Minutes for RAID working group 01/11/94
Accredited Standards Committee
X3, Information Processing Systems

Doc. No.: X3T10/94-011R0
Date: January 11, 1994
Project:
Ref. Doc.:
Reply to: G. Penokie

To: Membership of X3T9.2 and X3T10

From: George Penokie

Subject: Minutes of RAID Study Group Meeting
January 11, 1994

Agenda

1. Opening Remarks
2. Attendance and Membership
3. Approval of Agenda
4. Report on last RAID Working Group
5. SCSI-3 Addressing (93-161r2)
6. SDA States and Types (93-191r1)
7. SDA Commands and Mode Pages (93-192r1)
8. SCSI Disk Array Model (93-140r3)
9. Dual-Controller Issues
10. Additional Device Models
11. Error Handling for SCSI Controllers
12. Action Items
13. Meeting Schedule
14. Adjournment

Results of Meeting

1. Opening Remarks

George Penokie the RAID Study Group Chair, called the meeting to order at 9:10 am, Tuesday, 11 January 1994. He thanked Vitro for hosting the meeting.

As is customary, the people attending introduced themselves. A copy of the attendance list was circulated for attendance and corrections.

It was stated that the meeting had been authorized by X3T9.2 and would be conducted under the X3 rules. However, the meeting will report to the newly formed X3T10 committee. Ad hoc meetings take no final actions, but prepare recommendations for approval by the X3T10 task group. The voting rules for the meeting are those of the parent committee, X3T9.2 and/or X3T10. These rules are: one vote per company; and any participating company member may vote.

The minutes of this meeting will be posted to the SCSI BBS and the SCSI

Reflector and will be included in the next committee mailing.

George stated that this is the 15th meeting of the RAID study group and the 2nd joint RAID Advisory Board/RAID study group meeting. The purpose of the group is to deal with interface issues related to using RAIDs. The study group will assess the issues and then formulate a strategy for dealing with them.

2. Attendance and Membership

Attendance at working group meetings does not count toward minimum attendance requirements for X3T9.2 membership. Working group meetings are open to any person or company to attend and to express their opinion on the subjects being discussed.

The following people attended the meeting:

RAID Study Group Meeting Attenders

Name	S	Organization	Phone Number -or- Electronic Mail Address
Mr. Jerrie L. Allen		Amdahl	408-944-3712
Mr. Mark Skrondal		Amdahl	408-944-3954
Mr. John Woods		Amdahl	408-944-5016
Mr. Gerry Johnsen		Ciprico	612-551-4055
Mr. Steve O'Neil		CMD Technology	714-454-0800
Mr. Bill Galloway		ComPaq	713-374-6732
Mr. Bob Solomon		Data General - Clarion	508-898-5014
Mr. Doug Hagerman		Digital Equipment Corp.	hagerman@starch.enet.dec.com
Mr. Ralph Weber		Digital Equipment Corp.	weber@star.enet.dec.com
Mr. Doug Anderson		Distributed Proc. Tech.	407-830-5522
Mr. Howard Grill		Formation	609-234-5020
Mr. Kevin Pokorney		Fujitsu	303-682-6649
Mr. Bill Hutchison		Hewlett Packard Co.	hutch@.boi.hp.com
Mr. Andrew Hill		HI-Data	617-937-6770
Mr. Paul Boulay		Hitachi	408-986-9770
Mr. Giles Frazier		IBM Corp.	512-838-1802
Mr. George Penokie		IBM Corp.	gop@rchvmp3.vnet.ibm.com
Mr. Roy Kurz		Jaycor	619-535-3117
Mr. Charles Binford		NCR	316-636-8566
Mr. M.W. Jibbe		NCR	316-636-8810
Mr. Herb Silverman		Peer Protocols	714-476-1016
Mr. Kim Le		STK - Ampen'f	303-673-7322
Mr. Robert N. Snively		Sun Microsystems, Inc	bob.snively@eng.sun.com

23 people present

Status Key: P - Principal
A - Alternate
O - Observer
L - Liaison
S,V - Visitor

** Status not applicable at this meeting because X3T10 committee not yet formed.

3. Approval of Agenda

The agenda developed at the meeting was approved.

4. Report on last RAID Working Group

Last RAID Working Group was held in San Diego, CA. Several members of the RAID Advisory Board were present. Minutes of the San Diego meeting are ready for publishing, but a document number has not yet been assigned. When the document number is assigned the minutes will be published on the SCSI Reflector.

5. SCSI-3 Addressing (93-161r2) Penokie

George presented a revised version of the SCSI-3 Addressing proposal (X3T9.2/93-161r2). George apologized for not distributing copies of the new document.

