To: T10 Technical Committee
From: Greg Pellegrino, Compaq Computer Corporation (greg.pellegrino@compaq.com)
Date: 20-21 February 2001
Subject: Minutes of the SRP WG – February 20-21, 2001 – Denver, Colorado

Revision History

Revision 0: first revision

Attendance

Name	S	Organization	Electronic Mail Address
Mr. Brian Forbes	V	Brocade	bforbes@brocade.com
Mr. Greg Pellegrino	V	Compaq Computer Corp.	greg.pellegrino@compaq.com
Mr. Madhu Rangarajan	V	Dell Computer	madhusudhan_rangarajan@dell.com
Mr. Robert H. Nixon	P	Emulex	bob.nixon@emulex.com
Mr. Ralph O. Weber	P	ENDL Texas	roweber@acm.org
Mr. Sesidhar Baddela	V	Hewlett Packard Co.	sesidhar@cup.hp.com
Mr. George O. Penokie	P	IBM / Tivoli Systems	gpenokie@tivoli.com
Mr. John P. Scheible	A	IBM Corp.	Scheible@vnet.ibm.com
Mr. Giles Frazier	V	IBM Corp.	grf@us.ibm.com
Mr. Tom Brey	V	IBM Corp.	tbrey@us.ibm.com
Mr. Bill Futral	V	Intel	william.futral@intel.com
Mr. Cris Simpson	P	Intel Corp.	cris.simpson@intel.com
Mr. John Lohmeyer	P	LSI Logic Corp.	lohmeyer@t10.org
Mr. Keith Holt	V	LSI Logic Corp.	keith.holt@lsil.com
Mr. Xai Phan	V	McData	xai.phan@mcdata.com
Mr. Konrad Lei	V	McData Corp.	klie@mcdata.com
Mr. Edward A. Gardner	P	Ophidian Designs	eag@ophidian.com
Mr. Seth Abrahams	V	Sun Microsystems, Inc.	seth.abrahams@sun.com (by telephone)
Mr. Rick Casaly	V	Troika Networks	rcasaly@troikaetworks.com
Mr. William C. Terrell	P	Troika Networks, Inc.	terrell@TroikaNetworks.com
Mr. Roger Cummings	AV	Veritas Software	roger.cummings@veritas.com

21 People Present

Status Key:
- P - Principal
- A,A# - Alternate
- AV - Advisory Member
- L - Liaison
- V - Visitor

Results of meeting

The SCSI over RDMA protocol meeting opened at 1:00 pm Friday 20 February 2001. Intel was thanked for hosting. This protocol standard maps SCSI over InfiniBand™ Architecture, Virtual Interface (VI) Architecture, and similar transports supporting RDMA (remote direct memory access).

Agenda

The agenda was approved at the meeting.

1. Total Data Length – Keith Holt
2. Flow Control – Keith Holt
3. Names, Addresses, and Identifiers – SAM Mapping – George Penokie
4. Multi-Channel Connections – Cris Simpson
5. Transport of Initiator Id – George Penokie
6. Document Review – Ed Gardner
7. Optional: Definition of well-known LU's – George Penokie
8. SRP/IB Annex Review, joint review with IBTA AWG – Greg Pellegrino
9. Adjournment

Topics

1. Total Data Length – Keith Holt

Presentation of data length subject of email "SRP WG agenda request" as posted to T10 Reflector. Group was in general agreement with the emailed subject. Keith will write proposal for adding this field.

No compliance statements regarding accuracy of the data total length field should be added. The field as added to SRP information units will be a 4-byte field and be revelavent only for information units using indirect data descriptor lists. The resulting information units could have this format:

Direct Memory Descriptor Case	**Indirect Memory Descriptor Case**
Header	Header
CDB	CDB
Reserved	**Data total length (bi-out)**
Data-out Descriptors …	Data-out Descriptors …
Data-in Descriptors …	**Bi-Data-in total length (optional)**
Unused	Data-in Descriptors …

Where data total length is used for all data in single direction cases and becomes the data-out total length in bi-directional cases. A second field labeled above bi-data-in total length is used for data-in total length in bi-directional cases. This fields breaks 8-byte alignment of fields in the information unit.

