

T10/08-319r0

Voting Results on T10 Letter Ballot 08-318r0 on  
 Forwarding OSD-2 to First Public Review  
 Ballot closed: 2008/09/03 12:00 noon MDT

Organization	Name	S	Vote	Add'l Info
AMCC	Paul von Stamwitz	P	Yes	
Brocade	David Peterson	P	Yes	Cmnts
Dell, Inc.	Kevin Marks	P	Yes	Cmnts
EMC Corp.	David Black	A	Yes	Cmnts
Emulex	William Martin	P	Yes	
ENDL	Ralph O. Weber	P	Yes	Cmnts
FCI	Douglas Wagner	P	Yes	
Finisar Corp.	David Freeman	P	Yes	
Foxconn Electronics	Elwood Parsons	P	Abs	Cmnts
Fujitsu	Mike Fitzpatrick	P	Yes	
General Dynamics	Nathan Hastad	P	Yes	
Hewlett Packard Co.	Rob Elliott	P	Yes	Cmnts
Hitachi Global Storage Tech.	Dan Colegrove	P	Yes	
IBM Corp.	Kevin Butt	P	No	Cmnts
Intel Corp.	Mark Seidel	P	Yes	
Iomega Corp.	Robert Payne	P	Yes	
Kawasaki Microelectronics Am	Joel Silverman	P	Yes	
KnowledgeTek, Inc.	Dennis Moore	P	Yes	
Lexar Media, Inc.	John Geldman	P	Yes	
LSI Corp.	John Lohmeyer	P	Yes	
Marvell Semiconductor, Inc.	David Geddes	P	Yes	
Maxim Integrated Products	Gregory Tabor	P	Abs	Cmnts
Microsoft Corp.	Robert Griswold	A	Yes	
Molex Inc.	Jay Neer	P	Yes	
NetApp	Frederick Knight	P	Yes	
Nvidia Corp.	Mark Overby	P	Abs	Cmnts
PMC-Sierra	Tim Symons	P	Yes	
Quantum Corp.	Paul Suhler	P	Yes	
Samsung	Michael Rogers	A	Yes	
SanDisk Corporation	Avraham Shimor	P	Yes	
Seagate Technology	Gerald Houlder	P	No	Cmnts
STMicroelectronics, Inc.	Benoit MERCIER	P	Yes	
Sun Microsystems, Inc.	Dale LaFollette	P	Yes	
Symantec	Roger Cummings	P	No	Cmnts
TycoElectronics	Michael Fogg	P	Abs	Cmnts
Western Digital	Mark Evans	P	Yes	

Ballot totals: (29:3:4:0=36)

29 Yes

3 No

4 Abstain

0 Organization(s) did not vote

36 Total voting organizations

12 Ballot(s) included comments

This 2/3rds majority ballot passed.

29 Yes are more than half the membership eligible to vote  
 [greater than 18] AND

29 Yes are at least 22 (2/3rds of those voting YES or NO [32]).

Key:

- P Voter is principal member
- A Voter is alternate member
- Abs Abstain vote
- DNV Organization did not vote
- Cmnts Comments were included with ballot
- NoCmnts No comments were included with a vote that requires comments

[This report prepared by LB2 v2.3.]

\*\*\*\*\*

Comments attached to Yes ballot from David Peterson of Brocade:

The work of our organization is not affected by the subject matter of this standard.

\*\*\*\*\*

Comments attached to Yes ballot from Kevin Marks of Dell, Inc.:

Dell comment number 1

Page=1 Subtype=Text Author=Kevin\_Marks  
Comment=I see both SPC-3 and SPC-4 referenced though out. Which is it?

Dell comment number 2

Page=25 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
devices;  
s/b  
devices; and

Dell comment number 3

Page=26 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
commands;  
s/b  
commands; and

Dell comment number 4

Page=46 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
a)  
s/b  
1)

I think

Dell comment number 5

Page=59 Subtype=Highlight Author=Kevin\_Marks  
Comment=

(see 4.11.3.1);  
s/b  
(see 4.11.3.1); and

Dell comment number 6  
Page=81 Subtype=StrikeOut Author=Kevin\_Marks  
Comment=

Dell comment number 7  
Page=178 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
undefined;  
s/b  
undefined; and

Dell comment number 8  
Page=216 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
0000h.  
s/b  
0000h;

Dell comment number 9  
Page=260 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
represents;

needs and/or?

Dell comment number 10  
Page=261 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
represents;

needs and/or?

Dell comment number 11  
Page=262 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
represents;

needs and/or?

Dell comment number 12  
Page=263 Subtype=Highlight Author=Kevin\_Marks  
Comment=  
represents;

needs and/or?

Dell comment number 13  
Page=321 Subtype=Highlight Author=Kevin\_Marks  
Comment=

,  
s/b

;

\*\*\*\*\*

Comments attached to Yes ballot from David Black of EMC Corp.:

EMC may submit late letter ballot comments ... later.

\*\*\*\*\*

Comments attached to Yes ballot from Ralph O. Weber of ENDL:

ENDL 1  
PDF pg 25, pg 1, Global  
SPC-3 s/b SPC-4

ENDL 2  
PDF pg 28, pg 4, 2.5  
SAM-4 is close enough to published to move it to normative references

ENDL 3  
PDF pg 29, pg 5, 2.5, Global in 3.1  
Glossary entries that use 'object' in the SAM-3 sense are no longer correct.  
Reword them.

ENDL 4  
PDF pg 29, pg 5, Global  
SAM-3 s/b SAM-4

ENDL 5  
PDF pg 29, pg 5, 3.1.2  
SCSI tasks seems unlikely to appear in the SAM-4 definition of a device server.

ENDL 6  
PDF pg 30, pg 6, 3.1.26  
The definition of object needs to be updated to be consistent with the SAM-4 switch to UML.