George noted that the EXTENDED IDENTIFY message has been removed from the proposal. The only 64-bit LUN addressing now proposed for SIP SCSI uses the translation table. This is considered sufficient for all anticipated parallel SCSI implementations. RAID systems that require larger address ranges must be implemented based on SBP, Fibre Channel, or another serial protocol.

This generated questions about how the equivalent addresses are generated and used in the serial protocols. Where is the VolSel bit equivalent? It's the low-order bit of the Access Method field. Bob Snively was called to describe Fibre Channel addressing maps into the addressing levels in 93-161r2.

Bob noted that the 93-161r2 proposal cannot describe the full addressing capabilities available in Fibre Channel. However, it can describe the addressing needs all but the most outrageous RAID configurations. Particularly, it can address a large set of practical and useful configurations.

The concern was raised that serial busses already have 64-bit LUNs. So, the 8-bit addressing level fields. George will add words to 93-161. In parallel SCSI, LUN values in addressing level fields are one-to-one with normal usage LUN values. On the serial busses, some type of vendor specific mapping may be required.

George also agreed to add examples that use pure IDENTIFY message formats to access storage array components.

The issue of changing the mapping table resurfaced. George said that no protections are provided beyond those mode page rules that apply to all SCSI mode pages. Reserving LUN 0 is the only true protection.

George indicated that r3 probably will be the last revision of 93-161. He will make that revision and distribute it soon. The group accepted George's proposal to never distribute r2, because r3 is almost the same as r2. The next activity will be incorporating 93-161r3 in the SCSI-3 Controllers

Commands (SCC) document.

6. SDA States and Types (93-191r1) Penokie

George presented a new document defining states of an SDA and the types of devices needed within Disk Arrays (X3T9.2/93-191r1).

A significant discussion took place about the use of peripheral device type for items such as fans and LED panels. George has received written comments expressing displeasure with this usage and proposing alternatives. The group reviewed a couple of options based on the SCSI-2 document. Two fields of interest are the Peripheral Qualifier and the Device-Type Modifier. Device-Type Modifier appeared preferable, because it is essentially reserved in SCSI-2 and because its 7 bits wide.

George noted a couple of minor changes. The changes generated no discussion.

Doug Hagerman presented some comments on the SDA states in 93-191. Doug contends that the state information in 93-191 is not sufficient to cover the real situations that occur in SDAs. He thinks that reliable management of the SDA requires a careful definition of states, how the SDA gets into them, and how it gets out of them.

Doug has prepared a document listing all the states that he can think of. Doug's document also separates the states into lists for the SDA, Redundancy Group, P-LUI, and Spare. Doug's work is based on George's 93-191, and contains at least all the states found there.

The group discussed the meaning of each state. Examples were given to clarify meanings. George will incorporate the states definition style into 93-191. He also will add those states in Doug's document that are not already present in 93-191.

7. SDA Commands and Mode Pages (93-192r1)

George presented a new document defining new commands and mode pages needed to implement SCSI disk arrays (X3T9.2/93-192r1). Recent revisions to the document bring it into alignment to the disk array model.

The document does not define exact command formats. It is more an outline that George is using to validate the model. It lists the required inputs and outputs. George will model the commands on existing SCSI commands. However, he plans to avoid 6-byte command formats, because they do not have large enough parameter size byte counts.

George noted the changes in the maintenance operations proposals. Reporting functions for specific P-LUIs may require multiple command/response operations. While this increases the number of transactions required for some tasks, it greatly simplifies the operating model.

No one could find where the state of the P-LUI is reported. Efforts were made to locate it on two pages that were not duplicated. George will either locate

where P-LUI state is already reported or determine where reporting P-LUI state should be reported.

A later discussion in the SCSI Disk Array Model reveled that George thought the state of the P-extent was the same as the state of the P-LUI. But, what about fans? Do they have P-extents? At this point, George added a state field to the Report P-LUI service.

Concerns were raised about rebuild times and priorities. What commands does 93-192 define for setting these kinds of dials. Doug Hagerman suggested that mode pages are a better model for control knobs. George asked that people in the group who have worked in this area bring in a proposal based on mode pages.

8. SCSI Disk Array Model (93-140r3)

George presented a revised version of the SCSI-3 Disk Array Model (X3T9.2/93-140r3).

A half dozen minor wording changes in the glossary were not controversial. Doug asked for a clarification of the relationship between SCSI Disk Arrays (SDAs) and Disk Array Conversion Layers (DACLs). Is the DACL a component of an SDA? George said, "Yes."

The Report P-LUI Geographic Location Service has been made separate from the Report P-LUI Service. Further discussion resulted in the addition of a state field to the Report P-LUI Service return data.

