Comment Resolution Status

Document: **T10/01-328r5** Date:4 April 2002

To: T10 Committee Membership
From: Cris Simpson, Intel Corporation
Subject: Response to T10 Letter Ballot comments on SRP

This document contains **T10/1415-D revision 13**, the SRP Working Draft.

Comments with possible implementation effects (list may be incomplete):

HP01: Service Name persistence . . . . . .Pending
HP09: Security Protocol . . . . . Rejected
HP27: Identifer construction rules . . . . . .Accepted
IBTA: IOControllerProfile I/O Class field . . . . . .Closed
OD 3: Cross-channel reporting . . . . . Rejected
OD 4: Swap GUID and Extension fields in Port Identifiers . . . . . .Closed
OD6: Solicited Events . . . . . .Pending
OD 8: Buffer formats & codes . . . . . Rejected
Troika: Correct Type Code in SRP_LOGIN_REJ . . . . . .Closed

Table of Comments

Brocade comments:

Compaq comments:

HP comments:

IBM / Tivoli comments:

InfiniBand™ Trade Association comment:

Intel comments:

Ophidian Designs comments:

New editor comments:

Texas Instruments comment:

Troika Networks comment:

Woven Electronics comment:

Brocade comments:

Bro101 **Rejected 07 Jan 2002**

The word 'which' is used inappropriately in many places. Suggested Solution: Do a global search for the word which and replace it with one of the following corrections: A) the word 'that'. B) a new sentence construction that does not require the word. C) nothing. (Which can simply be removed in many cases.)

All occurences of "which" are correct both grammatically and according to the Chicago Manual of Style.

Bro102 page 67 line 13 **Closed**

The word 'must' is used inappropriately. Suggested Solution: The line 'At least one IB I/O controller must be present' should be replaced. I am not sure if this is a requirement that at one or more controllers shall be present. If so, wording like 'At least one IB I/O controller shall be present' is appropriate.

Proposed text:

At least one IB I/O controller acting as an SRP target port shall ~~must~~ be present.

Bro103 page viii line 3 **Closed**

X3.269 is not the proper name Suggested Solution: This value is not correct and should be marked as TBD or XXX or something like that. In any case, it is an NCITS document, not an X3 document.

IBM005 See IBM005. Proposed text:

This foreword is not part of American National Standard NCITS.***:200x ~~X3.269-199x~~.

Bro104 page viii line 8 **Closed**

"'by National' s/b 'by the National'" Suggested Solution: Correct as requested.

Insert "the" as requested.

Bro105 page ix line 7 **Closed**

CRS: Agree w/ comment. Added 'Working Draft' note to Master Page - s/b sufficiently clear, allow correcting text.

"'The working draft SCSI' s/b 'The SCSI'" Suggested Solution: This correction should be made now, even though the document is still a working draft, because it is clearly labeled in lots of places that it is a draft, but the text in it is intended to be the content of the standard.

IBM006 The document is a working draft until it is published by ANSI or NCITS. NCITS requires that we prominently label it a "working draft" until then. See IBM006.

Bro106 page 1 line 6 **Closed**

Accepted.

"'The working draft SCSI' s/b 'The SCSI'" Suggested Solution: This correction should be made now, even though the document is still a working draft, because it is clearly labeled in lots of places that it is a draft, but the text in it is intended to be the content of the standard.

The document is a working draft until it is published by ANSI or NCITS. NCITS requires that we prominently label it a "working draft" until then.

Bro107 page 3 lines 32-35 **Closed**

Accepted, corrected formats, added URL.

Global Engineering should be included here as well, since the drafts are not available from ANSI or NCITS. Suggested Solution: Include Global Engineering as a document source. Include www.t10.org as a document source for standards in development.

The following note will be added to the end of sub-clause 2.3 (copied from sam4r07):

NOTE 1 - For more information on the current status of a ~~the~~ document, contact the NCITS Secretariat at 202-737-8888 (phone), 202-638-4922 (fax) or via Email at ncits@itic.org. To obtain copies of these ~~this~~ documents, contact Global Engineering at 15 Inverness Way, East Englewood, CO 80112-5704 at 303-792-2181 (phone), 800-854-7179 (phone), or 303-792-2192 (fax).

Bro001 **Closed**

The draft now seems to equate 'SRP target port' and 'IB service', so an SRP target port is designated by a ServiceID. This implies there can be many ports per IOC. This is a significant change from prior drafts where the target port was equated with an IOC, and there was just a single ServiceID per port. It requires a different model for software (OSs or whatever) to manage which hosts have access to which devices in a multi-host environment. Previously, access control was needed only to the level of IOCs, the draft now implies a need to manage not only who can use which IOCs, but which devices within an IOC. Suggested Solution: No solution required if interpretation is correct and implications are understood

Accepted, no change requested.

It is true that annex B equates an SRP target port to an IB service, with the caveat that "IB service" is not clearly defined by the IB specification. It is better to say that annex B equates an SRP target port to an IB service entry. Note that multiple SRP target ports (multiple IB service entries) could all use the identical ServiceID, with the particular SRP target port determined by the SRP target port identifier supplied during login.

The intent to allow many SRP target ports per IB I/O controller has been around for quite some time. The only new item in this draft was the specific mechanism for determing the extension field of the SRP target port identifier from the service name. That mechanism was agreed to at a teleconference in late September.

Note that access control is needed not just to IB I/O controllers and SRP target ports, but also to individual logical units. SCSI provides such access controls.

Bro002 page 60 lines 22-23 **Rejected 28 Nov 2001**

The definition of 'IB channel adapter GUID' implies it is the Node GUID but doesn't say so; might as well be explicit Suggested Solution: 'An IB Node GUID that uniquely identifies an IB channel adapater'

EAG: The IB specification treats "channel adapter" and "node" as synonyms. However, use of "node" appears to have been denigrated, it only appears as the names of some attributes and components. For example, the definition of the NodeGUID component of the NodeInfo attribute is that it contains the GUID of a channel adapter, that is, a channel adapter GUID. The IB specification glossary defines channel adapter but does not define node.

Bro003 page 60 lines 39-40 **Closed**

Added:This value is present as the GUID attribute of the IOControllerProfile. (See Table B.7)

The definition of 'IB I/O controller GUID' implies it is the IOControllerProfile GUID but doesn't say so; might as well be explicit Suggested Solution: 'An IB IOControllerProfile GUID that uniquely identifies an IB channel adapater'

The definition of IB I/O controller GUID is correct, the GUID value does identify the I/O controller. Replacing "I/O controller" with "IOControllerProfile" replaces a somewhat obscure term (I/O controller) with a confusing acronym (IOControllerProfile). For example, one natural interpretation of "IOControllerProfile GUID" is that it is an identifier of the IOControllerProfile attribute for use in protocol operations (e.g. MADs), not an identifier of the I/O controller. Also, the GUID value (whatever it is called) does not identify an IB channel adapter as stated in your suggested solution.

However, annex B does not state that the IB I/O controller GUID is the value reported in IOControllerProfile. Adding that would be a useful clarification. Proposed changes to **page 64** lines 36-37.

Proposed text:

The IO CONTROLLER GUID field is ~~shall be~~ the IB I/O controller GUID value that identifies ~~of~~ the IB I/O controller containing the SRP target port. This shall be the value reported in the GUID component of the IB I/O controller's IOControllerProfile attribute.

Bro004 page 62 line 50 **Closed**

IB GIDs can have link-local scope and thus may not be 'globally' unique Suggested Solution: Change to 'unique within a subnet', or 'either unique within a subnet or globally unique'

Current text:

Each IB port is assigned one or more 16-bit IB LIDs by the IB subnet manager. Each IB port has one or more 128-bit IB GIDs. Each IB GID is globally unique, and may be formed in part from the IB port GUID. An IB GID conforms to the format of an IPv6 address. The IB subnet manager provides a service to determine one or more IB LIDs and IB GIDs corresponding to an IB port GUID or IB channel adapter GUID.

CRS: This text seems overly informative. How GIDs are formatted, formed, or resolved is not relevant to SRP - it just uses them.

Proposed text:

The IB subnet manager assigns one or more IB LIDs and one or more IB GIDs to each IB port.

Bro005 page 63 lines 16-17 **Closed**

(In Table B.1, GID row) replace 'worldwide' with 'varies' and a reference to the IB spec's Addressing chapter.

IB GIDs can have link-local scope and thus may not be unique 'worldwide' Suggested Solution: Change 'worldwide' to 'IB subnet or worldwide'

Bro006 page 63 lines 23-48 **Discussion needed**

Figure B.3's equating of 'SRP Target Ports' with 'IB consumers' is problematic. A 'target port' is a sort of service access point---somewhere where interested parties initially go to obtain service, but without any implication that that's where the service is actually provided. (In IB, it's the Connection Manager that receives the initial connection request, interprets the ServiceID contained therein, and performs some magic that results in the instantiation of a QP bound to some entity that actually provides the target services). This target-services-providing entity fits the definition of 'IB consumer'. But the mapping of ServiceIDs-cum-SRP target ports onto such entities is clearly a matter of implementation, and could be one-to-many, many-to-one, or many-to-many Suggested Solution: One possibility: to the left of the IB Consumers show a table/list of service IDs within each IB I/O unit and label these entries as SRP Target Ports; use arrows to show a mapping from the entries to the IB Consumers, with e.g. one Consumer mapped to two IDs and another mapped to one ID to show that the mappings are not always 1 to 1. A further refinement might be to use another set of arrow between the Consumers and the QPs to show that the this mapping is also not 1 to 1

This comment is correct, but it’s not immediately obvious how to incorporate it into a legible diagram. Note the further complication introduced by connection redirection. The IB consumer (IB QP and the software, etc. behind it) may be in an unrelated device / node / whatever, it need not be the same channel adapter or even an I/O unit. The sole purpose of the I/O unit and I/O controller is to obtain a service ID, connecting to that service ID may lead somewhere else altogether.

Notes from 28 Nov 2001 teleconference: show service entries in figures B.2 and B.3, in accompanying text explain that each service entry identifies an SRP target port.

The following is an attempt at a modified figure B.2 and the text describing service entries that identify SRP target ports

An IB I/O unit is an InfiniBand™ Architecture device that contains an IB channel adapter with one or more IB ports, IB QPs, and one or more IB I/O controllers. Figure 0.1 shows an example IB I/O unit.

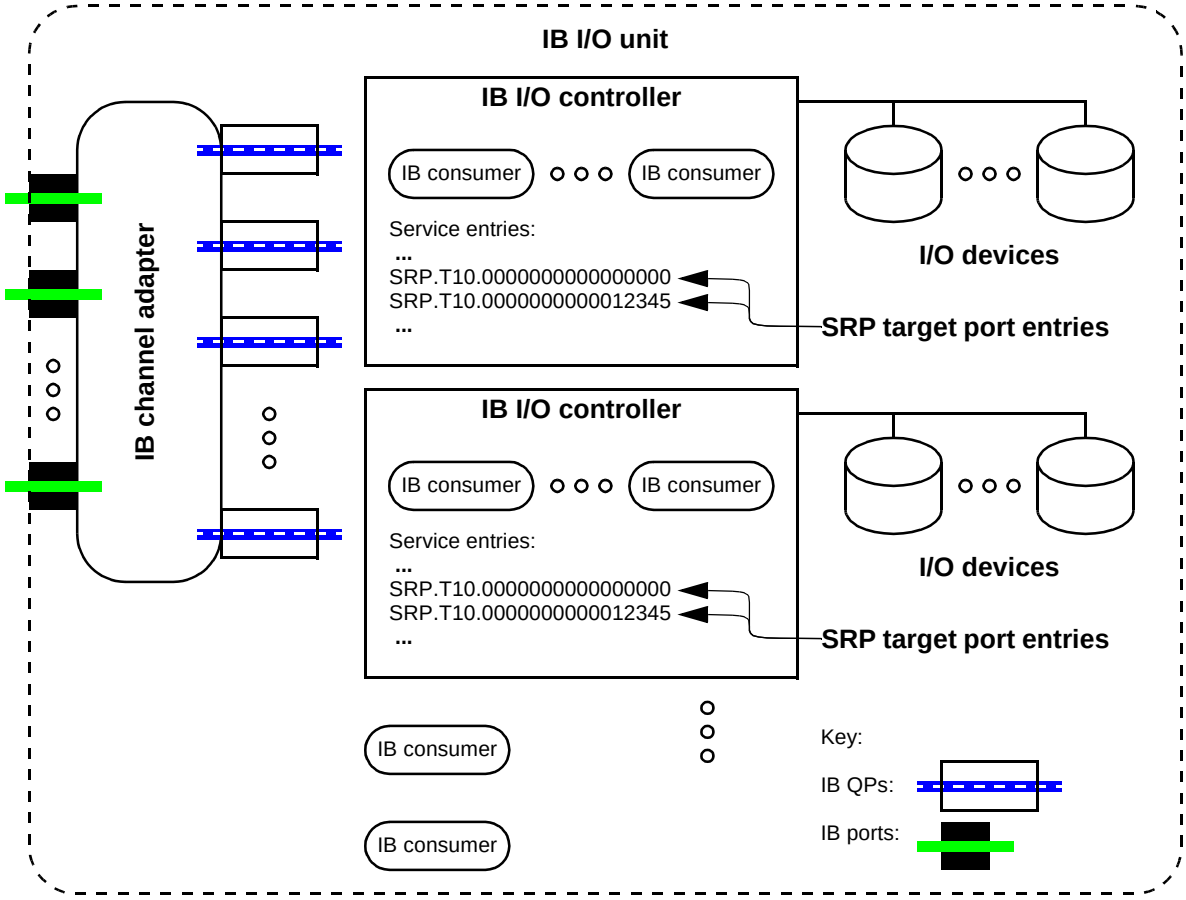


Figure 0.1 - IB I/O unit example

One or more service entries are associated with each IB I/O controller. A service entry contains a name and other information identifying an individual service provided by the IB I/O controller. A service entry may identify an SRP target port or a non-SRP service (e.g. a network interface service). Requirements for service entries that identify SRP target ports are described in table B.8.

Each IB port has a 64-bit globally unique identifier called an IB port GUID. Each IB channel adapter has a IB channel adapter GUID (which is shared by all IB ports on the IB channel adapter). Each IB I/O controller has an IB I/O controller GUID.

Bro007 page 64 line 16 **Closed**

Missing word Suggested Solution: 'used by the SRP initiator port'?

Proposed text:

The GUID field should an IB GUID available to the SRP initiator port, e.g. the IB channel adapter GUID for an IB channel adapter used by the SRP initiator port.

Bro008 page 64 line 21 **Accepted 28 Nov 2001**

Names of IB attributes are incomplete Suggested Solution: 'IOUnitInfo, IOControllerProfile, and ServiceEntries'

Proposed text:

SRP target ports shall be implemented in IB I/O units. The IB I/O unit shall include a device management agent to provide IOUnitInfo, IOControllerProfile~~,~~ and ServiceEntries attributes ~~and make available an IB I/O controller GUID~~.

Note the IOControllerProfile attribute contains the I/O controller GUID, there is no need to call it out separately. Indeed, calling it out separately risks confusion that it is somehow different from the GUID component of the IOControllerProfile attribute.

Bro009 page 64 lines 48-49 **Closed**

Done across document.

'I/O' is broken across lines (and pages) Suggested Solution: Make sure the slash in 'I/O' is non-breaking

Bro010 page 64 lines 48-49 **Accepted 28 Nov 2001**

The phrase 'processor unit or IB I/O controller' makes an incorrect distinction; target ports can only be found on IB I/O controllers by definition, whether or not the I/O controller embodies a processor unit Suggested Solution: Omit 'processor unit or'

Proposed text:

IB communications managers on each InfiniBand™ Architecture device manage InfiniBand™ Architecture connections using IB MADs transported over the IB general service interface. SRP initiator ports and SRP target ports shall use the active/passive (client/server) connection establishment protocol. The ~~processor unit or~~ IB I/O controller containing the SRP target port shall act as the server and the processor unit or IB I/O controller containing the SRP initiator port shall act as the client.

Bro011 page 65 lines 13-14 **Accepted 28 Nov 2001**

'IB I/O controllers acting as SRP target ports' could be construed as a 1-to- 1 correspondence between controllers and target ports Suggested Solution: 'IB I/O controllers making SRP target ports available' or 'IB I/O controllers hosting SRP target ports'?

Proposed text:

The IB service ID associated with each matching service name may be used in the communication management process to open InfiniBand™ Architecture connections to ~~IB I/O controllers acting as~~ an SRP target port~~s~~. The SRP target port identifier for each SRP target port is constructed as described in table B.3.

Bro012 page 69 line 1 **Rejected 28 Nov 2001**

'An IB I/O controller acting as an SRP target ports' could be construed as a 1-to-1 correspondence between controllers and target ports Suggested Solution: 'And IB I/O controller making SRP target ports available' or 'An IB I/O controller hosting SRP target ports'?

While the comment is correct, the 28 Nov 2001 teleconference agreed that it was better to delete the entire paragraph (**page 69** lines 1-3).

Bro013 page 69 lines 4-5 **Accepted 28 Nov 2001**

'IB I/O controllers acting as SRP target ports' could be construed as a 1-to- 1 correspondence between controllers and target ports Suggested Solution: 'IB I/O controllers making SRP target ports available' or 'IB I/O controllers hosting SRP target ports'?

Proposed text:

~~IB I/O controllers acting as SRP target ports shall include at least one ServiceName/ServiceID pair in the device management ServiceEntries attribute pair defined in InfinibandTM Architecture Specification Volume 1 Release 1.0.a as described in table B.8.~~

An IB I/O controller's ServiceEntries attribute contains one or more ServiceName/ServiceID pairs. ServiceName/ServiceID pairs that meet the requirements listed in table B.8 shall identify an SRP target port.

Compaq comments:

CPQ001a page a **Accepted 28 Nov 2001**

Remove: American National Standard for Information Systems.

See spc2r18 or spi4r07 (or other pre-public review versions) for style examples.

CPQ001b page a **Rejected 28 Nov 2001**

Change "working draft SCSI RDMA Protocol" to "SCSI RDMA Protocol".

The document is a working draft until it is published by ANSI or NCITS. NCITS requires that we prominently label it a "working draft" until then.

CPQ002 **Rejected 28 Nov 2001**

Update the PDF properties title and author

These do not form part of the printed standard. Maintaining the properties is additional manual effort with no benefit.

CPQ003 page c line 1 **Rejected 28 Nov 2001**

Remove revision history, line numbers, change bars, etc. from final version

IBM002 This is not a final version. See IBM002, IBM003.

IBM003

CPQ004 page 1 lines 21-24, page 2 line 25, page 3 lines 19-21 **Accepted 28 Nov 2001**

Delete CAM from figure 1 Delete these SCSI-2 standards from the example standards list: Serial Storage Architecture SCSI-2 Protocol SSA-S2P [ANSI X3.294:1996] Common Access Method: SCSI Common Access Method CAM [ISO/IEC 9316-421] [ANSI X3.232:1996]

The 28 Nov 2001 teleconference voted that this be accepted.

CPQ005 page 2 line 3 **Closed**

Change Fiber to Fibre

CPQ006a page 4 line 9 **Rejected 07 Jan 2002**

Add:

3.1.8 autosense data: Sense data (see 3.1.49) that is returned in the SRP_RSP IU payload. See SAM-2.

CPQ033 "Autosense" is a mechanism for delivering sense data, the data delivered by autosense is just ordinary "sense data". SAM-2 does not define "autosense data". See CPQ033

CPQ006b page 5 line 4 **Accepted 28 Nov 2001**

Add:

3.1.49 sense data: Data returned to an application client as a result of an autosense operation, asynchronous event report, or REQUEST SENSE command. See SPC-2.

Proposed text:

3.1.22a sense data: Data returned to an application client in the SENSE DATA field of an SRP_RSP response or an SRP_AER_REQ request. See SAM-2.

CPQ007 page 16 lines 28-31 **Accepted 17 Jan 2002**

This section should mention the SRP_CRED_REQ and SRP_CRED_RSP IUs, which are dedicated to flow control service.

Replace the paragraph on lines 28-31 of **page 16** with the following:

SRP uses a credit based flow control algorithm to limit the number of SRP requests that an SRP initiator port may send to an SRP target port. The algorithm uses a field, REQUEST LIMIT DELTA, that is present in most information units sent by an SRP target port to an SRP initiator port~~, and~~. The REQUEST LIMIT DELTA field is used to manipulate a state variable, REQUEST LIMIT, associated with each SRP initiator port. The value of the REQUEST LIMIT state variable determines whether or not the SRP initiator port may send new SRP requests.

Most information units containing a REQUEST LIMIT DELTA field do not generate a confirmation that the SRP initiator port has received the information unit and processed the contents of the REQUEST LIMIT DELTA field. The SRP_CRED_REQ request does generate a confirmation through the SRP_CRED_RSP response (see 6.10 and 6.11).

An SRP initiator port shall process the REQUEST LIMIT DELTA fields of information units received on the same RDMA channel in the order that they are received. An SRP initiator port shall process the REQUEST LIMIT DELTA field of a request before sending that request's response. E.g. an SRP initiator port shall process the REQUEST LIMIT DELTA field of an SRP_CRED_REQ request before sending the SRP_CRED_RSP (see 6.10 and 6.11).

The following rules specify the flow control algorithm for SRP requests sent by SRP initiator ports:

The second paragraph above directly addresses this comment. The third paragraph was added as a result of discussion during the 07 Jan 2002 teleconference.

CPQ008 page 18 **Closed**

intel0065 Table 2 Remove period from "NO DATA BUFFER DESCRIPTOR PRESENT."

CPQ009 page 18 line 32 **Closed**

intel0066 Table 2 There is no reference to note b. It probably needs to be in the 2h row buffer descriptor length cell, where "count" is used

Note that a reference is not always necessary, however in this case one is useful.

CPQ010 page 18 line 36 **Closed**

Table 2 Add a period at the end of note c.

CPQ011 page 20 line 43 to page 20 line 3 **Rejected 07 Jan 2002**

Add a fairly content-free table showing a direct data buffer containing a memory descriptor so this section has a visual reference like the indirect section.

Replace the text of this sub-clause (**page 20** line 43 to **page 20** line 3) with the following:

The DIRECT DATA BUFFER DESCRIPTOR format code value specifies that the corresponding data buffer descriptor field is sixteen bytes in length and contains a direct data buffer descriptor. The contents of the count field are reserved. SRP target ports are not required to check the contents of the count field. Table 3a shows the format of a direct data buffer descriptor.

Table 3a - Direct data buffer descriptor

Byte \ Bit	7	6	5	4	3	2	1	0
0	MEMORY DESCRIPTOR (see table 1)							
• • •								
15								

The MEMORY DESCRIPTOR field of a direct data buffer descriptor contains a single memory descriptor (see table 1). The memory descriptor identifies the data buffer, which is a single memory segment within a memory region's virtual address space. If a direct data buffer descriptor defines a data-out buffer, the SRP target port shall only issue RDMA Read operations using the memory descriptor contained in the direct data buffer descriptor. If a direct data buffer descriptor defines a data-in buffer, the SRP target port shall only issue RDMA Write operations using the memory descriptor contained in the direct data buffer descriptor. The SRP target port shall use the contents of the DATA LENGTH field of the memory descriptor as the length of the data-out buffer or data-in buffer.

CPQ012 page 21 line 30 **Accepted 07 Jan 2002**

Table 5 note a count should be defined with a note b similar to that in table 2

The comment is intended to refer to table 4, not table 5. The 07 Jan 2002 teleconference directed that this comment be accepted.

CPQ013 page 21 line 26 **Rejected 07 Jan 2002**

Table 4 If n is zero in 16*n+19, then the table shows byte 20 followed by byte 19. Remove the 20 and that numbering problem is eluded.

This is the common way of depicting variable length optional fields in many SCSI standards, including SPC-n, FCP-n and elsewhere in SRP.

CPQ014 page 26 line 41 **Closed**

Change: "maximum length" to "maximum length in bytes

CPQ015 page 26 lines 13-16 **Rejected**

I thought we decided that TAG fields don't have bits labeled (MSB)/(LSB).

EAG: Rejected : SAM-2 requires an arithmetic comparison of tag values (5.8.2 Overlapped Commands, pdf page 96 in sam2r21). Implementing an arithmetic comparison requires identifying the least and most significant bits. I believe that is the only requirement for this in all of SCSI, I would welcome its removal.

CRS: (Move from Rejected to DIscussion) I find no such requirement for 'arithmetic comparison' to detect duplicates. A bit-by-bit compare will do fine. However, SAM-2 5.8.2 requires arithmetic evaluation for ***reporting*** *overlapped tags:*

> *If the overlapped command condition was caused by an untagged task or a tagged task with a* ***tag value exceeding FFh****, then the sense key shall be set to ABORTED COMMAND and the additional sense code shall be set to OVERLAPPED COMMANDS ATTEMPTED. Otherwise, an additional sense code of TAGGED OVERLAPPED TASKS shall be returned with the ADDITIONAL SENSE CODE QUALIFIER field set to the value of the duplicate tag.*

I'd like to suggest a change in SAM-2 from "if tag value > FFh" to "if tag field size > one byte". With SRP's eight-byte tags, there doesn't appear to be any value to having one reporting scheme for tags 0-ffh, and another for 0100h-ffffffffffffffffh.

Latter tag reporting method obsoleted. Specifying MSB/LSB may have value for analyzers interpreting tags.

CPQ016 page 26 line 24 **Accepted 07 Jan 2002**

Table 9 The REQUIRED BUFFER FORMATS cell is missing the horizontal lines present in other multibyte cells

CPQ017 page 27 line 10 **Closed**

Table 10 Remove period from first Reserved. row

CPQ018 page 28 line 40 and page 28 line 44 **Closed**

Change (two places): maximum length to "maximum length in bytes"

CPQ019 page 30 line 46 **Closed**

Table 14 Capitalize Reserved

CPQ020 page 30 line 24 **Accepted 07 Jan 2002**

Table 13 The SUPPORTED BUFFER FORMATS cell is missing the horizontal lines present in other multibyte cells

CPQ021 page 34 lines 24-42 **Rejected 07 Jan 2002**

Table 17 Add period after Reserved or remove from other rows

The rule is that a period should appear after descriptions that are sentences or major fragments of sentences, but not after simple words (e.g. no period after "Reserved").

CPQ022 page 35 and page 36 **Closed**

CRS: Also changed table 19 to task management function codes

Rename TASK MANAGEMENT FLAGS to TASK MANAGEMENT FUNCTION. It doesn't really contain flags.

Rob Elliott will request the same change in other standards as they come up for review.

CPQ023 page 36 lines 8-17 **Rejected 07 Jan 2002**

Table 19 end each row with a period (or don't)

Each row that is a sentence ends with a period, which is correct. The row that is the isolated word "Reserved" does not end with a period, which is also correct. The period will be removed following "Restricted" in line 14.

CPQ024 page 36 **Closed**

Table 19 Change Codes to Code.

CPQ025 page 36 line 5 **Closed**

Table 19 Remove small caps from TABLE.

CPQ026 page 37 lines 38-44 **Rejected 07 Jan 2002**

Table 20 Per Patrick Fitzgerald at JNI, please require that DATA-OUT BUFFER DESCRIPTOR and DATA-IN BUFFER DESCRIPTOR start on 8-byte aligned boundaries. The ADDITIONAL CDB field is only 4 byte aligned.

This was discussed in several SRP working groups. It was raised as one of the potential issues with adding a total length field to indirect data buffer descriptors, since that field causes those descriptors to be a multiple of 4 bytes but not 8 bytes. Therefore it is impossible to align both descriptors in commands that contain both. We also discussed (in less length) the impact of wierd CDB sizes on buffer descriptor alignment. The unanimous concensus in all of these discussions was that there was no need to require 8 byte alignment of any buffer descriptor, 4 byte alignment was sufficient. Note that the first descriptor will in fact be 8 byte aligned for all common CDB lengths.

CPQ027 page 37 line 45 and page 37 line 48 **Closed**

Table 20 footnotes Change: length to: length in bytes

CPQ028 page 38 lines 20-36 **Accepted 07 Jan 2002**

Table 21 SAM-2 rev 20 still requires that untagged tasks be supported by all protocols. 01-318 will remove this requirement and make SRP legal.

Note that SPI-n also does not define a task attribute for untagged tasks when using information units, and now requires use of information units.

CPQ029 page 38 line 33 **Closed**

Table 21 Change a to an in the ACA row

CPQ030 page 38 line 20 **Closed**

Table 21 Remove small caps from TABLE

CPQ031 page 42 line 7 **Rejected 07 Jan 2002**

After: The STATUS field contains the status of a task that completes. See the SAM-2 standard for a list of status codes. Add this sentence and a table: Some of the status codes defined in SAM-2 are listed in table xx. Table xx - Some STATUS codes 00h GOOD 02h CHECK CONDITION 08h BUSY 18h RESERVATION CONFLICT 28h TASK SET FULL 30h ACA ACTIVE 40h TASK ABORTED This helps save the reader a reference to SAM-2 for the most popular fields.

The notion that anyone can understand or implement any SCSI protocol without referring to SAM-2 is fallacious. Encouraging anyone to avoid referring to SAM-2 will contribute to interoperability problems. Adding such a table will lead readers to infer that that table lists the only status codes they need to deal with. Redundantly defining status codes in multiple documents is a bad idea.

CPQ032 page 44 line 32 **Rejected 07 Jan 2002**

Remove from 2nd sentence of SENSE DATA paragraph: as specified by the SCSI Primary Commands-2 standard.

While the referenced text is redundant, a redundant reference is harmless. That sentence is copied verbatim from spi4r08.

CPQ033 page 44 lines 30-34 **Rejected 07 Jan 2002**

Reword the SENSE DATA paragraph to focus on the term autosense which is defined in SAM-2 rather than the REQUEST SENSE command in SPC-2. Change: The SENSE DATA field contains the information specified by the SCSI Primary Commands-2 standard for presentation by the REQUEST SENSE command. The proper sense data shall be presented when a SCSI status byte of CHECK CONDITION is presented by the SCSI Primary Commands -2 standard. to: The SENSE DATA field contains the autosense data (see SCSI Architecture Model - 2) when a SCSI STATUS byte of CHECK CONDITION is presented.

The present text is essentially identical to what every other autosense protocol specifies. While it might be desirable to formally define autosense data in SAM-2, then reference that from the protocol documents, that would need to start with the SAM-2 changes, not here. See CPQ006a.

CPQ006a

CPQ034 page 47 line 3 **Rejected 07 Jan 2002**

Change report an asynchronous event. to: report an asynchronous event (see SAM-2).

Rather than add a cross-reference here, add a glossary entry for "aynchronous event" that will cross-reference SAM-2.

CPQ035 page 47 line 3 **Closed**

Add sentence to first paragraph: Parameters managing the use of asynchronous event reporting are contained in the Control mode page (see SPC-2). This sentence is in SAM-2, but a direct reference from SRP seems helpful.

CPQ036 page 48 lines 3-8 **Rejected 07 Jan 2002**

IBM0126 Reword the SENSE DATA paragraph like in 6.9, but don't call it autosense here, call it "sense data for the event".

CPQ033 See CPQ033

CPQ037 page 50 line 15 and page 51 line 23 **Rejected 07 Jan 2002**

Table 29 Section 7.3 LUN should be LU (this is broken in SPC too) - the logical unit number is irrelevent here.

SRP references SPC-2 and SPC-3, it uses the names used in those documents.

CPQ038 Annex B **Accepted 07 Jan 2002**

Change (many places): Infiniband to: InfiniBand

The variable defining the reference to "InfiniBandTM Architecture Specification Volume 1 Release 1.0.a" will be corrected.

CPQ039 Annex B **Open**

There are too many TMs. There only needs to be one per page or one per the whole section.

There is no way to accomplish one per page without unreasonable manual effort. While there may be more TMs than necessary, including them is at worst harmless, at best legally necessary. Many other documents include a TM with every reference. I will not change this without either a legal opinion or direction from the ANSI editor.

7 Jan 2002 teleconference: Ed Gardner will obtain contact information for the ANSI editor (Harvey) from Ralph Weber, then confirm the proper style. This is the first T10 document that contains frequent references to a trademarked term.

CPQ040 Annex C **Closed**

Ralph Weber agreed to put alias formats for each protocol in SPC-3, so this annex can be removed.

HP comments:

CRS: Added numbers to all HP comments for easier cross-referencing.

HP01 **Pending**

Feb1: Add table in Annex A for port name, identifer, etc. SAM mappings. Expectation was for persistence already, text will make explicit.

Need a mandatory requirement to persistently report service names (DevMgtGetResp(ServiceEntries)) across IOU/IOC power cycles in order to persistently identify an SRP target port.

Description: Table B.8 describes the format of service name as SRP.T10:xxxxxxxxxxxxxxxx. Since the string xxxxxxxxxxxxxxxx in the service name identifies the 64 bit extension identifier value used to construct the SRP target port identifier, it is required that the service name reported by an IOU for a given SRP target port to be persistent across IOU/IOC power cycles. IB boot records contain SRP initiator port identifier, SRP target port identifier and logical unit name to locate an SRP boot LUN and the assumption is that the target port ID is persistent.

State that SRP port identifiers have the properties of names (see SAM-2: persistence, world-wide unique in context of SRP). Then the above falls out. *****

HP02 **Rejected Not reviewed**

These informal comments are the result of a newcomer's first in-depth reading of the SRP specification. I hope they will suggest avenues for further improvement, but they are not formulated at this time as specific requests for changes.

These comments derive from my work on iSCSI, and are in anticipation of development of iWARP, which will be an RDMA protocol for IP networks. IWARP is intended to provide a standard protocol-independent means of doing direct data placement into host memory, without the need for anonymous reassembly buffers. We anticipate that iSCSI and other Internet storage protocols such as CIFS and NFS will be adapted to iWARP. Inclusion of a formalized RDMA transport layer in the IP storage protocol stack places iSCSI on a path to converge with SRP.

Each protocol can learn from the other. Today, SRP, while meant to be generally applicable, is demonstrably applicable only to InfiniBand. ISCSI's applicability is similarly limited to IP networks. In the future, we may be able to engineer a single SCSI transport that works both with InfiniBand's RDMA service and with iWARP.

These are my personal comments, and are not meant to reflect an HP consensus. We at HP have not yet taken the time to form an internal consensus on SRP.

HP03 Page 1 Line 11 **Open**

It is not clear at the outset just what kind of standard SRP is. The text says that "the SCSI family of standards provides for many different transport protocols?" Is SRP a transport protocol? The text continues, "This standard defines the rules for exchanging information between SCSI devices using an RDMA communication service." So SRP is a mapping from SCSI to an abstract RDMA communication service? What then is the SCSI transport? Is it the combination of SRP and the underlying real RDMA communication service? The standard continues, "Other SCSI transport protocol standards?" So, perhaps SRP is a SCSI transport. A statement along these lines would help a lot: "SRP, in combination with a compatible underlying RDMA

communication service, is a SCSI transport. This document defines SRP and the requirements that SRP has for the underlying RMDA communication service."

HP04 Page 1 Line 19 **Pending**

Figure 1 shows the relationship of SCSI protocol standards, such as this one, to the other standards.

"Figure 1 shows the relationship of this standard to the other standards?" But it doesn't. The SRP standard is not identified in the figure. Despite the disclaimer, layering of the blocks does suggest a hierarchy, protocol stack and system architecture. But the figure does not indicate the applicability of SRP to the implementation of a SCSI transport, as far as I can tell.