ENDL 7  
PDF pg 31, pg 7, 3.1.50  
What SAM-3 calls a task, SAM-4 calls a command. The 'task' definition needs to be restructured to coincide with SAM-4.

ENDL 8  
PDF pg 38, pg 14, 4.1  
Task Priority s/b Command Priority [per SAM-4]

ENDL 9 Technical

PDF pg 48, pg 24, 4.7

After write access to an object has been denied, there appears to be no way to  
ue a command to restore that access.

ENDL 10

PDF pg 102, pg 78, 4.12.10, p3, s1

QUERY TASK s/b QUERY TASK, QUERY TASK SET, and QUERY ASYNCHRONOUS EVENT

ENDL 11 Technical

PDF pg 112, pg 88, 4.14, p 4, s 2

reported as deferred errors (see SPC-3) s/b handled as described in 4.11.3.

ENDL 12 Technical

PDF pg 117, pg 93, 4.16.1, p 1, s 1

SPC-4 subclause 4.5.1 defines exceptions that make this 'all errors' statement

invalid. Resolve the conflict between OSD-2 and SPC-4.

ENDL 13 Technical

PDF pg 117, pg 93, 4.16.1, last p,

I believe that 4.11.3 defines the error recovery mechanism for errors that occur after a command has completed, and the mechanism does not involve deferred error reporting.

ENDL 14

PDF pg 122, pg 98, 4.17, 2nd p before table 58

before the task containing that command enters the enabled task state s/b before the command enters the enabled command state [per SAM-4]

ENDL 15

PDF pg 124, pg 100, 5.1, p3

The structure of the operation code field is described in SPC-4, not SAM-4

ENDL 16

PDF pg 127, pg 103, 5.2.6.2, table 62

byte 47 s/b byte 51 [to match table 60]

ENDL 17 Technical

PDF pg 128, pg 104, 5.2.6.2, end of 2nd p on pg

[insert]If the attribute length field is set to a value that is less than 18,

the unused attribute value bytes shall be placed at the highest offsets in the

attribute value field.

ENDL 18

PDF pg 128, pg 104, 5.2.6.3, table 63

byte 47 s/b byte 51 [to match table 60]

ENDL 19 Technical

PDF pg 128, pg 104, 5.2.6.3, table 63

The attribute length field s/b 2 bytes not 4 bytes. For compatibility, insert

two reserved bytes before the field.

ENDL 20  
PDF pg 130, pg 106, 5.2.6.4, table 64  
byte 47 s/b byte 51 [to match table 60]

ENDL 21  
PDF pg 131, pg 107, 5.2.7, last p in subclause  
bit being set s/b bit being set to one

ENDL 22  
PDF pg 210, pg 186, 6.24, table 120 & text following it  
[field name] TASK TAG s/b COMMAND IDENTIFIER [per SAM-4]

ENDL 23  
PDF pg 211, pg 187, 6.24, table 121  
Task Tag Specified s/b Command Identifier Specified

ENDL 24  
PDF pg 211, pg 187, 6.24, table 121  
QUERY UNIT ATTENTION s/b QUERY ASYNCHRONOUS EVENT

\*\*\*\*\*

Comments attached to Abs ballot from Elwood Parsons of  
Foxconn Electronics:

Lack of expertise

\*\*\*\*\*

Comments attached to Yes ballot from Rob Elliott of  
Hewlett Packard Co.:

comment number 1  
Page=27 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
(global)

ISO/IEC 14776-413, SCSI Architecture Model - 3 (SAM-3) [ANSI INCITS  
402-2005]

Upgrade all SAM-3 references to SAM-4, since SAM-4 is published.

---

comment number 2  
Page=28 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
T10/1731-D  
s/b  
ANSI INCITS xxx-2008

since SPC-4 is published

---

comment number 3  
Page=28 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
T10/1683-D  
s/b  
ANSI INCITS xxx-2008

---

comment number 4  
Page=38 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
Task Priority  
s/b  
Command Priority

to match final version of sam4

---

comment number 5  
Page=38 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
response to an INQUIRY command  
s/b  
the standard INQUIRY data

---

comment number 6  
Page=39 Subtype=StrikeOut Subj=Cross-Out Author=RElliott  
Comment=  
Delete  
and

---

comment number 7  
Page=40 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
disc drives  
s/b  
disk drives

Only Seagate spells it "disc"

---

comment number 8  
Page=40 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
share directly access

does not parse

---

comment number 9

Page=40 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
possibly  
s/b  
possibly the

---  
comment number 10  
Page=40 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
initiator devices  
s/b  
SCSI initiator devices

also in figure 3

---  
comment number 11  
Page=40 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
application clients place  
s/b  
that application clients place

---  
comment number 12  
Page=41 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
OBSD (see 3.1.27) logical unit  
s/b  
OSD logical unit

That's the phrase used in the next sentence, among other places

---  
comment number 13  
Page=41 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
This OSD object  
s/b  
This type of OSD object

since there may be more than one. As written, it sounds like there is just one.

---  
comment number 14  
Page=41 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
This OSD object  
s/b  
This type of OSD object

since there may be more than one. As written, it sounds like there is just one.



---

comment number 15  
Page=41 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
This OSD object  
s/b  
This type of OSD object

since there may be more than one. As written, it sounds like there is just one.

---

comment number 16  
Page=41 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
it  
s/b  
the user object

---

comment number 17  
Page=42 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
assigned by  
s/b  
are assigned by

---

comment number 18  
Page=42 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
assigned by  
s/b  
are assigned by

---

comment number 19  
Page=42 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
Footnote a also applies to "well known collections and "collection or user object", since those rows have multiple Partition\_ID values to choose from as well.

