
American National Standard for Information Systems - Information Technology – SCSI RDMA Protocol

This is an internal working document of T10, a Technical Committee of Accredited Standards Committee NCITS (National Committee for Information Technology Standards). As such this is not a completed standard and has not been approved. The contents may be modified by the T10 Technical Committee. The contents are actively being modified by T10. This document is made available for review and comment only.

Permission is granted to members of NCITS, its technical committees, and their associated task groups to reproduce this document for the purposes of NCITS standardization activities without further permission, provided this notice is included. All other rights are reserved. Any duplication of this document for commercial or for-profit use is strictly prohibited.

T10 Technical editor: Edward A. Gardner
 Ophidian Designs
 1262 Hofstead Terrace
 Colorado Springs, CO 80907-4011
 USA

 Telephone: +1.719.593.8866
 Facsimile: +1.719.593.8866
 Email: eag@ophidian.com

Reference number
ANSI NCITS.***:200x

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

NCITS Secretariat

NCITS Secretariat
1250 Eye Street, NW Suite 200
Washington, DC 20005

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

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

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: gpenokie@tivoli.com

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

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

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

ABSTRACT

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

PATENT STATEMENT

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.

Revision History

Work remaining:

- a) [01-177] RDMA communication model;
- b) [01-172] SRP to SAM-2 mapping; and
- c) [01-028] SRP to Infiniband™ mapping.

Under consideration:

- d) Discussion of target / initiator port identifier structure

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.

Contents

Foreword	9
Introduction	10
1 Scope	11
2 Normative references	13
2.1 Approved references	13
2.2 References under development	13
3 Definitions, symbols, abbreviations and conventions	14
3.1 Definitions	14
3.2 Acronyms	14
3.3 Keywords	15
3.4 Conventions	15
3.5 Notation for procedures and functions	16
4 Structure and concepts	17
4.1 RDMA communication model	17
4.2 Information unit classes	17
4.3 SRP request flow control	17
4.4 Data buffers	18
4.4.1 Memory descriptors	18
4.4.2 Data buffer descriptors	19
5 SRP Information Units	23
5.1 Summary	23
5.10 SRP_TPAR_RSP response	38
5.12 SRP_AER_RSP response	40
6 SCSI mode parameters	41
6.1 SCSI mode parameter overview and codes	41
6.2 Disconnect-reconnect mode page	41
6.3 Logical unit control mode page	42
6.4 Port control mode page	42
Annex A SCSI RDMA protocol services	43
Annex B Alias entry designation formats	45

Tables

Table 1 - Memory descriptor	18
Table 2 - Data buffer descriptor formats	19
Table 3 - Indirect data buffer descriptor	20
Table 4 - SRP requests sent from initiators to targets	23
Table 5 - SRP responses sent from targets to initiators.	23
Table 6 - SRP requests sent from targets to initiators	23
Table 7 - SRP responses sent from initiators to targets.	23
Table 8 - SRP_LOGIN_REQ request	24
Table 9 - MULTI-CHANNEL ACTION code values	25
Table 10 - SRP_LOGIN_RSP response	26
Table 11 - MULTI-CHANNEL RESULT code values	27
Table 12 - SRP_LOGIN_REJ response.	28
Table 13 - SRP_LOGIN_REJ REASON codes.	28
Table 14 - SRP_LOGOUT request	29
Table 15 - SRP_LOGOUT REASON codes	29
Table 16 - SRP_TASK_MGMT request.	30
Table 17 - TASK MANAGEMENT FLAGS	31
Table 18 - SRP_CMD request	32
Table 19 - TASK ATTRIBUTE	33
Table 20 - SRP_RSP response	34
Table 21 - RESPONSE DATA field.	37
Table 22 - RSP_CODE values	37
Table 23 - SRP_TPAR_REQ request	38
Table 24 - SRP_TPAR_RSP response	38
Table 25 - SRP_AER_REQ request	39
Table 26 - SRP_AER_RSP response	40
Table 27 - SRP mode page codes.	41
Table 28 - Disconnect-reconnect mode page	41

Figures

Figure 1 - SCSI document relationships	11
Figure 2 - Memory descriptor mapping	18
Figure 3 - Indirect data buffer mapping without cached memory descriptors	21
Figure 4 - Indirect data buffer mapping with cached memory descriptors.	22

Foreword

This foreword is not part of American National Standard X3.269-199x.

Suggestions for improvement, requests for interpretation, addenda, or defect reports are welcome. They should be sent to the NCITS Secretariat, Information Technology Industrial Council (ITI), 1250 Eye Street, NW, Suite 200, Washington, DC 20005.

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

Karen Higginbottom, Chair

David Michael, Vice-chair

Monica Vago, Secretary

(NCITS Membership to be inserted)

NCITS 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 RDMA channels allows the large body of SCSI application and driver software to be successfully used on Infiniband™, the VI Architecture and other interfaces that support RDMA channel semantics.

The SCSI RDMA Protocol (SRP) standard is divided into 6 major clauses:

Clause 1 is the scope.

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

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

Clause 4 describes significant concepts of the SCSI RDMA Protocol.

Clause 5 describes the information units used to transfer SCSI commands, status and channel control information across RDMA channels.

Clause 6 defines the SCSI management features for SRP, including the SRP mode pages.

The SCSI RDMA Protocol (SRP) standard has **TBD** annexes.

Annex A is **TBD**

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

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.