HP05 Page 2 Line 28 **Open**

SRP is included in a list of transport protocols. So it is a transport protocol. But certainly it is not a complete transport protocol. A discussion of how SRP is used in combination with an underlying RDMA service and its transport protocol to form a SCSI transport protocol would be very instructive to the reader. This would involve a layering diagram-why not?

HP06 Page 8 Line 4 **Open**

It would be useful to say at the beginning of clause 4 that the purpose of clause 4 is to describe an abstract RDMA service that is suitable for supporting SRP. That is, to define SRP's requirements of an underlying RDMA service.

HP07 Page 8 Line 17 **Open**

"This clause describes various functions that may be provided?" Don't you mean to say that this clause describes various functions that must be provided by an RDMA service, in support of SRP? How the function is provided is immaterial, and of course it can be provided through further functional decomposition. Why mention it? Generally, this whole clause 4 seems to be descriptive of RDMA services in general, but not prescriptive in terms of SRP's requirements. It is difficult to separate descriptive information from requirements.

HP08 Page 8 Line 20 **Open**

"Annex B describes the mapping of these functions?" Is it the intention of SRP to work with other RDMA services besides InfiniBand? If so, it might be useful to mention that future revisions of the standard may include other Annexes that define the mapping of SRP to other RDMA services.

HP09 Page 10 Line 12 **Rejected**

CRS: Propose that this comment be rejected. WG agreed Feb 15.

SRP is deficient in not providing a security protocol for client (initiator) authentication. Is the notion of "other parameters required by the RDMA communication service" to be interpreted as suggesting that the RDMA service itself should provide authentication? Given that SCSI port names are conveyed by SRP, this doesn't seem possible. (The RDMA service will have its own names for its end nodes, but they're not related to SCSI/SRP port names.)

HP10 Page 11 Line 36 **Open**

"An RDMA communication service may require?" This sounds to vague and inclusive. What does SRP require of the RDMA service? That's all that should be defined in clause 4. It seems like SRP either will depend on the RDMA service's providing flow control for messages, or it will provide its own flow control. If SRP provides its own flow control, and doesn't depend on flow control from the RDMA service, then there is no reason to discuss flow control except maybe to mention that it is not required.

HP11 Page 12 Line 40 **Rejected**

4.5 Ordering and Reliability. Very glad to see this here. Wish it were in SAM-2.

HP12 Page 14 Line 24 **Open**

"Server address" probably should be "server identifier".

HP13 Page 15 Line 24 **Open**

Establishing multiple connections between an I,T port pair is an interesting concept, but may not be very useful, ultimately. The paragraph states that all such RDMA channels are associated with the single I_T nexus. While there is no ordering assumed between different RDMA channels (15-41), this channel independence cannot be maintained once the tasks are forwarded to the SCSI layer, where the RDMA channel allegiance of the task is forgotten, and only the I_T information is retained. Effectively, the tasks will merge from multiple transmission channels into a single queue as they transition from SRP to SCSI, and the original partial order will be replaced by a total order. Correct operation will result, but performance will suffer. Perhaps the only practical use of this construct is for the asynchronous transmission of task management requests, as in the given example.

HP14 Page 16 Line 28 **Open**

A request windowing scheme would be easier to describe than this request limit mechanism. Race conditions would not be an issue.

HP15 Page 20 Line 4 **Open**

Indirect data buffer descriptor. I don't see a good use for this facility in an IO application such as SRP, and I question its inclusion here. The channel adapter local to the memory that is to be read or written (typically the channel adapter of the Initiator) can use a scatter/gather list (SGL) to define an arbitrary virtual memory segment for an I/O buffer, and assign it a unique memory handle. This segment can then be read or written, starting at any offset, and in any order, by the target's RDMA mechanism's simply generating a series of RDMA reads or writes, always referring to the same memory handle, but using different offsets and lengths for each operation. (For example, a series of RDMA writes to increasing offsets, eventually filling the memory segment.) The direct data buffer descriptor format is sufficient for this operation, because the SGL provides for scatter/gather to bufflets that start and end at arbitrary addresses in physical memory (not just page-aligned addresses), just as a traditional DMA controller does.

The only motivation I can find for the indirect model is to reduce the number of SGLs (or mapped memory regions) that the initiator's channel adapter must deal with. Unfortunately, the

use of the indirect mechanism means that we must trust the target devices that share a memory region not to step on each other through misoperation or by deliberately generating invalid memory descriptors. While this is the truest form of remote DMA, because it leave the matter of address generation to the target device, it also leave the initiator exposed to target device misoperation, or worse.

I am not sufficiently familiar with IB HCA architecture to know whether such HCAs are limited to mapping only regions of contiguous pages, which would necessitate including the indirect data buffer descriptor method to support non-page-oriented IO.

HP16 Page 25 Line 1 **Open**

Login request. The statement that the login request "shall only be sent during RDMA channel establishment" seems to me overly restrictive on the RDMA model. Furthermore, I'm not sure I discern in clause 4 that the RMDA service must transport SRP login information during its own connection establishment, although this requirement is made clear in clause 5, line 14-13. It would seem quite natural to establish an RDMA connection first, and then log in SRP using the RDMA connection. (As an example, iSCSI establishes a TCP connection, and then logs it into a new or existing iSCSI session.)

HP17 Page 25 Line 1 **Open**

Login request. Need to resolve how security protocols are handled in the SRP world. The login request does not contain any provision for initiator port authentication to the target.

HP18 Page 25 Line 32 **Rejected**

SAM-2 r21 has no limits on port identifier size.

So port identifiers are 16 bytes. But SAM-2 rev. 17 allows 8 bytes only, and iSCSI allows 260 bytes or more (still in discussion). These differences need to be rationalized. It would be best if SCSI itself would adopt a naming convention for its ports, rather than delegating this crucial task to its many transports. If SCSI were to name its ports, then SRP would only have to convey the SCSI port identifier passed down the stack by SCSI, and not make provision for conveying an identifier defined by a lower-level transport.

HP19 Page 25 Line 32 **Rejected**

Feb1: WG agrees but sees no need to change.

The port identifier fields, at 16B, are too small to carry identifiers as used by iSCSI. This may prove problematical as we attempt to merge iSCSI and SRP for use with iWARP.

HP20 Page 54 Line 1 **Rejected**

A consumer may have many associated QPs.

SRP annex. Are Queue Pairs (QP) in one-to-one correspondence with IB consumers?

HP21 Page 54 Line 23 **Rejected**

That an IO Unit has a single CA is an IB decision - not within SRP's scope to define.

"An IB I/O unit?contains an IB channel adapter." Why restrict it to a single channel adapter? In Figure B.3 the analogous (but nameless) initiator unit-defined by the dashed lines-is shown with multiple channel adapters. An iSCSI device is conceived as having multiple channel adapters (known informally as channel groups and in the specification as portal groups). OTOH, since an IB I/O unit is not named (it has no GUID associated with it), is there any purpose to the architecture's defining it?

HP22 Page 54 Line 28 **Rejected**

Yes.

Figure B.2. Can I/O controllers be virtual objects?

HP23 Page 54 Line 28 **Rejected**

There's no reason to prohibit multiple consumers, and the term consumer is deliberately vague within the IB spec - it's the thing (e.g., a process) that reads/writes a QP . Multiple connections are independent of multiple consumers.

Figure B.2. What is the purpose of allowing multiple IB consumers per IB I/O controller? Is it so that multiple IB connections can be terminated within an IB I/O controller? (This relates to the question above about correspondence between QPs and IB consumers.)

HP24 Page 54 Line 28 **Rejected**

SCSI target ports contain the task router (SAM-2 4.7.2). There are no SRP restrictions on LU sharing.

Figure B.2. How are shared LUs modeled? Do SRP target ports contain the "task router" function described recently by Penokie? Can two IB I/O controllers have an underlying LU in common, or is this functionality restricted to two IB consumers within the same IB I/O controller?

HP25 Page 55 Line 9 **Closed**

Agreed. Should be removed.

Table B.1. IB port GUID is described as "Identifies an IB port within an IB channel adapter". This can be taken to mean that the naming scope for IB port is within a single channel adapter. I doubt that is the intention, since IB port GUIDs are globally unique. Similar comment for IB I/O controller GUID?with the further observation that IB I/O units themselves are not named, and so cannot form a naming scope. It seems to me that the first three lines of this table should read, "Identifies a _____", without qualification. It is incidental, isn't it, that an IB port is contained in an IB channel adapter (and an IB I/O controller is contained in an IB I/O unit)? The fact that the discovery process finds IB channel adapters, and then IB I/O controllers, and then IB consumers, utilizing the containment properties, seems irrelevant to describing the naming architecture, when globally unique names are used.

HP26 Page 55 Line 25 **Open**

Figure B.3. What is the object indicated by the dashed lines in the initiator model, analogous to the IB I/O unit in the target model?

HP27 Page 55 Line 25 **Accepted**

*CRS: Considering that targets can't derive any information (other than uniqueness) from an Init ID, the current text "*should be constructed...", "should an IB GUID available to the SRP initiator port...", *sounds like implementation suggestions.*

We implicitly define the size of the IID and TID in SRP_LOGIN_REQ, but it seems that we should do so explicitly - Annex A? We may want to say "this part is a GUID, this is an extension", and then in Annex B say what GUID it is.

Feb15: Add a SRP-general description to Clause 5. In Anx B, Shall be a GUID (EUI?), should be a CA GUID + extension.

Figure B.3. and 56-1. Table B.2. The rules for constructing initiator ports seem entirely too lax. The text says, "Initiator port identifier should be constructed?" And then the Table indicates that GUID, for example, is the channel adapter GUID. Is there no meaning associated with the initiator port ID? Is the only design goal that the 16B port ID be globally unique? Will any GUID do at all? If so, let's be explicit about this, and **let's not make any suggestions about the origin (and possible meaning) of the port name.**

OD4

But it would be a better model, I think, for the "GUID" used in the initiator port ID to be associated not with the IB channel adapter, but instead with the (unnamed) SRP initiator device. It is the SRP initiator device that is associated with a naming domain such as an operating system image. IB channel adapters will be shared among operating system images, and using them as a naming domain would require that the operating system images cooperate, or that the selection of port identifier extension be delegated to the virtual machine layer, both of which are undesirable.

While we're at it, let's decouple the naming of SRP ports entirely from IB. Although SCSI really should be the layer that names its ports, let's for the moment assume that SCSI continues to delegate port naming to its transport. But let's assume further that SRP accepts the responsibility to name its ports, and doesn't delegate it further to IB. SRP can then generate its own name for SRP initiator device, with an identifier extension to make a unique port name. Analogously, SRP can name the entities identified in the figure as SRP target devices. SRP could adopt a naming scheme that uses 16B "GUIDs" analogously to IB's, and it could draw from the same naming assignment authority that IB uses. But this is not the same as saying that IB defines SRP's port names, and in fact, the description of SRP port naming would be moved from the IB annex to the main SRP text.

This change would require that during the discovery process, the IB I/O unit return the full name of the SRP port from its Service Entries table, in step 3.

This approach to naming ports brings SRP much closer to iSCSI. What is unresolved is iSCSI's use of long text strings to name iSCSI devices vs. the use of more compact GUID numbers. The two mechanisms could be combined with the introduction of a name service that dereferences string IDs to GUIDs.

IBM / Tivoli comments:

IBM001 page a to page d **Rejected Not reviewed**

Page c - The page numbering in first part of the front matter is a,b,c, and d instead of roman numerals. This needs to be corrected.

As described by George Penokie, the T10 cover page and related front matter (e.g. revision history) will be torn off and discarded, leaving the remainder of the document as the ANSI standard. That is, page a through page d will be discarded, leaving page i as the first page of the ANSI standard. This is the required result. Any page numbering other than the current document would result in ambiguous page numbers (e.g. two page i's) or the first page of the resulting standard being something other than page i.

IBM002 page c **Rejected Not reviewed**

Page c - d - The Revision list needs to be removed before public review.

CPQ003 This draft is and was not intended for public review. See CPQ003.

IBM003 page c **Rejected Not reviewed**

All - All the line numbers need to be removed throughout the document.

CPQ003 This draft is and was not intended for public review. See CPQ003.

IBM004 page c **Rejected Not reviewed**

All - The printing date information at the bottom of every page needs to be removed.

This is a draft for T10 review, not the final standard.

IBM005 page viii line 3 **Accepted Not reviewed**

Page vii - Forward - the BSR number x3.269-199x is not correct for this standard. It should be 'NCITS.xxx-200x' until the actual number is assigned.

Bro103 See Bro103.

IBM006 page ix line 7 **Closed**

Page viii - LIne 7 - The statement ' The working draft SCSI RDMA Protocol (SRP) standard is divided into the following clauses:' should be 'The SCSI RDMA Protocol standard is divided into the following clauses:

Bro105 The document is a working draft until it is published by ANSI or NCITS. NCITS requires that we prominently label it a "working draft" until then. See Bro105.

IBM007 **Rejected Not reviewed**

All - The acronym SRP should be replaced with 'SCSI RDMA Protocol' in all cases in this document.

George Penokie has stated that T10 standards may either use an acronym or spell out the name of a standard, provided they are consistent. This standard consistently uses the acronym. See IBM010.

IBM010

IBM008a page 1 line 47, page 2 line 2 **Rejected Not reviewed**

Page 1-2 - The following standards should be removed from the list: FC-AL and FC-PH.

Both are approved standards that have not been withdrawn.

IBM008b page 2 line 4 **Rejected Not reviewed**

Page 2 - The following standard should be removed from the list: FC-PH-2.

That standard is not present in the list.

IBM008c page 2 line 10, page 2 line 20, page 2 line 31 **Rejected Not reviewed**

Page 2 - The following standards should be removed from the list: SPI-3, FCP and SPC.

All three are approved standards that have not been withdrawn. The follow-on projects for each of these have not been published or approved by INCITS (as of January 11, 2002).

IBM008d page 3 line 8 **Closed**

Page 3 - The following standard should be removed from the list: RMC.

IBM009 page 4 lines 4-6 **Rejected Not reviewed**

Page 4 - section 3.1.1 - The last sentence implies that SRP_LOGIN_RSP is the only use for accept data. I believe this is not correct. This should be stated to be an example of accept data.

Transporting an SRP_LOGIN_RSP is SRP's only use for accept data.

CRS: I don't read the def as being exclusive in any case.

IBM010 **Rejected Not reviewed**

All - The full name of a standard should always be used instead of the acronym. This should be change throughout the document.

IBM007

Duplicate comment. See IBM007.

IBM011 page 4 line 19 **Discussion needed**

CRS: There's a distinction between SRP the protocol and SRP the spec. SRP will always be the protocol, but SRP-2 will be the spec.

Page 4 - line 19 and others - when SRP is used and it is referring to this document then it should be changed to 'this standard'. Line 19 is one case where this appears to be true.

Change "SRP" to "this standard" in: **page ix** line 14, **page ix** line 15, **page 4** line 5, **page 4** line 19, **page 4** line 37, **page 5** line 2, **page 5** line 8, **page 5** line 11, **page 8** line 4, **page 10** line 11, **page 10** line 28, **page 10** line 35, **page 10** line 44, **page 11** line 2, **page 11** line 5, **page 12** line 41

, **page 12** line 44

, **page 10** line 44

Other changes:

page ix line 7: "The SCSI RDMA Protocol (SRP) standard" to "This standard".
page ix line 17: "features for SRP, including the SRP mode pages" to "features for this standard".
page 4 line 27: "the SRP" to "this standard".
page 12 line 42: "Use of SRP" to "Operation".

Self-references not changed:

All IU names, SRP request, SRP response, SRP information unit, SRP device, all occurences of SRP initiator/target port
page 14 line 3, **page 14** line 3, **page 16** line 28, **page 24** line 1, **page 24** line 4

IBM012 page 4 **Closed**

Replaced with definition from SAM2r22.

Page 4 - section 3.1.13 - The statement 'An externally addressable object...' should be 'An.addressable object...'. The term externally implies that the addressing is outside the standard.

IBM013 PDF Page 16 **Open**

Page 4 - section 3.1.15 - The last sentence implies that SRP_LOGIN_REQ is the only use for login data. If this is not correct. Then this should be stated to be an example of login data.

IBM014 PDF Page 16 **Open**

Page 4 - section 3.1.15 - The statement '...server agent or consumer...' should be '....server agent or server consumer...'

IBM015 PDF Page 16 **Open**

Page 4 - section 3.1 - The terms client consumer, server agent, and server consumer should be definitions is the glossary.

IBM016 PDF Page 17 **Open**

Page 5 - section 3.1.22 - The statement '...server agent or consumer...' should be '....server agent or server consumer...'

IBM017 PDF Page 17 **Open**

Pave 5 - section 3.2 - line 34 - The acronym for SRP implies that in almost all cases SRP should be changed to 'this standard'.

IBM018 page 8 **Closed**

Page 8 - line 5 - The statement 'by means of' should be change to 'using'.

IBM019 page 8 **Closed**

'established and destroyed'

Page 8 - line 44 - The statement 'established and disconnected' should be either 'established and removed' or 'connected and disconnected'. It this case I think the first option is better. The wording in the remaining document must then be make to match this change.

IBM020 page 8 **Pending**

Pages 8 - 11 - section 4.2 - This clause should be broken in subclauses and there should be references added between the steps in the figure and the text descriptions of those steps. This will help the reader relate the figures flow to the text.

IBM021 page 9 **Open**

Page 9 - lines 7-9 - The for example text should be change to (e.g., ...).

intel0016 **IBM022 page 9** **Open**

Page 9 - line 2 - The statement '...directed to a server and, if...' is not clear because there is a server agent and a server consumer. Which is this server supposed to be?

IBM023 page 9 **Open**

Page 9 - line5 - The statement '...identify the server with which...' is not clear because t there is a server agent and a server consumer. Which is this server supposed to be?

IBM024 page 9 **Open**

Page 9 - Figure 3 - line 40 - The arrow exiting to the right seems to dead end. Where does the flow go from there. All the other exit points are clear that one is not.

IBM025 page 10 **Open**

say "server identifier" identifies a server containing one or more target ports.

(T) Page 10 - line 12 - This states '...the server identifier shall identify one or more SRP target ports, and the login data...'. How is it possible for a single server identifier to identify more that one SRP port.? SCSI requires all target port identifiers be unique within a domain.

IBM026 PDF page 10 **Open**

Page 10 - at least lines 2-15 - The term 'server' is used by itself several times. There needs to be a qualifier on server so the reader does not assume that server equates to server agent and server consumer.

IBM027 page 10 **Open**

Page 10 - lines 28 - 29 - The statement 'With SRP the reject data includes an SRP_LOGIN_REJ response (see 6.4).' Is confusing in that it implies the SRP (which is this standard) has additional requirements than what was just specific in the sentence before. That does not compute and needs to be fixed.

IBM028 page 10 **Open**

Page 10 - lines 31 -32 - Is it possible for an RDMA channel to be successfully established and not operational? If not then the statement 'and is operational' should be deleted. If so then it needs to be explained how it is possible.

IBM029 page 10 **Open**

Page10 - line 34 - The statement '...server agent or consumer...' should be '....server agent or server consumer...'. This needs to be looked for throughout the document and corrected.

IBM030 page 10 **Open**

Page 10 - line 35 - The statement With SRP the accept data includes an SRP_LOGIN_RSP response (see 6.3).' Is confusing in that it implies the SRP (which is this standard) has additional requirements than what was just specific in the sentence before. That does not compute and needs to be fixed.

IBM031 page 10 **Open**

Page 10 - lines 44-45 - The statement 'With SRP the login data includes an SRP_LOGIN_REQ request (see 6.2)...' Is confusing in that it implies the SRP (which is this standard) has additional requirements than what was just specific in the sentence before. That does not compute and needs to be fixed.

IBM032 page 10 **Open**

Page 10 - lines 43 - 44 - The sentence 'The server agent is provided the login data from the client consumer's request in addition to RDMA communication service specific data.' is awkward. It would be better stated as 'The server agent receives the login data and RDMA communication service specific data from the client consumer's request.'.

IBM033 page 11 **Open**

Page 11 - line 2 - The statement 'With SRP the reject data shall contain an SRP_LOGIN_REJ response (see 6.4).' Is confusing in that it implies the SRP (which is this standard) has additional requirements than what was just specific in the sentence before. That does not compute and needs to be fixed.

IBM034 page 11 **Open**

Page 11 - lines 5 - 6 - The statement 'With SRP the accept data shall contain an SRP_LOGIN_RSP response (see 6.3)...' Is confusing in that it implies the SRP.(which is this standard) has additional requirements than what was just specific in the sentence before. That does not compute and needs to be fixed.

IBM035 page 11 **Closed**

Page 11 - line 11 - The term 'such' should be deleted.

IBM036 page 12 **Open**

Page 11 - lines 30-31 - The statement '...to deliver the message to the other consumer associated with the specified RDMA channel (the receiving consumer).' should be changed to '...to deliver the message to the receiving consumer.' There is no need to redefine what a receiving consumer is as that is done in the first paragraph of this section.

IBM037 page 12 **Open**

Pages 11 - 12 - section 4.4 - This clause should be broken in subclauses. For example at least an overview, one for read RDMA, and one for write RDMA. PDF Page 24

IBM038 page 12 **Closed**

Page 12 - line 5 - The statement 'as well' should be deleted.

IBM039 page 12 **Closed**

Page 12 - line 14 - The following statement 'Such information may be communicated by an application protocol.' Does not seem relevant to this standard and should be deleted.

IBM040 PDF Page 23 **Open**

Page 12 - lines 41-43 - This paragraph contains information that is not useful and should be deleted. It essentially states that RDMA communication has characteristics defined here and those not defined here are out side the scope of this standard. That is true but it is also true for every clause in this standard.

IBM041 page 13 **Closed**

Page 12 - line 45 - The statement 'or else' should be just 'or'.

IBM042 page 13 **Closed**

Feb15: As suggested by Bob Nixon: "without duplication"

Page 12 - line 46 - The term 'exactly' should be deleted. There is no difference between 'exactly once' and 'once'.

IBM043 page 12 **Open**

Page many - The terms Write and Read in RDMA Write and RDMA Read should not be capitalized.

IBM044 page 13 **Closed**

Page 13 - line 14 - The term 'satisfy' should be changed to 'meet'.

IBM045 page 14 **Closed**

Added (see SAM-2). Added comma after i.e. .

Page 14 - line 8 - The statement 'I_T nexus' is correct but there is no reference to where one would find out more about what it is. This needs to be added.

IBM046 page 14 **Open**

Page 14 - line 7 - The statement 'for its lifetime' is not clear. It should be stated as 'as long as it is established'. This ties it to the previous section. Note this assumes that the term established in 4.2 is not changed.

IBM047 page 14 **Open**

Page 14 - lines 24-28 - This whole paragraph does not look like it belongs here or anywhere and it should be deleted. It appears to be attempting to defines things that are either already defined in section 4 or don't need to be defined.

IBM048 page 14 **Closed**

While this does seem redundant, removing it would leave "tasks sent on that RDMA channel", which doesn't quite seem right, as we send IUs, not tasks.

At Feb15 mtg. decided to stay with orignal suggestion.

Page 14 - line 40 - The statement 'that were contained in SRP_CMD requests (see 6.8)' should be deleted as it is redundant with the statement 'outstanding SCSI tasks'.

IBM049 page 13, page 14 **Pending**

Accept.

We need to expand clause 4 discussion of Send (and other operations) to discuss completion, e.g., how long target waits after the Send before doing the disconnect.- DONE.

New text: An SRP target port should send an SRP_T_LOGOUT request (see 6.6) and wait for the RDMA communication service status indication (see 4.5.2) before requesting that an RDMA channel be disconnected. *George wants SHALL send, unless TP does not have a credit (as in case of IP not responding to a SRP_CRED_REQ).*

intel0042 (t) Page 14 - line 43 - The statement '...an SRP target port should send an...' gives inadequate guidance to a target implementor. This should be required to send the SRP_T_LOGOUT or

not send it. Or it should be specified when it is required to be sent and when it is not required to be sent.

IBM050 PDF page 15 **Pending**

*Accepted - Requested wording from George that isn't **too** broad here.*

(t) Page 15 - line 4 - I recommend adding into this list a statement that other SCSI related parameters (e.g., mode pages, logs) not be effected by the disconnect. This should avoid the hole the FC has dug for itself in this area.

IBM051 page 15 **Closed**

Page 15 - line 18 - The statement '...operation, if accepted, may allow...' should be '...operation may allow...'. The if accepted is redundant with may.

IBM052 page 15 **Closed**

Page 15 - line 36 - The term 'may' should be deleted.

IBM053 page 15 **Closed**

Page 15 - lines 36-40 - the format of the e.g is incorrect. It should be...'standards (e.g., ...).'.

IBM054 page 16 **Closed**

Page 15 - line 49 - The statement 'as well as' should be 'or'.

IBM055 PDF Page 28 **Open**

Page 16 - line 3 - The term 'initiation' should be 'start' or 'beginning'.

IBM056 PDF Page 28 **Open**

Page 16 - line 5 - The term 'all' should be 'the'.

IBM057 PDF page 16 **Pending**

To A, add SRP_CMD. Cover all SRP_ reqs.

Page 16 - lines 7-8 - I am not aware of a SCSI command that specifies that status not be returned. If there is such a thing then an e.g., would be helpful. If there is no such thing then this item should be deleted.

IBM058 PDF page 16 **Closed**

Page 16 - line 18 - What is the 'it' referring to? The 'it' needs to be replaced with whatever 'it' is.

IBM059 page 16 **Closed**

Page 16 - line 23 - The term 'might' should be 'may'.

IBM060 page 16 **Rejected**

Unable to see what could be confused. Willing to consider suggestions on text that wouldn't be extremely awkward.

Page 16 - line 23 - What is the 'it' referring to? The 'it' needs to be replaced with whatever 'it' is.

IBM061 PDF Page 28 **Open**

Page 16 - line 24 - The statement '...to at most one...' seems redundant. It should be '...to one...'.

IBM062 PDF Page 28 **Open**

Page 16 - lines 28-29 - The statement '...present in most information units...' is troublesome. There either needs to be a list of the IUs that have the field or a reference to a place that would tell my which IUs have or do not have the field.

IBM063 PDF Page 28 **Open**

Page 16 and others? - The when to use small caps rule is not being followed here. The rule is that small caps are only used when the field is being named (e.g., xxx field would have the xxx in small caps). When contents of the field is being called out it is not in small caps (e.g. request limit and request limit delta are both signed...').

IBM064 PDF Page 28 **Open**

Page 16 - line 49 - The sentence starting with 'An SRP port shall not specify a negative...' should be a separate item in the list.

IBM065 PDF page 16 **Closed**

Feb1: Change flow control to "target port buffer management". No change to non-commands.

(t) Page 16 - section 5.3 - This section on flow control seems overly complex for what appears to be actually needed. The only SRP request that even needs to have multiple outstanding requests in the command. All others should not be streamed but should be interlocked and some should be allowed to occur at any time. This all needs to be looked at to make sure the design point is what we really want.

IBM066 PDF Page 29 **Open**

Page 17 - Figure 4 - The way the arrows are pointing for the virtual address implies that it is not the address of the first byte of the memory segment. It currently implies that it is the space from the memory handle to the beginning of the memory segment which is the memory region. It is also not clear as to what the boundaries are of the memory region. The current drawing implies it is only the area above the memory segment. I do not believe that is correct so it needs to be fixed.

IBM067 PDF Page 29 **Open**

Page 17 - line 26 - There is no indication as to what kind of value the memory handle is. This would normally not be a problem except that the other two fields to explicitly indicate that they are unsigned integer values. I generally consider all fields to be unsigned integers but in this case there is doubt cast about that assumption.

IBM068 PDF Page 30 **Open**

Page 18 - line 1 - The statement 'A SRP...' should be 'An SRP...' This needs to be checked for throughout the document and corrected.

IBM069 PDF Page 30 **Open**

Page 18 - line 3 - The statement '...within its memory segment.' should be '...within the memory segment.'.

IBM070 PDF Page 30 **Open**

Page 18 - line 2 - The statement 'SRP target ports shall only issue the appropriate type of RDMA operation for a memory descriptor' appears to be restating what was stated in the previous sentence.and therefore should be deleted. The sentence would then read 'SRP target ports shall ensure that each RDMA operation...'.

IBM071 PDF Page 30 **Open**

Page 18 - line 3 - There needs to be a connection between the text above the a.b.c list and the list. Something like 'segment by using the following rules:'.

IBM072 PDF Page 30 **Open**

Page 18 - lines 15-17 - The sentences 'The format of each data buffer descriptor is specified by a format code value. Some data buffer descriptor format code values use the contents of a count field to further specify the data buffer descriptor format.' should be deleted as the information is a duplicate of what is in table 2.

IBM073 PDF page 18 **Open**

Page 18 - table 2 - line 35 - footnote c - There statement 'and and' should be just 'and' and there is not period at the end of the sentence.

IBM074 PDF Page 30 **Open**

Page 18 - table 2 - line 27 - The equation 20+16*count should be change to 20 + 16 x count. This change from * to x should be make throughout the document.

IBM075 PDF Page 30 **Open**

Page 18 - table 2 - footnote b - This should have a reference from the cell with 'count' in it.

IBM076 PDF Page 30 **Open**

Page 18 - lines 43-45 - The sentence 'An SRP initiator port shall not specify a data buffer descriptor format that was not indicated in the REQUIRED BUFFER FORMATS field value for that RDMA channel.' does not make sense. How can the initiator port be indicating the buffer formats in the REQUIRED BUFFER FORMATS field and at the same time not specifying the buffer formats in the REQUIRED BUFFER FORMATS field that were not indicated in the in the REQUIRED BUFFER FORMATS field. This is circular and needs to be fixed.

IBM077 PDF Page 30 **Open**

Page 18 - line 41 - There should be a reference to table 2 as follows 'data buffer descriptor formats (see table 2)'.

IBM078 PDF Page 30 **Open**

Page 18 - line 47 - The statement '...RDMA channel and...' should be '...RDMA channel request and...'.

IBM079 PDF Page 30 **Open**

Page 18 - line 49 - The statement '...RDMA channel and...' should be '...RDMA channel request and...'.

IBM080 PDF Page 30 **Open**

Page 18 - line 40 - There should be a reference to table 3 as follows 'The REQUIRED BUFFER FORMATS field (see table 3)...'.

IBM081 PDF Page 31 **Open**

Page 19 - line 4 - I believe the 'and' should be an 'or'. I don't believe a target port would do both IU at the same time.

IBM082 PDF Page 31 **Open**

Page 19 - lines 3-4 - There should be a reference to table 3 as follows 'The SUPPORTED BUFFER FORMATS field (see table 3)...'.

IBM083 PDF Page 31 **Open**

Page 19 - line 8 - The statement '...contents of the REQUIRED BUFFER...' should be '...contents of both the REQUIRED BUFFER....'.

IBM084 page 19, page 19 **Closed**

Two parts:

Init tells targ whether Init 'may use' IDBDs. Text implies that setting IBDB to zero in LOGIN_REQ is a promise that init will not send a CMD w/ an IDBD, but does not so state.

In request, reword to say initiator sets to specify whether it uses indirect format. Do not use should or shall.

edit010 *Added reference to what T_LOGOUT codes to report if detected..*

(t) Page 19 - line 18 and line 28 - Why is that when the IDBD bit and the DDBD bit is set to zero it is a should instead of a shall? This should be changed to a shall unless there is some good reason.

IBM085 page 20 **Closed**

Page 19 - note 2 - This note should note be a note. It should be part of the main text. It should also be restated as: 'The length of requests sent by an SRP initiator port, as determined by the data buffer descriptor formats, shall be limited to the MAXIMUM INITIATOR TO TARGET IU LENGTH field (see xxx) returned in the SRP_LOGIN_RSP response.

IBM086 PDF page 20 **Closed**

Accept.

(t) Page 19 - lines 39 - 40 - The sentence 'SRP target ports are not required to check the contents of the count field.' should be changed to 'SRP target ports shall ignore the contents of the count field.'.

IBM087 page 20 , page 34 **Closed**

EAG: Accept. Remove invalid count logout reason codes. Add incorrect IU length reason code.

intel0096 *CRS: Corrected text. Handling reason code under intel0096.*

(t) Page 19 - lines 44 - 45 - The sentence 'SRP target ports are not required to check the contents of the count field.' should be changed to 'SRP target ports shall ignore the contents of the count field.'.

IBM088 PDF Page 31 **Open**

Page 19 and others - line 39 and others - The term 'count field' is used in many places. First there are two of them so it should be 'count fields'. Second is not clear that these are the count fields in the SRP_CMD request. I recommend changing 'count field' to 'count fields in the SRP_CMD request' in all places in the main body text.

IBM089 PDF Page 32 **Open**

Page 20 - line 8 - The statement 'count field' should be 'DATA-OUT BUFFER DESCRIPTOR COUNT field (or DATA-IN BUFFER DESCRIPTOR COUNT field)'.

IBM090 PDF Page 32 **Accepted**

GOP: Says ref should be to FIGURE 5, not table 5.

Page 20 - line 12 - A reference to table 5 should be added to the end of the paragraph.

IBM091 PDF Page 32 **Open**

Page 20 - table 4 - line 29 - Footnote a - It's not clear which count field is being referred to. Is it the one in table 2 or the ones in the SRP_CMD_ request. This needs be fixed with the proper terminology and a reference to the correct place.

IBM092 PDF Page 32 **Open**

Page 20 - line 34 - The statement 'The DATA LENGTH field of the INDIRECT TABLE MEMORY DESCRIPTOR field value contains...' is not correct. It should be 'The DATA LENGTH field of the memory descriptors in the indirect table contains...'.

IBM093 PDF Page 32 **Open**

Page 20 - line 39 - The sentence 'SRP target port behavior when the TOTAL LENGTH field contains any other value is vendor specific.' should be moved to the end of the paragraph and restated as 'If the.TOTAL LENGTH field value is not equal to the to sum of the DATA LENGTH field values the SRP target port's behavior shall be vendor specific.'.

IBM094 PDF Page 32 **Open**

Page 20 - line 42 - It's not clear which count field is being referred to. Is it the one in table 2 or the ones in the SRP_CMD_ request. This needs be fixed with the proper terminology and a reference to the correct place.

IBM095 PDF Page 32 **Open**

Page 20 - line 47 - This should be the start of a new subclause. Something like 'SRP target port indirect data restrictions'. PDF Page 33

IBM096 PDF Page 32 **Open**

Page 21 - line 7 - This paragraph should be the start of a new subclause titled something like 'Examples of Indirect data buffers'.

IBM097 page 18 **Pending**

Although intended and supported, we don't show any examples or discuss it. Add a picture, some text, including that DBDs are not trequired to be same type. May need new heading.

Added text, no picture, didn't seem worth 1000 words.

(t) Page 20 and 21 - The possibility of having both a data-in and a data-out buffer is not described here. Why not? This needs to be fixed.

IBM098 PDF Page 32 **Open**

Page 21 - lines 12 and 13 - The term 'might' should be changed to 'may'. This should be done throughout this document.

IBM099 PDF Page 35 **Open**

Page 23 - line 48 - The statement 'A requestor shall provide a TAG value in each SRP request that is unique among all of the requestor's outstanding SRP requests with a particular responder. A responder shall copy the TAG value from each SRP request to the SRP request's SRP response. Responders are not required to check whether the TAG values of outstanding SRP requests are unique.'

should be

'Each SRP request shall contain a TAG value that is unique among all of the outstanding SRP requests from a particular SRP initiator port. Each SRP response shall contain a copy of the TAG value from the corresponding SRP request. Responders are not required to check whether the TAG values are unique.'

IBM0100 PDF Page 37 **Open**

Page 25 - line 2 - The term 'conveys' should be changed to 'sends'.