---

comment number 20  
Page=44 Subtype=Text Subj=Note Author=RElliott  
Comment=  
The "A CREATE USER" and "A REMOVE" cells should be left justified to match the column headers

---

comment number 21  
Page=45 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=

which  
s/b  
that

---  
comment number 22  
Page=45 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
REFERESH  
s/b  
REFRESH

---  
comment number 23  
Page=46 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
REFERESH  
s/b  
REFRESH

---  
comment number 24  
Page=47 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
device sever  
s/b  
device server

---  
comment number 25  
Page=59 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
conditions(  
add space

---  
comment number 26  
Page=77 Subtype=Text Subj=Note Author=RElliott  
Comment=  
With snapshots allowing partitions to share parts of there data, it is more likely that an error will affect multiple partitions. For c) A) c), it may be worth reporting all affected Partition\_IDs rather than just one, or the extent of damage will be unclear.

---  
comment number 27  
Page=108 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
SHAPSHOT  
s/b  
SNAPSHOT

---  
comment number 28  
Page=110 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=  
device sever  
s/b  
device server

---

comment number 29  
Page=110 Subtype=StrikeOut Subj=Cross-Out Author=RElliott  
Comment=  
Delete  
the

---

comment number 30  
Page=112 Subtype=Text Subj=Note Author=RElliott  
Comment=  
Should OSD add an FUA\_NV bit to distinguish between these data locations:

FUA: data is safe on the medium. (if the OSD controller is using disk drives as its medium, the disk drives can be moved to another OSD controller)

FUA\_NV: data is safe either in NV cache or on the medium. (disk drives cannot be moved without also moving the NV cache)

---

comment number 31  
Page=117 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
OSD logical units shall use descriptor format sense data (see SPC-3) to report all errors.

While the intent is admirable, this is a bit too aggressive (like SAT-2's attempt to mandate descriptor format for the ATA PASS-THROUGH commands).

1. OSD ought to honor the D\_SENSE bit in the Control mode page. If fixed format is requested, then return very abbreviated (almost useless) fixed-format data rather than return another format.

2. SAM-4/SPC-4 require that fixed format be used to report reset unit attention conditions and the codes needed to run MODE SELECT to turn on fixed-format, so old software just doesn't get confused

---

comment number 32  
Page=124 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
216  
s/b  
228

to match the value in table 59.

May be better to say  
"the value specified in table 59" to avoid this problem.

---

comment number 33

Page=126 Subtype=Text Subj=Note Author=RElliott

Comment=

The definition of DPO in 4.14 and 5.2.3 is a bit misleading - it sounds like a weak "should" version of FUA, advising the device server to not put data in the cache.

Really, DPO signals that the application client does not expect to read the data again, so the device server need not retain the data in the cache in hopes of a cache hit on a read.

To maintain reasonable write performance, however, it is important that the device server temporarily put the data in its cache, performing the write to medium when convenient. DPO should not cause a drop to the FUA level of performance.

---

comment number 34

Page=126 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

ADDITIONAL LENGTH

s/b

smallcaps

---

comment number 35

Page=134 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

0h

s/b

00h

to match width of other values in the table

---

comment number 36

Page=146 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

Service action codes values between...

s/b

For the operation code 7Fh, service action codes values between...

and join with the next sentence to share that "For..."

There are some non-7Fh opcodes in the table, and this comment does not apply to their service action values.

---

comment number 37

Page=205 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

REBUILD IN PROGRESS

was intended for SCC (RAID volumes) to report that they are rebuilding a RAID volume. OSD performing an object check seems a bit different. I think

a new additional sense code is worth adding for this reason. The progress indicator will make more sense if the operation in progress is clearly identified. The time for OBJECT STRUCTURE CHECK to complete is based on the amount of metadata; the time for a RAID rebuild is based on the amount of data.

Same comment applies to 6.22.3.

---

comment number 38

Page=206 Subtype=Text Subj=Note Author=RElliott

Comment=

With snapshots, several partitions might share the same data. It seems like OBJECT STRUCTURE CHECK might need to ensure that all commands to all such partitions are terminated in 6.22.3.

---

comment number 39

Page=211 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

QUERY UNIT ATTENTION

s/b

QUERY ASYNCHRONOUS EVENT

to match final SAM-4 terminology

---

comment number 40

Page=228 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

REFERESH

s/b

REFRESH

---

comment number 41

Page=240 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

partition(

add space

---

comment number 42

Page=249 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

device sever

s/b

device server

---

comment number 43

Page=264 Subtype=Highlight Subj=Highlight Author=RElliott

Comment=

The field that the Maximum CDB continuation length field establishes the

upper limit for is a 4-byte field. Should the length of this attribute be 4 bytes to match?

---

comment number 44  
Page=332 Subtype=Highlight Subj=Highlight Author=RElliott  
Comment=  
F1  
s/b  
F1h

---

comment number 45  
Page=335 Subtype=Text Subj=Note Author=RElliott  
Comment=  
Consider adding a footnote explaining how all the OSD-1 opcodes were made obsolete and replaced in OSD-2. Item d) on page 2 already explains that, but this table listing opcodes would be a good place to highlight it.

---

comment number 46  
Page=335 Subtype=Text Subj=Note Author=RElliott  
Comment=Add "(part n of 2)" to table B.1 header

\*\*\*\*\*

Comments attached to No ballot from Kevin Butt of IBM Corp.:

IBM comment number 1  
Page=29 Subtype=Text Author=satran  
Comment=  
Here you talk about a single capability while in 3.1.16 you say "not the first". In credential you may want to say also a capability or several capabilities

IBM comment number 2  
Page=36 Subtype=Text Author=satran  
Comment=  
the use of term object is somewhat confusing here (as we talk about storage objects). You may want to consider "modules" or something else

IBM comment number 3  
Page=41 Subtype=Text Author=satran  
Comment=  
LBAs are not defined here. You may want to remove the statement that reffers to LBAs (it contains no information)

IBM comment number 4  
Page=41 Subtype=Text Author=satran  
Comment=This is hard to parse (can be parsed ambiguously)

IBM comment number 5

Page=43 Subtype=Text Author=satran  
Comment=

has to clarify uniqueness within a partition. The "assigned by OSD LU" is a bit confusing (perhaps explain)

IBM comment number 6

Page=50 Subtype=Text Author=satran  
Comment=with

IBM comment number 7

Page=61 Subtype=Text Author=satran  
Comment=

I think that this formulation is problematic. I would like to see a clause that says that at least the effected changes are reflected in results (sense?)