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

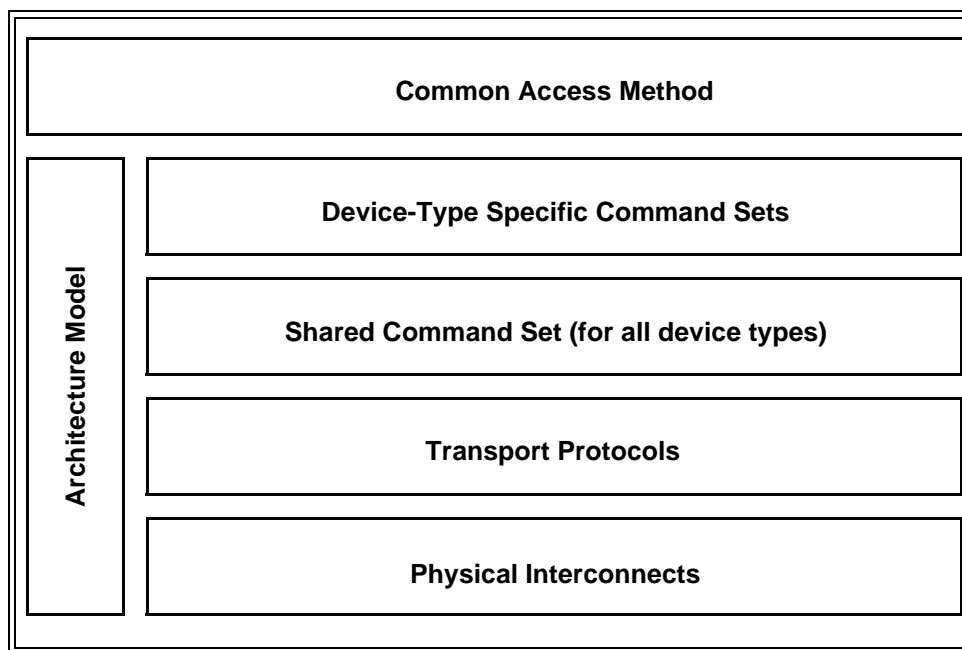


Figure 1 - SCSI document relationships

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.

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]

Fibre Channel Physical and Signalling Interface	FC-PH	[ISO/IEC 14165-111] [ANSI X3.230:1994]
Fiber Channel Physical Amendment 1		[ANSI X3.230/AM1:1996]
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]
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]
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]
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:		
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]
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] [ANSI X3.232:1996]

The term SCSI is used to refer to the family of standards described in this clause.

2 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.

2.1 Approved references

2.2 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) standard [T10/1157-D]

ISO/IEC 14776-312, SCSI Primary Commands - 2 (SPC-2) standard [T10/1236-D]

3 Definitions, symbols, abbreviations and conventions

3.1 Definitions

3.1.1 application client: An object that is the source of SCSI commands. Further definition of an application client may be found in SAM-2.

3.1.2 byte: An 8-bit construct.

3.1.3 command: A request describing a unit of work to be performed by a device server. A detailed definition of a command may be found in SAM-2.

3.1.4 command descriptor block (CDB): The structure used to communicate commands from an application client to a device server. A detailed definition of a command descriptor block may be found in SPC-2.

3.1.5 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).

3.1.6 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).

3.1.7 device server: An object within a logical unit that executes SCSI tasks according to the rules of task management. A detailed definition of a device server may be found in SAM-2.

3.1.8 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.

3.1.9 initiator: A SCSI device containing application clients that originate device service requests to be processed in a device server. A detailed definition of an initiator may be found in SAM-2.

3.1.10 initiator port identifier: A value by which a SCSI initiator port is referenced within a domain (see SAM-2).

3.1.11 logical unit: An externally addressable entity within a target that implements a SCSI device model and contains a device server. A detailed definition of a logical unit may be found in SAM-2.

3.1.12 logical unit number (LUN): An encoded 64-bit identifier for a logical unit. A detailed definition of a logical unit number may be found in SAM-2.

3.1.13 status: One byte of response information sent from a device server to an application client upon completion of each command. A detailed definition of status may be found in SAM-2.

3.1.14 target: A SCSI device that receives SCSI commands and directs such commands to one or more logical units. A detailed definition of a target may be found in SAM-2.

3.1.15 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 3.1.4)
LSB	Least significant bit
LUN	Logical Unit Number (see 3.1.12)
MSB	Most significant bit
NCITS	National Committee for Information Technology Standards
RDMA	Remote Direct Memory Access
SAM-2	SCSI Architecture Model - 2 (see 2.2)

SCSI	The architecture defined by the family of standards described in clause 1
SPC-2	SCSI Primary Commands - 2 (see 2.2)
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 error.

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.

4 Structure and concepts

4.1 RDMA communication model

EDITOR'S NOTE 1 - See document T10/01-177.

4.2 Information unit classes

Each SRP information unit is classified as a SRP request or a SRP response (see 5.1). 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. RDMA channel management requests may be issued by SRP targets as well as SRP initiators.

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.

The following list describes all circumstances where an SRP response shall not be returned for an SRP request:

- a) If an SRP request conveys a SCSI command and SAM-2 specifies that STATUS is not returned for that command, then an SRP response shall not be returned for the SRP request;
- b) An SRP response shall not be returned for SRP_LOGOUT; and
- c) If an SRP device becomes aware of a failure preventing further communication on an RDMA channel, it shall abort all outstanding SRP requests received on that RDMA channel without returning SRP responses.

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.