IBM0101 PDF Page 37 **Open**

Page 25 - line 42 - The statement '...wishes to send...' should be changed to '...sends...'.

IBM0102 PDF Page 37 **Open**

Page 25 - line 42 - The statement '....be 64 or larger.' should be '....be greater than or equal to 64.' or '...be greater than 63.'.

IBM0103 PDF Page 38 **Open**

Page 26 - lines 1-2 - The statement 'The MULTI-CHANNEL ACTION field identifies how an SRP target port treats any existing RDMA channel associated with the same I_T nexus. The MULTI-CHANNEL ACTION field is defined in table 10.' should be changed to 'The MULTI-CHANNEL ACTION field (see table 10) indicates how an SRP target port handles existing RDMA channels.associated with the same I_T nexus.'.

IBM0104 PDF Page 38 **Open**

Page 26 - table 10 - All the codes except for the 2 that are defined need to be listed as reserved. The row should have '02h - FFh' in the action column and 'reserved' in the description column.

IBM0105 PDF Page 39 **Open**

Page 27 - line 2 - The term 'conveys' should be changed to 'sends'.

IBM0106 PDF Page 40 **Open**

Page 28 - lines 1-2 - The statement 'MULTI-CHANNEL RESULT identifies how the SRP target port treated existing RDMA channels associated with the same I_T nexus. Table 12 defines this field.' should be changed to 'The MULTI-CHANNELRESULT field (see table 12) indicates

how an SRP target port handles existing RDMA channels associated with the same I_T nexus.'.

IBM0107 PDF Page 40 **Open**

Page 28 - table 12 - All the codes except for the 3 that are defined need to be listed as reserved. The row should have '03h - FFh' in the action column and 'reserved' in the description column.

IBM0108 PDF Page 42 **Open**

Page 30 - line 4 - The statement '...failed, rendering it non-operational.' should be changed to '...failed.'.

IBM0109 PDF Page 43 **Open**

Page 31 - line 4 - The statement '...failed, rendering it non-operational.' should be changed to '...failed.'.

Page 32 - line 2 - The term 'conveys' should be changed to 'sends'.

IBM0110 PDF Page 44 **Open**

Page 32 - lines 37-38 - The statement '...logical unit component of the nexus for the task management request.' should be changed to '...logical unit to which to send task management request.'.

IBM0111 PDF Page 46 **Open**

Page 34 - line 2 - The term 'conveys' should be changed to 'sends'.

IBM0112 PDF Page 46 **Open**

Page 34 - 35 - Table 20 - This table splits up a paragon and worse a sentence. This needs to be fixed.

IBM0113 PDF Page 46 **Open**

Page 34 - table 20 - The notation 'do' and 'di' are confusing when placed into a sentence (as in the footnotes). They should be changed to 'x' and 'y'.

IBM0114 PDF Page 48 **Open**

Page 36 - line 2 - The term 'conveys' should be changed to 'sends'.

IBM0115 PDF Page 48 **Open**

Page 36 - line 6 - The statement '...message capable of containing...' should be changed to '...message containing...'.

IBM0116 PDF Page 49 **Open**

Page 37 - The statement 'set to 1' should be 'set to one' and the statement 'set to 0' should be 'set to zero' in all cases throughout this document.

IBM0117 PDF Page 49 **Open**

Page 37 - line 44 - The statement 'are not reliable and' should be deleted as it contains no useful information.

IBM0118 PDF Page 50 **Open**

Page 38 - line 3 - Add a reference to the RSP_CODE values table (table 24) at the end of this paragraph.

IBM0119 page 42 **Closed**

(t) Page 38 - lines 15-17 - The statement 'If DOUNDER is set to 1, a transfer that did not use the entire data-out buffer was performed and the value of DATA-OUT RESIDUAL COUNT shall be equal to: data-out buffer length - highest offset of any data-out byte transferred - 1' needs to be changed to 'If DOUNDER is set to one and a transfer that did not fill the entire data-out buffer was performed the value of DATA-OUT RESIDUAL COUNT is defined as follows: DATA-OUT RESIDUAL COUNT = (data-out buffer length) - (highest offset of any data-out byte transmitted + 1)'

IBM0120 page 42 **Closed**

(t) Page 38 - lines 22-23 - The statement 'DATA-OUT RESIDUAL COUNT shall be equal to: data-out transfer length required by command - data-out buffer length' needs to be changed to 'The DATA-OUT RESIDUAL COUNT is defined as follows: DATA-OUT RESIDUAL COUNT = (Transfer length required by command) - (data-out buffer length)'

IBM0121 page 43 **Closed**

(t) Page 38 - lines 34-36 - The statement 'If DIUNDER is set to 1, a transfer that did not fill the entire data-in buffer was performed and the value of DATA-IN RESIDUAL COUNT shall be equal to: data-in buffer length - highest offset of any data-in byte transferred - 1' needs to be changed to

' If DIUNDER is set to one and a transfer that did not fill the entire data-in buffer was performed the value of DATA-IN RESIDUAL COUNT is defined as follows:

DATA-IN RESIDUAL COUNT = (data-in buffer length) - (highest offset of any data-in byte transmitted + 1)

IBM0122 page 43 **Closed**

(t) Page 38 - lines 41-43 - The statement 'DATA-IN RESIDUAL COUNT shall be equal to: data-in transfer length required by command - data-in buffer length' needs to be changed to "The DATA-IN RESIDUAL COUNT is defined as follows: DATA-IN RESIDUAL COUNT = (Transfer length required by command) - (data-in buffer length)'.

IBM0123 page 43 **Closed**

Page 39 - line 1 -The term 'certain' should be deleted.

IBM0124 page 44 **Pending**

(t) Page 39 - lines 30 - 41 - All this should be deleted and replaced with 'The SENSE DATA field contains the autosense data specified by the SCSI Primary Commands-2 standard. The proper SENSE DATA shall be presented when the SCSI status byte of CHECK CONDITION is presented as specified by the SCSI Primary Commands-2 standard. If no conditions requiring the presentation of SCSI sense data have occurred, the SENSE DATA field shall not be included in the SRP_RSP response and the RSPVALID bit shall be zero. SRP devices shall perform autosense.'

IBM0125 page 47 **Closed**

*It's not transport stuff in view here.Changed to : "*A target port sends an SRP_AER_REQ request (see table 27) to report an asynchronous event. "

Page 41 - line 2 - The term 'conveys' should be changed to 'sends'.

IBM0126 page 48 **Pending**

Added 'as': ' data as specified...'.

CPQ036

(t) Page 42 - lines 3-13 - All this should be deleted and replaced with the following "The SENSE DATA field contains sense data specified by the SCSI Primary Commands-2 standard.'. This is AER not a check condition they are different things. The only thing that should be stated here is that sense data is returned.

IBM0127 PDF Page 56 **Open**

Page 44 and others - line 16 and others - The term 'set to 0' and 'set to 1' should be 'set to one' and 'set to zero'.

IBM0128 PDF Page 56 **Open**

Page 44 - line 19 - The term 'all' should be deleted as it is redundant.

IBM0129 PDF Page 58 **Open**

Page 46 - figure A.2 and A.3 - line 15 and 43 - The statement '(SRP initiator' should be '(SRP initiator port)'.

IBM0130 PDF Page 58 **Open**

Page 46 and others- lines 22-26 and others - The 1,2,3 list should not have line spaces between numbered items. This should be fixed in all cases

IBM0131 PDF Page 64 **Open**

Page 52 - line 20 - The term 'executes' should be changed to 'processes'.

IBM0132 PDF Page 64 **Open**

Page 52 - line 32 - The statement '...a device or component...' should be 'an IB device or component...'.

IBM0133 PDF Page 65 **Open**

Page 53 - line 20 - There seems to be no definition of what a 'connection manager' is. This should, at least, be added to the glossary.

IBM0134 PDF Page 65 **Open**

Page 53 - section B.3.2 - The abbreviation IOC needs to be added to the list.

IBM0135 PDF Page 67 **Open**

Page 55 - lines 1-2 - The sentence 'The IB more IB LIDs and IB GIDs corresponding to an IB port GUID or IB channel adapter GUID.' does not seem to be a complete sentence and is not clear as to what it is trying to state. This needs to be fixed.

IBM0136 page 64 **Closed**

Accept.

(t) Page 56 - line 2 - Why is the should not a shall. I believe it should be changed to a shall.

IBM0137 PDF Page 68 **Open**

State that shall use IB GUID, but don't mention CA GUID or other specific GUID source.

Page 56 - line 15 - The statement '...field should an IB GUID...' should be '...field should be an IB GUID...'.

IBM0138 PDF Page 68 **Open**

Page 56 - line15 - The statement '...port, e.g. the...SRP initiator port.' should be '...port (e.g., the...SRP initiator port).'.

IBM0139 PDF Page 68 **Open**

Page 56 - lines 15-16 - The statement 'the IB channel adapter GUID for an IB channel adapter used the SRP initiator port.' is not very clear as to what it is. This needs to be fixed.

IBM0140 PDF Page 68 **Open**

Page 56 - line 20 - There is not clue as what a 'device management agent' is. This could be fixed by replacing 'device management agent' with the more generic term 'entity'.

IBM0141 PDF Page 68 **Open**

Page 56 - line 22 - The term 'indicated' is confusing in this sentence. A better term would be 'identified'.

IBM0142 PDF Page 68 **Open**

Page 56 - line 39 - The term 'indicated' is confusing in this sentence. A better term would be 'identified'.

IBM0143 PDF Page 68 **Open**

Page 56 - lines 41-42 - This sentence seems out of place here. I should be moved to right after figure B.3.

IBM0144 PDF Page 68 **Closed**

Page 56 - line 49 and page 57 - line 1 - The term ' IB I/O ' has been split across lines (and in this case across pages) at the /. This needs to be fixed so it will not happen. There is an option in frame that if selected will prevent this. It should be enabled for this document.

IBM0145 PDF Page 69 **Open**

Page 57 - line 34 - The 'it' at the beginning of the sentence should be replaced with whatever the 'it' is.

IBM0146 PDF Page 69 **Open**

Page 57 - line 46 and page 58 - line 1 - Why is the receive data-out mapped to RDMA requests and send data-in mapped to RDMA WRITE packets? One is a 'request' the other a 'packet' this seems strange shouldn't they be the same?

IBM0147 PDF Page 73 **Open**

Page 61 - table B.8 - line 31 - The statement '(binary zeros)' should be '(i.e., binary zeros)'.

IBM0148 PDF Page 73 **Open**

SRP does not define any format for the 3rd party device identifier for third party reservations. This needs to be added to comply with requirements in SPC-3.

IBM0149 **Open**

p50 line 11. "See 4x1" is a typo. I think this should be "See 4.4".

IBM0150 **Open**

p50 line 14. "Sever" should be "server".

IBM0151 **Open**

p50 line 35. "Sever" should be "server".

IBM0152 **Open**

p57 section B.6.5. The descriptions for data-in and data-out are not symmetrical. One is described in terms of an "RDMA READ Request" and the other in terms of "one or more RDMA WRITE packets". I think the rules are the same for both data-in and data-out (please let me know if I'm incorrect in that assumption). Describing them differently implies that they are somehow different, and generates unnecessary confusion. (This is the same as Tivoli comment number 146).

InfiniBand™ Trade Association comment:

IBTA page 68 **Closed**

by William Futral (Intel) The IBTA Application Working Group understands that the SRP document is out for review and would like to offer the following comment.

The value assigned to I/O Class field in Table B.7 of the SRP document needs to be changed as a result of a change made to the format of this component in the latest InfiniBand(TM) Identifiers Annex, which is a supplement to InfiniBand(TM) Architecture Specification Volume 1.

Attached is a PDF document that contains the new wording in the IBTA Annex (see T10/01-319).

A Class Category needs to be selected for the SRP protocol and inserted in the I/O class field in place of the 0xFF value currently stated. For example, if the Storage Class was selected, the value for I/O class in your Table b.7 would become 0x0100.

Bill Futral
Application Working Group Co Chair
InfiniBand Trade Association

Intel comments:

intel0001 Sect:1 page 1 **Closed**

Transport protocol s/b 'SCSI Protocol' Suggest shading box to clarify what we're doing inthis spec

intel0002 Sect:1 page 1 **Closed**

Remove 'Physical'

intel0003 page 4 **Open**

inconsistent use of 'the' before SRP - suggest no 'the'

intel0004 page 4 **Open**

Is it necessary to specify field size in definition?

intel0005 page 4 **Open**

'Application protocol' is not defined, thus what constitutes app proto data is unclear

intel0006 page 4 **Open**

Key feature is that data placement is under control of receiver

intel0007 page 4 **Open**

'path' is a poor term, implies routing

intel0008 page 4 (C) **Open**

'a transport protocol or service' - which is it? There appears to be an abstraction layering problem Using 'service' to define a service suggests we don't have a clean definition - we don't

intel0009 page 5 **Open**

rewrite as 'specific to an RDMA comm service'

intel0010 page 5 (C) **Open**

TP ID ' within an RDMA comm service' - another abstraction issue - what is a service?

intel0011 page 5 **Open**

Any reason to spec field size?

intel0012 Sect:3.3.9 page 6 **Closed**

reported as AN error

intel0013 page 8 (C) **Open**

Clause 4 alternates between being a generic overview of RDMA, including discussion of features not used by SRP (e.g., solicited events in 4.3), and being normative (numerous SHALLs), which seems out of place in a clause entitled '...model'

Suggest separating the architectural model from the normative.

intel0014 page 8 **Open**

Seems redundant to Line 10 above.

intel0015 page 9 (C) **Open**

Model is unclear: "A client consumer requests that the RDMA communication service establish an RDMA channel."

But RDMA_CS is defined as a protocol. The sense should be that the client requests a SERVICE PROVIDER establish a channel.

intel0016 page 9 **Closed**

IBM022

Duplicate of IBM 022, handled there.

"The request is directed to a server" - Ambiguous

There are several standard meanings for 'server' - a piece of HW, a process, etc.

intel0017 page 9 **Open**

Should we add 'and validate' to 'Determine'?

intel0018 page 10 (C) **Open**

We need a similar diagram for channel teardown.

intel0019 page 10 **Open**

(Many places in this clause) Some formatting is needed to set off model-specific terms such as "channel establishment failure response" - suggest bold or small caps. This would making parsing and understanding much easier.

intel0020 page 10 **Open**

Given the vague definition of RDMA CS, it's hard to tell what ' internal to the RDMA communication service' does or does not mean.

intel0021 page 10 **Open**

"An RDMA channel rejected response returns reject data" s/b "Rejection" data

intel0022 page 10, page 11 **Open**

'With SRP the reject data includes' - near duplicate of page 11, Line 2

intel0023 page 11 **Open**

'service specific data' s/b 'service-specific data' (global replace)

intel0024 page 10 **Open**

'requests that are acceptable to the RDMA communication service shall be passed to the server agent.' (SHALL in model clause.) What does it mean to be acceptable to the service? As there is no mapping of 'Server Agent' to any entity, on what is this requirement placed? Can this requirement be stated in SRP or Annex B -specific terms?

intel0025 page 11 **Open**

'reject(ion) data shall contain an SRP_LOGIN_REJ...' (SHALL) Do we need a subclause similar to '4.5 Ordering and Reliability' to capture size issues, so we can specify requirements on underlying interconnects? (e.g., Must be able to return _REJ as part of connection establishment protocol.)

intel0026 page 4 , page 11 **Open**

'accept data' s/b 'acceptance data'

intel0027 page 11 **Open**

It is unclear how an RDMA comm svc requests that a channel be disconnected.

intel0028 page 11 **Open**

Need to discuss the case of a channel being destroyed due to an error.

intel0029 page 11 **Open**

'A disconnect request causes an RDMA channel to become non-operational.' Is this a request by a consumer to the local CS provider, or to the remote client, server agent,...?

intel0030 page 11 **Open**

'may or may not' Since 'May' and 'May Not' are both defined to be equivalent to 'May or May Not', there appears to be no reason to include both. (global)

intel0031 page 11 **Open**

Suggest: 'The completion status of operations... is indeterminate.'

intel0032 page 11 **Open**

'disconnect request' s/b 'disconnection request' (global)

intel0033 page 11 **Open**

'An RDMA channel may allow its consumers to exchange messages.' One that did not would be useless for the present case, wouldn't it?

intel0034 page 11 **Open**

Now provided.

'may provide normal and solicited message reception notification,' Since not used by SRP, why included?

intel0035 page 11 **Open**

'providing the following to an RDMA communication service' Again, CS model issue - how do you provide this to a protocol?

intel0036 page 12 (C) **Open**

Sent mail 4 April asking EAG to clarify his intent on that statement.

'An RDMA communication service is not required to provide a way for a requesting consumer to determine whether the data has been written into the specified range of addresses in registered memory.' If the target does not know whether a write has completed, how does it know when to send status, and whether status is good or not?

intel0037 page 13 **Open**

'or else disconnect the RDMA channel.' 'destroy' is a better term to reflect the error case.

intel0038 page 13 **Open**

disconnect s/b destroy

intel0039 page 14 **Open**

NO, but should change match to "is identical to"

'An SRP target port shall not accept a new RDMA channel unless its SRP target port identifier matches the value in the SRP_LOGIN_REQ request.' As we have not defined 'match', do we need to explicitly allow wildcards?

intel0040 page 14 **Closed**

Addtional - spelling

intel0041 Sect:5.1.1 page 14 **Closed**

Mar 1: SHOULD

'Prior to requesting that an RDMA channel be disconnected, an SRP initiator port may send an SRP_I_LOGOUT' s/b SHALL send

intel0042 Sect:5.1.1 page 14 **Rejected**

IBM049

Mar 1: Stay w/ should

'Prior to requesting that an RDMA channel be disconnected, an SRP target port should send an SRP_T_LOGOUT request' s/b 'SHALL send'

intel0043 page 15 **Open**

'Following acceptance of a login specifying single RDMA channel operation that single RDMA channel' Add comma after 'operation'

intel0044 Sect:5.1.3 page 15, page 30 **Closed**

CRS: Either use
0001 0003h Unable to associate RDMA channel with specified I_T nexus. *or Propose new code -*

Mar 1: Add new code: RDMA Channel limit reached for this initiator (see 5.1.3)

CRS: Correct ref is 5.1.4.

'shall not accept such a login' What _REJ reason code is returned?

intel0045 Sect:5.1.3 page 15 **Closed**

identifoer

intel0046 page 15 **Open**

Break E.g. sentence into two or more sentences, or write as a note.

intel0047 page 16 **Open**

Suggest creating 5.3.1 Initiator Requests, and 5.3.2 Target Requests, to discuss separately. _Many_ reviewers have become confused with 'SRP target ports shall limit...' Add pointer to Table 7 and emphasis that these are target-initiated SRP requests, _not_ RDMA requests.

intel0048 page 16 **Closed**

'credit based' s/b 'credit-based'

intel0049 page 16 (C) **Open**

Results are are vendor-specific.

'An SRP initiator port shall not send an SRP request on any RDMA channel whose REQUEST LIMIT has a value less than or equal to zero.' What is Target Port response to this?

intel0050 page 17 (C) **Open**

'To ensure that task management requests may be sent, an SRP initiator port may choose to send commands only when REQUEST LIMIT is greater than one'

Since TargPort can remove an arbitrary number of credits at any time, Init Port can be prohibited from performing Task Mgmt or sending SRP_I_LOGOUT.

intel0051 page 17 (C) **Open**

'An SRP initiator port shall add...whenever it receives an information unit on that RDMA channel' What does 'receive' mean? Received at what layer? There may be a significant delay between receiving and reading.

intel0052 page 17 (C) **Open**

State that target shall not assume initiator has seen or responded to credit change until response is received. For changes that do not have responses (e.g. srp_rsp), there may be no way target can determine or assume initiator has responded.

When initiator disconnects channel, it shall send logout if positive credit balance. It shall simply disconnect (without logout) if zero or negative credit balance.

Consider sending logout as private data on disconnect? No, don't do that (Randy).

Target behavior is unpredictable if initiator exceeds credit limit.

Target Port maintains, implicitly or explicitly, a value representing its view of the number of free request contexts (Call this Target Request Limit TRL) When there are no requests outstanding, TRL will be equal to the initiator's REQUEST LIMIT (IRL).

The description in 5.3 only describes IRL, but TRL may differ from IRL, and there is no definition of when IRL is changed. Specifically, when TargPort sends SRP_CRED_REQ with a negative value, when does TP update TRL? It only makes sense to update upon receipt of SRP_CRED_RSP, but that is not stated.

Rewrite to describe with state variable at IP and at TP, and rules for updating.

intel0053 page 17 (C) **Accepted**

When TPort rcvs SRP_CRED_RSP.

When can TPort be sure that IPort has seen the REQ_LIMIT_DELTA in an SRP_RSP? (Receipt of transport ACK is not enough)

intel0054 page 17 **Open**

Cris suggested making limits with a guardband. Rob said make limit +2^30, which with worst case race condition means -2^31. Cris wants diagram with examples.

'An SRP target port shall not specify a negative value of REQUEST LIMIT DELTA that might cause REQUEST LIMIT to drop below -2^31' Given wrapping, it's impossible to drop below -2^31 in 32-bit 2's comp. Would -2^16 be negative enough?

intel0055 page 17 (c) **Closed**

Accept.

'An SRP target port shall account for all possible race conditions to meet these requirements.' Remove this sentence.

intel0056 page 17 **Open**

'memory segment' and 'memory region' need to be defined before use.

intel0057 Sect:5.4.1 page 17 **Rejected**

Byte addresses and offsets are deeply ingrained in the model.

'identifies the byte address' Isn't the interpretation of a VA up to the particular interconnect/transport?

intel0058 page 58 **Open**

(Memory Handle) 'The SRP initiator port shall use this value to locate the region.' It doesn't appear to be within our scope to define initiator memory controller implementations. Remove this sentence.

intel0059 page 17 **Open**

Drawing seems to indicate that memory addresses increase moving downward. Should be explicit.

intel0060 page 18 **Open**

'SRP target ports shall only issue the appropriate type of RDMA operation for a memory descriptor,' Add: 'depending on whether the descriptor was a DATA-IN or DATA-OUT descriptor'

intel0061 page 18 **Open**

'a) The RDMA operations VIRTUAL ADDRESS shall be greater' Should specify STARTING address.

Although VIRTUAL ADDRESS is a field name in Table 1, the field may have a different name in a particular interconnect's request format. Should not be in CAPS.

intel0062 page 18 **Open**

'Some data buffer descriptor format code values' s/b 'descriptor formats'

intel0063 page 18 **Open**

'use the contents of a count field to further specify the data buffer descriptor format.' specify -> describe

intel0064 page 18 (C) **Open**

'use the contents of a count field to further specify the data buffer descriptor format.'

'count' is essentially a pointer to another field someplace, but this is far from obvious when reading. Suggest we define a format for 'virtual fields', e.g, '*COUNT', or 'vCOUNT', which the reader could easily recognize. Clause 3 would contain a table allowing *COUNT to be looked up as 'SRP_CMD DATA_OUT BUFFER DESCRIPTOR COUNT or SRP_CMD DATA_IN BUFFER DESCRIPTOR COUNT, as appropriate'

intel0065 page 18 **Closed**

CPQ008 Remove period after PRESENT

intel0066 page 18 **Closed**

CPQ009 *(DUPLICATE OF CPQ 09)*

Note 'b' is not referenced above, probably s/b on 'count'

intel0067 Sect:5.4.1 Pg:18 Ln:43 **Open**

'initiator port may specify in SRP_CMD requests (see 6.8) sent on that RDMA channel. An SRP initiator port shall not specify a data buffer descriptor format that was not indicated in the REQUIRED BUFFER FORMATS field value for that RDMA channel. ' What is target response if it does?

intel0067a Sect:5.4.1 Pg:18 Ln:43 **Open**

'SRP target ports are not required to check SRP_CMD requests for data buffer descriptor formats that were not indicated in the REQUIRED BUFFER FORMATS field value.' Not clear - are they required to validate that they did a valid format?

intel0068 Sect:5.4.1 Pg:18 Ln:47 **Open**

'An SRP target port may accept an RDMA channel and' s/b 'channel establishment request'

intel0069 Sect:5.4.2.2 Pg:18 Ln:49 **Open**

shall reject the RDMA channel and return after channel, add 'establishment request'

intel0070 Sect:5.4.2.2 Pg:19 Ln:16 **Open**

indirect data buffer descriptor (IDBD) Use caps or formatting to set off these field names

intel0071 Sect:5.4.2.2 Pg:19 Ln:16 **Open**

if the SRP initiator port may specify the INDIRECT s/b 'if the TP will accept...'

intel0072 Sect:5.4.2.2 Pg:19 Ln:18 **Open**

does not use (Sense is that IP forebears use of indirect) shall not use?

intel0073 Sect:5.4.2.4 Pg:19 Ln:44 **Open**

'sixteen bytes' Previously defined in Table 2 - eschew multiple definitions

intel0074 Sect:5.4.2.4 Pg:19 Ln:48 **Open**

target port shall only issue RDMA Read operations using the memory descriptor contained in the direct data buffer descriptor. Statement does not have desired effect - limits what you can read, but does not limit accesses to READs. s/b 'shall issue only RDMA Reads when using'

intel0075 Sect:5.4.2.4 Pg:20 Ln:1 **Open**

shall issue only RDMA Writes...

intel0076 Sect:5.4.2.5 Pg:20 Ln:6 **Open**

format code value 'value' appears to be superfluous

intel0077 Sect:5.4.2.5 Pg:20 Ln:8 **Open**

'The length....sixteen bytes.' Drop sentence - redundant to Table 2

intel0078 Sect:5.4.2.5 Pg:20 Ln:10 **Open**

'An indirect data buffer is comprised of one or more memory segments' Need a real definition.

intel0079 Sect:5.4.2.5 Pg:20 Ln:11 **Open**

segments may or may not be contiguous. s/b 'may be discontiguous'

intel0080 Sect:5.4.2.5 Pg:20 Ln:11 **Open**

remove 'may be in a single memory region'

intel0081 Sect:5.4.2.5 Pg:20 Ln:12 **Open**

of the memory segments (ADD: listed in an IBDB)

intel0082 Sect:5.4.2.5 Pg:20 Ln:13 **Open**

may have any length As the length field is finite, so is the segment length

intel0083 Sect:5.4.2.5 Pg:20 Ln:29 **Open**

value contained in the data buffer descriptor\u2019s count field. Implies that the field is contained within the DBD

intel0084 Sect:5.4.2.5 page 21 **Discussion needed**

Mar 1: PMDL was Ed's original idea. Spec stability may be more important. Revisit after other 'count' changes made.

'count' Suggest replacing with 'PMDL Length'

intel0085 Sect:5.4.2.5 Pg:20 Ln:31 **Open**

DESCRIPTOR field value is a memory descriptor Suggest: DESCRIPTOR field contains a memory descriptor

intel0086 Sect:5.4.2.5 Pg:20 Ln:33 **Open**

concatenated together 'together' is redundant Stamp Out and Abolish Redundancy!

intel0087 Sect:5.4.2.5 Pg:20 Ln:35 **Open**

contains the number of memory descriptors in the indirect table times sixteen. Suggest: contains the length, in bytes, of the indirect table (16 bytes * number of descriptors in table)

intel0088 Sect:5.4.2.5 Pg:20 Ln:36 **Open**

MEMORY DESCRIPTOR field value contains any other drop 'value'

intel0089 Sect:5.4.2.5 Pg:20 Ln:43 **Open**

list of n memory descriptors Use bold or something to set off n

intel0090 Sect:5.4.2.5 Pg:20 Ln:47 **Open**

shall only issue s/b shall issue only

intel0091 Sect:5.4.2.5 Pg:21 Ln:1 **Open**

shall only issue s/b shall issue only (also Ln 4)

intel0092 Sect:5.4.2.5 Pg:21 Ln:12 **Open**

All four..., each might..., or several might be...

Awkward - generalize to: segments may be in different memory regions

intel0093 Sect:5.4.2.5 Pg:21 Ln:44 **Open**

value contains Drop: value (i.e.,) Add: in bytes

intel0094 Sect:6.1 page 24, page 45, page 47, page 49 **Closed**

Only possible Os are SRP_CRED_REQ and SRP_AER_REQ.

Mar 1: All are mandatory. Make sure spec says so.

Added statement, removed other text on IU pages about mandatory status.

Add M/O column, or statement that all are mandatory.

intel0095 Sect:6.1 Pg:23 Ln:24 **Open**

Add space between Tables 6 and 7 to clarify distinction between I>T and T>I requests.

intel0096 Sect:6.1 page 25, page 34 **Closed**

There is a code for bad type. Do we need one for "bad length for type"?

IBM087 *Mar1: Yes.*

shall send SRP_T_LOGOUT What reason code?

intel0097 Sect:6.1 Pg:23 Ln:48 **Open**

Need to define requestor, responder. Much reviewer confusion wrt Targ as requestor.

intel0098 Sect:6.1 Pg:24 Ln:2 (C) **Open**

Reject: If an initiator sends duplicate tags, target behavior is unpredictable. It is a non-goal to operate predictably in such conditions. Such behavior means the initiator is broken; it should be fixed. Add statement that target behavior is unpredictable. "If tag values are not unique, responder behavior is unpredictable".

Responders are not required to check whether the TAG values of outstanding SRP requests are unique. Since duplicate tags would likely cause a credit leak (one response for two requests), this could lead to deadlock, as InitReqLimit and TRL would be out of sync. We either need to require verification of uniqueness, or provide a ReqLimit re-sync mechanism.

intel0099 Sect:6.2 Pg:24 Ln:2 **Open**

shall only be sent during RDMA s/b: shall be sent only during RDMA

intel0100 Sect:6.2 Pg:24 Ln:41 **Open**

maximum length Add: in bytes

intel0101 Sect:6.3 Pg:27 Ln:4 **Open**

shall only be sent s/b: shall be sent only

intel0102 Sect:6.3 Pg:27 Ln:40 **Open**

maximum length Add: in bytes

intel0103 Sect:6.3 Pg:27 Ln:45 (C) **Open**

Accept, double check arithmetic (srp_aer_req vs. srp_rsp).

52 or larger AER_REQ requires 56

intel0104 Sect:6.4 Pg:29 Ln:3 **Open**

a(n) SRP target

intel0105 Sect:6.4 Pg:29 Ln:40 **Open**

too large Need a way to specify, so that Init does not have to guess

intel0106 page 34 **Open**

Need new subcluase for 'behavior'. To 4.3 or 4.5, add ACK/timrout wording.

"delay a vendor specific time" s/b Wait for transport ACK or timeout error

intel0107 Sect:6.6 Pg:31 Ln:3 (C) **Open**

An SRP_T_LOGOUT request may also be used to notify the SRP initiator port that an RDMA channel has failed, rendering it non-operational. If the channel has failed, it won't be able to carry this IU. We DO need a way to report failures.

intel0108 Sect:6.6 Pg:31 Ln:30 **Open**

There are no references in spec to reason codes 2,3, 6-9. Do we need some SHALLS pointing to them?

intel0109 Sect:6.6 Pg:31 Ln:45, **Open**

See also: **page 14**

delay a vendor... Reference: xport ack or timeout

intel0110 Sect:6.8 Pg:34 Ln:14 **Open**

COUNT Change to PMDL Length

intel0111 Sect:6.8 Pg:34 Ln:40 **Open**

Add ref a,b to notes below

intel0112 Sect:6.9 Pg:36 Ln:36 **Open**

Since SENSE DATA length is 7 bytes + a one-byte length field, at least the top two bytes s/b reserved. We may want to have this field be that one-byte length field, with 7 assumed, as in SPC.

intel0113 Sect:6.9 Pg:37 Ln:9 **Open**

length of the...buffer Ref 5.4 for length determination

intel0114 Sect:6.9 Pg:37 Ln:26 **Open**

indicates (that) the contents ....shall be ignored and (that) the

intel0115 Sect:6.9 Pg:37 Ln:26 **Open**

The(value of the) SENSE DATA LIST LENGTH field (be a multiple of four).

intel0116 Sect:6.9 Pg:37 Ln:26 (C) **Open**

reject: actual sense data length is in sense header.

SENSE DATA LIST LENGTH shall contain the length of the truncated SENSE DATA field. This is at odds with SPC-2, which returns the total length. How would you know that you had missed some Sense Data?

intel0117 Sect:6.9 Pg:37 Ln:47 **Open**

shall contain a length of 4 Also defined in Table 23 - refer to table instead

intel0118 Sect:6.9 Pg:38 Ln:17 **Open**

structure eqn as DOBL - (offset + 1) Much easier to understand (global change to all similar eqns) Formatting - more white space above and below, use bold font

intel0119 Sect:6.9 Pg:38 Ln:24 **Open**

may or may not not is the more worrisome case (more so for Ln 25)

intel0120 Sect:6.9 Pg:38 Ln:31 **Open**

Some commands may have a non-zero residual Add: e.g., INQUIRY

intel0121 Sect:6.9 Pg:38 Ln:45 **Open**

may not

intel0122 Sect:6.9 Pg:39 Ln:1 **Open**

certian (SRP) protocol errors

intel0123 Sect:6.9 Pg:39 Ln:18 **Open**

Drop NO FAILURE. Check SPI and FCP.

Would there ever be a case where a RSP of NO FAILURE was returned?

intel0124 Sect:6.9 Pg:39 Ln:31 **Open**

sense data shall be presented presented s/b returned Also Ln 32,33

intel0125 Sect:6.9 Pg:39 Ln:33 **Open**

whose Use whose wrt people only

intel0126 Sect:6.9 Pg:39 Ln:30 **Open**

SPC-2 Annnex C references SPC-3 - which?

intel0127 Sect:6.11 Pg:40 Ln:43 (C) **Open**

See comments on 5.3 for CRED_RSP issues

intel0128 Sect:6.12 Pg:41 Ln:31 **Open**

Don't need four bytes for SENSE data length (7 + 1 byte)

intel0129 Sect:6.12 Pg:41 Ln:43 **Open**

The (value of) the SENSE DATALen field (shall be a multiple of four.)

intel0130 Sect:6.12 Pg:41 Ln:44 (C) **Open**

Shall not be allowed.