IBM comment number 8

Page=80 Subtype=Text Author=satran  
Comment=

that statement is not completely correct. I would state it that "as a result of the communications between SM and client the client should be able to build a capability and should have the cap-key. The rest of the text calls a combination of those two a credential and uses for it a data-structure similar to those used in communications with the data server for ease of illustration"

IBM comment number 9

Page=81 Subtype=Text Author=satran  
Comment=mising "is"

IBM comment number 10

Page=82 Subtype=Text Author=satran

Comment=should mention that only cap-keys have to be confidential

IBM comment number 11

Page=98 Subtype=Text Author=satran  
Comment=

A flow diagram of the checks might help a lot the reader and perhaps replace the text

IBM comment number 12

Page=112 Subtype=Text Author=satran  
Comment=

should not say something about "no space" as opposed to quota exhaustion.

IBM comment number 13

Page=117 Subtype=Text Author=satran

Comment=are all references consistently to SPC3 or SPC4

IBM comment number 14

Page=126 Subtype=Text Author=satran  
Comment=mention table at 172?

IBM comment number 15

Page=132 Subtype=Text Author=satran

Comment=

In 5.28 and others can't the error report be more specific - NOT supported xxx - or have the form major, minor with even more detail

IBM comment number 16

Page=134 Subtype=Text Author=satran  
Comment=be updated

IBM comment number 17

Page=144 Subtype=Text Author=satran  
Comment=

that is better but I would add UNSUPPORTED to all field that are illegal due to lack of a feature

IBM comment number 18

Page=153 Subtype=Text Author=satran  
Comment=I would prefer this command removed

IBM comment number 19

Page=221 Subtype=Text Author=satran  
Comment=

It should be stated explicitly that an area of an object cleared with a clear command or having as content the default value of 0 in all bytes may be reported as a data hole!

\*\*\*\*\*

Comments attached to Abs ballot from Gregory Tabor of Maxim Integrated Products:

Maxim has no expertise in this topic and defers to those with informed opinions.

\*\*\*\*\*

Comments attached to Abs ballot from Mark Overby of Nvidia Corp.:

NVIDIA can provide no significant review of the document due to lack of technical knowledge of OSD.

\*\*\*\*\*

Comments attached to No ballot from Gerald Houlder of Seagate Technology:

Seagate comment number 1  
Page=22 Subtype=Highlight Author=IrenS  
Comment=Should be 2008.

Seagate comment number 2



Page=22 Subtype=Highlight Author=IrenS  
Comment=Grammar. Either drop "of" or reword the sentence.

Seagate comment number 3

Page=42 Subtype=Highlight Author=IrenS  
Comment=

The description of the root object lists several data manipulation commands that are not allowed (READ, WRITE, APPEND). There are now other data manipulation commands (CLEAR and PUNCH). Should these commands be added to the list? Or, maybe the description should be generalized, such as "The device server shall terminate all commands that are defined to manipulate user data (e.g., READ, WRITE, APPEND, ...) that are sent to the root object with CHECK CONDITION status, ..."

Seagate comment number 4

Page=43 Subtype=Highlight Author=IrenS  
Comment=

There are now other data manipulation commands (CLEAR and PUNCH). Should these commands be added to the list? Or, maybe the description should be generalized, such as "The device server shall terminate all commands that are defined to manipulate user data (e.g., READ, WRITE, APPEND, ...) that are sent to the partition object with CHECK CONDITION status, ..."

Seagate comment number 5

Page=43 Subtype=Highlight Author=IrenS  
Comment=

Seagate comment number 6

Page=43 Subtype=Highlight Author=IrenS  
Comment=

There are now other data manipulation commands (CLEAR and PUNCH). Should these commands be added to the list? Or, maybe the description should be generalized, such as "The device server shall terminate all commands that are defined to manipulate user data (e.g., READ, WRITE, APPEND, ...) that are sent to the collection object with CHECK CONDITION status, ..."

Seagate comment number 7

Page=45 Subtype=Highlight Author=310425  
Comment=Spelling error in "REFERESH"

Seagate comment number 8

Page=46 Subtype=Highlight Author=310425  
Comment=

Seagate comment number 9

Page=46 Subtype=Highlight Author=310425  
Comment=Spelling error in "REFERESH"

Seagate comment number 10

Page=46 Subtype=Highlight Author=310425  
Comment=Should be "other than TRACKING"

Seagate comment number 11

Page=46 Subtype=Highlight Author=310425  
Comment=Should be "an object"

Seagate comment number 12  
Page=47 Subtype=Highlight Author=310425  
Comment=Should be "server"

Seagate comment number 13  
Page=47 Subtype=Highlight Author=IrenS  
Comment=Wrong reference, should be 5.2.7, not 5.2.5

Seagate comment number 14  
Page=48 Subtype=Highlight Author=IrenS  
Comment=This notation is not defined until section 4.8.5

Seagate comment number 15  
Page=49 Subtype=Highlight Author=IrenS  
Comment=  
States that the denial of write access to an object with members means denial of the ability to create new members in that object. It should also deny the ability to remove member objects.

Seagate comment number 16  
Page=49 Subtype=Highlight Author=IrenS  
Comment=  
Once the object accessibility attribute is set to 1, can it be set to zero?  
Seems like this write protects the object AND ATTRIBUTES, preventing it from being set back to 0. That is, once an object is made read-only, it cannot be reverted back to read/write mode.