4.3 SRP request flow control

SRP request flow control allows an SRP device to limit the number of SRP requests that may be sent to it on an RDMA channel. SRP devices may use SRP request flow control to manage internal and RDMA channel related resources.

SRP responses are not subject to flow control; they may be sent at any time. An SRP device may limit the number of SRP responses it might receive by limiting the number of SRP requests it has outstanding.

SRP targets shall limit themselves to at most one outstanding SRP request per RDMA channel. Upon sending an SRP request, an SRP target shall not send another SRP request on the same RDMA channel until after it receives the SRP response for the previous SRP request.

SRP uses a credit based flow control algorithm to limit the number of SRP requests that an SRP initiator may send to a target. The algorithm uses a field, REQUEST LIMIT DELTA, that is present in most information units sent by an SRP target to an SRP initiator, 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 initiators shall implement a separate copy of REQUEST LIMIT for each RDMA channel;
- b) Upon successful completion of RDMA channel establishment an SRP initiator shall initialize the RDMA channel's REQUEST LIMIT to the value of REQUEST LIMIT DELTA received in SRP_LOGIN_RSP. Except for providing SRP_LOGIN_REQ when requesting RDMA channel establishment, the SRP initiator shall not send any SRP information units on the RDMA channel prior to initializing REQUEST LIMIT;
- c) An SRP initiator 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 shall not send an SRP request on 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 initiator may choose to send commands only when REQUEST LIMIT is greater than one;

- d) An SRP initiator shall decrement an RDMA channel's REQUEST LIMIT by one whenever it sends an SRP request on that RDMA channel;
- e) An SRP initiator 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
- f) An SRP target shall not specify a positive value of REQUEST LIMIT DELTA that might cause REQUEST LIMIT to exceed $2^{31}-1$. An SRP target shall not specify a negative value of REQUEST LIMIT DELTA that might cause REQUEST LIMIT to drop below -2^{31} . A target shall account for all possible race conditions to meet these requirements.

4.4 Data buffers

4.4.1 Memory descriptors

A memory descriptor is a 16-byte structure that identifies a memory segment (see table 1). Figure 2 illustrates the mapping of a memory descriptor to a memory segment.

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)							

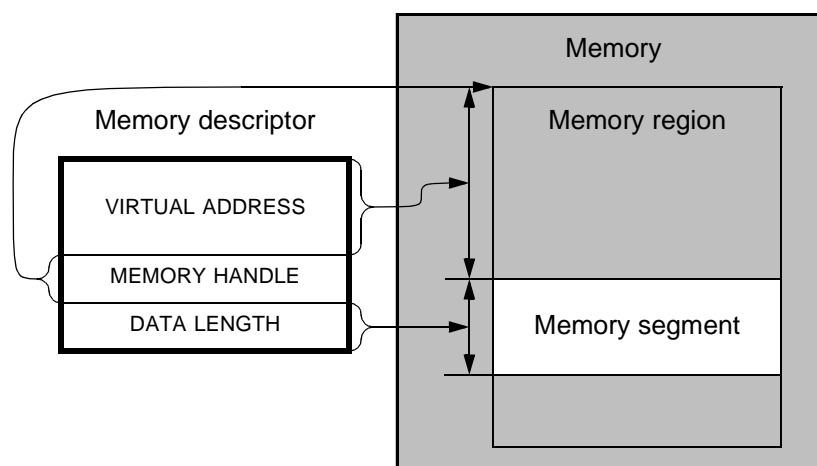


Figure 2 - Memory descriptor mapping

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.

The MEMORY HANDLE field identifies the region that contains the memory segment. The target shall supply this value with any RDMA operation that accesses the memory segment. The 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 target may use a memory descriptor for either RDMA Read operations or RDMA Write operations but not both. SRP targets 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:

- 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.

4.4.2 Data buffer descriptors

4.4.2.1 Overview

An SRP_CMD request (see 5.7) 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. For each buffer descriptor a bit, DOIND or DIIND, indicates whether the buffer descriptor is direct or indirect. A count field, DOCOUNT or DICOOUNT, indicates the number of memory descriptors present in the SRP_CMD request. DOIND and DOCOUNT define the length and format of a data-out buffer descriptor. DIIND and DICOOUNT define the length and format of a data-in buffer descriptor. Table 2 lists the permissible combinations and the resulting data buffer descriptor length and format.

Table 2 - Data buffer descriptor formats

DOIND DIIND	DOCOUNT DICOOUNT	Length (bytes)	Data buffer descriptor format
0	0	0	Buffer descriptor not present.
0	1	16	Direct data buffer descriptor (see 4.4.2.2).
0	$m > 1$		Reserved.
1	0		Reserved.
1	1	20	Indirect data buffer descriptor without cached memory descriptors (see 4.4.2.3).
1	$m > 1$	$16*m+4$	Indirect data buffer descriptor with $m-1$ cached memory descriptors (see 4.4.2.3).

Support for indirect data buffers is optional. SRP targets indicate whether they support indirect data buffers during RDMA channel establishment.

MAXIMUM INITIATOR TO TARGET IU, negotiated during RDMA channel establishment (see 5.3), imposes an implicit limit on the length of an indirect data buffer descriptor. The actual limit for any particular indirect data buffer descriptor also depends upon the CDB length and the number of buffer descriptors required by the command.