If no sense data is provided, What would the point be - to force Init to issue Req Sense Request? Should this be allowed?

intel0131 Sect:6.12 Pg:42 Ln:1 **Open**

SENSE DATA LIST LENGTH shall contain the length of the truncated SENSE DATA field. Appears to violate SPC-2.

intel0132 Sect:6.12 Pg:42 Ln:7 **Open**

presented s/b (returned in response to)

intel0133 Sect:7.2 Pg:43 Ln:21 **Open**

The following subclause defines the fields in the disconnect-reconnect mode Nope - same subclause

intel0134 Sect:7.2 page 50 **Closed**

Gray-out or mark as Reserved the fields that are reserved for SRP. There's a lot of noise for the two fields that are used...

intel0135 Sect:7.2 Pg:44 Ln:1 **Open**

SRP devices shall only use (the) disconnect-reconnect page parameter fields Use formatting for disconnect-reconnect

intel0136 Sect:7.2 Pg:44 Ln:1 **Open**

SRP devices shall only use ...fields defined below. What about the standard mode page header fields?

intel0137 Sect:7.2 Pg:44 Ln:7 **Open**

field shall not be implemented by SRP target ports Define in terms of behavior, not implementation. Appears to have been covered by para above.

intel0138 Sect:7.2 Pg:44 Ln:17 **Open**

If the EMDP bit is set to 0, the SRP target port shall generate (RDMA requests with) continuously increasing () addresses for a single SCSI command.

intel0139 Sect:7.2 Pg:44 Ln:19 **Open**

affect the order of frames within an RDMA. What's a frame? Within an RDMA what?

intel0140 Sect:7.2 Pg:44 Ln:24 **Open**

intel0141 Sect:7.2 Pg:44 Ln:24 **Open**

protocol specific s/b protocol-specific (also Ln 27)

intel0142 Sect:7.2 Pg:44 Ln:28 **Open**

LUN -> PORT

intel0143 Sect:A.1 Pg:45 Ln:11 **Open**

Top right box s/b Device Server?

intel0144 Sect:A.1 Pg:45 Ln:29 **Open**

four step, two step s/b four-step, two-step (global)

intel0145 Sect:A.1 Pg:46 Ln:16 **Open**

Need close paren after initiator

intel0146 Sect:A.1 Pg:46 Ln:43 **Open**

Close paren

intel0147 Sect:A.3 Pg:47 Ln:11 **Open**

See table A.1 for the definitions of the names used within Don't see names there - objects?

intel0148 Sect:A.4.1 Pg:48 Ln:44 **Open**

Use bold for EXECUTE COMMAND

intel0149 Sect:B.3.1.7 Pg:52 Ln:35 **Open**

IBTA uses caps for G S I

intel0150 Sect:B.3.1.2 Pg:52 Ln:23 **Open**

Do we need to define, spell out GUID?

intel0151 Sect:B.3.1.14 Pg:53 Ln:1 **Open**

Ports also present on switches.

intel0152 Sect:B.3.1.16 Pg:53 Ln:5 **Open**

Speel out QP, use IBTA definitition.

intel0153 Sect:B.3.2 Pg:53 Ln:20 **Open**

IBTA uses caps for R T U

intel0154 Sect:B.4 Pg:54 Ln:50 (C) **Open**

Each IB GID is globally unique, Not true - see IBA Vol 1 4.1.1

intel0155 Sect:B.4 Pg:55 Ln:17 **Open**

worldwide Varies - see IBA Vol 1, 4.1.1

intel0156 Sect:B.5 Pg:56 Ln:2 **Open**

An SRP initiator device is one or more IB consumers may consist of

intel0157 Sect:B.5 Pg:56 Ln:15 **Open**

The GUID field should (be) an IB GUID available to the SRP initiator port, Must it be a GUID, an IB GUID, ...?

intel0158 Sect:B.5 Pg:56 Ln:17 **Open**

The IDENTIFIER EXTENSION field shall be chosen by the SRP initiator port to ensure that all SRP initiator port identifiers are unique. Over what domain?

intel0159 Sect:B.5 Pg:56 Ln:36 **Open**

[containing] the SRP target port. providing?

intel0160 page 64 **Open**

What is distinction between fabric and components thereof?

Given SAM-2 ambiguity on what SvcDelSys is, it's hard to resolve this.

The service delivery subsystem contains queue pairs, IB channel adapters, IB ports, and the InfiniBand TM Architecture fabric.

Contains exclusively? How does this map to Clause 4 RDMA Comm Service?

intel0161 Sect:B.5 Pg:56 Ln:47 **Open**

general service interface IBTA uses caps

intel0162 Sect:B.5 Pg:56 Ln:48 **Open**

I/ (breaks across page) O Remove slash from FRAME list of characters for line breaks.

intel0163 Sect:B.6.2 Pg:57 Ln:13 **Open**

open IBA connections use establish instead

intel0164 Sect:B.6.3 Pg:57 Ln:25 **Open**

Port and CM Redirection or Port Redirection. Very hard to parse - use bold or underscores inside the names

intel0165 Sect:B.6.4 Pg:57 Ln:38 **Open**

SRP_LOGOUT IU list as T_LOGOUT, I_LOGOUT or define as a virtual field

intel0166 Sect:B.6.4 Pg:57 Ln:38 **Open**

CM disconnect request use caps -it's not generic

intel0167 Sect:B.6.4 Pg:57 Ln:38 **Open**

The sender may disconnect if its send queue has transitioned to (THE) error state. What do you mean by disconnect here - local action?

intel0168 Sect:B.6.4 Pg:57 Ln:42 **Open**

The receiver of an SRP_LOGOUT IU shall respond with an InfiniBand TM Architecture transport acknowledgement and disconnect. Destroy QP, send DREQ, ...?

intel0169 Sect:B.6.5 Pg:57 Ln:46 **Open**

to an ... RDMA READ Request. One or more requests.

intel0170 Sect:B.6.5 Pg:58 Ln:1 **Open**

WRITE packets WRITE requests

intel0171 Sect:B.7 Pg:58 Ln:37 **Open**

outcommands

intel0172 Sect:B.7 Pg:59 Ln:7 **Open**

Why list ChangeID and OptionROM to say we don't care about them?

intel0173 Sect:B.7 Pg:60 Ln:23 (c) **Rejected**

Too hard to do in a dynamic environment.

Send Message Depth Reserved -> Maximum Initiator Request Limit This allows initiators to efficiently allocate buffers

intel0174 Sect:B.7 page 68 Ln:24 (c) **Closed**

RDMA Read Depth reserved -> Maximum IOC-issued RDMA depth Allows inits to efffi-ciently allocate RDMA resources

intel0175 Sect:B.7 page 68 Ln:26 (c) **Closed**

Send Message Size rsvd -> MAXIMUM INITIATOR TO TARGET IU SIZE Eliminates need to guess this value

intel0176 Sect:B.7 page 68 Ln:46 (c) **Closed**

This field is expected to be marked obsolete in future versions of the InfiniBand TM Architecture Not for T10/ANSI to say

intel0177 Sect:B.7 Pg:61 Ln:13 (C) **Open**

Is :reserved a literal? If not, express as :zzzz, explain below that it is reserved.

intel0178 Sect:B.7 Pg:61 Ln:16 **Open**

No references to Table notes.

intel0179 Sect:B.7 Pg:61 Ln:16 **Open**

padded s/b extended

Ophidian Designs comments:

OD 1 Page 13, lines 5-7 **Open**

multiple RDMA writes on the same channel store data in order. Some RDMA communication services (e.g. iWARP) are unable to ensure strict ordering of overlapping RDMA Write operations during normal operation. While methods are available to ensure strict ordering, invoking them for all RDMA Writes would severely affect performance.

SAM-2 does not specify the result of multiple commands to overlapping buffers in most cases. It is unclear whether it specifies the result in any situation (see T10/01-309). Overlapping transfers, also called data overlay, within a single command is unusual enough that some SCSI protocols routinely prohibit it.

This requirement should be removed from SRP. It should be replaced with a statement that overlapping transfers may yield unpredictable results unless the RDMA client (SRP) takes special precautions. The nature of said special precautions, if any, are RDMA communication service specific. A section should be added to clause 5 discussing data overlay to specify that SRP target ports shall take said special precautions whenever data overlay occurs within a command.

OD 2 Page 13, line 13,. **Open**

RDMA read operations may complete in any order. While this states that RDMA Read operations may complete in any order, it is not clear what data they are required to return. See the first example in T10/01-309r0.

If T10/01-309r0 is accepted, this should be clarified to indicate that the data returned by RDMA Read operations need not reflect concurrent RDMA Writes that precede the RDMA Read.

If T10/01-309r0 is not accepted, this should be changed to require that RDMA Reads and RDMA Writes to overlapping locations are strictly ordered for memory access.

OD3 **Rejected**

Feb15: Previously discussed - defer to SRP-2.

Page 14, RDMA channel disconnection
Page 15, Multiple independent RDMA channel operation
Page 16, lines 9 and 10 (list items b and c)
Page 27, SRP_LOGIN_RSP response
Page 30, SRP_I_LOGOUT request
Page 31, SRP_T_LOGOUT request

One of the characteristics of a network or fabric communication service is that errors affecting a channel can rarely be reported using that channel. In the context of SRP, many errors that disconnect an RDMA channel will be reported to one consumer but not the other. The consumer receiving the report cannot use the same RDMA channel to notify the other consumer, as the channel is no longer operational.

It is nonetheless useful for both consumers to know that an RDMA channel has failed. When using multiple independent RDMA channels, the consumers could use one of the other channels to report a channel failure. SRP should be extended to support this. This should be mandatory behavior whenever multiple channels are used between the same SRP initiator port

and the same SRP target port. The following paragraphs summarize the changes to SRP to accomplish this.

The SRP_LOGIN_RSP response should return a channel handle. The channel handle shall be non-zero and unique among all channels in use on the same I_T nexus. Zero is valid if and only if the SRP target port only supports one channel per nexus. The channel handle should be a 16-bit field in bytes 28 and 29.

The SRP_I_LOGOUT and SRP_T_LOGOUT requests should specify an optional channel handle. The channel handle should be a 16-bit field in bytes 2 and 3. If the channel handle is zero, it specifies that the channel on which the request was sent is being logged out; no response is generated. This is identical to the behavior currently specified by SRP. If the channel handle is non-zero then the specified channel is being logged out. A response is generated to confirm the logout and to indicate that all outstanding requests on that channel have been discarded. Targets shall not use of a non-zero channel handle that specifies the channel on which the SRP_T_LOGOUT request is sent. Use of a non-zero channel handle that specifies the channel on which the SRP_I_LOGOUT request is sent results in target specific behavior.

Extend the discussion of RDMA channel disconnection (page 14) and multiple independent RDMA channel operation (page 15) to require that targets report disconnection using an alternate channel if one is available.

Amend the list of requests that do not have responses on page 16 to say that SRP_I_LOGOUT and SRP_T_LOGOUT do not have responses when the channel handle is zero, but do have responses when the channel handle is non-zero.

Note that this change cannot be straightforwardly added in an SRP-2. An initiator or target that ignores the channel handle field (because it was reserved in SRP) would logout the wrong channel.

OD4 page 64 tables B.2 and page 64 B.3 **Closed**

CRS: Agreed Jan to swap GUID, extension. Agreed Feb1 NOT to change to :: format.

eag: Write more detailed proposal.

HP27 State that initiator port identifier embeds no information -- totally opaque.

Closed here, handle under HP27 Page 55 Line 25.

Target port identifiers may embed information on how to locate the target.

SRP port identifiers for Infiniband are 128-bit identifiers with an embedded GUID (EUI-64). Infiniband GIDs are 128-bit identifiers with an embedded GUID (EUI-64). Unfortunately they are formatted incompatibly. Annex B specifies that the EUI-64 occupies the most significant bytes of an SRP port identifier while the EUI-64 occupies the least significant bytes of an InfiniBand GID or IPv6 formatted address. The bytes not occupied by the EUI-64 are also different.

Having conflicting formats of otherwise equivalent identifiers is guaranteed to lead to interoperability problems. Various people have stated (in SRP working groups) that they expect to identify SRP targets using IPv6 formatted identifiers. SRP should be changed to satisfy this.

A new informative annex should be added recommending that SRP port identifiers adhere to IPv6 address formatting conventions and use one of the three forms listed below. Annex B should require that InfiniBand SRP port identifiers be one of the three forms listed below.

1. The Link-Local prefix (FE80h:0:0:0::/64) concatenated with an EUI-64.

2. The Site-Local prefix (FEC0h:0:0::/48) concatenated with 16-bit locally administered value concatenated with an EUI-64.

3, Any value configured manually or by a system management agent.

OD 5 Pages 4 and 5, **Open**

Glossary terms, and their use throughout the document, Clause 4: When SRP was proposed and for much of its development no satisfactory glossary of RDMA terms was available. Available external documents used definitions specific to particular implementations. That has recently changed. See the message titled “iWARP Glossary” posted to the yahoo RDMA reflector on September 27, 2001 by Jim Wendt. It would be beneficial if SRP were changed to use the same terms and definitions.

OD6 Page 11 lines 20-22 **Pending**

Refer to 02-064r4.

Normal and solicited message reception:

OD6a SRP_Login_Req page 26, page 26,

OD6h SRP_Login_Rsp page 28, page 28, page 29

OD6b SRP_TSK_MGMT page 35, page 35 , page 35

OD6c SRP_CMD page 38, page 37, page 38

OD6d SRP_RSP page 40, page 40

OD6e SRP_T_LOGOUT page 33, page 33, page 33

OD6f SRP_CRED_REQ page 45, page 45

OD6g SRP_AER_REQ page 47, page 47

OD6i SRP_LOGIN_REJ

OD6j SRP_I_LOGOUT page 32

OD6k SRP_CRED_RSP page 46

OD6l SRP_AER_RSP page 49

This feature is described in the RDMA communication service model, yet not used by SRP. Interrupt mitigation is important in high end systems. Therefore this should be supported by SRP information units. A description of how to do so follows.

Define a bit to be included in all SRP information units. Recommend this be bit 0 of byte 1 and called noturg (notification urgency or not urgent, take your pick).

In initiator to target requests, noturg specifies the notification urgency for the response. The initiator may set it to any value.

In target to initiator responses, noturg specifies the notification urgency. The target shall copy it from the request.

In target to initiator requests, noturg shall be zero. Specify this individually in each request, not as a general rule, so that it may be changed for future requests.

In initiator to target responses, the target shall ignore noturg.

In Annex B, specify that the target shall send information units with solicited event notification enabled if noturg is zero. The target shall send information units with solicited event notification disabled if noturg is one. The initiator shall ignore noturg and send all information units with solicited event notification enabled.

OD 7 **Rejected 28 Nov 2001**

Rob, Cris: reject. Worry about it in the future if/when it's a problem.

RDMA communication service specific opcode. SRP currently requires RDMA Read support for practical operation. However there are RDMA communication services that do not support an RDMA Read. So-called Unreliable Connections on InfiniBand are on example. Note that these have the same reliability characteristics as most existing SCSI protocols (e.g. FCP). Various people have suggested that they would be the most natural service for storage access, except for the lack of RDMA Read. Some VI Architecture implementations also lack RDMA Read.

It is straightforward to emulate an RDMA Read. The target sends a request to the initiator identifying the data to be read. The initiator responds with an RDMA Write supplying the required data, then a response to indicate completion. All that is missing is SRP opcodes that could be used for the request and response.

This is one example of a need for an RDMA communication service specific operation. Others might be required in the future for as yet unanticipated reasons. The purpose of defining this now is to describe proper behavior for an initiator that does not recognize the request.

The following could be defined using a new pair of opcodes or as an extension to the existing SRP_CRED_REQ and SRP_CRED_RSP. I don't particularly care which is used.

Define a target to initiator request. It is formatted identically to SRP_CRED_REQ with the addition of an action code field and action code specific parameters. I recommend a 16-bit action code field. The action code specific parameters may be any length (including zero) provided the total request length is within the limit agreed to during login.

Define the corresponding initiator to target response. It is formatted identically to SRP_CRED_RSP with the addition of an action code, an action response code and action code specific parameters. The action code is an echo of the value in the request (could be omitted). The action response code indicates the outcome of the action. Define value zero to designate the action is not supported, all other values reserved. The action code specific parameters may be any length (including zero) provided the total request length is within the limit agreed to during login. If the response code indicates the action was not supported, the action code specific parameters shall be zero length.

OD 8 page 18, **Rejected**

Feb15: Rejected by WG.

Data buffer format code and count values: The combination of a data buffer format code and a data buffer format count is awkward. Their interpretation is interdependent. We really have a single 12-bit field. It would simplify the description (and probably the implementation) if we had a single encoded data buffer format field. The following is a suggested way to encode an 8-bit data buffer format code:

00h	NO DATA BUFFER DESCRIPTOR PRESENT
01h	DIRECT DATA BUFFER DESCRIPTOR
02h – 0Fh	Reserved

1xh INDIRECT DATA BUFFER DESCRIPTOR

10h INDIRECT DATA BUFFER DESCRIPTOR WITH NO PARTIAL MEMORY DESCRIPTOR LIST

11h INDIRECT DATA BUFFER DESCRIPTOR WITH 1 ENTRY PARTIAL MEMORY DESCRIPTOR LIST

12h INDIRECT DATA BUFFER DESCRIPTOR WITH 2 ENTRY PARTIAL MEMORY DESCRIPTOR LIST

etc.

1Fh INDIRECT DATA BUFFER DESCRIPTOR WITH 15 ENTRY PARTIAL MEMORY DESCRIPTOR LIST

20h to FFh Reserved

These values would occupy bytes 6 and 7 of SRP_CMD, byte 5 would be reserved.

New editor comments:

edit001 page 60 **Closed**

IB GID: A port address used for directing packets between IB subnets. An IB GID is a 128-bit value that conforms to the IPv6 address format. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a

*Some GIDs are subnet-local, and thus fail the '*directing packets between IB subnets'. *Suggest removing first sentence.*

edit002 Table B.6, 7 **Accepted**

Add a footnote to the table saying it means "does not change or override IB reqmts".

Should we remove the '**no requirement**' statements from B.6 and 7, and replace with a statement that if not mentioned, SRP places no requirements? This avoids the appearance of overriding IBA specs, which may place requirements on them.

edit003 page 63 **Closed**

*Change LID description from "*Address assigned by the IB subnet manager to each IB port"
to
"Local routing address assigned to each IB port by the IB subnet manager"

edit004 page 64 **Open**

This seems an odd place to hide architecture model mappings. Move to a more appropriate place or remove.

edit005 **Discussion needed**

Will be considered at Mar13 CAP meeting.

SPC-3 says "These [alias] associations shall be cleared under any event that resets the logical unit and *events designated by the SCSI protocol*." It appears that we need to have a list or a statement that there are no such events. Where would it go?

edit006 page 43 **Open**

RSP_CODE 06h is not covered in table.

edit007 **Open**

Sense Data Length

edit008 page 15 **Open**

Although the target port is required to abort requests upon disconnect, logout, etc., there is no specification of the **order** in which tasks are to be aborted. This may result in a race condition. For example, if the target port issues ABORT TASK requests in the order oldest-to-youngest, a newer task could begin execution once an older task was aborted. There could be undesirable side-effects if (e.g.,) the older task had been issued with the Ordered task attribute to ensure that the task completed before the younger task began execution.

Proposed: Tasks shall be aborted from youngest to oldest.

edit009 **Open**

Since a logical unit would not have knowledge that an initiator was accessing it over multiple RDMA channels, it appears that a deferred error could be reported on any channel of the I_T nexus. This appears to include errors for commands that were issued on channels that have since been disconnected.

edit010 page 19 **Pending**

Although Targ Port is not required to check data buffer format, we need to say how it handles the detection of a bad one (T_LOGOUT with codes XXX, as appropriate). Change from 'not required to check', to 'should check'?

edit011 page 19 C **Pending**

Need to specify what REJECT code Targport shall return.

edit012 page 13 **Open**

"If an RDMA communication service is unable to meet these requirements "
THESE is ambiguous. Move to 4.5.1, make in "in this subclause"?

edit013 page 13 **Open**

"Messages sent {by the same consumer} on the same RDMA channel shall be delivered to the receiving consumer in the order they were sent."

is there any reason to say "by the same consumer"?

edit014 page 4 **Open**

Need to define:
server
server agent

server consumer
application protocol
application protocol data
,

edit015 **page 53** **Open**

Server Agent != Target Port

edit016

edit017

edit018

edit019

edit020

edit021

Texas Instruments comment:

This has the appearance of a draft copy, not a final review copy. Change bars and line numbers should not be on a letter ballot document.

Troika Networks comment:

Troika Networks, Inc.: page 30 **Closed**

Table 13 changed to C2h.

The TYPE code value of 80h in tble 13 is incorrect according to table 6 and should be value C2h.

Woven Electronics comment:

Woven Electronics: **Rejected**

Can not Contribute

Working Draft

T10 Project 1415-D

Revision 13
4 April 2002

CPQ001a
American National Standard for Information Systems – Information Technology –
CPQ001b
SCSI RDMA Protocol (SRP)

T10 Technical editor:

Cris Simpson
Intel Corporation
2111 NE 25th Ave.
Hillsboro, Oegon 97124
USA

Telephone: +1.503.712.4333
Facsimile: +1.503.712.2200
Email: cris.simpson@intel.com

Reference number

ANSI INCITS.*:200x**

IBM001

POINTS OF CONTACT

T10 Chair

John Lohmeyer
LSI Logic
4420 Arrows West Drive
Colorado Springs, CO 80907-3444
USA

Telephone: +1.719.533.7560
Facsimile: +1.719.533.7183
Email: lohmeyer@t10.org

T10 Vice Chair

George O. Penokie
IBM / Tivoli Systems
3605 Highway 52 North
Rochester, MN 55901
USA

Telephone: +1.507.253.5208
Facsimile: +1.507.253.2880
Email: gop@us.ibm.com

INCITS Secretariat

INCITS Secretariat
1250 Eye Street, NW Suite 200
Washington, DC 20005
http://www.incits.org/

Telephone: +1.202.737.8888
Facsimile: +1.202.638.4922
Email: ncits@itic.org

T10 Web Site

www.t10.org

T10 Reflector

To subscribe send email to majordomo@T10.org with ‘subscribe’ in message body
To unsubscribe send email to majordomo@T10.org with ‘unsubscribe’ in message body
Internet address for distribution via T10 reflector: T10@T10.org

Document Distribution

NCITS Online Store
managed by Techstreet
1327 Jones Drive
Ann Arbor, MI 48105

Web: http://www.techstreet.com/ncits.html
Telephone: +1.734.302.7801 or
+1.800.699.9277
Facsimile: +1.734.302.7811

Global Engineering
15 Inverness Way East
Englewood, CO 80112-5704

Web: http://global.ihs.com
Telephone: +1.303.792.2181 or
+1.800.854.7179
Facsimile: +1.303.792.2192

Revision History

Revision 13 (4 April 2002)

a) In-progress changes resolving letter ballot comments.
b) Removed Annex C.

Revision 12 (11 March 2002)

CPQ003

In-progress changes resolving letter ballot comments.

IBM002

Revision 11 (14 February 2002)

IBM003

In-progress changes resolving letter ballot comments.

IBM004

Special thanks to Ed Gardner, Ophidian Designs, for his work as the original editor for SRP.

Revision 10 (3 October 2001)

a) [01-289r0] Comments from 24 September 2001 SRP teleconference.
b) [01-298r1] Comments from 28 September 2001 SRP teleconference.
c) Reformatted SRP to Infiniband™ annex.

Revision 09 (12 September 2001)

a) [01-230r2] SRP buffer descriptor rewrite;
b) [01-250r1] SRP operation overview;
c) [01-263r0] Comments from August 21 SRP teleconference; and
d) Uniform use of "SRP target port", "SRP initiator port", "RDMA channel" and "IB channel".

Revision 08 (8 August 2001)

a) [01-028r6] SRP Infiniband™ annex;
b) [01-193r1] SRP alias entry designation formats (with extensive editorial changes);
c) [01-205r1] SRP Initiator Logout proposal;
d) [01-177r2] SRP model for RDMA communication services; and
e) [01-172r4] SRP to SAM-2 protocol.

Revision 07 (17 July 2001)

a) [01-195] Changes from June 19-20 SRP working group minutes; and
b) Corrections described in June 21 T10 reflector message from Kamran_Tavakoli@adaptec.com.

Revision 06 (14 June 2001)

a) [01-171r0] SRP_LOGOUT_REJECT, as modified during the May 25 teleconference (see 01-178);
b) [01-173r1] SRP bidirectional residuals, as modified during the May 25 teleconference (see 01-178);
c) Other changes approved during the May 25 teleconference (see 01-178);
d) Reconciled SRP_AER_REQ format to match revised SRP_RSP;
e) Reconciled SRP_TASK_MGMT format to match current SRP_CMD; and
f) Editorial changes and minor corrections in response to comments received on previous revisions.

Revision 05 (23 May 2001)

Numerous editorial changes. No intentional technical changes.

Revision 04 (10 May 2001)

Added mode pages, residual count clarification, AER, scatter / gather revision, total transfer length, logout, target / initiator port identifiers in login. Removed VI terminology, target reset, multiple command IUs. Believed to contain all approved changes through May 3 working group other than those listed above.

Revision 03 (29 January 2001)

Added RDMA Communication Model description. Fixed editorial errors in command IUs (restored bytes 4 to 7, three dots).

Revision 02 (4 January 2001)

Incorporates 00-354r2, scatter/gather and IU format changes defined at November 29-30 SRP working group (see 01-009r0), name changed to SRP, partial changes to use non-VI terminology.

Revision 01 (7 July 2000)

First semi-complete draft. Based on 99-316r1, 00-172r0 and 00-240r0. Tags expanded from 16 to 32 bits. TRD COUNT renamed REQUESTLIMIT and expanded to 32 bits. SVP_CMD and SVP_RSP IUs expanded to accomodate these fields and provide additional reserved words. Defined IU maximum size negotiation. Changed order of data transfer descriptor to match the order in Infiniband RDMA transport header.

Revision 00 (17 May 2000)

Partial draft. IBM001

ANSI (r)
INCITS.***:200x

American National Standard for Information Systems – Information Technology –

SCSI RDMA Protocol (SRP)

Secretariat

InterNational Committee for Information Technology Standards

Approved mm dd yy

American National Standards Institute, Inc.

ABSTRACT

This standard describes the message format and protocol definitions required to transfer commands and data between a SCSI (Small Computer System Interface) initiator port and a SCSI target port using an RDMA communication service.

American National Standard

Approval of an American National Standard requires verification by ANSI that the requirements for due process, consensus, and other criteria for approval have been met by the standards developer. Consensus is established when, in the judgment of the ANSI Board of Standards Review, substantial agreement has been reached by directly and materially affected interests. Substantial agreement means much more than a simple majority, but not necessarily unanimity. Consensus requires that all views and objections be considered, and that effort be made towards their resolution.

The use of American National Standards is completely voluntary; their existence does not in any respect preclude anyone, whether he has approved the standards or not, from manufacturing, marketing, purchasing, or using products, processes, or procedures not conforming to the standards.

The American National Standards Institute does not develop standards and will in no circumstances give interpretation on any American National Standard. Moreover, no person shall have the right or authority to issue an interpretation of an American National Standard in the name of the American National Standards Institute. Requests for interpretations should be addressed to the secretariat or sponsor whose name appears on the title page of this standard.

CAUTION NOTICE: This American National Standard may be revised or withdrawn at any time. The procedures of the American National Standards Institute require that action be taken periodically to reaffirm, revise, or withdraw this standard. Purchasers of American National Standards may receive current information on all standards by calling or writing the American National Standards Institute.

CAUTION: The developers of this standard have requested that holders of patents that may be required for the implementation of the standard, disclose such patents to the publisher. However, neither the developers nor the publisher have undertaken a patent search in order to identify which, if any, patents may apply to this standard. As of the date of publication of this standard and following calls for the identification of patents that may be required for the implementation of the standard, no such claims have been made.

No further patent search is conducted by the developer or the publisher in respect to any standard it processes. No representation is made or implied that licenses are not required to avoid infringement in the use of this standard.

Contents

Tables

Figures

Foreword

This foreword is not part of American National Standard INCITS.***-200x. Bro103

Suggestions for improvement, requests for interpretation, addenda, or defect reports are welcome. They should be sent to the INCITS Secretariat, c/o Information Technology Industry Council , 1250 Eye Street, NW, Suite 200, Washington, DC 20005. IBM005

This standard was processed and approved for submittal to ANSI by the InterNational Committee for Information Technology Standards (INCITS). Committee approval of this standard does not necessarily imply that all committee members voted for approval. At the time it approved this standard, INCITS had the following members: Bro104

Karen Higginbottom, Chair
David Michael, Vice-chair
Monica Vago, Secretary

(INCITS Membership to be inserted)

INCITS technical committee T10 on Lower-Level Interfaces, which developed this standard, had the following members:

John B. Lohmeyer, Chair
George O. Penokie, Vice-Chair
Ralph Weber, Secretary

(T10 Membership to be inserted)

Introduction

The Small Computer System Interface (SCSI) command set is widely used and applicable to a wide variety of device types. The transmission of SCSI command set information across an RDMA communication service allows the large body of SCSI application and driver software to be successfully used on the InfiniBand™ Architecture, the VI Architecture and other interfaces that support RDMA communication service semantics.

Bro105 The SCSI RDMA Protocol (SRP) standard is divided into the following clauses:

IBM006 Clause 1 is the scope.

IBM011 Clause 2 enumerates the normative references that apply to this standard.

Clause 3 describes the definitions, symbols, abbreviations, and conventions used in this standard.

Clause 4 describes the RDMA communication service model.

IBM011 Clause 5 describes significant concepts of SRP.

IBM011 Clause 6 describes the information units used by SRP.

IBM011 Clause 7 defines the SCSI management features for SRP, including the SRP mode pages.

Annex A through Annex C form an integral part of this standard.

Draft American National Standard **INCITS.***:200x**

American National Standard for Information Systems –
Information Technology –

SCSI RDMA Protocol (SRP)

Bro106

1 Scope

The SCSI family of standards provides for many different transport protocols that define the rules for exchanging information between different SCSI devices. This standard defines the rules for exchanging information between SCSI devices using an RDMA communication service. Other SCSI transport protocol standards define the rules for exchanging information between SCSI devices using other interconnects.

The set of SCSI standards specifies the interfaces, functions and operations necessary to ensure interoperability between conforming SCSI implementations. This standard is a functional description. Conforming implementations may employ any design technique that does not violate interoperability.

HP04

Figure 1 shows the relationship of SCSI protocol standards, such as this one, to the other standards and related projects in the SCSI family of standards as of the publication of this standard.

CPQ004

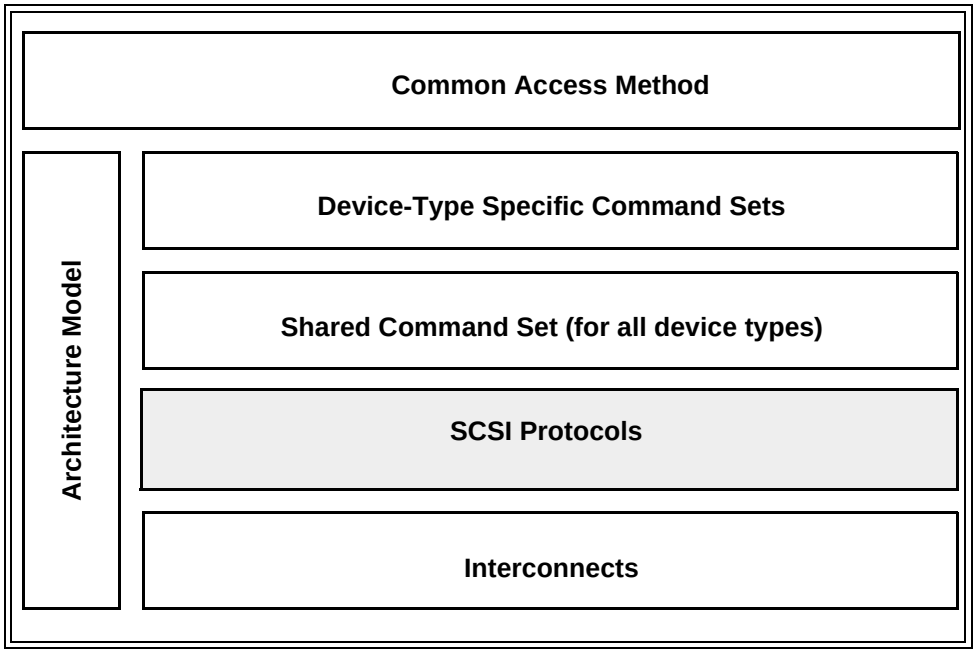


Figure 1 - SCSI document relationships

intel0001

Figure 1 is intended to show the general relationship of the documents to one another. Figure 1 is not intended to imply a relationship such as a hierarchy, protocol stack or system architecture. It indicates the applicability of a standard to the implementation of a given transport.

intel0002

At the time this standard was generated, examples of the SCSI general structure included:

Physical Interconnects:

Fibre Channel Arbitrated Loop	FC-AL	[ANSI X3.272:1996]
Fibre Channel Arbitrated Loop -2	FC-AL-2	[ISO/IEC 14165-122] [ANSI NCITS.332:1999]

IBM008a

IBM008a	Fibre Channel Physical and Signalling Interface	FC-PH	[ISO/IEC 14165-111] [ANSI X3.230:1994]
CPQ005	Fibre Channel Physical Amendment 1		[ANSI X3.230/AM1:1996]
IBM008b	Fibre Channel 3rd Generation Physical Interface	FC-PH-3	[ISO/IEC 14165-113] [ANSI X3.303-1998]
	Fibre Channel Framing and Signaling Interface	FC-FS	[T11/1331-D]
	High Performance Serial Bus		[ANSI IEEE 1394:1995]
	SCSI Parallel Interface - 2	SPI-2	[ISO/IEC 14776-112] [ANSI X3.302:1999]
IBM008c	SCSI Parallel Interface - 3	SPI-3	[ISO/IEC 14776-113] [ANSI NCITS.336:2000]
	SCSI Parallel Interface - 4	SPI-4	[ISO/IEC 14776-114] [T10/1365-D]
	Serial Storage Architecture Physical Layer 1	SSA-PH	[ANSI X3.293:1996]
	Serial Storage Architecture Physical Layer 2	SSA-PH-2	[ANSI NCITS.307:1998]
	Transport Protocols:		
	Serial Storage Architecture Transport Layer 1	SSA-TL-1	[ANSI X3.295:1996]
	Serial Storage Architecture Transport Layer 2	SSA-TL-2	[ANSI NCITS.308:1998]
IBM008c	SCSI-3 Fibre Channel Protocol	FCP	[ISO/IEC 14776-221] [ANSI X3.269:1996]
	SCSI-3 Fibre Channel Protocol - 2	FCP-2	[ISO/IEC 14776-222] [T10/1144-D]
	Serial Bus Protocol - 2	SBP-2	[ISO/IEC 14776-232] [ANSI NCITS.325:1999]
CPQ004	Serial Storage Architecture SCSI-2 Protocol	SSA-S2P	[ANSI X3.294:1996]
	Serial Storage Architecture SCSI-3 Protocol	SSA-S3P	[ANSI NCITS.309:1998]
	SCSI on Scheduled Transfer	SST	[T10/1380-D]
	SCSI RDMA Protocol	SRP	[T10/1415-D]
	Shared Command Sets:		
IBM008c	SCSI-3 Primary Commands	SPC	[ISO/IEC 14776-311] [ANSI X3.301:1997]
	SCSI Primary Commands - 2	SPC-2	[ISO/IEC 14776-312] [T10/1236-D]
	SCSI Primary Commands - 3	SPC-3	[ISO/IEC 14776-313] [T10/1416-D]
	Device-Type Specific Command Sets:		
	SCSI-3 Block Commands	SBC	[ISO/IEC 14776-321] [ANSI NCITS.306:1998]
	SCSI Block Commands - 2	SBC-2	[T10/1417-D]
	SCSI-3 Stream Commands	SSC	[ISO/IEC 14776-331] [ANSI NCITS.335:2000]
	SCSI Stream Commands - 2	SSC-2	[T10/1434-D]
	SCSI-3 Medium Changer Commands	SMC	[ISO/IEC 14776-351] [ANSI NCITS.314:1998]
	SCSI Medium Changer Commands - 2	SMC-2	[T10/1383-D]
	SCSI-3 Multimedia Command Set	MMC	[ANSI X3.304:1997]
	SCSI Multimedia Command Set - 2	MMC-2	[ISO/IEC 14776-362] [ANSI NCITS.333:2000]

SCSI Multimedia Command Set - 3	MMC-3	[T10/1363-D]
SCSI-3 Controller Commands	SCC	[ISO/IEC 14776-341] [ANSI X3.276:1997]
SCSI Controller Commands - 2	SCC-2	[ISO/IEC 14776-342] [ANSI NCITS.318:1998]
SCSI Reduced Block Commands	RBC	[ISO/IEC 14776-326] [ANSI NCITS.330:2000]
~~SCSI Reduced MultiMedia Commands~~	~~RMC~~	~~[T10/1364-D]~~ IBM008d
SCSI-3 Enclosure Services Commands	SES	[ISO/IEC 14776-371] [ANSI NCITS.305:1998]
SCSI Specification for Optical Card Reader/Writer	OCRW	[ISO/IEC 14776-381]
Object-based Storage Devices Commands	OSD	[T10/1355-D]
Architecture Model:		
SCSI-3 Architecture Model	SAM	[ISO/IEC 14776-411] [ANSI X3.270:1996]
SCSI Architecture Model - 2	SAM-2	[ISO/IEC 14776-412] [T10/1157-D]
Common Access Method:		
SCSI Common Access Method	CAM	[ISO/IEC 9316-421] CPQ004 [ANSI X3.232:1996]

The term SCSI is used to refer to the family of standards described in this clause.