Format may be changed from that captured in the minutes. Discussion participants agreed that such a proposal makes sense and provides some guidelines of its content.

2. Flow Control – Keith Holt

Presentation of flow control subject of email "SRP WG agenda request" as posted to T10 Reflector. Keith requested a change of the SRP text for flow control. The agreed resolution was that no change of the SRP specification would occur.

Adding a picture to the SRP document showing flow control was suggested. The SRP specification should indicate that flow control advertises a place to land a command not the ability to queue commands for execution.

3. Names, Addresses, and Identifiers – SAM Mapping – George Penokie

Discussion of T10/01-084r0, Names, Addresses, Identifiers, Oh my!, authored by George Penokie, as posted on 27 February 2001 and selectively distributed via email 19 February 2001.

Summary of document:

a. Identifier is same as Address. Binary representation. Unique within a SCSI Domain. It may change. Alpa is an example.
b. Name is a label of an object. May be any format. It Shall be world wide unique. Should never change. WWID is an example.
c. A name could be an identifier. An identifier cannot be a name.
d. SCSI Objects
    i. Initiator port
    ii. Target port
    iii. Logical unit
    iv. Initiator device (new)
    v. Target device (new)
e. Nexus concept
    i. Document represents what is now, may change.
    ii. I, T, and L are routing and task identification information.

iii. Q is task identification

f. Identifiers and names
    i. Document represents what is now.

George will propose to add content of this proposal to SAM.

Rob Elliott's comments made via email prior to WG meeting.

a. Tracking persistent reservations by initiator device name breaks Fibre Channel
    a. If Compaq really cares about this they should bring it up outside SRP WGs, within the context of SAM.

The IB annex should not use the device term.

4. Multi-Channel Connections – Cris Simpson

Cris presented thoughts concerned with multiple channels in a SRP connection. He will create document T10/01-071r0, InfiniBand™ Multi-Channel Connections for SRP, as a proposal in the next WG meeting.

Summary of presentation:

g. An SRP connection could have one control channel and one or more data channels.
h. Could use send-only QPs for control channel.
    i. Divides service levels between control and data (bulk transfers).
i. All data channels must connect to the same HCA on the initiator side.
j. Data channels and control channels may be on different HCAs on the initiator side.
k. Two party – (1+N) channel
    i. Initiator will have to tell target which channel to use.
    ii. Option: could use same HCA on host so that target can chose which channel to use as all are valid.
        1. bounded by currently established channels
l. Three party – (1+1) channel
    i. InfiniBand could support sending a command on one QP pair describing a memory region on another QP pair. The QP pair for data need not be on the same initiator.

5. Transport of Initiator Id – George Penokie

Discussion of T10/01-034r1, Transport of Initiator Identification Across Large Networks, authored by George Penokie, as posted on 16 February 2001.

Resolution: identifiers should be exchanged at login, not for each command. Dismiss proposal. George won't pursue it.

7. Definition of well-known LU's – George Penokie

Discussion of T10/01-068r0, Definition of Well Known Logical Units, authored by George Penokie, as posted on 16 February 2001.

Comments:

m. Table 5 – length 1, code 1 undefined. Should be specified as reserved.
n. Table 6 – field at byte n should be W-LUN not LUN.
o. Replace most (all?) uses of know with known.
p. Wording in section 3.3 should be modified (changes captured by George in document).
q. Check spelling of defination throughout doc
r. All commands for a W-LUN are required if that W-LUN is supported.

George will pursue this proposal – flesh out before submitting again.

6. Document Review – Ed Gardner

Review of T10/1415-D revision 0.3, SCSI over RDMA Protocol, authored by Ed Gardner, as posted on 29 January 2001.