4.4.2.2 Direct data buffers

The presence of a direct data buffer descriptor is indicated by DOIND or DIIND containing 0 and DOCOUNT or DICOUNT containing 1. A direct data buffer descriptor contains a single memory descriptor (see table 1). The memory descriptor identifies the data buffer, which is a single contiguous memory segment within a memory region's virtual address space. If a direct data buffer descriptor defines a data-out buffer, the SRP target 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 shall only issue RDMA Write operations using the memory descriptor contained in the direct data buffer descriptor. The SRP target shall use the DATA LENGTH field of the memory descriptor as the length of the data-out buffer or data-in buffer.

4.4.2.3 Indirect data buffers

The presence of an indirect data buffer descriptor is indicated by DOIND or DIIND containing 1 and DOCOUNT or DICOUNT containing 1 or larger. An indirect data buffer contains one or more memory segments, which may or may not be contiguous. The memory segments may be in a single memory region or spread among several memory regions. The data buffer is the concatenation of the memory segments. Each memory segment may have any length, including a length of zero bytes. Table 3 shows the format of an indirect data buffer descriptor.

Table 3 - 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									
20		CACHED MEMORY DESCRIPTOR LIST							
...									
19+16*m									

INDIRECT TABLE MEMORY DESCRIPTOR identifies a memory segment that contains an indirect table. An indirect table is a list of one or more memory descriptors. The memory segments identified by the memory descriptors in the indirect table, concatenated together, comprise the indirect data buffer. The DATA LENGTH field of INDIRECT TABLE MEMORY DESCRIPTOR contains the number of memory descriptors in the indirect table times 16. Target behavior when INDIRECT TABLE MEMORY DESCRIPTOR contains any other value is vendor specific.

TOTAL LENGTH contains the sum of the DATA LENGTH fields of the memory descriptors in the indirect table. Target behavior when TOTAL LENGTH contains any other value is vendor specific. The target shall use either the TOTAL LENGTH field or the sum of the DATA LENGTH fields as the length of the data-out buffer or data-in buffer.

CACHED MEMORY DESCRIPTOR LIST is only present when DOCOUNT or DICOUNT contain a value larger than 1. It is a list of m-1 memory descriptors, where m is the value contained in DOCOUNT or DICOUNT. CACHED MEMORY DESCRIPTOR LIST contains copies of the first m-1 memory descriptors in the indirect table. Target behavior when CACHED MEMORY DESCRIPTOR LIST contains any other value is vendor specific.

An SRP target shall only issue RDMA Read operations to the indirect table.

If an indirect data buffer descriptor defines a data-out buffer, the SRP target shall only issue RDMA Read operations using the memory descriptors contained in the indirect table or the CACHED MEMORY DESCRIPTOR LIST.

If an indirect data buffer defines a data-in buffer, the SRP target shall only issue RDMA Write operations using the memory descriptors contained in the indirect table or the CACHED MEMORY DESCRIPTOR LIST.

Figure 3 illustrates an indirect data buffer without cached memory descriptors. 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 4.4.1 and figure 2. Figure 3 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.

In this example DOIND or DIIND would contain 1 and DOCOUNT or DICOUNT would contain 1. The 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 identifying memory segments 1 through 3. The INDIRECT TABLE MEMORY DESCRIPTOR field of the data buffer descriptor identifies the memory segment containing the indirect table. The DATA LENGTH field of INDIRECT TABLE MEMORY DESCRIPTOR would contain 192, the length of the indirect table. The TOTAL LENGTH field of the data buffer descriptor would contain the sum of the DATA LENGTH fields of the memory descriptors in the indirect table (the sum of DATA LENGTH 1, DATA LENGTH 2 and DATA LENGTH 3). This sum is the total length of the data buffer.