2 Normative references

2.1 Normative references

The following standards contain provisions that, by reference in the text, constitute provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent editions of the standards listed below.

Copies of the following documents may be obtained from ANSI: approved ANSI standards, approved and draft international and regional standards (ISO, IEC, CEN/CENELEC, ITUT), and approved and draft foreign standards (including BSI, JIS, and DIN). For further information, contact ANSI Customer Service Department at +1.212.642.4900 (telephone), +1.212.302.1286 (facsimile) or via the World Wide Web at http://www.ansi.org.

Additional availability contact information is provided below as needed.

2.2 Approved references

ISO/IEC 14776-312, SCSI Primary Commands - 2 (SPC-2) [ANSI NCITS.351:200x]

2.3 References under development

At the time of publication, the following referenced standards were still under development. For information on the current status of the document, or regarding availability, contact the relevant standards body or other organization as indicated.

ISO/IEC 14776-412, SCSI Architecture Model - 2 (SAM-2) [T10/1157-D]

ISO/IEC 14776-313, SCSI Primary Commands - 3 (SPC-3) [T10/1416-D]

NOTE 1 - For more information on the current status of a document, contact the INCITS Secretariat at +1.202.737.8888 (phone), +1.202.638.4922 (fax) or via Email at ncits@itic.org. To obtain copies of these Bro107

documents, contact Global Engineering at 15 Inverness Way, East Englewood, CO 80112-5704 at +1.303.792.2181 (phone), 1.800.854.7179 (phone), or +1.303.792.2192 (fax), or at http://global.ihs.com .

3 Definitions, symbols, abbreviations and conventions

edit014

3.1 Definitions

IBM009 intel0026 **3.1.1 acceptance data:** Application protocol data communicated from a server consumer to the client consumer when a new RDMA channel is accepted (see 4.2). SRP uses acceptance data to communicate the SRP_LOGIN_RSP response (see 6.3).

IBM011 **3.1.2 application client:** An object that is the source of SCSI commands (see SAM-2).

CPQ006a **3.1.3 byte:** An 8-bit construct.

3.1.4 channel attributes: Information provided during RDMA channel establishment that identifies the type and characteristics of the desired RDMA channel (see 4.2). The format and interpretation of channel attributes are RDMA communication service specific.

3.1.5 command: A request describing a unit of work to be performed by a device server (see SAM-2).

3.1.6 command descriptor block (CDB): The structure used to communicate commands from an application client to a device server (see SPC-2).

IBM011 **3.1.7 consumer:** An object that communicates with other consumers using an RDMA communication service (see 4.1). Within SRP, a consumer is either an SRP target port or an SRP initiator port.

3.1.8 data-in buffer: The buffer identified by the application client to receive data from the device server during the execution of a command (see SAM-2).

3.1.9 data-out buffer: The buffer identified by the application client to supply data that is sent from the application client to the device server during the execution of a command (see SAM-2).

3.1.10 device server: An object within a logical unit that executes SCSI tasks according to the rules of task management (see SAM-2).

IBM011 intel0003 **3.1.11 information unit:** An organized collection of data specified by the SRP to be transferred as login data, reject data, accept data or a message on an RDMA channel.

3.1.12 initiator port identifier: A value by which a SCSI initiator port is referenced within a domain (see SAM-2).

IBM012 **3.1.13 logical unit:** ~~An externally addressable object within a target that implements a SCSI device model and contains a device server (see SAM-2).~~ A target-resident object that implements a device model and processes SCSI commands sent by an application client.

intel0004 **3.1.14 logical unit number (LUN):** A 64-bit identifier for a logical unit (see SAM-2).

intel0005 IBM011 **3.1.15 login data:** Application protocol data communicated from a client consumer to a server agent or consumer during RDMA channel establishment (see 4.2). SRP uses login data to communicate the SRP_LOGIN_REQ request (see 6.2).

intel0006 **3.1.16 message:** A communication sent by one consumer to another using an RDMA channel (see 4.3).

intel0007 **3.1.17 RDMA channel:** A communication path between two consumers of an RDMA communication service (see 4.1).

intel0008 **3.1.18 RDMA communication service:** A transport protocol or service that provides messages and RDMA operations between pairs of consumers (see clause 4).

3.1.19 RDMA operation: Either an RDMA Read operation or an RDMA Write operation.

3.1.20 RDMA Read operation: An operation by which a requesting consumer may fetch data from memory registered by the other consumer associated with an RDMA channel (see 4.4).

3.1.21 RDMA Write operation: An operation by which a requesting consumer may store data into memory registered by the other consumer associated with an RDMA channel (see 4.4).

3.1.22 reject data: Application protocol data communicated from a server agent or consumer to the client consumer when a new RDMA channel is rejected (see 4.2). SRP uses reject data to communicate the SRP_LOGIN_REJ response (see 6.4). IBM011 CPQ006b

3.1.23 server identifier: Information provided to an RDMA communication service by a client consumer that identifies a server with which to establish an RDMA channel (see 4.2). The format and interpretation of a server identifier are RDMA communication service specific. intel0009

3.1.24 SRP initiator port: A SCSI initiator port that uses SRP to communicate with an SRP target port. IBM011

3.1.25 SRP initiator port identifier: A value by which an SRP initiator port is identified to an SRP target port.

3.1.26 SRP target port: A SCSI target port that uses SRP to communicate with an SRP initiator port. IBM011

3.1.27 SRP target port identifier: A value by which an SRP target port is identified within an RDMA communication service. intel0010

3.1.28 status: One byte of response information sent from a device server to an application client upon completion of each command (see SAM-2). intel0011

3.1.29 target port identifier: A value by which a SCSI target port is referenced within a domain (see SAM-2).

3.2 Acronyms

CDB	Command Descriptor Block (see 3.1.6)
INCITS	InterNational Committee for Information Technology Standards
LSB	Least significant bit
LUN	Logical Unit Number (see 3.1.14)
MSB	Most significant bit
NCITS	National Committee for Information Technology Standards (now INCITS)
RDMA	Remote Direct Memory Access
SAM-2	SCSI Architecture Model - 2 (see 2.3)
SCSI	The architecture defined by the family of standards described in clause 1
SPC-2	SCSI Primary Commands - 2 (see 2.3)
SRP	SCSI RDMA Protocol (this standard)

3.3 Keywords

3.3.1 expected: A keyword used to describe the behavior of the hardware or software in the design models assumed by this standard. Other hardware and software design models may also be implemented.

3.3.2 ignored: A keyword used to describe an unused bit, byte, word, field or code value. The contents or value of an ignored bit, byte, word, field or code value shall not be examined by the receiving SCSI device and may be set to any value by the transmitting SCSI device.

3.3.3 invalid: A keyword used to describe an illegal or unsupported bit, byte, word, field or code value. Receipt of an invalid bit, byte, word, field or code value shall be reported as an error.

3.3.4 mandatory: A keyword indicating an item that is required to be implemented as defined in this standard.

3.3.5 may: A keyword that indicates flexibility of choice with no implied preference (equivalent to "may or may not").

3.3.6 may not: Keywords that indicate flexibility of choice with no implied preference (equivalent to "may or may not").

3.3.7 obsolete: A keyword indicating that an item was defined in prior SCSI standards but has been removed from this standard.

3.3.8 optional: A keyword that describes features that are not required to be implemented by this standard. However, if any optional feature defined by this standard is implemented, then it shall be implemented as defined in this standard.

3.3.9 reserved: A keyword referring to bits, bytes, words, fields and code values that are set aside for future standardization. A reserved bit, byte, word or field shall be set to zero, or in accordance with a future extension to this standard. Recipients are not required to check reserved bits, bytes, words or fields for zero values. Receipt of reserved code values in defined fields shall be reported as an error.

intel0012

3.3.10 restricted: A keyword referring to bits, bytes, words, and fields that are set aside for use in other SCSI standards. A restricted bit, byte, word, or field shall be treated as a reserved bit, byte, word or field for the purposes of the requirements defined in this standard.

3.3.11 shall: A keyword indicating a mandatory requirement. Designers are required to implement all such mandatory requirements to ensure interoperability with other products that conform to this standard.

3.3.12 should: A keyword indicating flexibility of choice with a strongly preferred alternative; equivalent to the phrase “it is strongly recommended”.

3.4 Conventions

Certain words and terms used in this standard have a specific meaning beyond the normal English meaning. These words and terms are defined either in 3.1 or in the text where they first appear.

Names of commands, statuses, sense keys, additional sense codes and additional sense code qualifiers are in all uppercase (e.g., REQUEST SENSE).

Names of fields and state variables are in small uppercase (e.g. ALLOCATION LENGTH). When a field or state variable name contains acronyms, uppercase letters may be used for readability (e.g. NORMACA). Normal case is used when the contents of a field or state variable are being discussed. Fields or state variables containing only one bit are usually referred to as the NAME bit instead of the NAME field.

Normal case is used for words having the normal English meaning.

Numbers that are not immediately followed by lower-case b or h are decimal values.

Numbers immediately followed by lower-case b (e.g. 0101b) are binary values.

Numbers or upper case letters immediately followed by lower-case h (e.g. FA23h) are hexadecimal values.

Lists sequenced by letters (e.g., a-red, b-blue, c-green) show no ordering relationship between the listed items. Numbered lists (e.g., 1-red, 2-blue, 3-green) show an ordering between the listed items.

If a conflict arises between text, tables or figures, the order of precedence to resolve the conflicts is text; then tables; and finally figures. Not all tables or figures are fully described in the text. Tables show data format and values.

Notes do not constitute any requirements for implementors.

3.5 Notation for procedures and functions

In this standard, the model for functional interfaces between objects is the callable procedure. Such interfaces are specified using the following notation:

[Result =] Procedure Name (IN ([input-1] [,input-2] …]), OUT ([output-1] [,output-2] …))

Where:

Result:	A single value representing the outcome of the procedure or function.
Procedure Name:	A descriptive name for the function to be performed.
Input-1, Input-2, …:	A comma-separated list of names identifying caller-supplied input data objects.
Output-1, Output-2, …:	A comma-separated list of names identifying output data objects to be returned by the procedure.
"[…]":	Brackets enclosing optional or conditional parameters and arguments.

This notation allows data objects to be specified as inputs and outputs. The following is an example of a procedure specification:

Found = Search (IN (Pattern, Item List), OUT ([Item Found]))

Where:

Found = Flag
Flag, which, if set, indicates that a matching item was located.

Input Arguments:

Pattern = … /* Definition of Pattern object */
Object containing the search pattern.

Item List = Item<NN> /* Definition of Item List as an array of NN Item objects*/
Contains the items to be searched for a match.

Output Arguments:

Item Found = Item … /* Item located by the search procedure */
This object is only returned if the search succeeds.

intel0013

4 RDMA communication service model

4.1 Overview

IBM011

SRP is designed to operate using an RDMA communication service. An RDMA communication service provides communication between pairs of consumers ~~by means of~~ using messages for control information and RDMA operations for data transfers.

IBM018

Figure 2 shows an example system that uses an RDMA communication service. Communication is provided by RDMA channels. An RDMA channel provides communication between two consumers. A single pair of consumers may communicate using many RDMA channels if sufficient resources are available. Some environments may use multiple special purpose RDMA channels between a single pair of consumers (e.g., a pair of consumers may use certain RDMA channels for messages and other RDMA channels for RDMA operations).

The RDMA communication service in figure 2 is comprised of adapters and other unspecified components (e.g. wires, fabric switches). The components of an RDMA communication service are implementation specific. Components such as adapters may or may not be present.

This clause describes various functions that may be provided by an RDMA communication service. A specific implementation of an RDMA communication service may or may not provide these functions. Any of these functions may be mapped to a sequence of several functions provided by the RDMA communication service. Annex B describes the mapping of these functions to those provided by the InfiniBand™ Architecture.

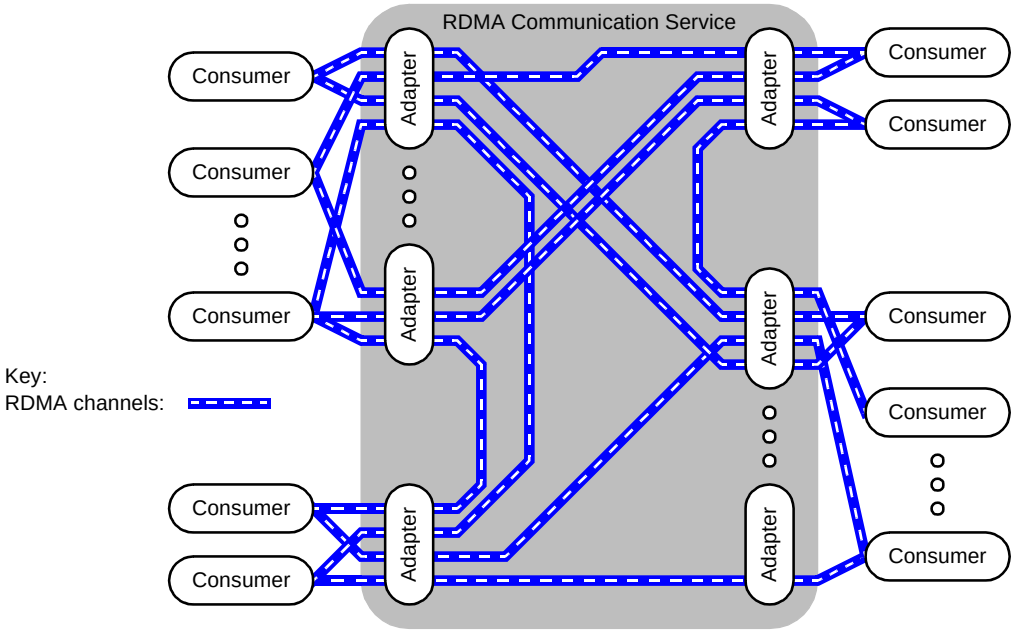


Figure 2 - RDMA communication service example

IBM020

4.2 RDMA Channels

4.2.1 Introduction

IBM019

An RDMA channel provides communication between a pair of consumers using messages, RDMA operations or both. An RDMA channel is a dynamic association~~connection~~, established and ~~disconnected~~ destroyed upon request. Establishing an RDMA channel may require obtaining resources to support the RDMA channel, either within the RDMA channel's consumers or within the RDMA communication service or both. Multiple RDMA channels may be established between the same pair of consumers if sufficient resources are available. The resources associated with an RDMA channel may be released after the RDMA channel is disconnected.

intel0014

4.2.2 Establishment

Figure 3 shows an example of the process by which an RDMA channel is established. A client consumer requests that the RDMA communication service establish an RDMA channel. The request is directed to a server and, if successful, resolved to a server consumer. The resulting RDMA channel provides communication between the client consumer and the server consumer. intel0015 IBM022

A client consumer provides a server identifier to identify the server with which to establish an RDMA channel. The format and interpretation of a server identifier are specific to the RDMA communication service. A server identifier may specify an individual server consumer or multiple server consumers. For example, a server identifier may identify an adapter as shown in figure 2, specifying all consumers that implement a specific application protocol and are accessible through that adapter. IBM023 IBM021

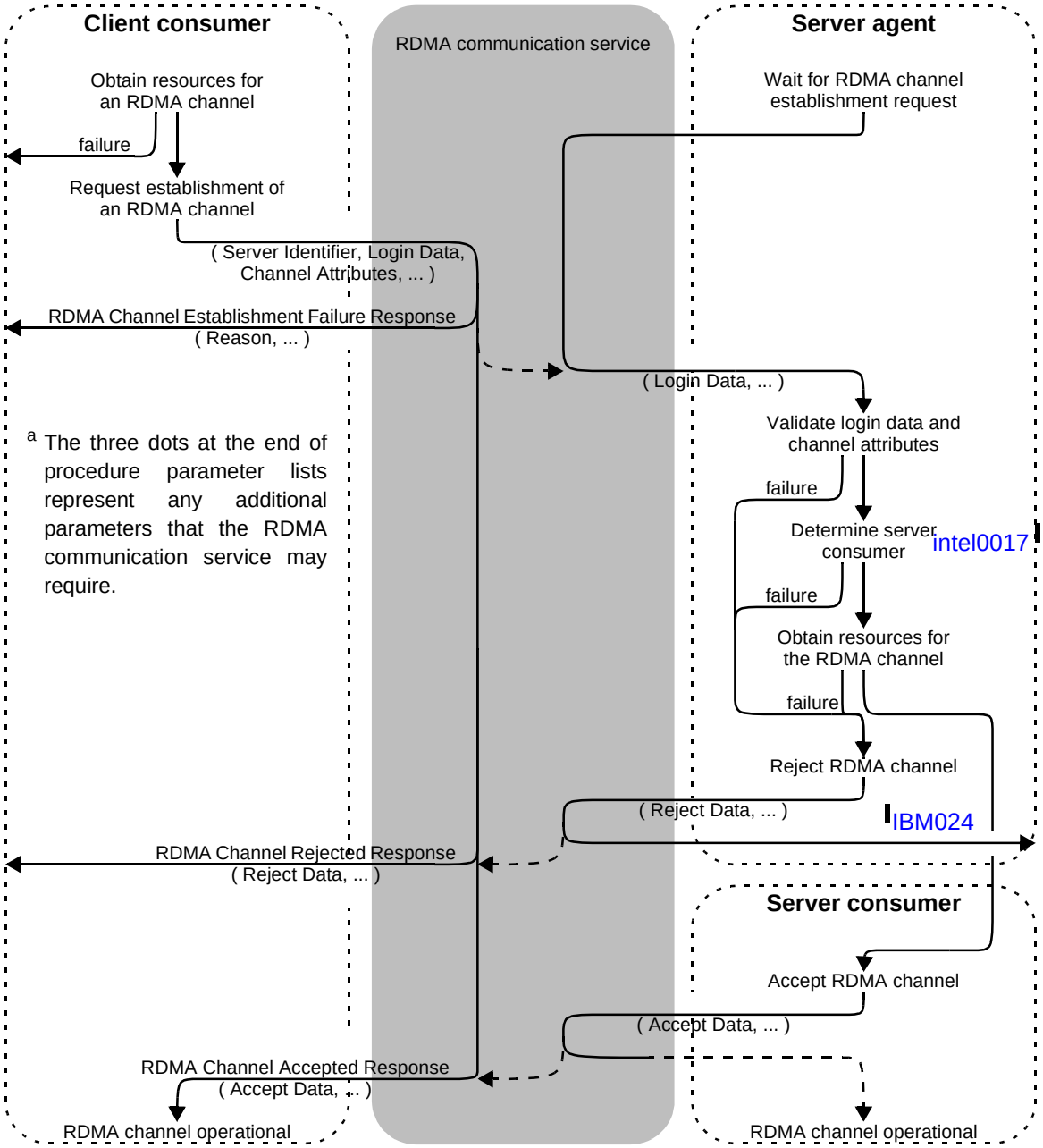


Figure 3 - Example RDMA channel establishment

intel0018 In the example shown in figure 3 the recipient of an RDMA channel establishment request, identified by a server identifier, is called a server agent. The server agent may use application protocol and server specific knowledge to determine whether an RDMA channel establishment request may be accepted and the server consumer to which it shall be assigned. The actions required of a server agent and their order are specific to the RDMA
IBM026 communication service and server. A server agent may or may not be a distinct object. Some or all of the actions that figure 3 shows being performed by a server agent may be performed by a server consumer or by the RDMA communication service.

An RDMA communication service may require that the client consumer obtain resources before requesting that an RDMA channel be established. After obtaining those resources, the client consumer may request that the RDMA communication service establish an RDMA channel. The request includes a server identifier, login data,
IBM011 channel attributes, and any other parameters required by the RDMA communication service. With SRP the
IBM025 client consumer shall be an SRP initiator port, the server identifier shall identify one or more SRP target ports, and the login data shall contain an SRP_LOGIN_REQ request (see 6.2).

The RDMA communication service returns one of three responses to the client consumer for an RDMA channel establishment request:

intel0019
a) An RDMA channel establishment failure response;
b) An RDMA channel rejected response; or
c) An RDMA channel accepted response.

intel0020 An RDMA channel establishment failure response indicates that the RDMA channel was not established for some reason internal to the RDMA communication service. An RDMA channel establishment failure response may return an RDMA communication service specific reason code to identify the cause of the failure as well as other RDMA communication service specific data.

intel0021 An RDMA channel rejected response indicates that the request was communicated to the server but rejected by the server agent or server consumer. An RDMA channel rejected response returns reject data, which is
IBM011 application protocol data provided by the server agent or server consumer. Reject data may include a reason for rejecting the request or other application protocol information. With SRP the reject data includes an
IBM027 SRP_LOGIN_REJ response (see 6.4). An RDMA channel rejected response may also return RDMA communication service specific data.

intel0022 An RDMA channel accepted response indicates that the RDMA channel has been successfully established and
IBM028 is operational. The client consumer may use the RDMA channel in accordance with the application protocol. An
IBM029 RDMA channel accepted response returns accept data, which is application protocol data provided by the
IBM011 server agent or consumer. Accept data may include application protocol parameters governing how the RDMA
IBM030 channel should be used. With SRP the accept data includes an SRP_LOGIN_RSP response (see 6.3). An RDMA channel accepted response may also return RDMA communication service specific data.

An RDMA communication service may require that a server agent register itself prior to receiving connection establishment requests. In figure 3 this is shown as a registration request (e.g., subroutine call) that returns control to the server agent when an RDMA channel establishment request is received. The way that a server agent registers with an RDMA communication service is specific to that service or the server.

intel0024 RDMA channel establishment requests that are acceptable to the RDMA communication service shall be
IBM032 passed to the server agent. The server agent is provided the login data from the client consumer's request in
IBM011 addition to RDMA communication service specific data. With SRP the login data includes an SRP_LOGIN_REQ
IBM031 request (see 6.2).

The server agent determines whether the RDMA channel establishment request may be accepted and determines the server consumer to be associated with the RDMA channel. If the request is not accepted the server agent or server consumer instructs the RDMA communication service to reject the RDMA channel. The

server agent or server consumer provides reject data and any RDMA communication service or server specific data that is required. With SRP the reject data shall contain an SRP_LOGIN_REJ response (see 6.4). IBM011 intel0022

If the RDMA channel establishment request is accepted, the server agent or server consumer instructs the RDMA communication service to accept the RDMA channel. The server agent or server consumer provides accept data and any RDMA communication service or server specific data that is required. With SRP the accept data shall contain an SRP_LOGIN_RSP response (see 6.3). IBM033 IBM011 IBM034 intel0025 intel0026

4.2.3 Disestablishment

An RDMA channel may be disconnected by a request from either of the RDMA channel's consumers or from the RDMA communication service. The consumers may each be notified that the RDMA channel has been disconnected, allowing the consumers to recover any resources associated with the RDMA channel. The time to deliver ~~such~~ a notification may vary depending upon the RDMA communication service, the consumer being notifed, and the specific circumstances of the disconnection request. intel0027 IBM035

A disconnect request causes an RDMA channel to become non-operational. Operations in progress on an RDMA channel at the time of a disconnect request and operations requested subsequent to a disconnect request may or may not complete. intel0028 intel0029 intel0030 intel0031

4.3 Messages intel0032

An RDMA channel may allow its consumers to exchange messages. A message is sent by one consumer associated with an RDMA channel (the sending consumer) to the other consumer associated with the RDMA channel (the receiving consumer). A message contains a payload of some number of data bytes. An RDMA communication service may provide normal and solicited message reception notification, which may be used to distinguish between more urgent and less urgent messages. intel0033 intel0034

A sending consumer requests that a message be sent by providing the following to an RDMA communication service: intel0035

a) the message's payload length;
b) the message's payload data;
c) the RDMA channel to use; and
d) whether to use normal or solicited message reception notification.

The RDMA communication service attempts to deliver the message to the other consumer associated with the specified RDMA channel (the receiving consumer). If delivery succeeds, the RDMA communication service notifies the receiving consumer that a message has been received, providing the message's length, payload, and the RDMA channel on which the message was received. The RDMA communication service may also provide an indication of whether the sending consumer specified normal or solicited message reception notification. IBM036

An RDMA communication service may require that receiving consumers provide message receive buffers to RDMA channels before messages are sent to them, and that the provided message receive buffers be large enough to hold any messages that arrive. Sending a message on an RDMA channel when no receive buffer has been provided, or when the provided receive buffer is too small for the message, may result in behavior that is not specified by this standard.

NOTE 2 - Such behavior may include (but is not limited to) disconnecting the RDMA channel, discarding or truncating the message, or delaying delivery of the message until a suitable message receive buffer becomes available. The RDMA communication service may or may not provide an error indication.

An RDMA communication service may or may not provide a way for a sending consumer to determine whether a message has been delivered to the receiving consumer.

4.4 RDMA operations

IBM037 IBM043 An RDMA channel may provide RDMA Write operations, RDMA Read operations, or both between its consumers.

A consumer may allow RDMA access by registering some or all of its memory with an RDMA communication service. The RDMA communication service returns a memory handle to identify the registered memory. The consumer may specify that the memory handle is usable for memory access on only a specified RDMA channel or on a group of RDMA channels. The consumer may impose other access restrictions allowed by the RDMA communication service ~~as well~~ (e.g. read-only acess). IBM038

A consumer that has registered memory and obtained a memory handle may communicate the memory handle to another consumer. This may be done using an application protocol contained in message payloads. The other consumer may then use the memory handle to request RDMA operations that access the memory registered by the first consumer.

The registered memory identified by a memory handle is represented as a memory address space. Accessible locations are identified by addresses. An RDMA communication service is not required to provide a way to determine, from a message handle, which memory locations are accessible, the number of locations that are accessible, or the type of access allowed. ~~Such information may be communicated by an application protocol.~~ IBM039

An RDMA Write operation allows a requesting consumer to store data into memory registered by another consumer. A requesting consumer provides the following to an RDMA communication service when it requests an RDMA Write operation:

a) An RDMA channel to use for the operation;
b) A memory handle that is usable for access on that RDMA channel;
c) A range of addresses within the memory address space identified by the memory handle; and
d) Data to be written into the specified range of addresses.

An RDMA communication service is not required to provide a way for a requesting consumer to determine whether the data has been written into the specified range of addresses in registered memory. An RDMA communication service is not required to provide a way for the consumer that registered the memory to determine whether an RDMA Write operation is in progress or has completed. intel0036

An RDMA Read operation allows a requesting consumer to fetch data from memory registered by another consumer. A requesting consumer provides the following to an RDMA communication service when it requests an RDMA Read operation:

a) An RDMA channel to use for the operation;
b) A memory handle that is usable for access on that RDMA channel;
c) A range of addresses within the memory address space identified by the memory handle; and
d) A buffer into which to place the data read from the specified range of addresses.

The RDMA communication service notifies the requesting consumer after data has been successfully obtained from the specified range of addresses and placed in the requestor's buffer. An RDMA communication service is not required to provide a way for the consumer that registered the memory to determine whether an RDMA Read operation is in progress or has completed.

4.5 Ordering and Reliability

4.5.1 Overview

IBM011 IBM011 SRP operates using an RDMA communication service having the characteristics described in this subclause. Use of SRP with an RDMA communication service having different characteristics is outside the scope of this standard.

4.5.2 Reliability

An RDMA communication service shall deliver each message sent on an RDMA channel to the receiving consumer or ~~else~~ disconnect the RDMA channel. Each delivered message shall be delivered to the receiving consumer ~~exactly~~ once, without duplication; the RDMA communication service shall discard any duplicates that may result from retransmission or other mechanisms. Each delivered message shall be delivered to the receiving consumer complete and error-free. intel0037 IBM041 IBM042

The RDMA communication service shall provide to the sending consumer an indication of the completion status of each RDMA communication service request. This status shall be one of: IBM049

a) successful - The request completed without error.
b) error - The request was not completed due to an error. The RDMA communication service may provide additional information about the error. This status should be returned immediately when the RDMA channel does not exist or has experienced an error.
c) timeout - No indication was received, completion status of request is unknown, RDMA communication service has experienced an error. The length of time after which a timeout indication is returned is specific to the RDMA communication service.

4.5.3 Ordering edit013

Messages sent by the same consumer on the same RDMA channel shall be delivered to the receiving consumer in the order they were sent. The data for all RDMA Write operations requested on an RDMA channel by a consumer prior to that same consumer sending a message on the same RDMA channel shall be available to the receiving consumer (e.g. stored into registered memory) before the message is delivered to the receiving consumer. If multiple RDMA Write operations requested on an RDMA channel by a consumer store data into the same registered memory location, the location's resulting contents shall be the data stored by the last RDMA Write operation.

Messages sent on different RDMA channels or by different consumers may be delivered in any order. The data for RDMA Write operations may be stored into registered memory in any order relative to the delivery of messages sent on other RDMA channels or by other consumers. RDMA Write operations requested on different RDMA channels may store data into the same registered memory location in any order.

RDMA Read operations may complete in any order. edit012

If an RDMA communication service is unable to meet ~~satisfy~~ these requirements on an RDMA channel, it shall disconnect the RDMA channel. IBM044 intel0038

5 Structure and concepts

IBM011 ## 5.1 Overview of SRP operation

5.1.1 RDMA channel establishment and login

IBM011 SRP initiator ports login with SRP target ports when a new RDMA channel is established for use with SRP. The
IBM045 login process associates an RDMA channel with a specific SRP initiator port and SRP target port (i.e., an
IBM046 I_T nexus (see SAM-2)) and negotiates parameters that govern the use of that RDMA channel for its lifetime.

SRP initiator ports and SRP target ports shall be determined by their role during RDMA channel establishment. An object that requests RDMA channel establishment as a client consumer (see 4.2) shall be an SRP initiator port. An object that accepts RDMA channel establishment as a server consumer (see 4.2) shall be an SRP target port.

Login occurs during RDMA channel establishment. An SRP initiator port shall provide an SRP_LOGIN_REQ request (see 6.2) as the login data when establishing a new RDMA channel. If an SRP target port accepts a new RDMA channel it shall provide an SRP_LOGIN_RSP response (see 6.3) as the accept data. If an SRP target port does not accept a new RDMA channel it shall provide an SRP_LOGIN_REJ response (see 6.4) as the reject data parameter when rejecting the new RDMA channel.

The SRP_LOGIN_REQ request (see 6.2) contains an SRP initiator port identifier and an SRP target port
intel0039 identifier. An SRP target port shall not accept a new RDMA channel unless its SRP target port identifier matches the value in the SRP_LOGIN_REQ request. If an SRP target port accepts a new RDMA channel, it shall treat all communication on that RDMA channel as being with the SRP initiator port identified by the SRP initiator port identifier specified in the SRP_LOGIN_REQ request.

IBM047 It is the SRP initiator port's responsibility to specify a server address that directs RDMA channel establishment to a server where the specified SRP target port may be accessed (see 4.2). It is the server agent's responsibiliy
intel0040 to direct RDMA channel establishment to a server consumer that is the specified SRP target port. Additional information on locating a server consumer matching a specified SRP target port identifier may be found in Annex B for SRP initiator ports and SRP target ports that use the InfiniBand™ Architecture.

5.1.2 RDMA channel disconnection

~~Prior to requesting that an RDMA channel be disconnected, an SRP initiator port may should send an SRP_I_LOGOUT request (see 6.5) to notify the SRP target port of the disconnection.~~

intel0041 An SRP initiator port should send an SRP_I_LOGOUT request (see 6.5) and wait for the RDMA communication service status indication (see 4.5.2) before requesting that an RDMA channel be disconnected.

After requesting that an RDMA channel be disconnected, after being notified that an RDMA channel has been disconnected, or upon receiving an SRP_T_LOGOUT request (see 6.6), an SRP initiator port shall:

a) Discard any outstanding request received from an SRP target port on that RDMA channel, without returning a response;
b) Not send any further messages on that RDMA channel;
c) Discard any subsequent messages received on that RDMA channel; and
IBM048 d) For any outstanding SCSI tasks ~~that were contained in SRP_CMD requests (see 6.8)~~ sent on that RDMA channel, indicate to the application client that the task has terminated with a service delivery system failure.

~~Prior to requesting that an RDMA channel be disconnected, an SRP target port should send an SRP_T_LOGOUT request (see 6.6) to notify the SRP initiator port of the disconnection.~~

intel0042 An SRP target port should send an SRP_T_LOGOUT request (see 6.6) and wait for the RDMA communication
IBM049 service status indication (see 4.5.2) before requesting that an RDMA channel be disconnected.

After requesting that an RDMA channel be disconnected, after being notified that an RDMA channel has been disconnected, or upon receiving an SRP_I_LOGOUT request (see 6.5), an SRP target port shall:

a) Abort all outstanding SCSI tasks that were contained in SRP_CMD requests (see 6.8) received on that RDMA channel, without returning a response; edit008
b) Discard any other outstanding requests received from an SRP initiator port on that RDMA channel, without returning a response;
c) Not send any further messages on that RDMA channe;
d) Discard any subsequent messages received on that RDMA channell; and
e) Not alter previously established conditions, including MODE SELECT parameters, reservations, ACA, and CA as a result of the disconnection. IBM050

5.1.3 Single RDMA channel operation

An SRP initiator port may specify single RDMA channel operation during login. If an SRP target port accepts such a login, it shall:

a) Attempt to send an SRP_T_LOGOUT request (see 6.6) on any established RDMA channel that specified the same SRP initiator port identifier. The reason code shall indicate that the RDMA channel was disconnected due to a MULTI-CHANNEL ACTION code in a new SRP_LOGIN_REQ request (see 6.2);
b) Request disconnection of any established RDMA channel (see 5.1.2) that specified the same SRP initiator port identifier; and
c) Reject any other RDMA channel establishment requests it has received that specified the same SRP initiator port identifier and that the SRP target port has not yet accepted.

Following acceptance of a login specifying single RDMA channel operation that single RDMA channel shall be used for all communication between the specified SRP initiator port and SRP target port. Subsequent logins specifying other modes of operation ~~, if accepted,~~ may allow communication using multiple RDMA channels. intel0043 IBM051

5.1.4 Multiple independent RDMA channel operation

An SRP initiator port may specify multiple independent RDMA channel operation during login. An SRP target port shall not accept such a login if doing so would require disconnecting an established RDMA channel with the same SRP initiator port, and shall return the SRP_T_LOGOUT request reason code RDMA CHANNEL LIMIT REACHED FOR THIS INITIATOR. intel0044

Following acceptance of a login specifying multiple independent RDMA channel operation one or more RDMA channels may be used for communication between the same SRP initiator port and the same SRP target port. All such RDMA channels are associated with the single I_T nexus defined by the SRP initiator port identifier and the SRP target port identifier. intel0045

When multiple independent RDMA channels are used operation of each SRP request is confined to a single RDMA channel. The sender of an SRP request chooses an RDMA channel to use for sending the SRP request. The sender of an SRP response shall use the same RDMA channel as the SRP request for sending the SRP response. All RDMA operations associated with the SRP request shall also use the same RDMA channel as the SRP request.