Comments:

s. Document is very VI centric. Rip out VI, SVP references.
t. Reference SAM-2 and SPC-2 not SAM and SPC.
    i. Ed noted that this results in referencing no approved documents.
    ii. The term approved means ANSI published.
u. Drop all ANSI blah blah blah blahs and reference only standards
    i. Be consistent in use of either standard acronyms or full document names.
    ii. Acronym is fine as long as the acronym is defined
v. Reference IBTA standards.
w. Architecture misspelled.
x. Refer to ISO/IEC 14776-312.
y. Pg 13
    i. Immediate data definition too VI oriented. Move to VI annex?
z. Protocol specific terms defined in the protocol specific annexes.
aa. General terms defined in main document.
bb. Capitalization
    i. SCSI terminology shall not be capitalized.
    ii. Ed: Terminology defined in reference documents will be copied for terminology and capitalization used in reference documents.
cc. Acronyms and definition descriptions should be separate sections. Or expand acronym first time used.
dd. 3.1.23 drop “unlinked” in definition.
ee. I/O has slash.
ff. Work Queue term will be removed from document.
gg. 3.2
    i. Remove NIC
    ii. Definition of SCSI-3 presents a family of standards. Not definition of 3.2. Find better definition.
    iii. Including definition of “may not” was discussed, resolved to leave.
hh. Change use of IU with information unit spelled out.
ii. 3.4
    i. Missing list of notations – section 3.5 in SPI-4. This is the notation used for procedure calls.
    ii. Add “e.g.” for binary and hex examples.
    iii. Delete discussion of exceptions – there are none.
    iv. Make “notes” a separate paragraph.
    v. Delete mention of tables.
    vi. Delete “not all tables…”.
jj. Add procedure calls to document.
kk. 4.1 and 4.1.1 are prime examples of the hanging paragraph crime.
    i. Hanging paragraph is paragraph in a section prior to the start of subsections.
    ii. Delete “note that”. “Note” may only be used in the context of a note.
    iii. When referring to the standard use “this standard” not “SRP”, otherwise trouble when writing SRP-2.
ll. One space between sentences
mm. 4.1.1
    i. “Process” definition.
    ii. “Connection” definition, add glossary definition – refer to section.
    iii. Proper term is not “bullsh__” but “nugatory”. This term is not currently used in the document.

iv. "Connection" in main body may be different than in IB Annex
v. 4.1.1.1
    1. George dislikes term "paradigm".
    2. Paragraph 1, last sentence - replace allocates with chooses.
    3. Paragraph 2 – remove "assumed to be dynamic objects".
    4. Paragraph 2 – replace "as used by SRP" with "this standard…".
    5. Figure 1 –
        a. Suggested to indicate IU proper names.
        b. Sequence implication for process behavior troublesome.
    6. Note 3 should be incorporated into text. It's important so make it early in this section.
    7. Reword "shall be able to " phrase.
vi. 4.1.1.2
    1. First sentence, delete "itself".
    2. Last paragraph, first sentence pg 18 – replace "does not assume" with "does not require".
    3. Be clear on disconnection versus command submission/completion. Does last sentence in last full paragraph page 18 say this?
    4. Explain that SRP uses connection establishment for login.
    5. Connection establishment should not be in the RDMA model section.
vii. 4.1.2 Messages
    1. Qualify "messages" term.
    2. Dump "mechanisms" term.
    3. Resolve "sending /receiving process" terms with "sender/receiver" terms.
    4. Dump "as used within".
    5. Question on "A memory handle is only valid on a specific connection." statement. May be valid across connections – valid on specified connections.
    6. B item in list – replace "offset" term with something more specific.
    7. Page 20, paragraph 2, sentence 2 - George has rewording that he will provide to Ed.
    8. Page 20, paragraph 2 – state that some transport may employ reliable delivery.
viii. 4.1.4
    1. Replace semi-colon with parenthetical comment with i.e. with end part.
    2. Paragraph 2 – reword "assumes" to "requires…"
    3. C. item of list.
        a. "in their entirety" phrase not needed.
        b. Replace "may be assumed" with "are present".
ix. 4.2
    1. Delete "note that".

nn. Ed is now properly sensitized to the "assume" term.
oo. Long running protocol definition model.
    i. Lower interface – establish protocol entities (Send, RDMA, work queues).
    ii. Upper interface – describe how the entities are used (login, command delivery).
    iii. Implementation – SAM mapping.
pp. Nuke waffle words.
qq. "Allocate" will be replaced throughout document with better/different terms.
rr. Need SAM-2 to SRP mapping.
ss. Suggestion on how to restructure document.
    i. Section for purely transport establishment
        1. Describe underlying protocol requirements required to map SRP onto.