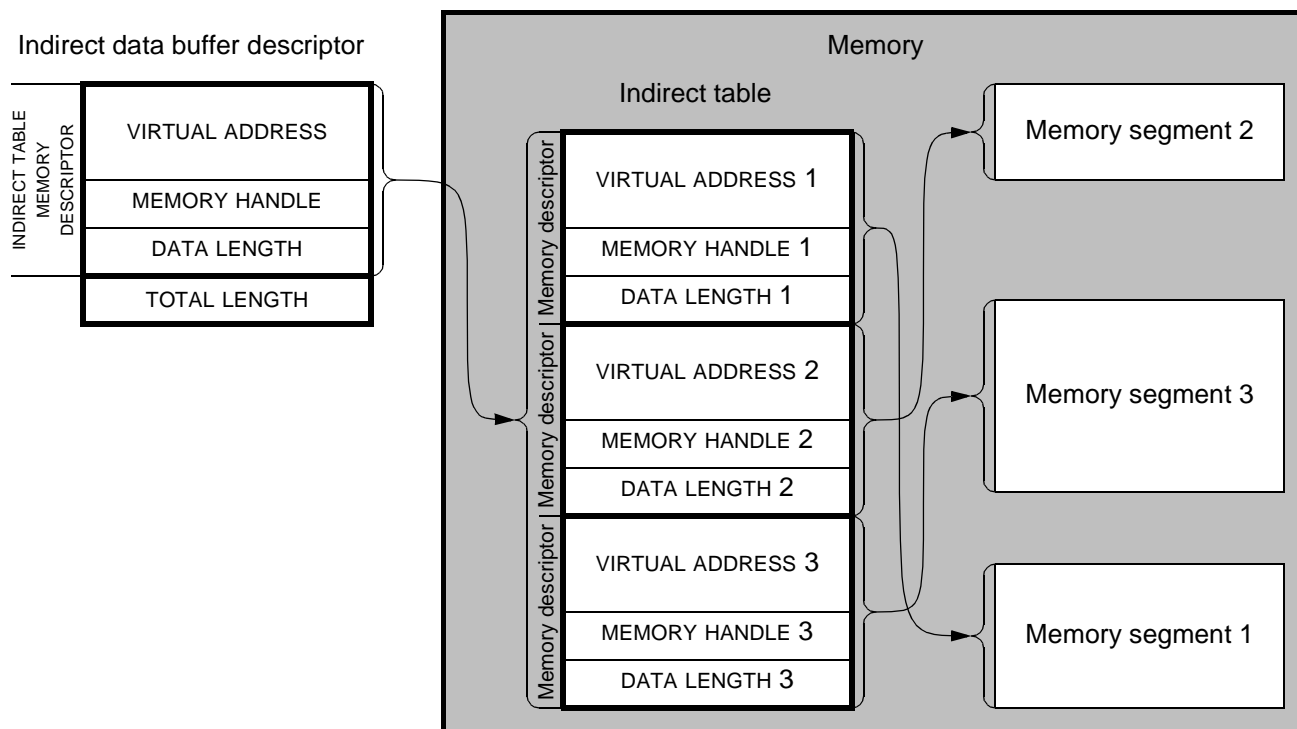


Figure 3 - Indirect data buffer mapping without cached memory descriptors

Figure 4 illustrates the same example except with cached memory descriptors. The data buffer, indirect table, INDIRECT TABLE MEMORY DESCRIPTOR and TOTAL LENGTH are all identical to the previous example. DOIND and DIIND would still contain 1, but DOCOUNT and DICOOUNT would now contain 3, indicating that CACHED MEMORY DESCRIPTOR LIST contains two memory descriptors. These are identical to the first two memory descriptors in the indirect table. The third memory descriptor is only present in the indirect table.

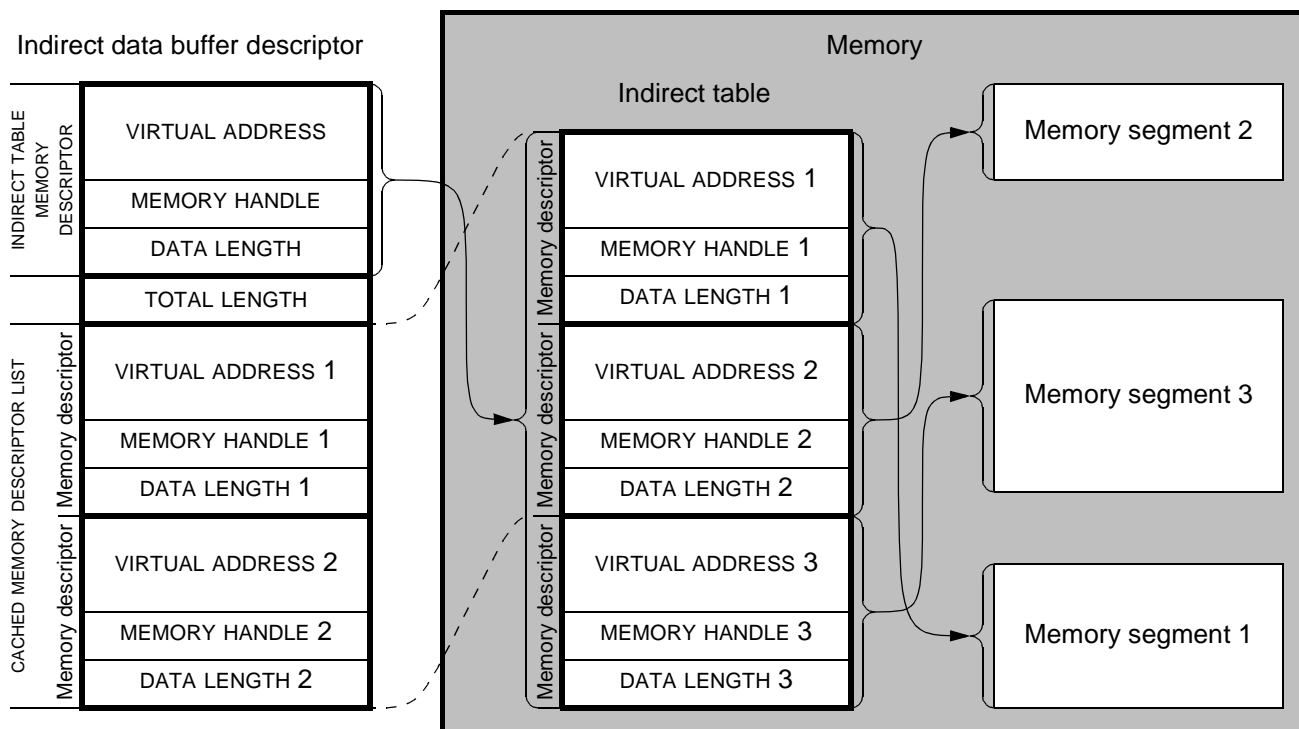


Figure 4 - Indirect data buffer mapping with cached memory descriptors

5 SRP Information Units

5.1 Summary

The information units used by SRP and their characteristics are shown in table 4, table 5, table 6 and table 7.