Seagate comment number 17  
Page=50 Subtype=Highlight Author=310425  
Comment=Should be "associated with each"

Seagate comment number 18  
Page=50 Subtype=Highlight Author=IrenS  
Comment=  
It is inconsistent that this paragraph indicates that multi-object commands can retrieve or store attributes to multiple objects, while table 6 indicates that only the GET and SET MEMBER OBJECTS command can do this.

Seagate comment number 19  
Page=54 Subtype=Highlight Author=IrenS  
Comment=  
The error conditions described in subparagraphs "a" and "b" refer to an "invalid attribute length". This is not relevant to the error being described. The error is the attribute number.

Seagate comment number 20  
Page=68 Subtype=Highlight Author=IrenS  
Comment=Leave a space between sentences.

Seagate comment number 21  
Page=70 Subtype=Highlight Author=IrenS  
Comment=

Seagate comment number 22  
Page=70 Subtype=Highlight Author=IrenS  
Comment=  
Shouldn't this require CREATE and WRITE permission bits for destination?

Seagate comment number 23  
Page=70 Subtype=Highlight Author=IrenS  
Comment=  
Shouldn't this require CREATE and WRITE permission bits for destination?

Seagate comment number 24  
Page=72 Subtype=Highlight Author=IrenS  
Comment=Why is this APPEND and not WRITE?

Seagate comment number 25  
Page=73 Subtype=Highlight Author=IrenS  
Comment=  
Looks like READ and WRITE are swapped here. Main partition should be WRITE and snapshot partition should be READ for the RESTORE PARTITION command.

Seagate comment number 26  
Page=74 Subtype=Highlight Author=IrenS  
Comment=  
The first rows indicate that the GET\_ATTR bit is required to get attribute from the Current Command Page. The description of GET\_ATTR in 4.11.2.2.1 specifically says that GET\_ATTR is NOT required to access the Current Command Page attributes (see page 39, 3rd paragraph).

Seagate comment number 27  
Page=74 Subtype=Highlight Author=IrenS  
Comment=

Seagate comment number 28  
Page=74 Subtype=Highlight Author=IrenS  
Comment=

Seagate comment number 29  
Page=77 Subtype=Highlight Author=IrenS  
Comment=Bad grammar: "established affected...".

Seagate comment number 30  
Page=79 Subtype=Highlight Author=IrenS  
Comment=  
The OBJECT STRUCTURE CHECK command should be included in the list of allowed commands.

Seagate comment number 31  
Page=81 Subtype=Highlight Author=IrenS  
Comment=Bad grammar, missing word(s).

Seagate comment number 32  
Page=85 Subtype=Highlight Author=IrenS  
Comment=This should be "authorized", not "unauthorized".

Seagate comment number 33

Page=88 Subtype=Highlight Author=IrenS  
Comment=Grammar: drop the word "contains"

Seagate comment number 34  
Page=96 Subtype=Highlight Author=IrenS  
Comment=Grammar: repeated words "in the in the".

Seagate comment number 35  
Page=104 Subtype=Highlight Author=IrenS  
Comment=Replace the word "show" with "shown"

Seagate comment number 36  
Page=117 Subtype=Highlight Author=IrenS  
Comment=  
This is no longer true. OSD-2 (section 4.11.3) defines a new exception management mechanism that should take care of deferred errors.

Seagate comment number 37  
Page=117 Subtype=Text Author=IrenS  
Comment=Marked set by IrenS

Seagate comment number 38  
Page=124 Subtype=Highlight Author=IrenS  
Comment=  
States that Additional CDB Length should be 216. The table says it should be 228.

Seagate comment number 39  
Page=126 Subtype=Highlight Author=IrenS  
Comment=Can we use a defined verb (shall/should) instead of "is"?

Seagate comment number 40  
Page=127 Subtype=Highlight Author=IrenS  
Comment=Byte offset is incorrect. Should start at 52.

Seagate comment number 41  
Page=128 Subtype=Highlight Author=IrenS  
Comment=Byte offset is incorrect. Should start at 52.

Seagate comment number 42  
Page=130 Subtype=Highlight Author=IrenS  
Comment=Byte offset is incorrect. Should start at 52.

Seagate comment number 43  
Page=142 Subtype=Highlight Author=IrenS  
Comment=  
The CDB Continuation Descriptor Length is a constant (16). It currently refers to "n", which is undefined in this table. The paragraph that describes the descriptor length can be specific - it must be 16.

Seagate comment number 44  
Page=143 Subtype=Highlight Author=IrenS  
Comment=

Seagate comment number 45

Page=143 Subtype=Highlight Author=IrenS  
Comment=

The CDB Continuation Descriptor Length is a constant (20). It currently refers to "n", which is undefined in this table. The paragraph that describes the descriptor length can be specific - it must be 20.

Seagate comment number 46

Page=143 Subtype=Highlight Author=IrenS  
Comment=

We suggest the removal of COPY USER OBJECTS command from this specification as agreed in the OSD twg. This feature will be revisited for future versions of OSD.

Seagate comment number 47

Page=147 Subtype=Highlight Author=IrenS  
Comment=REMOVE COLLECTION should have footnote "b".

Seagate comment number 48

Page=153 Subtype=Highlight Author=IrenS  
Comment=

We suggest the removal of COPY USER OBJECTS command from this specification as agreed in the OSD twg. This feature will be revisited for future versions of OSD.

Seagate comment number 49

Page=156 Subtype=Highlight Author=IrenS  
Comment=

Indicates that attribute list type Fh should be used. The referenced section says that list type Fh is obsolete. It should be type Eh.

Seagate comment number 50

Page=160 Subtype=Highlight Author=IrenS  
Comment=

Wrong reference. Should be 5.2.7, not 5.2.5. Check all other places as well.

Seagate comment number 51

Page=169 Subtype=Highlight Author=IrenS  
Comment=Wrong reference. Should be 5.2.7, not 5.2.5

Seagate comment number 52

Page=194 Subtype=Highlight Author=IrenS  
Comment=

This only really applies to the get attributes parameters. "Set" should be removed.