2. Alternatively, describe relationship between transport and SCSI.

ii. Section for purely SRP establishment.
iii. Section relating two previous sections.

A schedule for SRP work was discussed, arrived at

a. Features from previous working groups – AEN, LOGOUT_IU, identifiers in LOGIN REQ/REP IUs – need to be incorporated. Plan to incorporate next week. Target review for March WG meeting.
b. Cris offered to tackle document restructuring.
c. George will do SAM-2 mapping section. This will be similar to SPI and SBC work already in progress.
d. Greg offered to work on SAM-2 mapping section as much of the work being done for IB Annex should be moved to main body. George stated that he is already doing this for SPI, natural to do it for SRP as well.
e. Letter ballot – July or August??

8. SRP/IB Annex Review – Greg Pellegrino

Review of T10/01-028r2, SRP InfiniBand™ Annex, authored by Greg Pellegrino, as posted on 19 February 2001. The InfiniBand™ Trade Association Application Working Group attending this joint meeting.

Introduction. Statement that this is an open meeting, e.g. not covered by IBTA NDA.

Discussion of glossary definitions.
Reference IB specification as source of all definitions that are copied from there.
Delete definition of automatic path migration, as there are no references to it in the annex.
Channel adapter needs more context, e.g. "Target" in TCA conflicts with SCSI usage.
Management datagram: is this definition needed (e.g. does annex discuss MADs). Delete "for communication". Is use of "packet" proper?
Proper reference for IPv6 addresses is RFC 2373, not 2460.

A.3 overview
Bill Futral, others: do not distinguish HCA and TCA, only use CA.
Figure 1: IB volume 1 discusses "Processor Node". Greg has received objections to use of "node", has substituted "unit". Discussion of whether this can be generic unit (remove "processor"). "Consumer" is defined as entity that issues verbs, which requires HCA, therefore can an IO unit have consumers? Further discussion of IB definition of HCA and TCA, suggestion that simply referring to CA will simplify diagrams and avoid conflicts.
Figure 2 and 3 are good. However, colors must be changed to ensure legible reproduction on black and white printers.
The CA acronym causes problems with SCSI Contingent Allegiance. Better to always spell out channel adapter. General recommendation from George Penokie to not use acronyms, e.g. say IO Controller, not IOC.
Discussion of whether third party operations can identify devices by LID rather than GID. Agreement that GID is necessary (although possibly not sufficient) since LIDs may be dynamic and/or not universally valid within a subnet, not to mention multi-subnet problems. Third party operations should not drive definition of target identifiers, as third-party target descriptors provide a way to define additional data as needed.
Switched to "Problems to Solve to map SRP on IB" (document number not assigned yet). Switched back.
The SRP target port corresponds to an IOC. The TCA's port identification may be advisory information on how to reach an SRP target port or IOC, but is not part of the SRP target port identifier.

Each IOC (each IOC GUID) identifies a single SRP target port. That is, there shall be only one SRP service ID (ServiceEntry) per IOC.
The iSCSI WWUI format is derived from a document that Jim Hafner has submitted to the reflector. Per the outspoken comment in the room, if this is desired, it should be submitted as a separate proposal against SAM-2 (rather than multiple conflicting proposals in different protocols).
Revisited question of piggy-backing SRP_LOGIN_REQ in private data of CM:REQ. Concern that future additions might expand SRP_LOGIN_REQ beyond available space. Agreed to defer such discussion until such time as a problem appears (e.g. some proposal expands SRP_LOGIN_REQ beyond available private data space).
Releasing a connection. On a reliable connection, sending a LOGOUT IU will (eventually) either result in a transport ACK or a transport error. Both are a completion of the send operation. The sender should / shall disconnect after either.
Objection was noted to tying the IOC GUID to LUN 0 VPD data; Greg chose to ignore discussion of that until a later meeting.

9. Adjournment

This meeting adjourned at 5:00 pm Wednesday.