While each SRP request is confined to a single RDMA channel, SCSI tasks and task management functions may be conveyed on independent RDMA channels associated with the same I_T nexus. SCSI tasks and task management functions ~~may~~ interact as specified by SAM-2, SPC-2 and other SCSI command standards (e.g., a SCSI task sent on one RDMA channel may be aborted by an ABORT TASK sent on a different RDMA channel associated with the same I_T nexus, and reservations obtained or released on any RDMA channel apply to SCSI tasks sent on all RDMA channels that are associated with the same I_T nexus.) IBM052 IBM053 intel0046

An RDMA communication service may or may not provide any ordering relationship between SRP requests, SRP responses and RDMA operations that use different RDMA channels. If ordering is important for a sequence of SRP requests, they should be sent using the same RDMA channel.

5.2 Information unit classes

Each SRP information unit is classified as a SRP request or a SRP response (see 6.1). SRP requests convey SCSI commands, task management requests and RDMA channel management requests. SRP responses convey SCSI command and task management service responses and RDMA channel management responses.

IBM054 RDMA channel management requests may be issued by SRP target ports or ~~as well as~~ SRP initiator ports.

In normal operation, SRP requests and SRP responses occur in pairs. Each SRP request elicits a single corresponding SRP response from the SRP device receiving the SRP request. An SRP request communicates the initiation of a remote procedure call; the corresponding SRP response communicates the remote procedure call's completion.

IBM057 An SRP response shall not be returned:

a) for an SRP_CMD request if the associated task is aborted;
b) for an SRP_T_LOGOUT request (see 6.6);
c) for an SRP_I_LOGOUT request (see 6.5); and
d) for outstanding SRP requests received on an RDMA channel when an SRP device becomes aware of a failure preventing further communication on that RDMA channel. In this case, the device shall abort all outstanding SRP requests received on that RDMA channel.

In all other cases an SRP device shall return a single SRP response for each SRP request it receives.

SRP responses shall be sent on the RDMA channel on which the corresponding SRP request was received.

IBM065 ## 5.3 SRP target port buffer management

SRP target port buffer management allows an SRP target device to limit the number of SRP requests that may

IBM058 be sent ~~to it~~ on an RDMA channel. SRP devices may use SRP target port buffer management to manage internal and RDMA channel-related resources.

IBM059 SRP responses are not subject to buffer management; they may be sent at any time. An SRP device may limit

IBM060 the number of SRP responses it ~~might~~may receive by limiting the number of SRP requests it has outstanding.

intel0047 SRP target ports shall limit themselves to at most one outstanding SRP request per RDMA channel. Upon

CPQ007 sending an SRP request, an SRP target port shall not send another SRP request on the same RDMA channel until after it receives the SRP response for the previous SRP request.

intel0048 SRP uses a credit-based buffer management algorithm to limit the number of SRP requests that an SRP initiator port may send to an SRP target port. The algorithm uses a field, REQUEST LIMIT DELTA, that is present in most

IBM011 information units sent by an SRP target port to an SRP initiator port, and a state variable, REQUEST LIMIT. The following rules specify the algorithm:

a) REQUEST LIMIT and REQUEST LIMIT DELTA are both signed two's complement 32-bit integers. SRP initiator ports shall implement a separate copy of REQUEST LIMIT for each RDMA channel;
b) Upon successful completion of RDMA channel establishment an SRP initiator port shall initialize the RDMA channel's REQUEST LIMIT to the value of REQUEST LIMIT DELTA received in the SRP_LOGIN_RSP response (see 6.3). Except for providing an SRP_LOGIN_REQ request (see 6.2) when requesting RDMA channel establishment, the SRP initiator port shall not send any SRP information units on the RDMA channel prior to initializing REQUEST LIMIT;
c) An SRP initiator port may send an SRP request on an RDMA channel when the value of the RDMA channel's REQUEST LIMIT is greater than zero. An SRP initiator port shall not send an SRP request on

intel0049 any RDMA channel whose REQUEST LIMIT has a value less than or equal to zero. To ensure that task

management requests may be sent, an SRP initiator port may choose to send commands only when REQUEST LIMIT is greater than one; intel0050

d) An SRP initiator port shall decrement an RDMA channel's REQUEST LIMIT by one whenever it sends an SRP request on that RDMA channel;

e) An SRP initiator port shall add (two's complement addition) the value of REQUEST LIMIT DELTA to an RDMA channel's REQUEST LIMIT whenever it receives an information unit on that RDMA channel; and intel0051 intel0052

f) An SRP target port shall not specify a positive value of REQUEST LIMIT DELTA that might cause REQUEST LIMIT to exceed 2^{30}. An SRP target port shall not specify a negative value of REQUEST LIMIT DELTA that might cause REQUEST LIMIT to drop below -2^{31}. intel0053 intel0054 intel0055

5.4 Data buffers

5.4.1 Memory descriptors

A memory descriptor is a 16-byte structure that identifies a memory segment (see table 1). Figure 4 illustrates the mapping of a memory descriptor to a memory segment. intel0056

Table 1 - Memory descriptor

Bit / Byte	7	6	5	4	3	2	1	0
0	(MSB)							
• • •				VIRTUAL ADDRESS				
7								(LSB)
8	(MSB)							
• • •				MEMORY HANDLE				
11								(LSB)
12	(MSB)							
• • •				DATA LENGTH				
15								(LSB)

The VIRTUAL ADDRESS field contains an unsigned integer value that identifies the byte address within the memory region of the first byte of the memory segment. intel0057

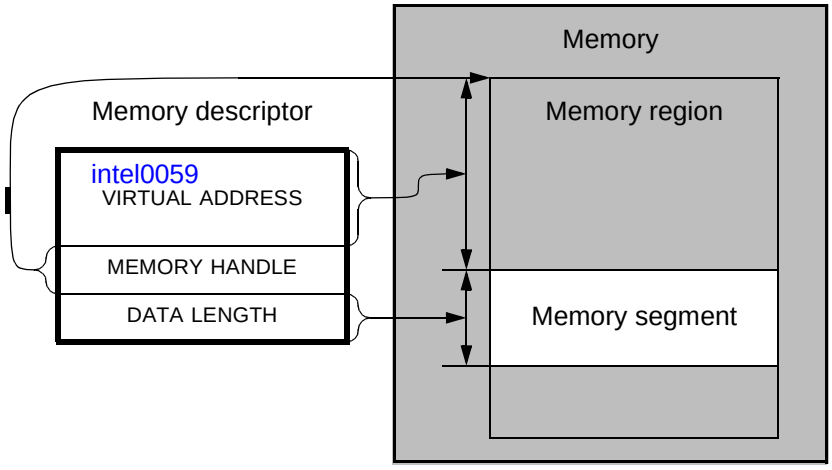


Figure 4 - Memory descriptor mapping

The MEMORY HANDLE field identifies the region that contains the memory segment. The SRP target port shall intel0058 supply this value with any RDMA operation that accesses the memory segment. The SRP initiator port shall use this value to locate the region.

The DATA LENGTH field contains an unsigned integer value that identifies the length of the memory segment in bytes. The interpretation of a memory descriptor where the sum of the VIRTUAL ADDRESS and DATA LENGTH fields exceeds 2^{64} is vendor specific.

A SRP target port may use a memory descriptor for either RDMA Read operations or RDMA Write operations intel0060 but not both. SRP target ports shall only issue the appropriate type of RDMA operation for a memory descriptor, and shall ensure that each RDMA operation is wholly contained within its memory segment:

intel0061

a) The RDMA operation's VIRTUAL ADDRESS shall be greater than or equal to the memory descriptor's VIRTUAL ADDRESS and less than the sum of the memory descriptor's VIRTUAL ADDRESS and DATA LENGTH; and
b) The sum of the RDMA operation's VIRTUAL ADDRESS and DATA LENGTH shall be greater than the memory descriptor's VIRTUAL ADDRESS and less than or equal to the sum of the memory descriptor's VIRTUAL ADDRESS and DATA LENGTH.

5.4.2 Data buffer descriptors

5.4.2.1 Overview

IBM097 An SRP_CMD request (see 6.8) may contain a data-out buffer descriptor, a data-in buffer descriptor, both or neither, depending upon the data transfer(s) requested by the SCSI command. The format of each data buffer intel0062 descriptor is specified by a format code value. In an SRP_CMD request with both data-in and data-out buffer intel0064 descriptors, there is no requirement that both buffer descriptors be of the same format. Some data buffer descriptor format code values use the contents of a count field to further specify the data buffer descriptor intel0063 format. Table 2 defines data buffer descriptor format code values.

Table 2 - Data buffer descriptor formats

OD 8

CPQ008

Data buffer descriptor format code	Reference	format code value[a]	buffer descriptor length (bytes)[c]
NO DATA BUFFER DESCRIPTOR PRESENT	5.4.2.3	0h	0
DIRECT DATA BUFFER DESCRIPTOR	5.4.2.4	1h	16
INDIRECT DATA BUFFER DESCRIPTOR	5.4.2.5	2h	20+16*count[b]

[a] The format code value for a data-out buffer descriptor is specified by the DATA-OUT BUFFER DESCRIPTOR FORMAT field of an SRP_CMD request (see 6.8). The format code value for a data-in buffer descriptor is specified by the DATA-IN BUFFER DESCRIPTOR FORMAT field of an SRP_CMD request (see 6.8).

CPQ009 [b] The count field for a data-out buffer descriptor is the DATA-OUT BUFFER DESCRIPTOR COUNT field of an SRP_CMD request (see 6.8). The count field for a data-in buffer descriptor is the DATA-IN

CPQ010 BUFFER DESCRIPTOR COUNT field of an SRP_CMD request (see 6.8).

IBM073 [c] The length of a data buffer descriptor is determined from its format code value and ~~and~~ the contents of its count field.

5.4.2.2 Supported data buffer descriptor formats

The REQUIRED BUFFER FORMATS field of the SRP_LOGIN_REQ request (see 6.2) indicates the data buffer descriptor formats that an SRP initiator port may specify in requests sent on an RDMA channel. An SRP initiator port shall set the REQUIRED BUFFER FORMATS field to indicate all data buffer descriptor formats that the SRP

initiator port may specify in SRP_CMD requests (see 6.8) sent on that RDMA channel. An SRP initiator port shall not ~~specify~~issue an SRP_CMD request (see 6.8) indicating a data buffer descriptor format that was not indicated in the REQUIRED BUFFER FORMATS field value for that RDMA channel. SRP target ports are not required to check SRP_CMD requests for data buffer descriptor formats that were not indicated in the REQUIRED BUFFER FORMATS field value. If a target port does detect that an initiator has specified a descriptor format not indicated in the REQUIRED BUFFER FORMATS field, the target port shall send an SRP_T_LOGOUT request (see 6.6) with the reason code 0000 0006h (Unsupported format code value specified in DATA-OUT BUFFER DESCRIPTOR FORMAT field) or the reason code 0000 0007h, (Unsupported format code value specified in DATA-IN BUFFER DESCRIPTOR FORMAT field), as appropriate. edit010

An SRP target port may accept an RDMA channel and return an SRP_LOGIN_RSP response (see 6.3) if the SRP target port is able to support all of the data buffer descriptor formats indicated in the REQUIRED BUFFER FORMATS field on that RDMA channel. An SRP target port shall reject the RDMA channel and return an SRP_LOGIN_REJ response (see 6.4) with reason code 0001 0004h (One or more requested data buffer descriptor formats are not supported) if the SRP target port is unable to support one or more of the data buffer descriptor formats indicated in the REQUIRED BUFFER FORMATS field on that RDMA channel. edit011

An SRP target port shall indicate the data buffer descriptor formats that it supports in the SUPPORTED BUFFER FORMATS field of the SRP_LOGIN_RSP response (see 6.3) and the SRP_LOGIN_REJ response (see 6.4). All SRP target ports shall support the DIRECT DATA BUFFER DESCRIPTOR format, ~~SRP target ports may or may not~~ and may support other data buffer descriptor formats.

Table 3 defines the contents of the REQUIRED BUFFER FORMATS field and the SUPPORTED BUFFER FORMATS field.

Table 3 - Supported data buffer descriptor formats

Byte \ Bit	7	6	5	4	3	2	1	0
0	Reserved							
1	Reserved					IDBD	DDBD	Reserved

An ~~indirect data buffer descriptor (IDBD) bit shall be set to one in an SRP_LOGIN_REQ request (see 6.2) if the~~ SRP initiator port ~~may specify~~ sets the indirect data buffer descriptor (IDBD) bit to one in a SRP_LOGIN_REQ request (see 6.2) if it requires that the target port support the INDIRECT DATA BUFFER DESCRIPTOR format. ~~The indirect data buffer descriptor (IDBD) bit should be set to zero in an SRP_LOGIN_REQ request if the SRP initiator port does not use the INDIRECT DATA BUFFER DESCRIPTOR format.~~ IBM084

The target port shall set the indirect data buffer descriptor (IDBD) bit ~~shall be set~~ to one in an SRP_LOGIN_RSP response (see 6.3) or in an SRP_LOGIN_REJ response (see 6.4) if the SRP target port supports the INDIRECT DATA BUFFER DESCRIPTOR format. The indirect data buffer descriptor (IDBD) bit shall be set to zero in an SRP_LOGIN_RSP response or in an SRP_LOGIN_REJ response if the SRP target port does not support the INDIRECT DATA BUFFER DESCRIPTOR format.

~~The direct data buffer descriptor (DDBD) bit shall be set to one in an SRP_LOGIN_REQ request (see 6.2) if the SRP initiator port may specify the DIRECT DATA BUFFER DESCRIPTOR format. The direct data buffer descriptor (DDBD) bit should be set to zero in an SRP_LOGIN_REQ request if the SRP initiator port does not use the DIRECT DATA BUFFER DESCRIPTOR format.~~ IBM084

An SRP initiator port sets the direct data buffer descriptor (DDBD) bit to one in a SRP_LOGIN_REQ request (see 6.2) if it requires that the target port support the INDIRECT DATA BUFFER DESCRIPTOR format.

The target port shall set the direct data buffer descriptor (DDBD) bit ~~shall be set~~ to one in an SRP_LOGIN_RSP response (see 6.3) or in an SRP_LOGIN_REJ response (see 6.4).

IBM085 The length of requests sent by an SRP initiator port, as determined by the data buffer descriptor formats, shall be limited to the MAXIMUM INITIATOR TO TARGET IU LENGTH field returned in the SRP_LOGIN_RSP response (see 6.3).

~~NOTE 3 - The MAXIMUM INITIATOR TO TARGET IU LENGTH field value returned in the SRP_LOGIN_RSP response (see 6.3) limits the length of requests that may be sent by an SRP initiator port. This limit may restrict the data buffer descriptor formats that the SRP initiator port may specify independent of the REQUIRED BUFFER FORMATS field value.~~

5.4.2.3 No data buffer descriptor present

IBM086 The NO DATA BUFFER DESCRIPTOR PRESENT format code value specifies that the corresponding data buffer descriptor field is not present. The contents of the count field are reserved. SRP target ports shall ignore ~~are not required to check~~ the contents of the count field.

5.4.2.4 Direct data buffer descriptor format

CPQ011 The DIRECT DATA BUFFER DESCRIPTOR format code value specifies that the corresponding data buffer descriptor field is sixteen bytes in length and contains a direct data buffer descriptor. The contents of the count

IBM087 field are reserved. SRP target ports shall ignore ~~are not required to check~~ the contents of the count field.

A direct data buffer descriptor contains a single memory descriptor (see table 1). The memory descriptor identifies the data buffer, which is a single memory segment within a memory region's virtual address space. If a direct data buffer descriptor defines a data-out buffer, the SRP target port shall only issue RDMA Read operations using the memory descriptor contained in the direct data buffer descriptor. If a direct data buffer descriptor defines a data-in buffer, the SRP target port shall only issue RDMA Write operations using the memory descriptor contained in the direct data buffer descriptor. The SRP target port shall use the contents of

CPQ011 the DATA LENGTH field of the memory descriptor as the length of the data-out buffer or data-in buffer.

5.4.2.5 Indirect data buffer descriptor format

The INDIRECT DATA BUFFER DESCRIPTOR format code value specifies that the corresponding data buffer descriptor field contains an indirect data buffer descriptor. The length of the data buffer descriptor field is twenty bytes plus the contents of the count field multiplied by sixteen bytes.

An indirect data buffer is comprised of one or more memory segments. The memory segments may or may not be contiguous. The memory segments may be in a single memory region or spread among several memory regions. The indirect data buffer is the concatenation of the memory segments. Each memory segment may have any length, including a length of zero bytes.

Table 4 shows the format of an indirect data buffer descriptor.

Table 4 - Indirect data buffer descriptor

Byte \ Bit	7	6	5	4	3	2	1	0
0								
• • •				INDIRECT TABLE MEMORY DESCRIPTOR				
15								
16	(MSB)							
• • •				TOTAL LENGTH				
19								(LSB)
20								
• • •				PARTIAL MEMORY DESCRIPTOR LIST				
16*n+19								

[a] The value n is the value contained in the data buffer descriptor's count field.

CPQ013

CPQ012

The INDIRECT TABLE MEMORY DESCRIPTOR field value is a memory descriptor (see table 1) that specifies a memory segment containing an indirect table. An indirect table is a list of one or more memory descriptors. The memory segments specified by the memory descriptors in the indirect table, concatenated together, comprise the indirect data buffer. The DATA LENGTH field of the INDIRECT TABLE MEMORY DESCRIPTOR field value contains the number of memory descriptors in the indirect table times sixteen. SRP target port behavior when the DATA LENGTH field of the INDIRECT TABLE MEMORY DESCRIPTOR field value contains any other value is vendor specific.

The TOTAL LENGTH field value is the sum of the DATA LENGTH field values of the memory descriptors in the indirect table. SRP target port behavior when the TOTAL LENGTH field contains any other value is vendor specific. The SRP target port shall use either the TOTAL LENGTH field value or the sum of the DATA LENGTH field values as the length of the data-out buffer or data-in buffer.

The PARTIAL MEMORY DESCRIPTOR LIST field is only present when the data buffer descriptor's count field contains a non-zero value. The PARTIAL MEMORY DESCRIPTOR LIST field contains a list of n memory descriptors that are copies of the first n memory descriptors in the indirect table. The value n is the value contained in the associated count field. SRP target port behavior when the PARTIAL MEMORY DESCRIPTOR LIST field contains any other value is vendor specific. intel0084

An SRP target port shall only issue RDMA Read operations to the indirect table.

If an indirect data buffer descriptor specifies a data-out buffer, the SRP target port shall only issue RDMA Read operations using the memory descriptors contained in the indirect table or the PARTIAL MEMORY DESCRIPTOR LIST field value.

If an indirect data buffer descriptor specifies a data-in buffer, the SRP target port shall only issue RDMA Write operations using the memory descriptors contained in the indirect table or the PARTIAL MEMORY DESCRIPTOR LIST field value.

Figure 5 illustrates an indirect data buffer descriptor that does not contain a PARTIAL MEMORY DESCRIPTOR LIST field. Memory is shown containing four memory segments: the indirect table, memory segment 1, memory segment 2 and memory segment 3. The mapping of each memory descriptor to its memory segment has been shown as a single arrow. For details of this mapping see 5.4.1 and figure 4. Figure 5 does not show the memory regions in which the memory segments reside. All four segments might be in a single memory region, each

might be in a separate memory region, or several might be in one memory region and the remainder in one or more other memory regions.

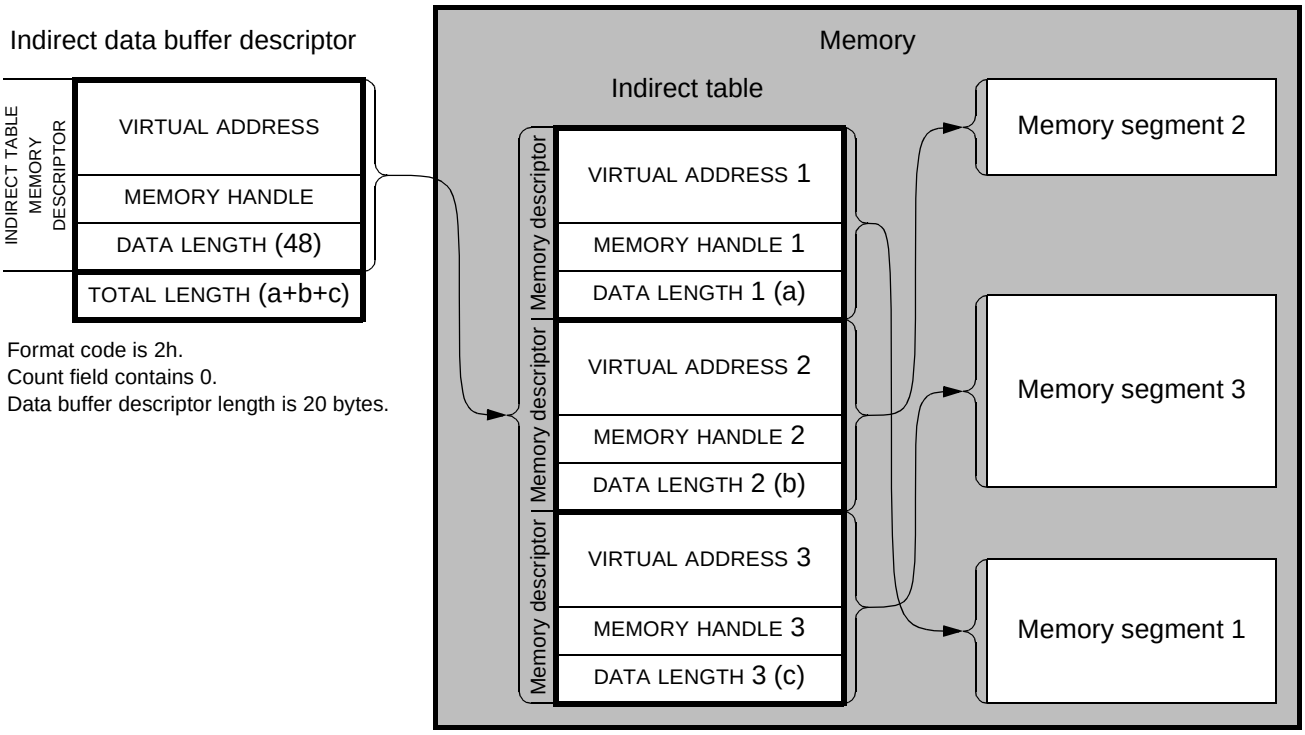
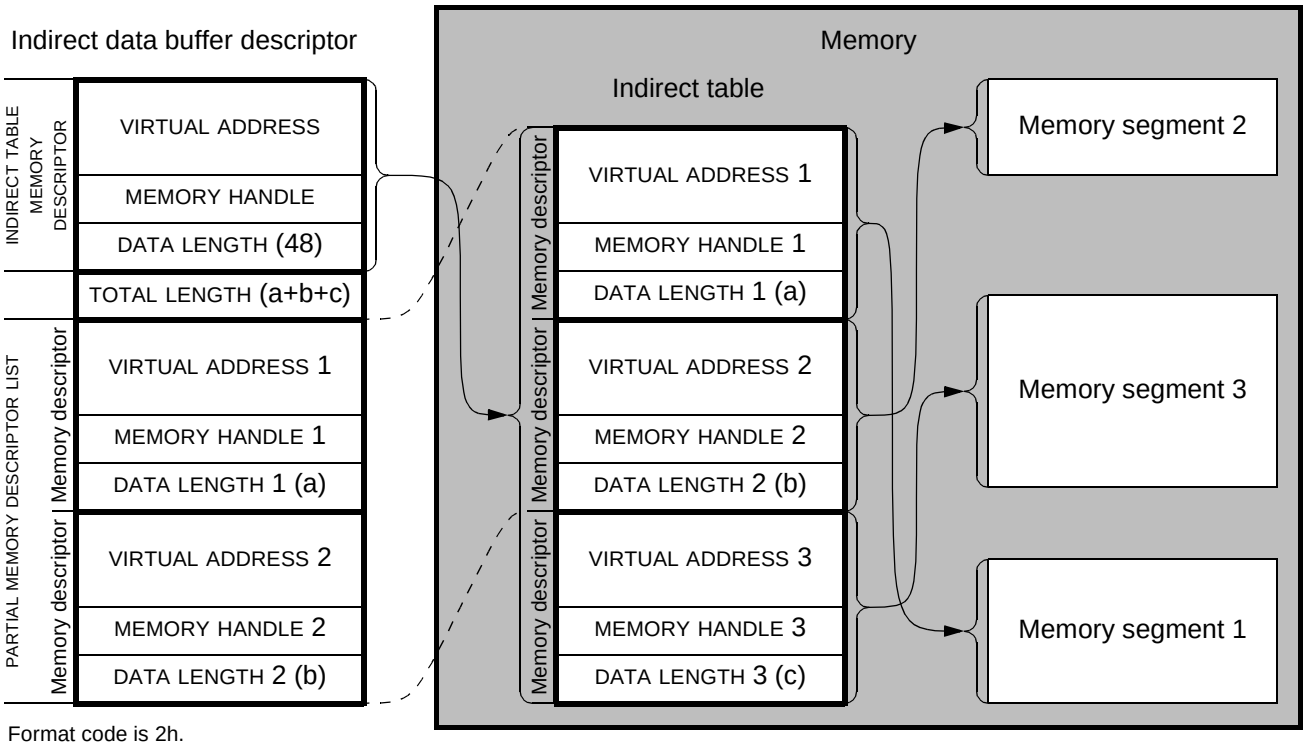


Figure 5 - Example indirect data buffer descriptor with no PARTIAL MEMORY DESCRIPTOR LIST field

In the example shown in figure 5 the data buffer descriptor format code value is 2h and the count field contains zero. The indirect data buffer descriptor is 20 bytes long. The data buffer is comprised of three memory segments: memory segment 1, memory segment 2 and memory segment 3. A separate memory segment contains the indirect table, a list of three memory descriptors specifying memory segments 1 through 3. The INDIRECT TABLE MEMORY DESCRIPTOR field value of the indirect data buffer descriptor specifies the memory segment containing the indirect table. The DATA LENGTH field of the INDIRECT TABLE MEMORY DESCRIPTOR field value contains 48 (i.e. the length of the indirect table). The TOTAL LENGTH field of the data buffer descriptor contains the sum of the DATA LENGTH field values of the memory descriptors in the indirect table (i.e. the sum of DATA LENGTH 1, DATA LENGTH 2 and DATA LENGTH 3). This sum is the total length of the data buffer.

Figure 6 illustrates the same example as in figure 5 except with a PARTIAL MEMORY DESCRIPTOR LIST field. The data buffer, indirect table, INDIRECT TABLE MEMORY DESCRIPTOR field value and TOTAL LENGTH field value are all identical to the example in figure 5. The data buffer descriptor format code is 2h, the same as in figure 5. However the count field contains the value 2, indicating that the PARTIAL MEMORY DESCRIPTOR LIST field is present and contains two memory descriptors. Those two memory descriptors are copies of the first two memory

descriptors in the indirect table. The third memory descriptor is only present in the indirect table. The indirect data buffer descriptor is 52 bytes long.



Format code is 2h.
Count field contains 2.
Data buffer descriptor length is 52 bytes.

Figure 6 - Example indirect data buffer descriptor with a PARTIAL MEMORY DESCRIPTOR LIST field

IBM011

6 SRP Information Units

6.1 Summary

IBM011 The information units used by SRP and their characteristics are shown in table 5, table 6, table 7 and table 8.

intel0094 All SRP initiator ports shall support sending the information units listed in table 5 and table 8, and shall support receiving the information units listed in table 6 and table 7.

All SRP target ports shall support sending the information units listed in table 6 and table 7, and shall support receiving the information units listed in table 5 and table 8.

Table 5 - SRP requests sent from SRP initiator ports to SRP target ports

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_LOGIN_REQ	6.2	00h	64	Login request
SRP_TSK_MGMT	6.7	01h	64	SCSI task management function
SRP_CMD	6.8	02h	48 minimum	SCSI command
SRP_I_LOGOUT	6.5	03h	16	SRP initiator port logout notification

Table 6 - SRP responses sent from SRP target ports to SRP initiator ports

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_LOGIN_RSP	6.3	C0h	52	Login successful response
SRP_RSP	6.9	C1h	36 minimum	SCSI status or service response
SRP_LOGIN_REJ	6.4	C2h	32	Login failure response

Table 7 - SRP requests sent from SRP target ports to SRP initiator ports

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_T_LOGOUT	6.6	80h	16	SRP target port logout or RDMA channel failure notification
SRP_CRED_REQ	6.10	81h	52	SRP target port credit adjustment request
SRP_AER_REQ	6.12	82h	56 minimum	Asynchronous event report request

Table 8 - SRP responses sent from SRP initiator ports to SRP target ports

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_CRED_RSP	6.11	41h	64	Response to SRP target port credit adjustment request
SRP_AER_RSP	6.13	42h	16	Asynchronous event report response

Byte 0 of each SRP information unit contains a TYPE code. The TYPE code value uniquely identifies the information unit and its format. The length of an information unit is indicated by its TYPE code and selected fields within the information unit. If an SRP target port receives an SRP information unit with an invalid TYPE code, or

whose length is incorrect for the information unit's type code, the SRP target port shall send an SRP_T_LOGOUT request (see 6.6) and disconnect the RDMA channel. intel0096

Bytes 8 through 15 of each information unit contain a TAG value, which provides a mechanism for matching SRP requests with their corresponding SRP responses. A requestor shall provide a TAG value in each SRP request that is unique among all of the requestor's outstanding SRP requests with a particular responder. A responder shall copy the TAG value from each SRP request to the SRP request's SRP response. Responders are not required to check whether the TAG values of outstanding SRP requests are unique.

6.2 SRP_LOGIN_REQ request

OD6a An SRP_LOGIN_REQ request (see table 9) conveys SRP ~~protocol~~ login parameters from an SRP initiator port to an SRP target port. The SRP_LOGIN_REQ request shall ~~only~~ be sent as login data during RDMA channel establishment .

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	Reserved	AESOLNT	CRSOLNT	LOSOLNT	Reserved		MULTI-CHANNEL ACTION	
27	Reserved							
28	Reserved							
• • •								
31								
32	INITIATOR PORT IDENTIFIER							
• • •								
47								
48	TARGET PORT IDENTIFIER							
• • •								
63								

CPQ015 CPQ016 OD6a

The TAG field is defined in 6.1.

CPQ014 The REQUESTED MAXIMUM INITIATOR TO TARGET IU LENGTH field specifies the maximum length in bytes of any information unit that the SRP initiator port wishes to send on this RDMA channel. This value shall be 64 or larger.

The REQUIRED BUFFER FORMATS field is defined in 5.4.2.2.

OD6a The asynchronous event solicited notification bit (AESOLNT) specifies whether an SRP_AER_REQ request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. See 6.12.

The credit request solicited notification bit (CRSOLNT) specifies whether an SRP_CRED_REQ request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. See 6.10.

The logout solicited notification bit (LOSOLNT) specifies whether an SRP_T_LOGOUT request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. See 6.6.

The MULTI-CHANNEL ACTION field identifies how an SRP target port treats any existing RDMA channel associated with the same I_T nexus. The MULTI-CHANNEL ACTION field is defined in table 10.

Table 10 - MULTI-CHANNEL ACTION code values

MULTI-CHANNEL ACTION	Description
00h	Single RDMA channel operation (see 5.1.3)
01h	Multiple independent RDMA channel operation (see 5.1.4)
02h	Reserved
03h	Reserved

CPQ017

The INITIATOR PORT IDENTIFIER field and the TARGET PORT IDENTIFIER field specify the I_T nexus that shall be associated with this RDMA channel.

6.3 SRP_LOGIN_RSP response

An SRP_LOGIN_RSP response (see table 11) indicates successful RDMA channel establishment and conveys SRP ~~protocol~~ login parameters from an SRP target port to an SRP initiator port. An SRP_LOGIN_RSP response shall ~~only~~ be sent as acceptance data ~~to indicate successful~~ during RDMA channel establishment (see 4.2). OD6h

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	Reserved			SOLNTSUP	Reserved		MULTI-CHANNEL RESULT	
27	Reserved							
28	Reserved							
• • •								
51								

The REQUEST LIMIT DELTA field is defined in 5.3.

The TAG field shall contain the same value as the TAG field in the SRP_LOGIN_REQ request (see 6.2).

CPQ018 MAXIMUM INITIATOR TO TARGET IU LENGTH specifies the maximum length in bytes of any information unit that the SRP target port is able to receive on this RDMA channel. This value shall be 64 or larger and greater than or equal to the value of REQUESTED MAXIMUM INITIATOR TO TARGET IU LENGTH specified in the SRP_LOGIN_REQ request (see 6.2). The SRP initiator port shall not send any information unit on this RDMA channel longer than this value.

CPQ018 MAXIMUM TARGET TO INITIATOR IU LENGTH specifies the maximum length in bytes of any information unit that the SRP target port may send on this RDMA channel. This value shall be 52 or larger. The SRP target port shall not send any information unit on this RDMA channel longer than this value.

The SUPPORTED BUFFER FORMATS field is defined in 5.4.2.2.

MULTI-CHANNEL RESULT identifies how the SRP target port treated existing RDMA channels associated with the same I_T nexus. Table 12 defines this field.

Table 12 - MULTI-CHANNEL RESULT code values

MULTI-CHANNEL RESULT	Description
00h	No existing RDMA channels were associated with the same I_T nexus.
01h	One or more existing RDMA channels were terminated.
02h	One or more existing RDMA channels continue to operate independently.
03h	Reserved

The solicited notification supported bit (SOLNTSUP) indicates whether the SRP target port supports solicited message reception notification for messages sent from the SRP target port to an SRP initiator port (see 4.3). If the SOLNTSUP bit is one, the SRP target port supports solicited message reception notification. If the SOLNTSUP bit is zero, the SRP target port only supports normal message reception notification. OD6h

6.4 SRP_LOGIN_REJ response

OD6i

An SRP_LOGIN_REJ response (see table 13) indicates ~~is sent by a SRP target port to notify the SRP initiator port~~ that an RDMA channel could not be established. An SRP_LOGIN_REJ response shall be sent as rejection data (see 4.2).

Table 13 - SRP_LOGIN_REJ response

Troika

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (C2h)							
1								
2				Reserved				
3								
4	(MSB)							
• • •				REASON				
7								(LSB)
8	(MSB)							
• • •				TAG				
15								(LSB)
16								
• • •				Reserved				
23								
24	SUPPORTED BUFFER FORMATS							
25								
26								
• • •				Reserved				
31								

CPQ020

The REASON field indicates the reason that the RDMA channel could not be established. This field is defined in table 14.

Table 14 - SRP_LOGIN_REJ response reason codes

REASON code	Description
0001 0000h	Unable to establish RDMA channel, no reason specified
0001 0001h	Insufficient RDMA channel resources
0001 0002h	REQUESTED MAXIMUM INITIATOR TO TARGET IU LENGTH value too large
0001 0003h	Unable to associate RDMA channel with specified I_T nexus
0001 0004h	One or more requested data buffer descriptor formats are not supported
0001 0005h	SRP target port does not support multiple RDMA channels per I_T nexus
0001 0006h	RDMA channel limit reached for this initiator
all other values	Reserved

intel0044

CPQ019

The TAG field shall contain the same value as the TAG field in the SRP_LOGIN_REQ request (see 6.2).

The SUPPORTED BUFFER FORMATS field is defined in 5.4.2.2.

