					TASK ATTRIBUTE		
30		Reserved							
31		ADDITIONAL CDB LENGTH = n						Reserved	
32		CDB							
...									
47									
48		ADDITIONAL CDB							
...									
47+4×n									
48+4×n		DATA-OUT BUFFER DESCRIPTOR							
...									
47+4×n+do ^a									
48+4×n+do ^a		DATA-IN BUFFER DESCRIPTOR							
...									
47+4×n+do+di ^b									
p		PADDING							
l									
p+q									
r									
l									
r+s		IMMEDIATE DATA							

Table Footnote

a) The value 'do' is the length in bytes of the DATA-OUT BUFFER DESCRIPTOR field, determined from the format code value contained in the DATA-OUT BUFFER DESCRIPTOR FORMAT field and the count value contained in the DATA-OUT BUFFER DESCRIPTOR COUNT field (see 5.6.2).

b) The value 'di' is the length in bytes of the DATA-IN BUFFER DESCRIPTOR field, determined from the format code value contained in the DATA-IN BUFFER DESCRIPTOR FORMAT field and the count value contained in the DATA-IN BUFFER DESCRIPTOR COUNT field (see 5.6.2).

If the SRP_CMD IU specifies a short write with immediate data, the value of the DATA-OUT BUFFER DESCRIPTOR COUNT field shall be zero.

The PADDING field contains enough bytes to align the IMMEDIATE DATA field at the offset specified by the value of the IMMEDIATE DATA OFFSET field (see SRP_LOGIN).

The IMMEDIATE DATA field contains data-out beginning at offset zero from the data-out buffer. The target port may determine the length of the immediate data by subtracting the value of the IMMEDIATE DATA OFFSET field (see SRP_LOGIN) from the length of the SRP_CMD information unit. If the target port is unprepared to handle the immediate data of a large write (see defn) when received, it may issue an RDMA read to fetch the data from the data-out buffer at a later time.

Editor's Note 1: Model will need an explicit requirement that received data length be returned. Implicit now.

Table 9 - SRP_LOGIN_REQ request

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (00h)							
1	Reserved							
...								
7								
8	(MSB)	TAG						
...								
15	(LSB)							
16	(MSB)	REQUESTED MAXIMUM INITIATOR TO TARGET IU LENGTH						
...								
19	(LSB)							
20	Reserved							
...								
23								
24	REQUIRED BUFFER FORMATS							
25								
26	IMMDATA	AESOLNT	CRSOLNT	LOSOLNT	Reserved		MULTI-CHANNEL ACTION	
27	IMMEDIATE DATA OFFSET							
28	Reserved							
...								
31								
32	INITIATOR PORT IDENTIFIER							
...								
47								
48	TARGET PORT IDENTIFIER							
...								
63								

ADD to defin for REQUESTED MAXIMUM INITIATOR TO TARGET IU LENGTH:

This length includes any immediate data.

When IMMDATA is set to one, the initiator port is requesting that the target port enable support for immediate data.

If IMMDATA is set to one, the IMMEDIATE DATA OFFSET field indicates the offset, in bytes, at which the immediate data, if present, begins within an SRP_COMMAND information unit.

Editor's Note 2: Is 256 a big enough offset?

Table 11 - SRP_LOGIN_RSP response

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (C0h)							
1	Reserved							
2								
3								
4	(MSB)	REQUEST LIMIT DELTA						
...								
7	(LSB)							
8	(MSB)	TAG						
...								
15	(LSB)							
16	(MSB)	MAXIMUM INITIATOR TO TARGET IU LENGTH						
...								
19	(LSB)							
20	(MSB)	MAXIMUM TARGET TO INITIATOR IU LENGTH						
...								
23	(LSB)							
24	SUPPORTED BUFFER FORMATS							
25								
26	IMMDSUPP	Reserved		SOLNTSUP	Reserved		MULTI-CHANNEL RESULT	
27	Reserved							
28	Reserved							
...								
51								

The target port sets the IMMDSUPP field to one to indicate that the target port has accepted the immediate data parameters present in the SRP_LOGIN request, and sets it to zero to indicate that the target port has accepted the login request, but has rejected the request to support immediate data.

Editor's Note 3: Things are sub-optimal (huge wasted IU receive buffers) if target accepts a big max Init to Targ IU size, but rejects immediate data. Init should logout and try again. Should target reject instead?

Editor's Note 4: Need to provide a way to determine, *a priori*, what value of MAX_IMMED_LEN the target port will support.