Seagate comment number 53

Page=200 Subtype=Highlight Author=IrenS  
Comment=

This only really applies to the get attributes parameters. "Set" should be removed.

Seagate comment number 54

Page=200 Subtype=Highlight Author=IrenS

Comment=What does Flash have anything to do with this command?

Seagate comment number 55  
Page=215 Subtype=Highlight Author=IrenS  
Comment=Wrong reference. Should be 5.2.7, not 5.2.5

Seagate comment number 56  
Page=229 Subtype=Highlight Author=IrenS  
Comment=Wrong reference. Should be 5.2.7, not 5.2.5

Seagate comment number 57  
Page=236 Subtype=Highlight Author=IrenS  
Comment=  
Not clear on what this paragraph is trying to address. The command tracking attributes page is removed by this command, which certainly qualifies as modifying them.

Seagate comment number 58  
Page=237 Subtype=Highlight Author=IrenS  
Comment=Wrong reference. Should be 5.2.7, not 5.2.5

Seagate comment number 59  
Page=240 Subtype=Highlight Author=IrenS  
Comment=Wrong reference. Should be 5.2.7, not 5.2.5

Seagate comment number 60  
Page=261 Subtype=Highlight Author=IrenS  
Comment=Missing the following entry: "P+7h"

Seagate comment number 61  
Page=267 Subtype=Highlight Author=IrenS  
Comment=Default isolation method attribute should be 110h, not 111h.

Seagate comment number 62  
Page=267 Subtype=Highlight Author=IrenS  
Comment=Supported isolation methods attribute should be 111h, not 112h.

Seagate comment number 63  
Page=272 Subtype=Highlight Author=IrenS  
Comment=missing "C0h reserved".

Seagate comment number 64  
Page=272 Subtype=Highlight Author=IrenS  
Comment=Should be 2FFh, not 1FFh.

Seagate comment number 65  
Page=278 Subtype=Highlight Author=IrenS  
Comment=Grammar: missing "is" between "object" and "the"

Seagate comment number 66  
Page=323 Subtype=Highlight Author=IrenS  
Comment=  
On behalf of Panasas:  
For purposes of capacity accounting, it is helpful to know how much space a given command has consumed. can we add (if we don't already have) a specific attribute on the "current command" page or some such

thing which returns "capacity delta" for the given command [type = int64]? i realize this is a little late in the game, but this was pointed out to me pretty recently at panasas, and should be pretty simple and non-intrusive. i believe there are some commands which already return this as part of the standard response code, so having a common dropping area for this would be cleaner as well.

\*\*\*\*\*

Comments attached to No ballot from Roger Cummings of Symantec:

1) since at least Revision 3, the description of atomicity in section 4.9.2 Atomicity includes:

#### 4.9.2 Atomicity

The atomicity guarantees described in this standard apply to the following (in priority order):

- 1) Commands being processed for which GOOD status has not yet been returned; and
- 2) Commands for which GOOD status has been returned, with the caveat that an application client has no way to distinguish between atomicity effects and effects caused by media failures after it receives the GOOD status.

} If a media error is detected while a command is being processed, the  
 } atomicity guarantees affect how much data may have been transferred  
 } before the error was detected.

If data transfer errors (e.g., media errors, NVRAM failures, power loss) cause a data loss after GOOD status has been returned, then the atomicity guarantees provide one of the boundaries for which data may have been lost. Detecting and recovering from such errors is described in 4.11.3.

i don't like the marked paragraph since it seems to imply that data must be read or written by the device sequentially, and that's not what higher performance disks do today.

for example, if you send a disk a full track write, it will typically start writing with whatever sector happens to be passing under the head (once the head reaches the appropriate track). it doesn't wait for the first sector of the write to appear.

bad block re-vectoring has a similar effect. if the 3rd sector in a write has been re-vectorred, it will be read or written after the other sectors in the read or write. so an i/o error may not be detected until all sectors other than the failed one have been transferred.

we can say that the atomicity guarantees affect the boundaries of how

much data is affected by a media error detected when an operation is in progress (presumably the media error will occur at an atomic i/o boundary), but that's about it.

note that neither of the last two paragraphs really requires any particular behavior.

2) in section 5.2.4 Capability, the choice of a word may be poor:

The capability is defined in 4.11.2.2. Any security method other than NOSEC (see 4.12.4) is design to return an error unless ...  
 ^^^^^^

"design" seems like a poor choice of word. i suggested "required", but according to Ralph Weber that's a bad idea:

> RE Q1: 'required' is a standards word that will cause T10 readers  
 > to expect to see exactly how the security method other than NOSEC  
 > returns an error ... inline ... in the same paragraph as the word  
 > 'required'. Sorry, but 'designed' is the best word I can think of  
 > to put in the cited text.

perhaps "is design to" could be changed to "will"?

if Ralph (or anyone else) continues to object to making a change, then i can live with "designed". but please fix the tense.

3) in Section 7.1.3.9 Partition Information attributes page there's a typo. it says:

If any object in the partition [is] the result of object duplications (see 4.13), the value of the used capacity attribute may increase for reasons that are not obvious consequences of the commands being processed as described in 4.13.5.

i think there's an "is" missing where i added it in brackets. probably "object duplications", in the same phrase, should be singular rather than plural.

4) there are a number of questions that i raise, below, about the COPY USER OBJECTS command. i think the consensus of the group was that we would simply drop the command from the OSD-2 spec. regardless, i raise the individual issues here ...

4.1) in section 6.4 COPY USER OBJECTS the command requires that the target object cannot previously exist (i.e. it's like create and write).

given that the command can copy multiple objects to the target object, and can pick out multiple ranges within each object, this seems like an odd restriction. it's not just a snapshot command, it's more of a general purpose copy command.



in it's current form my company doesn't have any particular use for the Copy User Objects command, regardless of whether or not it creates the target object, but if we work out the issues with OSD to OSD communication such that a Copy User Objects command can copy from one OSD to another, then i would very much like a Copy User Objects command that worked with pre-existing objects.

do we need a Copy User Objects command that works with existing objects? (my inclination would be to change this to require an existing object, but we could also add another copy command that works with existing objects.)

