



BROCADE

July 14, 2001

T10/01-221r0

To: John Lohmeyer, Chair, T10 Standards Committee, NCITS
From: Lino Costantino, Howard Green, Robert Snively
Date: July 14, 2001
Subject: Comments on SRP and InfiniBand Annex

1 Introduction

Brocade has done a preliminary review of the SRP document (revision 6) and the proposed SRP IB Annex (T10/01-028r6) and has identified the following comments that should be addressed before the document is sent out for ballot.

2 SRP comments, concerns, and questions

2.1 Editorial issues for compliance with T10 conventions

Replacement of the word “can”:

In 4.3, item c, the word “can be” should be replaced with “are”.

Replacement of the word “which”:

In most places where it used, the word “which” is grammatically incorrect and should be replaced with the word “that”. In some cases the sentence should be reconstructed to eliminate the word “which” entirely. See a high quality writer’s manual for details and guidance. A global search and correction is required.

Correction of hanging paragraphs:

The hanging paragraph in clause 2 should be corrected.

2.2 Editorial error in 4.4.2.3, page 22, length definition

Section 4.4.3.2 on page 22 indicates:

“The DATA LENGTH field of INDIRECT TABLE MEMORY DESCRIPTOR would contain 192, the length of the indirect table.

The text appears to indicate that the value 192 should actually be 48.

2.3 Correction of “indirect not supported” error

In table 10, there is an explicit negotiation for the support of indirect data buffer descriptors. The “1” state of the INDSUP bit indicates that. The implication is that support of indirect data buffer descriptors is negotiated. In that case, there should be no rejection of an SRP Login Request because of the non-support of indirect data buffer descriptors. In Table 13 (describing the reason codes for rejection of a login request), the “Indirect data buffer descriptors not supported.” entry should be deleted.

2.4 Addition of login reject error code for “no SCSI Target process”

If an SRP_LOGIN_REQ IU arrives at a physical port, and there's no “target port” behind the physical port that has a Target Port Identifier matching that in the SRP_LOGIN_REQ, the login

Brocade Communications Systems, Inc.
1745 Technology Drive San Jose, CA 95131
T 408.487.8000 F 408.487.8101
www.brocade.com

should be rejected. An SRP_LOGIN_REJ reason code needs to be added to Table 13 to indicate “No SCSI Target process available”.

There is a corresponding case for SRP_LOGOUT if an existing target port process has disappeared and is therefore being logged out. An SRP_LOGOUT reason code of “No SCSI Target process available” should be added to Table 15.

3 SRP Annex for IB comments, concerns, and questions

3.1 Editorial issues for compliance with T10 conventions

Replacement of the word “must”

A global search for the word “must” should be made and it should be replaced with the word “shall”. Examples include Table A.5,

3.2 Correction of GUID

In section A.4 on page 6.

“A GUID is an IPv6 address.” should be restated as “A GUID conforms to the IPv6 address format.”

3.3 Treatment of redundant controllers

How are redundant IO controllers handled? Can two physical IO controllers connected to the same LUNs have the same GUID? How are the two controller distinguished? Would it be by the path?

This must be explained more completely in the model.

3.4 Definition of processor unit creates confusion

The model presented in the annex isn't too far off in the end, but the explanation can be quite confusing. One part of the problem is the “processor unit” thing it defines. The choice of this terminology (which is unique to the appendix) is very unfortunate, as it immediately suggests that a processor unit is somehow distinct and different from an IO unit. Which leads to contradictions like when the text states that targets have to be IO units, but that it's OK to have targets on processor units, and initiators on IO units, and... To make matters worse, the only way to emplace a target in a processor unit is to turn it into an IO unit! All this feels like there may be a fundamental lack of faith in the abstractions involved.

The better way to do this would be to define something like a “platform” (which used to be in IB, but unfortunately was removed), where a platform has a some set of CAs, and each CA has one or more ports. An initiator is necessarily always on some platform. An IO unit is a specialized function or structure defined by IB and imposed upon or added to a platform, and in such a way as to not preclude other forms of communication. A target always is implemented as an IO controller within an IO unit. Because the IO unit is, after all, a platform, a target can, if it chooses, act in the role of an initiator without conflict.

3.5 Unclear explanation of the concept of port

The use of the “{initiator,target} port identifier” notion is confusing. As I understand it, one of these “port” identifiers identifies a service access point belonging to a target or initiator, and is bound to physical IB ports in a many-to-many relationship. I like this model just fine; it feels like exactly the right thing to do, not least because it's what iSCSI does. But iSCSI manages to explain without ambiguity by avoiding the use of the term “port” (which, of course, has high religious significance in the TCP world). The problem here is just that the over-overloading of

“port” is misleading. I think there needs to be a statement about this, that a “port identifier” does *not* identify a physical IB port.

3.6 Clarification of disconnect-reconnect page parameters

To what do the disconnect-reconnect page parameters apply in the IB context? Is it the “port” *per se*, or the initiator's use of the port (i.e., the specific connection)?

3.7 Incomplete definition of IB mapping for the memory descriptor format

The Annex needs to define an IB-specific mapping for the memory descriptor format defined in Table 1 (p. 19) in the SRP spec. The problem involves the “memory handle” field in the memory descriptor. In IB, there's no memory handle as such (it's implicitly defined by the QP at the initiator) but there *is* a per-memory-area RDMA Key (R_Key) that has to be included in the RDMA Extended Transport Header (RETH) of each IB RDMA packet. Now, if one happens to stick the R_Key into the memory descriptor's memory handle field, one winds up with a structure that looks remarkably like the RETH.

3.8 Concern about completeness of IB specific mapping

Are there other SRP things that need to be specialized for IB? Do we need, for example, an IB-specific Disconnect-Reconnect page interpretation?