6.5 SRP_I_LOGOUT request

An SRP_I_LOGOUT request (see table 15) is sent by an SRP initiator port to notify the SRP target port that the SRP initiator port is disconnecting the RDMA channel. An SRP_I_LOGOUT request may also be used to notify the SRP target port that an RDMA channel has failed, rendering it non-operational. An SRP_I_LOGOUT request shall be sent as a 16-byte message with normal message reception notification (see 4.3).

OD6j

Table 15 - SRP_I_LOGOUT request

Byte \ Bit	7	6	5	4	3	2	1	0
0	TYPE (03h)							
1	Reserved							
2								
7								
8	(MSB)	TAG						
• • •								
15								(LSB)

The TAG field is defined in 6.1.

After sending an SRP_I_LOGOUT request an SRP initiator port may delay a vendor specific time to allow the SRP_I_LOGOUT request to be delivered to the SRP target port. The SRP initiator port shall then request that the RDMA channel be disconnected and perform the actions specified in 5.1.2.

Upon receiving an SRP_I_LOGOUT request an SRP target port shall perform the actions specified in 5.1.2. The SRP target port shall not send an SRP response to an SRP_I_LOGOUT request.

6.6 SRP_T_LOGOUT request

An SRP_T_LOGOUT request (see table 16) is sent by a SRP target port to notify the SRP initiator port that the SRP target port is disconnecting the RDMA channel. An SRP_T_LOGOUT request may also be used to notify the SRP initiator port that an RDMA channel has failed, rendering it non-operational. An SRP_T_LOGOUT request shall be sent as a 16-byte message. OD6e

Table 16 - SRP_T_LOGOUT request

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (80h)							
1	Reserved							SOLNT
2								
3								
4	(MSB)							
• • •				REASON				
7								(LSB)
8	(MSB)							
• • •				TAG				
15								(LSB)

OD6e

The solicited notification (SOLNT) bit indicates whether the SRP initiator port specified normal or solicited message reception notification for SRP_T_LOGOUT requests during login (see 6.2). The SOLNT bit shall contain the value that was specified in the LOSOLNT bit of the SRP_LOGIN_REQ request. OD6e

If the solicited notification (SOLNT) bit is one and the SRP target port supports solicited message reception notification (see 6.3), the SRP target port shall send the SRP_T_LOGOUT response with solicited message reception notification (XXXsee 4.3). Otherwise the SRP target port should send the SRP_T_LOGOUT response with normal message reception notification. An SRP initiator port shall not validate the SOLNT bit against whether an SRP_RSP response was actually received with normal or solicited message reception notification.

The REASON field indicates the reason for disconnecting the RDMA channel. This field is defined in table 17.

Table 17 - SRP_T_LOGOUT request reason codes

REASON code	Description
0000 0000h	No reason specified.
0000 0001h	Inactive RDMA channel (reclaiming resources).
0000 0002h	Invalid information unit TYPE code received by SRP target port.
0000 0003h	Valid response type code with no corresponding SRP target port request outstanding.

Table 17 - SRP_T_LOGOUT request reason codes

REASON code	Description
0000 0004h	RDMA channel disconnected due to MULTI-CHANNEL ACTION code in new SRP_Login_Req.
0000 0006h	Unsupported format code value specified in DATA-OUT BUFFER DESCRIPTOR FORMAT field
0000 0007h	Unsupported format code value specified in DATA-IN BUFFER DESCRIPTOR FORMAT field
0000 0008h	Invalid length for IU type
all other values	Reserved

el0096 IBM087

CPQ021

The TAG field is defined in 6.1.

intel0106 After sending an SRP_T_LOGOUT request an SRP target port may delay a vendor specific time to allow the SRP_T_LOGOUT request to be delivered to the SRP initiator port. The SRP target port shall then request that the RDMA channel be disconnected and perform the actions specified in 5.1.2.

Upon receiving an SRP_T_LOGOUT request an SRP initiator port shall perform the actions specified in 5.1.2. The SRP initiator port shall not send an SRP response to an SRP_T_LOGOUT request.

6.7 SRP_TSK_MGMT request

An SRP_TSK_MGMT request conveys a SCSI task management request (table 18). An SRP_TSK_MGMT request shall be sent with normal message reception notification (see 4.3). OD6b

Table 18 - SRP_TSK_MGMT request

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (01h)							
1	Reserved					UCSOLNT	SCSOLNT	Reserved
• • •	Reserved							
7								
8	(MSB)							
• • •	TAG							
15								(LSB)
16								
• • •	Reserved							
19								
20	(MSB)							
• • •	LOGICAL UNIT NUMBER							
27								(LSB)
28	Reserved							
29	Reserved							
30	TASK MANAGEMENT FUNCTION							
31	Reserved							
32	(MSB)							
• • •	TAG OF TASK TO BE MANAGED							
39								(LSB)
40								
• • •	Reserved							
47								

OD6b CPQ022

The unsuccessful completion solicited notification bit (UCSOLNT) specifies whether an SRP_RSP response reporting unsuccessful completion of the task management request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. See 6.9. OD6b

The successful completion solicited notification bit (SCSOLNT) specifies whether an SRP_RSP response reporting successful completion of the task management request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. See 6.9.

The TAG field is defined in 6.1.

The LOGICAL UNIT NUMBER field specifies the address of the logical unit component of the nexus for the task management request. The structure of the LOGICAL UNIT NUMBER field shall be as defined in the SCSI Architecture Model-2 standard. This field is reserved if the task management request is not directed to either an I_T_L or I_T_L_Q nexus.

CPQ022 The TASK MANAGEMENT FUNCTION field is defined in table 19. If TASK MANAGEMENT FUNCTION contains a reserved or restricted value, the task manager shall return an SRP_RSP response (see 6.9) containing GOOD status. The RSP_CODE field shall be set to TASK MANAGEMENT FUNCTION NOT SUPPORTED.

CPQ025

Table 19 - TASK MANAGEMENT FUNCTION codes

CPQ024

Code	Description
01h	The task manager shall perform an ABORT TASK function (see SAM-2).
02h	The task manager shall perform an ABORT TASK SET function (see SAM-2).
04h	The task manager shall perform a CLEAR TASK SET function (see SAM-2).
08h	The task manager shall perform a LOGICAL UNIT RESET function (see SAM-2).
20h	Restricted.
40h	The task manager shall perform a CLEAR ACA function (see SAM-2).
All other values	Reserved

CPQ023

If TASK MANAGEMENT FLAGS specifies that an ABORT TASK function shall be performed, the TAG OF TASK TO BE MANAGED field specifies the TAG value from the SRP_CMD request (see 6.8) that contained the task to be aborted. The TAG OF TASK TO BE MANAGED field shall be ignored if TASK MANAGEMENT FLAGS specifies any other function.

6.8 SRP_CMD request

An SRP_CMD request conveys a SCSI command (see table 20). An SRP_CMD request shall be sent as a

Table 20 - SRP_CMD request

Byte \ Bit	7	6	5	4	3	2	1	0
0	TYPE (02h)							
1	Reserved					UCSOLNT	SCSOLNT	Reserved
• • •	Reserved							
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	Reserved							
• • •								
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								
48+4*n+do	DATA-IN BUFFER DESCRIPTOR							
• • •								
47+4*n+do+di								

[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 5.4.2).

[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 5.4.2).

OD6c

CPQ026

CPQ027

CPQ027

message whose length is 48 bytes plus the lengths of the ADDITIONAL CDB, DATA-OUT BUFFER DESCRIPTOR and DATA-IN BUFFER DESCRIPTOR fields. An SRP_CMD request shall be sent with normal message reception notification (see 4.3).

OD6c

OD6c The unsuccessful completion solicited notification bit (UCSOLNT) specifies whether an SRP_RSP response reporting unsuccessful completion of the task management request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. XXXSee 6.9.

The successful completion solicited notification bit (SCSOLNT) specifies whether an SRP_RSP response reporting successful completion of the task management request should use normal or solicited message reception notification. This bit shall be set to one to request solicited notification, or set to zero to request normal notification. XXXSee 6.9.

The DATA-OUT BUFFER DESCRIPTOR FORMAT field specifies the format of the DATA-OUT BUFFER DESCRIPTOR field (see 5.4.2).

The DATA-IN BUFFER DESCRIPTOR FORMAT field specifies the format of the DATA-IN BUFFER DESCRIPTOR field (see 5.4.2).

The DATA-OUT BUFFER DESCRIPTOR COUNT field provides additional information to specify the format of the DATA-OUT BUFFER DESCRIPTOR field (see 5.4.2).

The DATA-IN BUFFER DESCRIPTOR COUNT field provides additional information to specify the format of the DATA-IN BUFFER DESCRIPTOR field (see 5.4.2).

The TAG field is defined in 6.1.

The LOGICAL UNIT NUMBER field specifies the address of the logical unit of the I_T_L_Q nexus for the current task. The structure of the logical unit number field shall be as defined in the SCSI Architecture Model-2 standard. If the addressed logical unit does not exist, the task manager shall follow the SCSI rules for selection of invalid logical units as defined in the SCSI Primary Commands-2 standard.

The TASK ATTRIBUTE field is defined in table 21.

CPQ028

Table 21 - TASK ATTRIBUTE

CPQ030

Codes	Description
000b	Requests that the task be managed according to the rules for a simple task attribute. (See SAM-2)
001b	Requests that the task be managed according to the rules for a head of queue task attribute. (See SAM-2)
010b	Requests that the task be managed according to the rules for an ordered attribute. (See SAM-2)
011b	Reserved
100b	Requests that the task be managed according to the rules for an automatic contingent allegiance task attribute. (See SAM-2)
101b-111b	Reserved

CPQ029

The ADDITIONAL CDB LENGTH field contains the length in 4-byte words of the ADDITIONAL CDB field.

The CDB and ADDITIONAL CDB fields together contain the CDB to be interpreted by the addressed logical unit. Any bytes between the end of the CDB and the end of the two fields shall be reserved.

The contents of the CDB shall be as defined in the SCSI command standards.

The DATA-OUT BUFFER DESCRIPTOR field specifies the buffer that shall be used for data-out transfers (see 5.4.2).

The DATA-IN BUFFER DESCRIPTOR field specifies the buffer that shall be used for data-in transfers (see 5.4.2).

6.9 SRP_RSP response

An SRP_RSP response (see table 22) conveys an SRP response to an SRP_TSK_MGMT request (see 6.7) or an SRP_CMD request (see 6.8) received by a SRP target port. SRP_RSP responses that contain neither RESPONSE DATA nor SENSE DATA shall be sent as a 36 byte message. SRP_RSP responses that contain either RESPONSE DATA or SENSE DATA shall be sent as the minimum length message capable of containing those fields.

Table 22 - SRP_RSP response

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (C1h)							
1	Reserved							SOLNT
2	Reserved							
3								
4	(MSB)	REQUEST LIMIT DELTA						
• • •								
7								(LSB)
8	(MSB)	TAG						
• • •								
15								(LSB)
16	Reserved							
17								
18	Reserved		DIUNDER	DIOVER	DOUNDER	DOOVER	SNSVALID	RSPVALID
19	STATUS							
20	(MSB)	DATA-OUT RESIDUAL COUNT						
• • •								
23								(LSB)
24	(MSB)	DATA-IN RESIDUAL COUNT						
• • •								
27								(LSB)
28	(MSB)	SENSE DATA LIST LENGTH = n						
• • •								
31								(LSB)
32	(MSB)	RESPONSE DATA LIST LENGTH = m						
• • •								
35								(LSB)
36	(MSB)	RESPONSE DATA (m bytes long)						
• • •								
35+m								(LSB)
36+m	(MSB)	SENSE DATA (n bytes long)						
• • •								
35+m+n								(LSB)

OD6d

OD6d

The solicited notification (SOLNT) bit indicates whether the SRP initiator port specified normal or solicited message reception notification for this response. If the STATUS field is non-zero or if the RSP_CODE field is present

and non-zero, then the SOLNT bit shall contain the value that was specified in the UCSOLNT bit of the corresponding SRP_CMD or SRP_TSK_MGMT request. Otherwise the SOLNT bit shall contain the value that was specified in the SCSOLNT bit of the corresponding SRP_CMD or SRP_TSK_MGMT request.

If the solicited notification (SOLNT) bit is one and the SRP target port supports solicited message reception notification (see 6.3), the SRP target port shall send the SRP_RSP response with solicited message reception notification (see 4.3). Otherwise the SRP target port should send the SRP_RSP response with normal message reception notification. An SRP initiator port shall not validate the SOLNT bit against whether an SRP_RSP response was actually received with normal or solicited message reception notification.

The REQUEST LIMIT DELTA field is defined in 5.3.

The TAG field shall contain the same value as the TAG field in the SRP_TSK_MGMT request (see 6.7) or SRP_CMD request (see 6.8) for which this SRP_RSP response is a response.

DOUNDER, when set to 1, indicates that the DATA-OUT RESIDUAL COUNT field is valid and contains the count of data bytes that were expected to be transferred from the data-out buffer, but were not transferred. The application client should examine the DATA-OUT RESIDUAL COUNT field in the context of the command to determine whether or not an error condition occurred.

DOOVER, when set to 1, indicates that the DATA-OUT RESIDUAL COUNT field is valid and contains the count of data bytes that could not be transferred from the data-out buffer because the length of the data-out buffer was not sufficient. The application client should examine the DATA-OUT RESIDUAL COUNT field in the context of the command to determine whether or not an error condition occurred.

DOUNDER and DOOVER, when both set to 0, indicate that the DATA-OUT RESIDUAL COUNT field is not valid; the SRP initiator port shall ignore its contents. The SRP target port shall not set both DOUNDER and DOOVER to 1.

DIUNDER, when set to 1, indicates that the DATA-IN RESIDUAL COUNT field is valid and contains the count of data bytes that were expected to be transferred to the data-in buffer, but were not transferred. The application client should examine the DATA-IN RESIDUAL COUNT field in the context of the command to determine whether or not an error condition occurred.

DIOVER, when set to 1, indicates that the DATA-IN RESIDUAL COUNT field is valid and contains the count of data bytes that could not be transferred to the data-in buffer because the length of the data-in buffer was not sufficient. The application client should examine the DATA-IN RESIDUAL COUNT field in the context of the command to determine whether or not an error condition occurred.

DIUNDER and DIOVER, when both set to 0, indicate that the DATA-IN RESIDUAL COUNT field is not valid; the SRP initiator port shall ignore its contents. The SRP target port shall not set both DIUNDER and DIOVER to 1.

SNSVALID, when set to 0, indicates the contents of the SENSE DATA LIST LENGTH field shall be ignored and the SENSE DATA field is not present. SNSVALID, when set to 1, indicates the contents of the SENSE DATA LIST LENGTH field specify the number of bytes in the SENSE DATA field.

If sense data is provided, SNSVALID shall be set to 1 and the SENSE DATA LIST LENGTH field shall specify the number of bytes in the SENSE DATA field. The SENSE DATA LIST LENGTH field shall only contain lengths that are multiples of four.

If returning all the sense data provided would cause the SRP_RSP response to be longer than the value of the MAXIMUM TARGET TO INITIATOR IU LENGTH field indicated in the SRP_LOGIN_RSP response (see 6.3) when the RDMA channel was established, the SRP target port shall return an SRP_RSP response whose length is the value from the MAXIMUM TARGET TO INITIATOR IU LENGTH field truncated to a multiple of four bytes. The SENSE DATA field shall be truncated as needed to achieve this length. SENSE DATA LIST LENGTH shall contain the length of the truncated SENSE DATA field.

If no sense data is provided, SNSVALID shall be set to 0. The SRP initiator port shall ignore the SENSE DATA LIST LENGTH field and shall assume a length of zero.

RSPVALID set to 0 indicates the contents of the RESPONSE DATA LIST LENGTH field shall be ignored and the RESPONSE DATA field is not present. RSPVALID set to 1 indicates the contents of the RESPONSE DATA LIST LENGTH field specify the number of bytes in the RESPONSE DATA field. RSPVALID set to 1 also indicates that the contents of the STATUS field are not reliable and shall be ignored by the SRP initiator port.

If response data is provided, RSPVALID shall be set to 1 and the RESPONSE DATA LIST LENGTH field shall specify the number of bytes in the RESPONSE DATA field. The RESPONSE DATA LIST LENGTH field shall contain a length of 4. Other lengths are reserved for future standardization.

If no response data is provided, RSPVALID shall be set to 0. The SRP initiator port shall ignore the RESPONSE DATA LIST LENGTH field and shall assume a length of zero.

Response data shall be provided in any SRP_RSP response that is sent in response to an SRP_TSK_MGMT request (see 6.7). The information in the RSP_CODE field shall indicate the completion status of the task management function.

Response data shall not be provided in any SRP_RSP response that returns a non-zero status code in the STATUS field.

CPQ031 The STATUS field contains the status of a task that completes. See the SAM-2 standard for a list of status codes.

If either DOUNDER or DOOVER is set to 1, the DATA-OUT RESIDUAL COUNT field contains a count of the number of residual data bytes that were not transferred from the data-out buffer for this SCSI command. Upon successful completion of an SRP I/O operation, the residual data-out byte count is normally zero and the DATA-OUT RESIDUAL COUNT value is not valid. Some commands may have a non-zero residual data-out byte count that is not an error. SRP target ports are not required to check the data-out length implied by the contents of the CDB for overrun or underrun before processing a SCSI command.

IBM0119 ~~If DOUNDER is set to 1, a transfer that did not use the entire data-out buffer was performed and the value of DATA-OUT RESIDUAL COUNT shall be equal to:~~If DOUNDER is set to one and a transfer that did not fill the entire data-out buffer was performed, the value of DATA-OUT RESIDUAL COUNT is defined as follows:

DATA-OUT RESIDUAL COUNT = (data-out buffer length) - (highest offset of any data-out byte transmitted + 1)

~~data-out buffer length - highest offset of any data-out byte transferred - 1~~

A condition of DOUNDER set to 1 may not be an error for some devices and some commands.

If DOOVER is set to 1, the transfer was truncated because the data-out transfer required by the SCSI command was longer than the data-out buffer. Those bytes that could not be transferred without exceeding the length of
IBM0120 the data-out buffer shall not be transferred. ~~DATA-OUT RESIDUAL COUNT shall be equal to:~~ The DATA-OUT RESIDUAL COUNT is defined as follows:

DATA-OUT RESIDUAL COUNT = (Transfer length required by command) - (data-out buffer length)

~~data-out transfer length required by command - data-out buffer length~~

If DOOVER is set to 1, the termination state of the SRP I/O operation is not certain. Data may or may not have been transferred from the data-out buffer and the SCSI status byte may or may not provide correct command completion information.

If either DIUNDER or DIOVER is set to 1, the DATA-IN RESIDUAL COUNT field contains a count of the number of residual data bytes that were not transferred to the data-in buffer for this SCSI command. Upon successful completion of an SRP I/O operation, the residual data-in byte count is normally zero and the DATA-IN RESIDUAL COUNT value is not valid. Some commands may have a non-zero residual data-in byte count that is not an error. SRP target ports are not required to check the data-in length implied by the contents of the CDB for overrun or underrun before processing a SCSI command.

~~If DIUNDER is set to 1, a transfer that did not fill the entire data-in buffer was performed and the value of DATA-IN RESIDUAL COUNTshall be equal to:~~ If DIUNDER is set to one and a transfer that did not fill the entire data-in buffer was performed, the value of DATA-IN RESIDUAL COUNT is defined as follows: IBM0121

DATA-IN RESIDUAL COUNT = (data-in buffer length) - (highest offset of any data-in byte transmitted + 1)

~~data-in buffer length - highest offset of any data-in byte transferred - 1~~

A condition of DIUNDER set to 1 may not be an error for some devices and some commands.

If DIOVER is set to 1, the transfer was truncated because the data-in transfer required by the SCSI command was longer than the data-in buffer. Those bytes that could not be transferred without exceeding the length of the data-in buffer shall not be transferred. ~~DATA-IN RESIDUAL COUNT shall be equal to:~~The DATA-IN RESIDUAL COUNT is defined as follows: IBM0122

DATA-IN RESIDUAL COUNT = (Transfer length required by command) - (data-in buffer length)

~~data-in transfer length required by command - data-in buffer length~~

If DIOVER is set to 1, the termination state of the SRP I/O operation is not certain. Data may or may not have been transferred to the data-in buffer and the SCSI status byte may or may not provide correct command completion information.

The DATA-OUT RESIDUAL COUNT, DATA-IN RESIDUAL COUNT, SENSE DATA LIST LENGTH and RESPONSE DATA LIST LENGTH fields shall always be present in the SRP_RSP response, regardless of whether their contents are valid.

The RESPONSE DATA field (see table 23) contains information describing ~~certain~~ protocol failures detected during processing of an SRP request received by the SRP target port. The RESPONSE DATA field shall be present if the SRP target port detects any of the conditions described by a non-zero RSP_CODE value (see table 24). IBM0123

Table 23 - RESPONSE DATA field

Bit / Byte	7	6	5	4	3	2	1	0
0	Reserved							
1	Reserved							
2	Reserved							
3	RSP_CODE							

The RSP_CODE field is defined in table 24.

Table 24 - RSP_CODE values

Codes	Description
00h	NO FAILURE or TASK MANAGEMENT FUNCTION COMPLETE.
01h	Reserved
02h	REQUEST FIELDS INVALID
03h	Reserved
04h	TASK MANAGEMENT FUNCTION NOT SUPPORTED
05h	TASK MANAGEMENT FUNCTION FAILED
07h-FFh	Reserved

edit006

~~The SENSE DATA field contains the information specified by the SCSI Primary Commands-2 standard for presentation by the REQUEST SENSE command. The proper sense data shall be presented when a SCSI status byte of CHECK CONDITION is presented as specified by the SCSI Primary Commands-2 standard. The SENSE DATA field shall contain the data that would be presented by a REQUEST SENSEcommand whose ALLOCATION LENGTH parameter contains the value:~~

~~MAXIMUM TARGET TO INITIATOR IU LENGTH - 36 - RESPONSE DATA LIST LENGTH~~

~~MAXIMUM TARGET TO INITIATOR IU LENGTH is the value specified in the SRP_LOGIN_RSP response (see 6.3) when the RDMA channel was established.~~

CPQ032 ~~NOTE 4 - The value 36 is the length of the fixed portion of an SRP_RSP response. It is the length of an SRP_RSP response excluding the SENSE DATA field and RESPONSE DATA field.~~

CPQ033 The SENSE DATA field contains the autosense data specified by the SCSI Primary Commands-2 standard. The proper sense data shall be presented when the SCSI status byte of CHECK CONDITION is presented as

IBM0124 specified by the SCSI Primary Commands-2 standard. If no conditions requiring the presentation of SCSI sense data have occurred, the SENSE DATA field shall not be included in the SRP_RSP response and the RSPVALID bit shall be zero. SRP devices shall perform autosense.

6.10 SRP_CRED_REQ request

An SRP target port may use SRP_CRED_REQ requests (see table 25) to adjust an SRP initiator port's REQUEST LIMIT value (see 5.3). ~~All SRP initiator ports shall support receiving SRP_CRED_REQ requests.~~ An SRP_CRED_REQ request~~s~~ shall be sent as a 16 byte message. intel0094

Table 25 - SRP_CRED_REQ request

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (81h)							
1	Reserved							SOLNT
2	Reserved							
3								
4	(MSB)							
• • •	REQUEST LIMIT DELTA							
7								(LSB)
8	(MSB)							
• • •	TAG							
15								(LSB)

OD6f

The solicited notification (SOLNT) bit indicates whether the SRP initiator port specified normal or solicited message reception notification during login (see 6.2) for SRP_CRED_REQ requests. The SOLNT bit shall contain the value that was specified in the CRSOLNT bit of the SRP_LOGIN_REQ request. OD6f

If the solicited notification (SOLNT) bit is one and the SRP target port supports solicited message reception notification (see 6.3), the SRP target port shall send the SRP_CRED_REQ request with solicited message reception notification (XXXsee 4.3). Otherwise the SRP target port should send the SRP_CRED_REQ request with normal message reception notification. An SRP initiator port shall not validate the SOLNT bit against whether an SRP_CRED_REQ request was actually received with normal or solicited message reception notification.

The REQUEST LIMIT DELTA field is defined in 5.3.

The TAG field is defined in 6.1.

6.11 SRP_CRED_RSP response

OD6k

An SRP_CRED_RSP response (see table 26) is the response to an SRP_CRED_REQ request (see 6.10) received by an SRP initiator port. All SRP initiator ports shall support generating SRP_CRED_RSP responses. SRP_CRED_RSP responses shall be sent as a 16-byte message with normal message recption notification (see 4.3)..

Table 26 - SRP_CRED_RSP response

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (41h)							
1	Reserved							
7								
8	(MSB)	TAG						
• • •								
15								(LSB)

The TAG field shall contain the same value as the TAG field in the SRP_CRED_REQ request (see 6.10) for which this SRP_CRED_RSP response is a response.

6.12 SRP_AER_REQ request

CPQ034

A target port sends an SRP_AER_REQ request (see table 27) to report ~~conveys a SRP target port request to report~~ an asynchronous event. An SRP_AER_REQ request~~s~~ shall be sent as the minimum length message capable of carrying the fields. Parameters managing the use of asynchronous event reporting are contained in the Control mode page (see SPC-2). ~~All SRP initiator ports shall support receiving SRP_AER_REQ requests and all SRP target ports shall support generating SRP_AER_REQ requests~~.

CPQ035 OD6g IBM0125

Table 27 - SRP_AER_REQ request

intel0094

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (82h)							
1	Reserved							SOLNT
2	Reserved							
3								
4	(MSB)							
• • •	REQUEST LIMIT DELTA							
7								(LSB)
8	(MSB)							
• • •	TAG							
15								(LSB)
16	Reserved							
• • •								
19								
20	(MSB)							
• • •	LOGICAL UNIT NUMBER							
27								(LSB)
28	(MSB)							
• • •	SENSE DATA LIST LENGTH = n							
31								(LSB)
32	Reserved							
• • •								
35								
36	(MSB)							
• • •	SENSE DATA (n bytes long)							
35+n								(LSB)

OD6g

The solicited notification (SOLNT) bit indicates whether the SRP initiator port specified normal or solicited message reception notification during login (see 6.2) for SRP_AER_REQ requests. The SOLNT bit shall contain the value that was specified in the CRSOLNT bit of the SRP_LOGIN_REQ request.

If the solicited notification (SOLNT) bit is one and the SRP target port supports solicited message reception notification (see 6.3), the SRP target port shall send the SRP_AER_REQ request with solicited message reception notification (XXXsee 4.3). Otherwise the SRP target port should send the SRP_AER_REQ request with normal message reception notification. An SRP initiator port shall not validate the SOLNT bit against whether an SRP_AER_REQ request was actually received with normal or solicited message reception notification.

The REQUEST LIMIT DELTA field is defined in 5.3.

The TAG field is defined in 6.1.

The SENSE DATA LIST LENGTH field shall specify the number of bytes in the SENSE DATA field. The SENSE DATA LIST LENGTH field shall only contain lengths that are multiples of four. If no sense data is provided, the SENSE DATA LIST LENGTH field shall be set to zero.

If including all the sense data provided would cause the SRP_AER_REQ request to be longer than the value of the MAXIMUM TARGET TO INITIATOR IU LENGTH field indicated in the SRP_LOGIN_RSP response (see 6.3) when the RDMA channel was established, the SRP target port shall send an SRP_AER_REQ request whose length is the MAXIMUM TARGET TO INITIATOR IU LENGTH field value truncated to a multiple of four bytes. The SENSE DATA field shall be truncated as needed to achieve this length. SENSE DATA LIST LENGTH shall contain the length of the truncated SENSE DATA field.

CPQ036 IBM0126 ~~The SENSE DATA field contains the information specified by the SCSI Primary Commands-2 standard for presentation by the REQUEST SENSE command. The proper sense data shall be presented when a SCSI status byte of CHECK CONDITION is presented as specified by the SCSI Primary Commands-2 standard. The SENSE DATA field shall contain the data that would be presented by a REQUEST SENSE command whose ALLOCATION LENGTH parameter contains the value:~~

~~MAXIMUM TARGET TO INITIATOR IU LENGTH - 36~~

~~MAXIMUM TARGET TO INITIATOR IU LENGTH is the value specified in SRP_LOGIN_RSP response (see 6.3) when the RDMA channel was established.~~

~~NOTE 5 - The value 36 is the length of the fixed portion of an SRP_AER_REQ request. It is the length of an SRP_AER_REQ request excluding the SENSE DATA field.~~

The SENSE DATA field contains sense data as specified by the SCSI Primary Commands-2 standard.

6.13 SRP_AER_RSP response

An SRP_AER_RSP response (see table 28) conveys an SRP initiator port's SRP response to an SRP_AER_REQ request (see 6.12). An SRP_AER_RSP response shall be sent as a 16-byte message with normal message recption notification (see 4.3). OD6l

~~All SRP initiator ports shall support generating SRP_AER_RSP responses and all SRP target ports shall support receiving SRP_AER_RSP responses~~. intel0094

Table 28 - SRP_AER_RSP response

Bit / Byte	7	6	5	4	3	2	1	0
0	TYPE (42h)							
1								
2				Reserved				
7								
8	(MSB)							
• • •				TAG				
15								(LSB)

The TAG field shall contain the same value as the TAG field in the SRP_AER_REQ request (see 6.12) for which this SRP_AER_RSP response is a response.

7 SCSI mode parameters

7.1 SCSI mode parameter overview and codes

This subclause describes the block descriptors and the pages used with MODE SELECT and MODE SENSE commands that influence, control and report the behavior of the SRP interface. All mode parameters not defined in this standard shall influence the behavior of the SCSI devices as specified in the appropriate command set document. The mode pages are addressed to the device server of a logical unit. The mode pages associated with SRP are listed in table 29.

Table 29 - SRP mode page codes

CPQ037

Page code	Description	Subclause
02h	Disconnect-reconnect page	7.2
18h	Protocol specific LUN page	7.3
19h	Protocol specific port page	7.4

7.2 Disconnect-reconnect mode page

The disconnect-reconnect page (see table 30) provides the application client the means to tune the performance of the service delivery subsystem. The following subclause defines the fields in the disconnect-reconnect mode page of the MODE SENSE or MODE SELECT command that are used by SRP target ports.

intel0134

Table 30 - Disconnect-reconnect mode page

Bit / Byte	7	6	5	4	3	2	1	0
0	PS	RESERVED	PAGE CODE (02h)					
1	PAGE LENGTH (0EH)							
2	BUFFER FULL RATIO							
3	BUFFER EMPTY RATIO							
4	BUS INACTIVITY LIMIT							
5								
6	PHYSICAL DISCONNECT TIME LIMIT							
7								
8	CONNECT TIME LIMIT							
9								
10	(MSB) MAXIMUM BURST SIZE							
11								(LSB)
12	EMDP	FAIR ARBITRATION			DIMM	DTDC		
13	RESERVED							
14	FIRST BURST SIZE							
15								

The application client passes the fields used to control the SRP interface to a device server by means of a MODE SELECT command. The device server then communicates the field values to the SRP target port. The field values are communicated from the device server to the SRP target port in a vendor specific manner.

SRP devices shall only use disconnect-reconnect page parameter fields defined below. If any other fields within the disconnect-reconnect page of the MODE SELECT command contain a non-zero value, the device server shall return CHECK CONDITION status for that MODE SELECT command. The device server shall set the sense key to ILLEGAL REQUEST and set the additional sense code to ILLEGAL FIELD IN PARAMETER LIST.

The BUFFER FULL RATIO field, BUFFER EMPTY RATIO field, BUS INACTIVITY LIMIT field, PHYSICAL DISCONNECT TIME LIMIT field and CONNECT TIME LIMIT field shall not be implemented by SRP target ports.

The MAXIMUM BURST SIZE field indicates the maximum size of an RDMA Read or RDMA Write operation that the device server shall perform. This value is expressed in increments of 512 bytes (e.g., a value of 1 means 512 bytes, two means 1024 bytes, etc.). The device server may round this value down as defined in SPC-2. A value of 0 indicates there is no limit on the amount of data transferred per data transfer operation. This value shall be implemented by all SRP devices. The application client and device server may use the value of this parameter to adjust internal maximum buffering requirements. A router between an SRP device and another protocol device (e.g. FCP) may intercept and adjust this value to reflect its own maximum buffering capabilities.

The ENABLE MODIFY DATA POINTERS (EMDP) bit indicates whether or not the SRP target port may use the random buffer access capability to order RDMAs for a single SCSI command. If the EMDP bit is set to 0, the SRP target port shall generate continuously increasing RDMA addresses for a single SCSI command. If the EMDP bit is set to 1, the SRP target port may issue RDMAs for a single SCSI command in any order. The EMDP bit does not affect the order of frames within an RDMA. The EMDP function shall be implemented by all SRP devices.

The FAIR ARBITRATION field, DISCONNECT IMMEDIATE (DIMM) bit, DATA TRANSFER DISCONNECT CONTROL (DTDC) field, and FIRST BURST SIZE field shall not be implemented by SRP target ports.

7.3 Protocol specific LUN page

CPQ037

The protocol specific LUN page shall not be implemented by SRP target ports.

7.4 Protocol specific port page

The protocol specific LUN page shall not be implemented by SRP target ports.

Annex A
(normative)

SRP interface protocol and services

A.1 Service interface protocol

This standard describes a SCSI device's behavior in terms of functional levels, service interfaces between levels and peer-to-peer protocols. For a full description of the model used in this standard see SAM-2. Figure A.1 shows the model as it appears from the point of view of this standard.

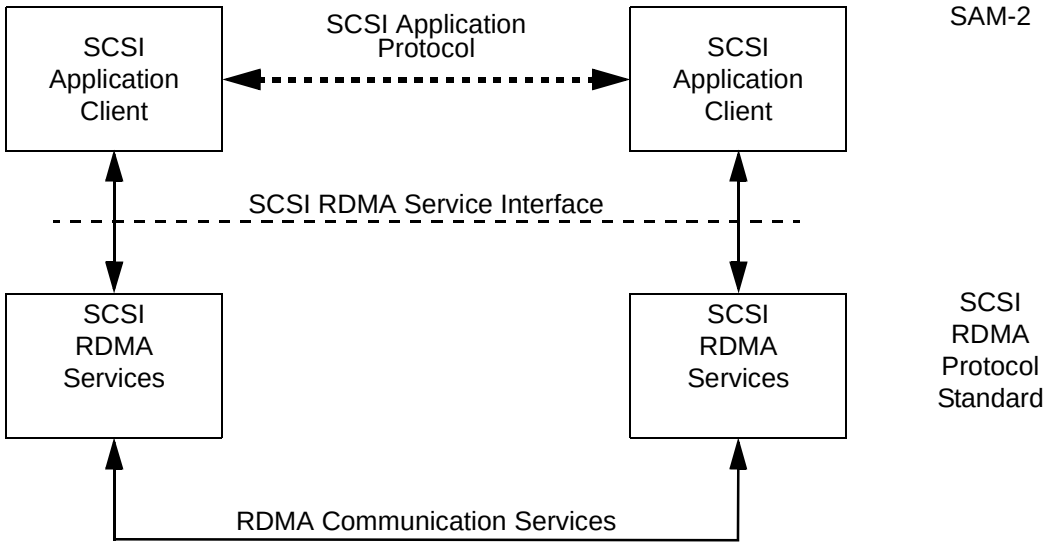


Figure A.1 - SRP reference model

Services between service levels are either four step confirmed services or two step confirmed services. A four step confirmed service consists of a service request, indication, response, and confirmation. A two step confirmed service consists of a service request and confirmation.