4.2) in section 6.4 COPY USER OBJECTS, the command allows the user to specify the "duplication method". while i can imagine how a copy-on-write implementation of copying an entire object would work, i have a more difficult time imagining copy-on-write for a duplicated object that's composed of a bunch of pieces from a bunch of other source objects.

assuming a given OSD implementation can support a copy-on-write copy of one object to another, but cannot support copy-on-write for an object composed of pieces of 27 other objects, what should the implementation report for the "Supported object duplication method attributes" attribute on the root attributes page (Section 7.1.2.8)?

should it report that it only supports byte-by-byte copies for "Copy User Objects", or should it report that it supports many space efficient types, but then fail the "Copy User Objects" command if any kind of range based-copying is requested?

one possible solution is to have a "clone user object" that only makes a writable clone of an entire existing user object, and another command for "Copy Data", that simply copies data from one or more ranges in one or more user objects to a target object.

4.3) in section 5.4.5 Copy user object source, the specification of the CPY\_ATTR bit (set to copy user attributes), says:

... If the CPY\_ATTR bit is set to one, all application client settable attributes (see 7.2.1) are copied from this source user object to the destination user object. ...

in the context of COPY USER OBJECTS, this is confusing. "Object Logical Length" is a client settable attribute. how is it treated? presumably, in the case where only part of the original object is copied, or multiple objects copied, we won't change the target object logical length to be the same as one of the source objects.

it seems to me that Object Logical Length should not be copied even if CPY\_ATTR bit is set. if there's a single source object that's copied in it's entirety then it should be the same length as the source object anyway.

the draft does describe what happens for the "Reserved Data Space"

attribute (it is not copied, it is added to what already exists for the duplicated object).

5) in section `_4.31.4.1 Time of duplication source object management_`, in Table 44, the entry for "DO NOT CARE" says:

The duplicated object may have any contents of the source object, including contents that were not in effect at either the beginning or the end of the duplication.

it's not clear to me how loose this is intended to be. is there a requirement that the duplicated object be a snapshot or not? in my view, a "snapshot" would represent the state of the source object at \*some\* point in the past, but i don't think this statement makes that a requirement. for example, on UNIX/LINUX systems if you copy a file that's being written to in a random fashion (say start of file, end of file, start of file, end of file, ...) using the cp command:

```
# cp busyfile busyfile_copy
```

the copy will consist of data that was in the source object at some point in time, but not all of the data will have been present at the same point in time (you'll get some of the writes to the end of the file, but not the corresponding writes to the beginning of the file).

i suggest that we clarify this and (possibly) add another Time of Duplication method such that one is "snapshot corresponds to the contents of the source object at some point in time while the copy operation was in progress", and DO NOT CARE is "snapshot may not correspond to any point in time".

6) in the case where the object is duplicated using one of the above methods (either DO NOT CARE or the new one that i suggested), is it reasonable to expect that the timestamps in the duplicated object will indicate the exact time that the copy occurred?

we should probably specify the behavior one way or the other. my preference, is for the object time stamp to represent the exact time that the copy occurred (assuming the copy is consistent).

7) in Table 36, one of the Row labels is inadequate:

The source partition attribute in the Snapshots Information attributes page (see 7.1.3.30) indicates the `_primary_` partition from which this partition is descended. [For snapshots, this indicates a primary or clone partition, and for clones this indicates a snapshot partition.]

clones may have a primary or snapshot partion as their source and snapshots may have a primary or clone partion. i've added words to indicate that.

8) Following Table 86, the text indicates that The IMMED\_TR bit is

described in 5.2.5, but it is not. it it described in 5.2.7.

9) in 6.7 "Create Clone", one paragraph says:

The get and set attributes parameters are described in 5.2.6. The format of the Data-In Buffer and Data-Out Buffer when attributes are being retrieved or set is described in 4.15. The destination Partition\_ID assigned by the CREATE SNAPSHOT [CREATE CLONE] command may be obtained from the Current Command attributes page (see 7.1.3.31).

i've indicated, in brackets, where it should say CREATE CLONE.

10) in 6.7.3 "Processing after the IMMED\_TR bit takes effect, if any", there's a grammar error where it says:

The membership and attributes of the snapshot/clone tracking well known collection for the destination partition should be maintained [so as] to [facilitate] restarting of an interrupted CREATE CLONE command with the minimum of repeated work ...

i added some words, in brackets, to try and fix it. the same wording exists in the CREATE SNAPSHOT section.

11) in 6.10.2, for "Create Snapshot", there's a typo:

The destination snapshot partition shall be added as the newest entry in the history change [chain] as described in 6.30.5.

i think the same error is repeated elsewhere.

12) the DETACH CLONE, REFRESH SNAPSHOT OR CLONE, and RESTORE PARTITION FROM SNAPSHOT command have checks against create completion time and referesh completion time. i assume the reason for these checks is to try and interlock the operations (so a clone can't be detached while it's being created, refreshed, or restored).

unfortunately, the checks don't really accomplish this. for example, the DETACH CLONE command specifies that the command will fail if ...

c) The create completion time attribute is undefined (see 3.1.51) and the refresh completion time attribute is undefined

this serves to protect partition while it's being created (because both timestamps are undefined), but does nothing to interlock with a refresh operation (since the create completion time attribute remains unchanged throughout the operation).

for the most part, these checks would accomplish what i assume they're intended to do if we made the following changes:

- 1) set Refresh Completion Time and Restore Completion Time attributes to the creation time when a partition is created
- 2) continue the current behavior of making Refresh Completion Time undefined while the refresh is in progress
- 3) make restore completion time undefined while a restore is in progress
- 3) change the checks for "refresh completion time is undefined \*and\* creation completion time attribute is undefined" to or instead, and add checks against the refresh completion time attribute

this would also require that primary partitions have a restore completion time attribute that is set when they are created, which may not be possible.

as a suggestion, instead of using the sundry completion times for the purpose of interlocking, it would probably be simpler to simply check if a tracking well known collection exists in the partition being detached, refreshed, or created and fail the operation if it exists.

or we could even leave the mechanism out entirely, and simply specify that if one of the other operations is in progress, the DEATCH CLONE operation will fail (no specification of how this is determined).

there is a generic mechanism described in 4.6.6.6 for to insure only one multi-object command is running on a partition (or is it collection), but the snapshot and clone commands are not multi-object commands.