Questions were raised about how to get current state of a volume. George described how one or more Report Service requests are used to learn the state of the volume. More Report requests can be directed to the P-LUIs (based on the volume report), if more detailed state information is needed.

A discussion of polling and log-based methods followed. George requested that someone bring a proposal for a logging page. This would cover polled/logged mechanisms for saving and reporting SDA state exceptions.

The change-by-change review of the SDA Model produced no major comments. Most of the comments were editorial in nature.

9. Dual-Controller Issues

George described his concerns about adding dual controller definitions to the SCSI Disk Array (SDA) Model and/or SCSI-3 Controller Commands (SCC) document. George has no experiential knowledge of dual controller configurations. So, he has no personal basis for leading standards development in that area.

Also, George wants to expedite delivery of the information that already has been developed. So, he is concerned about adding uncertainty to that schedule.

George noted that dual controller work can be added in a second pass of the

SDA Model and SCC. This did not satisfy the gut-level interests of many people present.

The group concluded that another member of the group would have to lead the dual controller discussion. No one volunteered at the meeting. On the other hand, no one proposed terminating any work on dual controllers immediately.

10. Additional Device Models

Doug Hagerman presented the first draft of his document containing minimalist device models for things like power supplies and fans. The intent is to make more components of a disk array addressable entities, so that they can be added to redundancy sets.

The document contains device models for power supplies, fans, consoles, and caches. Doug's goal is to test the validity of the draft device models.

George stated a general philosophy that anything that is an addressable entity can be added to redundancy sets. But, making a device type is not needed for making an addressable entity. For example, P-extents are not covered by any device model. But, since they are addressable, they can be included in redundancy sets.

This philosophy may lead to ending the current problem with using too many device types. George will be thinking about that.

Doug preceded to review the lists of devices and commands. The purpose was identifying the correct list of devices. For each device, Doug sought to list the commands that make sense for that particular device. Also, Doug wanted to validate the functions of the various commands.

11. Error Handling for SCSI Controllers

Doug Hagerman presented the first draft of his document describing error handling in SDAs. Doug started by noting that SDAs have much more diverse failure modes than just plain disks.

Doug initiated a discussion of Contingent Allegiance in SDAs. If a leaf disk in an SDA goes into Auto Contingent Allegiance (ACA) due to an operation on the array, can a physical path (different) be used to clear the ACA. Doug proposed that the disk views the various path to it in the same way that it would view paths from different initiators. That is, the ACA only can be cleared by operations on the path where the CHECK CONDITION status was returned. All other operations on all other paths see an ACA ACTIVE status until the ACA is cleared on the proper path.

Doug proposes no new status codes. But, George noted that ACA ACTIVE is missing for the document's list. Discussion of the ASC/ASCQ list was more complex.

George strongly defended the current SCSI method for naming and describing ASC/ASCQ assignments. Doug and several others want to define ASC/ASCQs in the

SCC document (as device model features). George prefers defining ASC/ASCQs in the SCSI-3 Primary Commands (SPC) and then using them in the SCC.

Doug next began working through the list of ASC/ASCQ descriptions needed. Each listed ASC/ASCQ was discussed in terms of:

- what it means,
- does an equivalent ASC/ASCQ already exist,
- is the proposed description too detailed?

Significant problems were uncovered for reporting the failing part in the REQUEST SENSE data. The FRU field is only one byte, which is too small by at least one byte (and at most seven bytes). There are no reserved or otherwise available fields in the REQUEST SENSE data. Reformatting the REQUEST SENSE data will wreak havoc on the entire SCSI world. Creating a new command to get the data presents at least all the problems already present for REQUEST SENSE.

The only solution that seemed even slightly credible is some kind of state changes log page. Then a new ASC/ASCQ would signal the host that the log page should be read.

12. Action Items

a) Penokie Revise the SCSI-3 Addressing (93-161) document and present this proposal to the next plenary for a vote.
b) Penokie Revise the SDA States and Types (93-191) document. Include the states description found in Doug Hagerman's.
c) Penokie Revise the SDA Commands and Mode Pages (93-192) document.
d) Penokie Revise the SCSI Disk Array Model (93-140) document and present this proposal to the next plenary for a vote.
e) Hagerman Revise "Error Handling for SCSI Controllers." Obtain an X3T10 document number for it.
f) Hagerman Revise "Additional Device Models." Obtain an X3T10 document number for it.

13. Meeting Schedule

The next meeting of the RAID Study Group is planned for Feb 15, 1994 at the Red Lion Hotel in Austin, Tx. The meeting is expected to run from 9:00am-5:00pm. This meeting will be a joint RAID Advisory Board Host Interface Group and X3T9.10 RAID Study Group meeting.

14. Adjournment

The meeting was adjourned at 04:44 pm. on Tuesday, January 11 1994.