Figure A.2 shows the service and protocol interactions for a four step confirmed service.

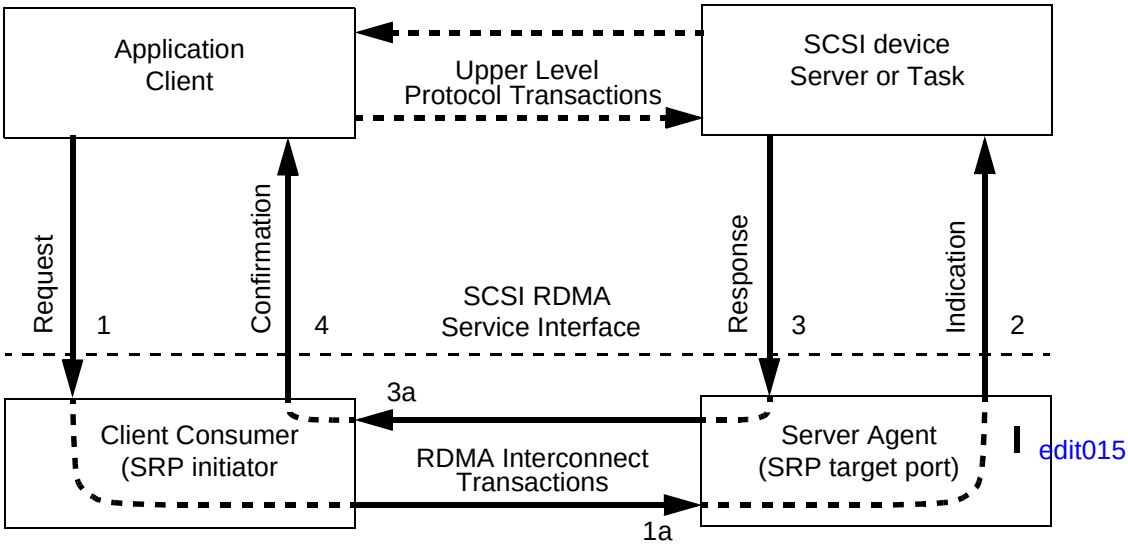


Figure A.2 - Model for a four step confirmed service

The SCSI RDMA four step confirmed service protocol consists of the following interactions:

1. A request to the client consumer to invoke a service;
2. An indication from the server agent notifying the SCSI device server or task manager of an event;
3. A response from the SCSI device server or task manager in reply to an indication;
4. A confirmation from the client consumer upon service completion.

Only application clients shall request a four step confirmed service be invoked.

Figure A.3 shows the service and protocol interactions for a two step confirmed service.

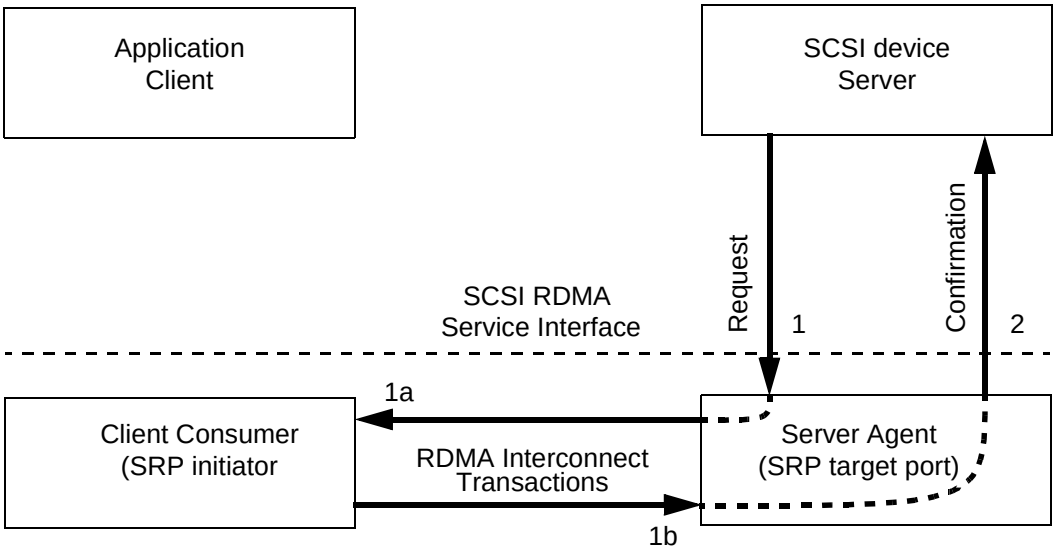


Figure A.3 - Model for a two step confirmed service

The SCSI RDMA two step confirmed service interface consists of the following interactions:

1. A request to the server agent to invoke a service;

2. A confirmation from the server agent upon service completion.

Only SCSI device servers shall request a two step confirmed service be invoked.

A.2 SRP services

SRP provides services to enable an application client to request and manage tasks (see SAM-2) and to enable a device server to receive commands and move data to and from an application client. The SRP services are described in terms of the services the SRP initiator port and SRP target port provide.

A.3 Procedure objects

See table A.1 for a list of the procedure objects used when passing services across the SRP service interface. See table A.1 for the definitions of the names used within this standard and the equivalent SAM-2 names of the procedure objects, the name of the standard where the objects are defined, the standard where the binary contents of the objects are defined, and the routing of the objects. The routing shows:

a) the source of the object
b) the final destination of the object, and
c) the routing of the object.

Table A.1 - SAM-2 procedure objects

Procedure object	Standard where object format is defined	Object routing
application client buffer offset	SAM-2	DS → targ→ init
data-out buffer size	SAM-2	AC → init
data-in buffer size	SAM-2	AC → init
command descriptor block	SAM-2/cmd[a]	AC → init → targ → DS
data-in buffer	cmd[b]	DS → targ→ init → AC
data-out buffer	cmd[b]	AC → init → targ → DS
device server buffer	cmd[b]	DS → targ→ init
I_T_L_x nexus	this standard	AC → init → targ → DS **or** AC → init → targ → TM **or** DS → targ→ init
request byte count	SAM-2	DS → targ
service response	this standard[c]	DS → targ→ init → AC **or** targ → DS
autosense request	SAM-2	AC → init → targ → DS
sense data	SPC-2	DS → targ→ init → AC
status	SAM-2	DS → targ→ init → AC
task attribute	this standard	AC → init → targ → DS

Key: AC=application client, cmd=SCSI command standards, DS=device server, init=SRP initiator port, SAM-2=SAM-2, TM=task manager, targ=SRP target port

[a] The portions not defined in SAM-2 are defined in the SCSI command standards (e.g., SPC-2).

[b] Parameter lists are defined within one of the SCSI command standards (e.g., SPC-2). SCSI standards do not define non-parameter list information.

[c] The SERVICE DELIVERY OR TARGET FAILURE value of the service response is not defined in SCSI.

A.4 Application client SCSI command services

A.4.1 Application client SCSI command services overview

The SCSI command services shall be requested by the application client using a procedure call defined as:

Execute Command (IN (I_T_L_x nexus, command descriptor block, [task attribute], [data-in buffer size], [data-out buffer], [data-out buffer size], autosense request), OUT ([data-in buffer], [sense data], status, service response))

A.4.2 Send SCSI command service

The send SCSI command service is a four step confirmed service (see figure A.2) that provides the means to transfer a command data block to a device server.

Processing the execute command procedure call for a send SCSI command service shall be composed of the four step confirmed service shown in table A.2.

Table A.2 - Processing of execute command procedure call for a send SCSI command service

Step (step number)[a]	Source to Destination	Protocol service name	SCSI Protocol Service Interface procedure calls
request (1)	application client to client consumer	send SCSI command request	Send SCSI command (IN (I_T_L_x nexus, command descriptor block, [task attribute], [data-in buffer size], [data-out buffer], [data-out buffer size], autosense request))
information unit transfer (1a)	client consumer to server agent	SRP_CMD request or SRP_TSK_MGMT request	See 6.7 and 6.8
indication (2)	server agent to device server	send SCSI command indication	SCSI command received (IN (I_T_L_x nexus, command descriptor block, [task attribute], autosense request))
If the send SCSI command requires a data transfer see A.5.2 for data-out delivery services and A.5.3 for data-in delivery services			
response (3)	device server to server agent	send SCSI command response	Send command complete (IN (I_T_L_x nexus, [sense data], status, service response))
information unit transfer (3a)	server agent to client consumer	SRP_RSP response	See 6.9
confirmation (4)	client consumer to application client	send SCSI command confirmation	Command complete received (IN (I_T_L_x nexus, [data-in buffer], [sense data], status, service response))
[a] See figure A.2 for step number			

A.5 Device server SCSI command services

A.5.1 Device server SCSI command services overview

The SCSI data buffer movement services shall be requested from the device server using a procedure call defined as:

> Move data buffer (IN (I_T_L_x nexus, device server buffer, application client buffer offset, request byte count)).

Either data-in delivery, data-out delivery, both data-in and data-out delivery, or neither data delivery may be used while processing one command. If both are used, the device server shall combine the data-in and data-out service responses into one service response.

A.5.2 Data-out delivery service

The data-out delivery service is a two step confirmed service (see figure A.3) that provides the means to transfer a parameter list or data from an SRP initiator port to a device server.

Processing the execute command procedure call for a data-out delivery service shall be composed of the two step confirmed service shown in table A.3.

Table A.3 - Processing of execute command procedure call for a data-out delivery service

Step (step number)[a]	Source/Destination	Protocol service name	SCSI Protocol Service Interface procedure call
request (1)	device server to server agent	data-out delivery request	Receive data-out (IN (I_T_L_x nexus, application client buffer offset, request byte count, device server buffer))
data-out transfer (1a and 1b)	server agent to client consumer	RDMA data-out transfer	See 4x1.
confirmation (2)	sever agent to device server	data-out delivery confirmation	Data-out received (IN (I_T_L_x nexus))

[a] See figure A.3 for step number

A.5.3 Data-in delivery service

The data-in delivery service is a two step confirmed service (see figure A.3) that provides the means to transfer a parameter list or data from a device server to an SRP initiator port.

Processing the execute command procedure call for a data-in delivery service shall be composed of the two step confirmed service shown in table A.4.

Table A.4 - Processing of execute command procedure call for a data-in delivery service

Step (step number)[a]	Source to Destination	Protocol service name	SCSI Protocol Service Interface procedure call
request (1)	device server to server agent	data-in delivery request	Send data-in (IN (I_T_L_x nexus, device server buffer, application client buffer offset, request byte count))
data-in transfer (1a and 1b)	server agent to client consumer	RDMA data-in transfer	See 4.4.
confirmation (2)	sever agent to device server	data-in delivery confirmation	Data-In delivered (IN (I_T_L_x nexus))

[a] See figure A.3 for step number.

A.6 Task management services

A.6.1 Task management functions overview

The task management services shall be requested from the application client using a procedure call defined as:

Function name (IN (nexus), service response)

A.6.2 Task management functions

This standard handles task management functions as a four step confirmed service that provides the means to transfer task management functions to a task manager.

The task management functions are defined in the SAM-2. This standard defines the actions taken by the SRP services to carry out the requested task management functions.

A.6.3 ABORT TASK

The SRP services request the SRP initiator port issue an SRP_TSK_MGMT request (see 6.7) with a TASK MANAGEMENT FLAGS field set to indicate an ABORT TASK function to be sent to the selected SCSI device.

A.6.4 ABORT TASK SET

The SRP services request the SRP initiator port issue an SRP_TSK_MGMT request (see 6.7) with a TASK MANAGEMENT FLAGS field set to indicate an ABORT TASK SET function to be sent to the selected SCSI device.

A.6.5 CLEAR ACA

The SRP services request the SRP initiator port issue an SRP_TSK_MGMT request (see 6.7) with a TASK MANAGEMENT FLAGS field set to indicate a CLEAR ACA function to be sent to the selected SCSI device.

A.6.6 CLEAR TASK SET

The SRP services request the SRP initiator port issue an SRP_TSK_MGMT request (see 6.7) with a TASK MANAGEMENT FLAGS field set to indicate a CLEAR TASK SET function to be sent to the selected SCSI device.

A.6.7 LOGICAL UNIT RESET

The SRP services request the SRP initiator port issue an SRP_TSK_MGMT request (see 6.7) with a TASK MANAGEMENT FLAGS field set to indicate a LOGICAL UNIT RESET function to be sent to the selected SCSI device.

A.6.8 TARGET RESET

SRP does not support use of the TARGET RESET task management function.

A.6.9 WAKEUP

SRP does not support use of the WAKEUP task management function.

Annex B

(normative)

SRP for the InfiniBand™ Architecture

B.1 Overview

This annex specifies requirements for mapping SRP onto the InfiniBand™ Architecture, a transport that implements a superset of the RDMA communication service (see clause 4). See Infiniband™ Architecture Specification Volume 1 Release 1.0.a for a description of the InfiniBand™ Architecture.

B.2 Normative references

Infiniband™ Architecture Specification Volume 1 Release 1.0.a. Infiniband℠ Trade Association (www.infinibandta.org).

IETF RFC 2373, IP Version 6 Addressing Architecture. R. Hinden and S. Deering. Internet Engineering Task Force (www.ietf.org).

B.3 Definitions and abbreviations

B.3.1 Definitions

B.3.1.1 IB channel adapter: A device that terminates an InfiniBand™ Architecture link and executes transport-level functions. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

Bro002 **B.3.1.2 IB channel adapter GUID:** An IB GUID that uniquely identifies an IB channel adapter.

B.3.1.3 IB communication manager: The software, hardware, or combination of the two that supports the InfiniBand™ Architecture communication management mechanisms and protocols. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.4 IB consumer: An object that communicates with other IB consumers using the InfiniBand™ Architecture. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

edit001 **B.3.1.5 IB GID:** ~~A port address used for directing packets between IB subnets.~~ A~~n IB GID is a~~ 128-bit value that conforms to the IPv6 address format. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.6 IB GUID: A value that uniquely identifies a device or component. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.7 IB general service interface: An interface providing management services other than IB subnet management. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.8 IB I/O controller: The part of an IB I/O unit that provides I/O services. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

Bro003 **B.3.1.9 IB I/O controller GUID:** An IB GUID that uniquely identifies an IB I/O controller. This value is present as the GUID field of the IOControllerProfile attribute. (See Table B.7)

B.3.1.10 IB I/O unit: One or more IB I/O controllers attached to the IB fabric through a single IB channel adapter. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.11 IB LID: A port address used for directing IB packets within an IB subnet. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.12 IB MAD: An IB packet used to manage an InfiniBand™ Architecture network. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.13 IB packet: The indivisible unit of InfiniBand™ Architecture data transfer and routing, consisting of one or more headers, a packet payload, and one or two CRCs. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.14 IB port: A location on an IB channel adapter to which a link connects. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.15 IB port GUID: An IB GUID that uniquely identifies an IB port.

B.3.1.16 IB QP: An interface used for communication. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.17 IB service ID: A value that allows an IB communication manager to associate an incoming connection request with the entity providing the service. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.18 IB subnet: A set of IB ports connected via IB switches that have a common IB subnet ID and are managed by a common IB subnet manager. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.19 IB subnet manager: Entity that configures and controls an IB subnet. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

B.3.1.20 IPv6 address: A 128-bit address constructed in accordance with IETF RFC 2373 for Internet Protocol version 6. See IETF RFC 2373.

B.3.2 Abbreviations

CM:Ready to use	IB connection manager Ready to Use message
CM:Reject	IB connection manager Reject message
CM:Request	IB connection manager Request message
CM:Response	IB connection manager Response message
CRC	Cyclic Redundancy Check
GID	Global ID
GUID	Globally unique identifier
IB	InfiniBand™ Architecture
IPv6	Internet Protocol version 6
LID	Local ID
MAD	Management datagram
QP	Queue pair

B.4 InfiniBand™ Architecture overview

InfiniBand™ Architecture devices contain IB consumers and one or more IB channel adapters. Each IB channel adapter contains one or more IB ports. Associated with each IB channel adapter are IB QPs that interface between IB consumers and the IB channel adapter. Figure B.1 shows an example InfiniBand™ Architecture device.

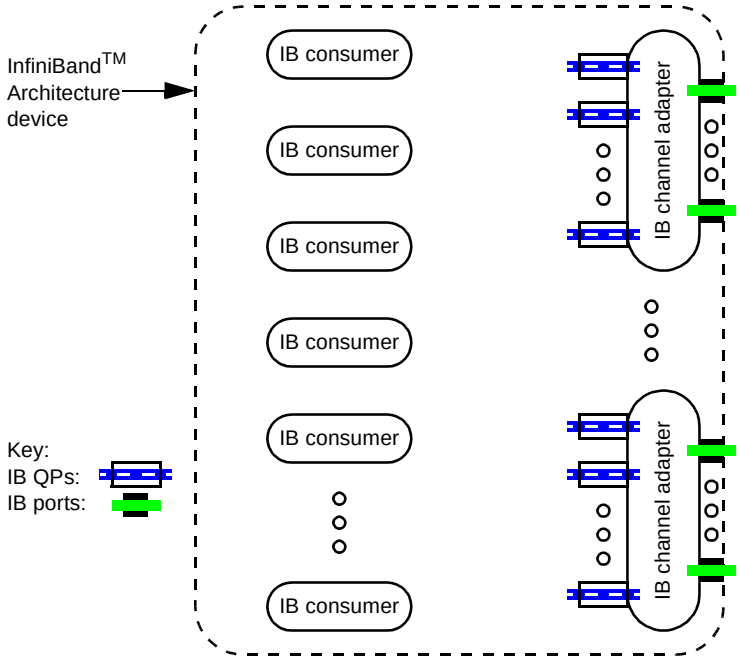


Figure B.1 - InfiniBand™ Architecture device example

An IB I/O unit is an InfiniBand™ Architecture device that contains an IB channel adapter with one or more IB ports, IB QPs, and one or more IB I/O controllers. Figure B.2 shows an example IB I/O unit.

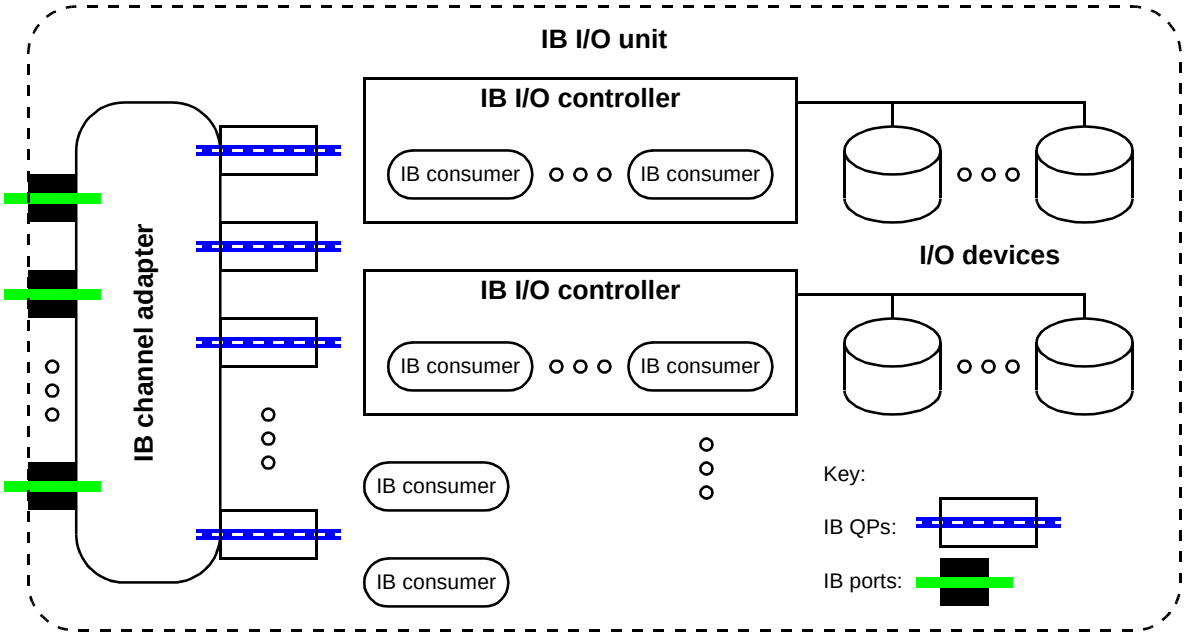


Figure B.2 - IB I/O unit example

Each IB port has a 64-bit globally unique identifier called an IB port GUID. Each IB channel adapter has a IB channel adapter GUID (which is shared by all IB ports on the IB channel adapter). Each IB I/O controller has an IB I/O controller GUID.

Bro004 The IB subnet manager assigns one or more IB LIDs and one or more IB GIDs to each IB port.

Table B.1 summarizes the InfiniBand™ Architecture names (IB GUIDs) and addresses (IDs) relevant to SRP.

Table B.1 - InfiniBand™ Architecture names and addresses

Name	Scope of uniqueness	Size	Description
IB port GUID	worldwide	64 bits	Identifies an IB port
IB channel adapter GUID	worldwide	64 bits	Identifies a IB channel adapter
IB I/O controller GUID	worldwide	64 bits	Identifies an IB I/O controller
IB LID	IB subnet	16 bits	Local routing address assigned to each IB port by the IB subnet manager
IB GID	varies[a]	128 bits	Address assigned by the IB subnet manager; (e.g., IB subnet prefix plus the IB port GUID)
[a)]Refer to Infiniband™ Architecture Specification Volume 1 Release 1.0.a			

HP05 HP05 edit003 Bro005

B.5 SCSI architecture mapping

Figure B.3 illustrates how SCSI initiator devices, SRP initiator ports, SRP target ports, and SCSI target devices map to InfiniBand™ Architecture objects. Bro006

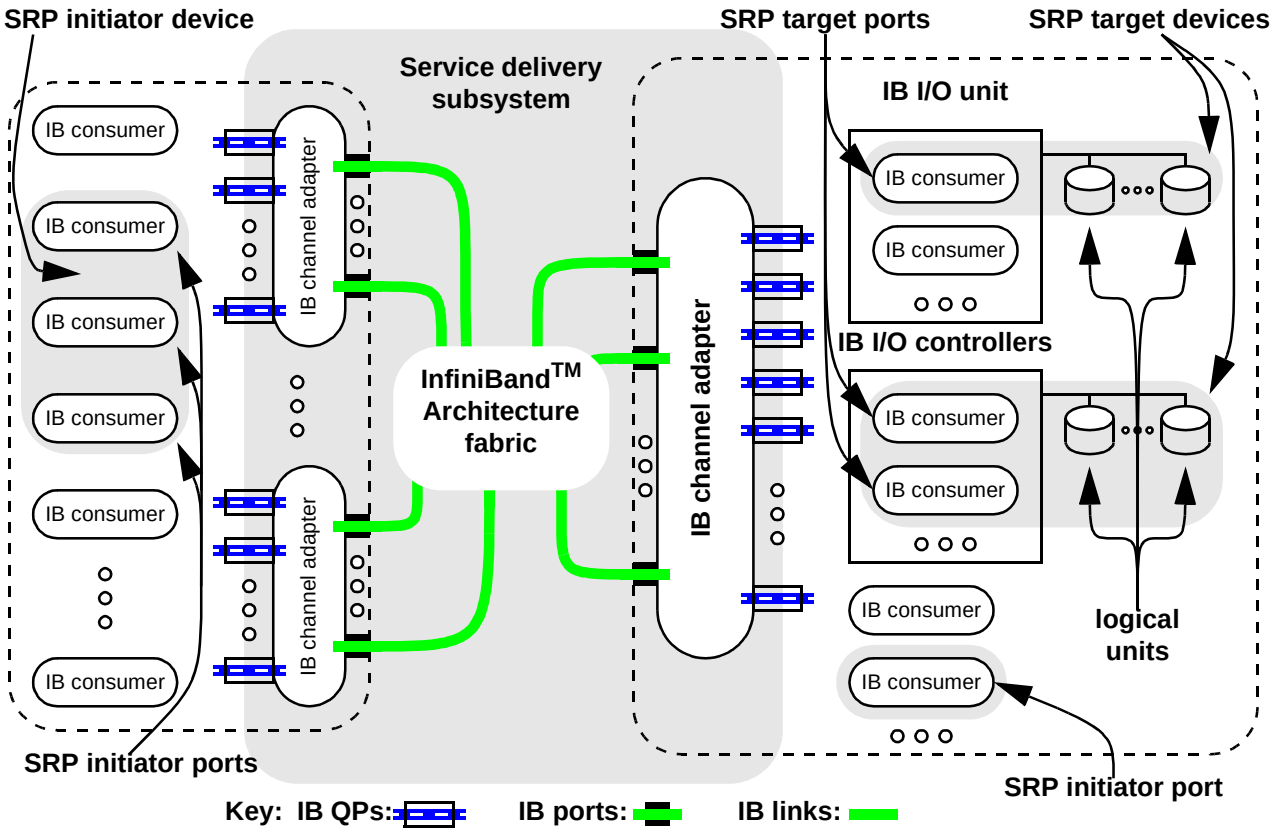


Figure B.3 - SCSI architecture mapping

IBM0136 An IB consumer in any InfiniBand™ Architecture device may be an SRP initiator port. An SRP initiator device is one or more IB consumers. The SRP initiator port identifier ~~should~~ shall be constructed as shown in table B.2.

Table B.2 - InfiniBand™ Architecture SRP initiator port identifier

Bit / Byte	7	6	5	4	3	2	1	0
0	(MSB)							
• • •				IDENTIFIER EXTENSION				
7								(LSB)
8	(MSB)							
• • •				GUID (e.g., IB channel adapter GUID)				
15								(LSB)

OD4

The IDENTIFIER EXTENSION field shall be chosen by the SRP initiator port to ensure that all SRP initiator port identifiers are unique.

Bro007 The GUID field should an IB GUID available to the SRP initiator port, e.g. the IB channel adapter GUID for an IB channel adapter used by the SRP initiator port.

Bro008 SRP target ports shall be implemented in IB I/O units. The IB I/O unit shall include a device management agent to provide IOUnit, IOController, and ServiceEntries attributes and make available an IB I/O controller GUID.

An SRP target port is indicated by a ServiceEntries attribute of an IB I/O controller. The SRP target port identifier shall be constructed as shown in table B.3.

Table B.3 - InfiniBand™ Architecture SRP target port identifier

Bit / Byte	7	6	5	4	3	2	1	0
0	(MSB)							
• • •				IDENTIFIER EXTENSION				
7								(LSB)
8	(MSB)							
• • •				IO CONTROLLER GUID				
15								(LSB)

OD4

The IDENTIFIER EXTENSION field shall be the value from the ServiceEntries attribute that indicates the SRP target port (see table B.8).

Bro003 The IO CONTROLLER GUID field shall be the IB I/O controller GUID of the IB I/O controller containing the SRP target port.

edit004 intel0160 The service delivery subsystem contains queue pairs, IB channel adapters, IB ports, and the InfiniBand™ Architecture fabric.

B.6 Communication management

B.6.1 Communication management overview

IB communications managers on each InfiniBand™ Architecture device manage InfiniBand™ Architecture connections using IB MADs transported over the IB general service interface. SRP initiator ports and SRP target ports shall use the active/passive (client/server) connection establishment protocol. The processor unit or IB I/O

Bro009 Bro010

controller containing the SRP target port shall act as the server and the processor unit or IB I/O controller containing the SRP initiator port shall act as the client.

B.6.2 Discovering SRP target ports

To discover the IB service ID of an SRP target port in an IB I/O unit, an SRP initiator port may use this sequence:

1. Retrieve the IOUnitInfo attribute from an IB I/O unit using a DevMgtGet IB MAD to determine the presence and slot number of each IB I/O controller attached to the IB I/O unit.
2. Retrieve the IOControllerProfile attributes from each IB I/O controller, each of which includes a ServiceEntries table.
3. Search the ServiceEntries table for service names matching the rules described in table B.8.

The IB service ID associated with each matching service name may be used in the communication management process to open InfiniBand™ Architecture connections to IB I/O controllers acting as SRP target ports. The SRP target port identifier for each SRP target port is constructed as described in table B.3. Bro011

B.6.3 Establishing a connection

To establish an InfiniBand™ Architecture connection, the client places the IB service ID in an IB communication management CM:Request message. The server associates the request with the appropriate SRP target port. The PrivateData field of the CM:Request message shall include an SRP_LOGIN_REQ request (see 6.2).

The SRP target port may choose to refuse the connection based on the SRP_LOGIN_REQ request content by returning a CM:Reject message with the reason code set to Consumer Reject. The PrivateData field of the CM:Reject message shall include an SRP_LOGIN_REJ response (see 6.4).

The SRP target port may choose to redirect the connection to a different endpoint (e.g. another IB port) by returning a CM:Reject message with the reason code set to either Port and CM Redirection or Port Redirection. The SRP initiator port should retry the connection establishment using the new endpoint. See Infiniband™ Architecture Specification Volume 1 Release 1.0.a.

If the server accepts the connection request and SRP login, the server returns a CM:Response message. The PrivateData field of the CM:Response message shall include an SRP_LOGIN_RSP response (see 6.3). The SRP initiator port may choose to refuse the connection based on the SRP_LOGIN_RSP response content by returning a CM:Reject message with a Reason code set to Consumer Reject. In this case, the PrivateData field of the CM:Reject message is reserved.

If the client accepts the connection reply and the SRP login response, it replies with a CM:Ready To Use message indicating both an InfiniBand™ Architecture and an SRP connection are open. It may start using the connection for communication.

B.6.4 Releasing a connection

Either the SRP initiator port or SRP target port may send an SRP_LOGOUT IU with a SEND operation. The sender shall send a CM disconnect request upon receipt of an InfiniBand™ Architecture transport level acknowledgement to the SRP_LOGOUT IU. The sender may disconnect if its send queue has transitioned to an error state. The receiver of an SRP_LOGOUT IU shall respond with an InfiniBand™ Architecture transport acknowledgement and disconnect.

B.6.5 Data-out and data-in operations

An SRP target port shall map a receive data-out SCSI protocol service interface procedure call to an InfiniBand™ Architecture RDMA READ Request. An SRP target port shall map a send data-in SCSI protocol

service interface procedure call to one or more InfiniBand™ Architecture RDMA WRITE packets. Table B.4 specifies the value of the InfiniBand™ Architecture RDMA header fields.

Table B.4 - InfiniBand™ Architecture RDMA header fields

InfiniBand™ Architecture RDMA Extended Transport Header field	Value
Virtual Address	VIRTUAL ADDRESS[a] + application client buffer offset[b]
Remote Key	MEMORY HANDLE[c]
DMA Length	request byte count[d]

[a] The contents of the VIRTUAL ADDRESS field in the memory descriptor (see table 1).

[b] The application client buffer offset parameter to the receive data-out (see table A.3) or send data-in (see table A.4) SCSI protocol service interface procedure call.

[c] The contents of the MEMORY HANDLE field in the memory descriptor (see table 1).

[d] The request byte count parameter to the receive data-out (see table A.3) or send data-in (see table A.4) SCSI protocol service interface procedure call.

B.7 InfiniBand™ Architecture protocol requirements

SRP target ports and SRP initiator ports shall support the Reliable Connection transport service type.

SRP target ports shall implement the device management class of general management services.

SRP initiator ports and SRP target ports shall support the transport functions described in table B.5.

Table B.5 - Transport operation support requirements

Transport functions	SRP initiator port	SRP target port
Send to	Mandatory	Mandatory
Send from	Mandatory	Mandatory
RDMA write to	Mandatory	Not used
RDMA write from	Not used	Mandatory
RDMA read to	Mandatory for data-outcommands	Not used
RDMA read from	Not used	Mandatory for data-out commands
RDMA Write with immediate data (to or from)	Not used	Not used
ATOMIC (to or from)	Not used	Not used

IB I/O units containing an IB I/O controller acting as an SRP target port shall report the device management IOUnit attributes defined in Infiniband™ Architecture Specification Volume 1 Release 1.0.a as described in table B.6.

Table B.6 - IOUnit attributes for SRP target ports

Field	SRP requirement
Change ID	No requirement
Max Controllers	At least one
Option ROM	No requirement
Controller List	At least one IB I/O controller acting as an SRP target port shall be present

edit002

Bro102

IB I/O controllers acting as SRP target ports shall report the device management IOControllerProfile attributes defined in Infiniband™ Architecture Specification Volume 1 Release 1.0.a as described in table B.7.

Table B.7 - IOControllerProfile attributes for SRP target ports

Field	SRP requirement
GUID	No requirement
Device ID	No requirement
Vendor ID	No requirement
Device Version	No requirement
Subsystem Vendor ID	No requirement
Subsystem ID	No requirement
I/O Class	0100h
I/O Subclass	609Eh
Protocol	0108h
Protocol Version	0001h
Service Connections	At least one
Initiators Supported	At least one
Send Message Depth	Reserved
RDMA Read Depth	Maximum IOC-issued RDMA depth[a]
Send Message Size	MAXIMUM INITIATOR TO TARGET IU SIZE[b]
RDMA Transfer Size	Reserved
Controller Operations Capability Mask:	
0: ST; Send Messages To IOCs	Shall be set to one.
1: SF; Send Messages From IOCs	Shall be set to one.
2: RT; RDMA Read Requests To IOCs	No requirement
3: RF; RDMA Read Requests From IOCs	Shall be set to one if an SRP target port supports data-out commands. No requirement otherwise.
4: WT; RDMA Write Requests To IOCs	No requirement
5: WF; RDMA Write Requests From IOCs	Shall be set to one.
6: AT; Atomic Operations To IOCs	No requirement
7: AF; Atomic Operations From IOCs	No requirement
Controller Services Capability Mask	Reserved[a]
Service Entries	At least one
ID String	No requirement

[a] The largest number of RDMA Read requests that this IO Controller may have outstanding on one channel.

[b] This value shall be no less than the largest value, in bytes, of MAXIMUM INITIATOR TO TARGET IU SIZE that this IO Controller shall return in the SRP_LOGIN_RSP information unit.

IBTA

intel0176

intel0174

intel0175

An IB I/O controller acting as an SRP target port shall register with its IB communications manager, specifying one of the service name strings described in table B.8. This string is assigned an “IO SERVICE ID” type IB service ID by the IB communications manager. Bro012

IB I/O controllers acting as SRP target ports shall include at least one ServiceName/ServiceID pair in the device management ServiceEntries attribute pair defined in Infiniband™ Architecture Specification Volume 1 Release 1.0.a as described in table B.8. Bro013

Table B.8 - ServiceEntries attribute pair for SRP target ports

Field	Length (bits)	SRP requirement
ServiceName_n	320	’SRP.T10:xxxxxxxxxxxxxxxx’ or ’SRP.T10:xxxxxxxxxxxxxxxx:reserved’
ServiceID_n	64	Assigned by the IB I/O controller

[a] A service name shall be recognized as identifying an SRP target port if and only if it satisfies all of the rules described in this table.

[b] The string ’SRP.T10’ and the colons shall appear exactly as shown (e.g. capital letters only).

[c] The string ’xxxxxxxxxxxxxxxx’ in the service name shall be sixteen hexadecimal digits. Only the characters 0 to 9 and A to F (capital letters only) are permitted. If any other character appears the service name shall not be recognized as identifying an SRP target port.

[d] The string ’xxxxxxxxxxxxxxxx’ in the service name identifies the 64-bit extension identifier value used to construct the SRP target port identifier (see table B.3)

[e] The string ’reserved’ shall either be ignored by SRP initiator ports or treated in accordance with a future revision of this standard.

[f] If the service name does not completely fill ServiceName_n field (i.e. it is less than 40 bytes), it shall be padded with null characters (binary zeros).