13) similarly, in 6.10 the CREATE SNAPSHOT command doesn't seem to have any interlocking with the REFRESH CLONE and RESTORE PARTITION FROM SNAPSHOT.

i think it should be an error to create a snapshot of a partition that is currently being modified by one of those other command.

14) there may be a similar race condition with REMOVE PARTITION and RESTORE PARTITION FROM SNAPSHOT. what happens (i.e. what error is returned) if a REMOVE PARTITION is executed on a partition while RESTORE PARTITION FROM SNAPSHOT is running with the same partition as the target?

15) in 7.1.3.30 "Snapshots Information attributes page", duplication method does not appear as an attribute in the Snapshots Information attributes page. i think it should.

adding this is also required in some later comments.

16) in 6.30 "REFRESH SNAPSHOT OR CLONE" and 6.35 "RESTORE PARTITION FROM SNAPSHOT", the duplication method can be defined, which means that it can be different than the duplication method used originally to create the partition.

for refresh, we should require that the same duplication method be used

to refresh the partition as was used to create. for restore, we should require that the same duplication method used to create the snapshot be used to restore the source partition.

(if someone believes that the REFRESH SNAPSHOT OR CLONE should also support changing the duplication method, then i might withdraw this objection. however, i think it would be better to have a separate command for that purpose since this one will both change the duplication method and, potentially, change the contents).

17) in 6.30 "REFRESH SNAPSHOT OR CLONE", the mechanism whereby an incomplete command can be completed by the REFRESH SNAPSHOT OR CLONE command seems to have a flaw or two.

While the snapshot/clone tracking well known collection is, roughly speaking, supposed to contain the set of objects that still need processing, the text of 6.30.2 says:

The snapshot/clone tracking well known collection (see 4.6.6.5.3) shall be update in the destination partition to include at least the following:

- a) Every user object and collection in the source partition shall have their User\_Object\_ID (see 4.6.5) or Collection\_Object\_ID (see 4.6.6) inserted as a member of the TRACKING collection (see 4.6.6.3); and
- b) The Command Tracking attributes page (see 7.1.3.20) shall be initialized to include at least the following:
  - A) The percent complete attribute shall be set to zero;
  - B) The active command status attribute shall be set to 88ABh (i.e., REFRESH SNAPSHOT OR CLONE command in progress); and
  - C) The ended command status attribute shall be set to FFFFh.

this doesn't make sense if we're completing an operation that was already in progress. in particular, adding all of the objects back to the collection seems to preclude processing only the objects that weren't completed.

18) in 6.35 "RESTORE PARTITION FROM SNAPSHOT", i don't see a mechanism to prevent a partition from being restored from two different snapshots at the same time. we should create such a mechanism, or at least define the error to be returned if this is attempted.

19) in 6.35 "RESTORE PARTITION FROM SNAPSHOT" and 6.30 REFRESH SNAPSHOT OR CLONE, i don't see any clauses which require objects be removed from the target partition. i would assume that after the command runs, the target partition is an exact duplicate of the original partition, which should include removing any objects that were created after the snapshot was created.

20) in 6.35.4 "Command completion", there's a typo:

If the RESTORE PARTITION FROM SNAPSHOT command processing [is] complete (i.e., if the percent complete attribute in the Command Tracking attributes page (see 7.1.3.20) of the snapshot/clone tracking well known collection

i've added the missing "is" in brackets.

21) in 7.1.3.8 Root information attributes page, after Table 163, there is a paragraph which states:

If it is defined (see 3.1.14), the maximum snapshots count attribute (number 1C1h) shall contain the non-zero number that is the largest value allowed in any snapshots count attribute in any Snapshots Information attributes page (see 7.1.3.30). If the maximum snapshots count attribute is defined, the following commands shall be supported:

- a) The CREATE SNAPSHOT command (see 6.10);
- b) The REFRESH SNAPSHOT command (see 6.30); and
- c) The RESTORE PARTITION FROM SNAPSHOT command (see 6.35).

however, there is a separate attribute, 311h Support for snapshot refreshing that is supposed to control whether or not snapshots can be refreshed (see text above Table 167):

If it is defined (see 3.1.14), the support for snapshot refreshing attribute (number 311h) (see table 167) shall indicate how the REFRESH SNAPSHOT command (see 6.30) is supported. If the support for snapshot refreshing attribute is undefined (see 3.1.51), then the REFRESH SNAPSHOT command is not supported.

there are two obvious ways to fix this. because i think a proliferation of optional features is a bad idea, i'd suggest that we change the above text to require that the attribute be defined.

however, if we feel that this feature needs to be optional (separate from CREATE SNAPSHOT), then RESTORE PARTITION FROM SNAPSHOT should also be optional even when CREATE SNAPSHOT is defined. but we should lump REFRESH SNAPSHOT and RESTORE PARTITION FROM SNAPSHOT together (i.e. support both or neither).

\*\*\*\*\*

Comments attached to Abs ballot from Michael Fogg of TycoElectronics:

Tyco Electronics is only involved in the Physical Layer

\*\*\*\*\* End of Ballot Report \*\*\*\*\*