Table 4 - SRP requests sent from initiators to targets

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_LOGIN_REQ	5.2	00h	64	Login request
SRP_TSK_MGMT	5.6	01h	64	SCSI task management function
SRP_CMD	5.7	02h	48 minimum	SCSI command

Table 5 - SRP responses sent from targets to initiators

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_LOGIN_RSP	5.3	C0h	52	Login response
SRP_RSP	5.8	C1h	36 minimum	SCSI status or service response
SRP_LOGIN_REJ	5.4	C2h\	32	Login failure

Table 6 - SRP requests sent from targets to initiators

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_LOGOUT	5.5	80h	16	Logout or channel failure notification
SRP_TPAR_REQ	5.9	81h	52	Target parameter request
SRP_AER_REQ	5.11	82h	56 minimum	Asynchronous event report request

Table 7 - SRP responses sent from initiators to targets

Information unit	Reference	TYPE value	Length (bytes)	Description
SRP_TPAR_RSP	5.10	41h	64	Response to target parameter request
SRP_AER_RSP	5.12	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 a target receives an SRP information unit with an invalid TYPE code, or whose length is incorrect for the information unit's type code, the target shall send an SRP_LOGOUT and disconnect the RDMA channel.

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.

5.2 SRP_LOGIN_REQ request

An SRP_LOGIN_REQ request (see table 8) conveys SRP protocol login parameters from an SRP initiator to an SRP target. The SRP_LOGIN_REQ request shall only be sent during channel establishment.

Table 8 - 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							
...									
19	(LSB)								
20	Reserved								
...									
23									
24	Reserved							INDREQ	
25	Reserved								
26	Reserved						MULTI-CHANNEL ACTION		
27	Reserved								
28	Reserved								
...									
31									
32	INITIATOR PORT IDENTIFIER								
...									
47									
48	TARGET PORT IDENTIFIER								
...									
63									

The TAG field is defined in 5.1.

REQUESTED MAXIMUM INITIATOR TO TARGET IU specifies the maximum length of any information unit that the initiator wishes to send on this channel. This value shall be 64 or larger.

INDREQ shall be set to zero if the initiator will not use indirect data buffer descriptors on this channel. INDREQ shall be set to one if the initiator may make use of target support for indirect data buffer descriptors.

MULTI-CHANNEL ACTION identifies how the target shall treat any existing RDMA channel associated with the same I_T nexus specified by INITIATOR PORT IDENTIFIER and TARGET PORT IDENTIFIER. Table 9 defines this field.

Table 9 - MULTI-CHANNEL ACTION code values

MULTI-CHANNEL ACTION	Description
00h	Terminate existing channels before attempting to establish the new channel. For each existing channel associated with the same I_T nexus, abort all outstanding tasks received on that channel and send an SRP_LOGOUT.
01h	Independent operation. Allow any existing channel to continue operation independent of the new channel.
02h	Reserved.
03h	Reserved

INITIATOR PORT IDENTIFIER and TARGET PORT IDENTIFIER specify the I_T nexus that shall be associated with this channel.

5.3 SRP_LOGIN_RSP response

An SRP_LOGIN_RSP information unit (see table 10) conveys SRP protocol login parameters from an SRP target to an SRP initiator. The SRP_LOGIN_RSP shall only be sent to indicate successful channel establishment.

Table 10 - 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							
...									
19									(LSB)
20	(MSB)	MAXIMUM TARGET TO INITIATOR IU							
...									
23									(LSB)
24	Reserved							INDSUP	
25	Reserved								
26	Reserved						MULTI-CHANNEL RESULT		
27	Reserved								
28	Reserved								
...									
51									

The REQUEST LIMIT DELTA field is defined in 4.3.

The TAG field shall contain the same value as the TAG field in the SRP_LOGIN_REQ request.

MAXIMUM INITIATOR TO TARGET IU specifies the maximum length of any information unit that the target is able to receive on this channel. This value shall be 64 or larger and greater than or equal to the value of REQUESTED MAXIMUM INITIATOR TO TARGET IU specified in the SRP_LOGIN_REQ request. The initiator shall not send any information unit on this channel longer than this value.

MAXIMUM TARGET TO INITIATOR IU specifies the maximum length of any information unit that the target may send on this channel. This value shall be 52 or larger. The target shall not send any information unit on this channel longer than this value.

INDSUP shall be set to zero if the target does not support indirect data buffer descriptors on this channel. INDSUP shall be set to one if the target does support indirect data buffer descriptors on this channel. The target shall set INDSUP to zero if the value of INDREQ in the SRP_LOGIN_REQ was zero.

MULTI-CHANNEL RESULT identifies how the target treated existing RDMA channels associated with the same I_T nexus. Table 11 defines this field.

Table 11 - MULTI-CHANNEL RESULT code values

MULTI-CHANNEL RESULT	Description
00h	No existing channels were associated with the same I_T nexus.
01h	One or more existing channels were terminated.
02h	One or more existing channels continue to operate independently.
03h	Reserved

5.4 SRP_LOGIN_REJ response

An SRP_LOGIN_REJ response (see table 12) is sent by a target to notify the initiator that an RDMA channel could not be established.

Table 12 - SRP_LOGIN_REJ response

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (80h)							
1								
2	Reserved							
3								
4								
...	REASON							
7								
8								
...	TAG							
15								
16								
...	Reserved							
31								

The REASON field indicates the reason that the RDMA channel could not be established. This field is defined in table 13.

Table 13 - SRP_LOGIN_REJ REASON codes

REASON code	Description
0001 0000h	Unable to establish RDMA channel, no reason specified.
0001 0001h	Insufficient channel resources.
0001 0002h	REQUESTED MAXIMUM INITIATOR TO TARGET IU value too large.
0001 0003h	Unable to associate RDMA channel with specified I_T nexus.
0001 0004h	Indirect data buffer descriptors not supported.
0001 0005h	Target does not support multiple channels per I_T nexus.
all other values	reserved

The TAG field shall contain the same value as the TAG field in the SRP_LOGIN_REQ request.

5.5 SRP_LOGOUT request

An SRP_LOGOUT request (see table 14) is sent by a target to notify the initiator that the target is disconnecting the RDMA channel. An SRP_LOGOUT request may also be used to notify the initiator that an RDMA channel has failed, rendering it non-operational.

Table 14 - SRP_LOGOUT request

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (80h)							
1								
2	Reserved							
3								
4	(MSB)	REASON						
...								
7								(LSB)
8	(MSB)	TAG						
...								
15								(LSB)

The REASON field indicates the reason for disconnecting the RDMA channel. This field is defined in table 15.

Table 15 - SRP_LOGOUT 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 target.
0000 0003h	Valid response type code with no corresponding target request outstanding.
0000 0004h	Channel terminated due to MULTI-CHANNEL ACTION code in new SRP_LOGIN_REQ.
all other values	Reserved

The TAG field is defined in 5.1.

After sending an SRP_LOGOUT request a target may delay a vendor specific time to allow the SRP_LOGOUT request to be delivered to the initiator. The target shall then disconnect the RDMA channel. The target shall not send any further information units on the RDMA channel nor shall it request any RDMA operations after sending an SRP_LOGOUT. The target shall discard all information units it may receive on the RDMA channel after sending an SRP_LOGOUT.

After receiving an SRP_LOGOUT request an initiator shall disconnect the RDMA channel. The initiator shall not send any information units on the RDMA channel after receiving an SRP_LOGOUT. The initiator shall not send an SRP response to an SRP_LOGOUT request.

5.6 SRP_TASK_MGMT request

An SRP_TASK_MGMT information unit conveys a SCSI task management request (table 16).

Table 16 - SRP_TASK_MGMT request

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (01h)							
1	Reserved							
...								
7								
8	(MSB)	TAG						
...								
15	(LSB)							
16	Reserved							
...								
19								
20	(MSB)	LOGICAL UNIT NUMBER						
...								
27	(LSB)							
28	Reserved							
29	Reserved							
30	TASK MANAGEMENT FLAGS							
31	Reserved							
32	(MSB)	TAG OF TASK TO BE MANAGED						
...								
39	(LSB)							
40	Reserved							
...								
47								

The TAG field is defined in 5.1.

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.

The TASK MANAGEMENT FLAGS field is defined in table 17. If TASK MANAGEMENT FLAGS contains a reserved or restricted value, the task manager shall return an SRP_RSP response containing GOOD status. The RSP_CODE field shall be set to TASK MANAGEMENT FUNCTION NOT SUPPORTED.

TABLE 17 - TASK MANAGEMENT FLAGS

Codes	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

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 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.

5.7 SRP_CMD request

An SRP_CMD request conveys a SCSI command (see table 18). An SRP_CMD request shall be sent as a message whose length is 48 bytes plus the lengths of the ADDITIONAL CDB, DATA-OUT BUFFER DESCRIPTOR and DATA-IN BUFFER DESCRIPTOR fields.

Table 18 - SRP_CMD request

Byte	Bit	7	6	5	4	3	2	1	0
0		TYPE (02h)							
1		DOIND	DIIND	Reserved					
2		DOCOUNT							
3		DICOUNT							
4		Reserved							
...									
7									
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									

The DOIND bit and DOCOUNT field specify the presence and length of the DATA-OUT BUFFER DESCRIPTOR field (see 4.4.2). DOIND shall contain zero if the value of INDSUP was zero in the SRP_LOGIN_RSP.

The DIIND bit and DICOOUNT field specify the presence and length of the DATA-IN BUFFER DESCRIPTOR field (see 4.4.2). DIIND shall contain zero if the value of INDSUP was zero in the SRP_LOGIN_RSP.

The TAG field is defined in 5.1.

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 19.

TABLE 19 - TASK ATTRIBUTE

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 a automatic contingent allegiance task attribute. (See SAM-2)
101b-111b	Reserved

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 actual CDB to be interpreted by the addressed logical unit. Any bytes between the end of the actual CDB and the end of the two fields shall be reserved.

The contents of the actual CDB shall be as defined in the SCSI command standards.

The DATA-OUT BUFFER DESCRIPTOR and DATA-IN BUFFER DESCRIPTOR fields are defined in 4.4.2.

5.8 SRP_RSP response

An SRP_RSP response (see table 20) conveys an SRP response to an SRP_TSK_MGMT or SRP_CMD request received by a target. 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 20 - SRP_RSP response

Bit Byte	7	6	5	4	3	2	1	0	
0	TYPE (C1h)								
1	Reserved								
2									
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)								

The REQUEST LIMIT DELTA field is defined in 4.3.

The TAG field shall contain the same value as the TAG field in the SRP_TSK_MGMT or SRP_CMD request for which this SRP_RSP 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 initiator shall ignore its contents. The target 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 initiator shall ignore its contents. The target 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 MAXIMUM TARGET TO INITIATOR IU specified in SRP_LOGIN_RSP when the RDMA channel was established, the target shall return an SRP_RSP information unit whose length is MAXIMUM TARGET TO INITIATOR IU 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 initiator shall ignore the SENSE DATA LIST LENGTH field and shall assume a length of zero.

RSPVALID, when 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, when set to 1, indicates the contents of the RESPONSE DATA LIST LENGTH field specify the number of bytes in the RESPONSE DATA field.

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 initiator shall ignore the RESPONSE DATA LIST LENGTH field and shall assume a length of zero.

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. Targets 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.

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

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 the data-out buffer shall not be transferred. DATA-OUT RESIDUAL COUNT shall be equal to:

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. Targets 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 COUNT shall be equal to:

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:

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 21) contains information describing certain protocol failures detected during processing of an SRP request received by the target. The RESPONSE DATA field shall be present if the target detects any of the conditions described by a non-zero RSP_CODE value (see table 22).

Table 21 - 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 22.

Table 22 - 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

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 - 32 - RESPONSE DATA LIST LENGTH

MAXIMUM TARGET TO INITIATOR IU is the value specified in SRP_LOGIN_RSP when the channel was established.

5.9 SRP_TPAR_REQ request

An SRP_TPAR_REQ request (see table 23) conveys a target request to negotiate SRP protocol parameters. SRP_TPAR_REQ requests shall be sent as a 52 byte message.

Table 23 - SRP_TPAR_REQ request

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (81h)							
1								
2	Reserved							
3								
4								
...	REQUEST LIMIT DELTA							
7								
8								
...	TAG							
15								
16								
...	Reserved							
51								

The REQUEST LIMIT DELTA field is defined in 4.3.

The TAG field is defined in 5.1.

5.10 SRP_TPAR_RSP response

An SRP_TPAR_RSP response (see table 24) conveys an SRP response to an SRP_TPAR_REQ request received by an initiator. SRP_TPAR_RSP responses shall be sent as a 64 byte message.

Table 24 - SRP_TPAR_RSP response

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (41h)							
1								
7	Reserved							
8								
...								
15	TAG							
16								
...								
63	Reserved							

The TAG field shall contain the same value as the TAG field in the SRP_TPAR_REQ request for which this SRP_TPAR_RSP is a response.

5.11 SRP_AER_REQ request

An SRP_AER_REQ request (see table 25) conveys a target request to report an asynchronous event. SRP_AER_REQ requests shall be sent as the minimum length message capable of carrying the fields. All SRP initiators shall support receiving SRP_AER_REQ and all SRP targets shall support generating SRP_AER_REQ.

Table 25 - SRP_AER_REQ request

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (82h)							
1	Reserved							
2								
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)						

The REQUEST LIMIT DELTA field is defined in 4.3.

The TAG field is defined in 5.1.

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 MAXIMUM TARGET TO INITIATOR IU specified in SRP_LOGIN_RSP when the RDMA channel was established, the target shall send an SRP_AER_REQ request whose length is MAXIMUM TARGET TO INITIATOR IU 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.

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 - 32 - RESPONSE DATA LIST LENGTH

MAXIMUM TARGET TO INITIATOR IU is the value specified in SRP_LOGIN_RSP when the channel was established.

5.12 SRP_AER_RSP response

An SRP_AER_RSP response (see table 26) conveys an initiator's SRP response to an SRP_AER_REQ request. An SRP_AER_RSP response shall be sent as a 16 byte message.

All SRP initiators shall support generating SRP_AER_RSP and all SRP targets shall support receiving SRP_AER_RSP.

Table 26 - SRP_AER_RSP response

Bit Byte	7	6	5	4	3	2	1	0
0	TYPE (42h)							
1	Reserved							
2								
7								
8	(MSB)	TAG						
...								
15								(LSB)

The TAG field shall contain the same value as the TAG field in the SRP_AER_REQ request for which this SRP_AER_RSP is a response.

6 SCSI mode parameters

6.1 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 27.

Table 27 - SRP mode page codes

Page code	Description	Subclause
02h	Disconnect-reconnect page	6.2
18h	Logical unit control page	6.3
19h	Port control page	6.4

6.2 Disconnect-reconnect mode page

The disconnect-reconnect page (see table 28) 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 targets.

Table 28 - 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						(LSB)	
11									
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 target. The field values are communicated from the device server to the target 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 sense key shall be set to ILLEGAL REQUEST and the additional sense code set to ILLEGAL FIELD IN PARAMETER LIST.

The BUFFER FULL RATIO field, BUFFER EMPTY RATION field, BUS INACTIVITY LIMIT field, PHYSICAL DISCONNECT TIME LIMIT field and CONNECT TIME LIMIT field are reserved for SRP devices.

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 target may use the random buffer access capability to order RDMA's for a single SCSI command. If the EMDP bit is set to 0, the target shall generate continuously increasing RDMA addresses for a single SCSI command. If the EMDP bit is set to 1, the target may issue RDMA's 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 are reserved for SRP devices.

6.3 Logical unit control mode page

The logical unit control mode page contains those parameters that select logical unit operation options. The implementation of any parameter and its associated functions is optional. The page follows the MODE SENSE and MODE SELECT rules specified by the SPC-2 standard.

This page is not currently defined for SRP devices.

6.4 Port control mode page

The port control mode page contains those parameters that select target port operation options. The page shall not be implemented by logical units other than LUN 0. The implementation of any parameter and its associated functions is optional. The page follows the MODE SENSE and MODE SELECT command rules specified by SPC-2.

This page is not currently defined for SRP devices.

Annex A

(normative)

SCSI RDMA protocol services

EDITOR'S NOTE 2 - See document T10/01-172

Annex B

(normative)

Alias entry designation formats

EDITOR'S NOTE 3 - See document T10/01-193

