ENDL
TEXAS

Date: 15 June 2001
To: T10 Technical Committee
From: Ralph O. Weber
Subject: SPC-2 Public Review Comments Resolution

During the SPC-2 (ANSI NCITS:351) first public review comments were received from the ANSI Editor and:

- IBM — T10/01-161r0
- Congruent Software — T10/01-174r0
- Veritas —T10/01-175r0

This document describes how those comments will be resolved and the changes that will be made between SPC-2 revisions 19 and 20 to implement the resolutions. It is the editor's opinion that at least one of the changes made to resolve the public review comments is substantive, necessitating a second public review.

Note: PDF page references are to SPC-2 revision 19.

1. IBM Comments

1.1 Subclause 5.5.3.2 (PDF page 46)

Comment

The following paragraph:

> The capability of preserving persistent reservations and registration keys across power cycles requires the use of a nonvolatile memory within the SCSI device. Any SCSI device that supports the persist through power loss capability of persistent reservation and has nonvolatile memory that is not ready shall allow the following commands into the task set:

Should be changed to:

> The capability of preserving persistent reservations and ~~registration~~ reservation keys across power cycles requires the use of a nonvolatile memory within the SCSI device. Any SCSI device that supports the persist through power loss capability of persistent reservation and has nonvolatile memory that is not ready shall allow the following commands into the task set:

This change is required because there is no such thing as a 'registration key' defined in SPC-2. The correct term in 'reservation key'.

Response

While 'reservation key' is the correct term, a 'reservation key' is one property of a 'registration'. It is equally correct to replace 'registration keys' with 'registrations' and such a replacement is better since the reader can more clearly see that 'reservations and registrations' are to different things whereas 'reservations and reservation keys' might easily be thought of as redundant names for the same thing.

Resolution

The cited paragraph will be changed to:

> The capability of preserving persistent reservations and ~~registration reservation keys~~ registrations across power cycles requires the use of a nonvolatile memory within the SCSI device. Any SCSI device that supports the persist through power loss capability of persistent reservation and has nonvolatile memory that is not ready shall allow the following commands into the task set:

1.2 Clause 5.5.3.4 (PDF page 48)

Comment

The following paragraph:

> If a PERSISTENT RESERVE OUT with a REGISTER AND IGNORE EXISTING KEY service action is sent when an established registration key exists, the registration shall be superseded with the specified service action reservation key. If a PERSISTENT RESERVE OUT with a REGISTER AND IGNORE EXISTING KEY service action is sent when there is no established registration key, a new registration shall be established.

Should be changed to:

> If a PERSISTENT RESERVE OUT with a REGISTER AND IGNORE EXISTING KEY service action is sent when an established ~~registration~~ reservation key exists, the registration shall be superseded with the specified service action reservation key. If a PERSISTENT RESERVE OUT with a REGISTER AND IGNORE EXISTING KEY service action is sent when there is no established ~~registration~~ reservation key, a new registration shall be established.

This change is required because there is no such thing as a 'registration key' defined in SPC-2. The correct term in 'reservation key'.

Response

While 'reservation key' is the correct term, a 'reservation key' is one property of a 'registration'. It is equally correct to replace 'registration key' with 'registration' and such a replacement is better since it improves the consistency of usage for 'registration' throughout the paragraph. Also to add clarity, 'the registration' will be changed to 'that registration'.

Resolution

The cited paragraph will be changed to:

> If a PERSISTENT RESERVE OUT with a REGISTER AND IGNORE EXISTING KEY service action is sent when an established ~~registration reservation key~~ registration exists, ~~the~~ that registration shall be superseded with the specified service action reservation key. If a PERSISTENT RESERVE OUT with a REGISTER AND IGNORE EXISTING KEY service action is sent when there is no established ~~registration reservation key~~ registration, a new registration shall be established.

1.3 Clause 7.21.4 (PDF page 176)

Comment

The following paragraph:

> Superseding reservations is mandatory if the RELEASE(10) command is implemented. An application client that holds a current logical unit reservation may modify that reservation by issuing another RESERVE command to the same logical unit. The superseding RESERVE command shall release the previous reservation state when the new reservation request is granted. The current reservation shall not be modified if the superseding reservation request cannot be granted. If the superseding reservation cannot be granted because of conflicts with a previous reservation, other than the reservation being superseded, then the device server shall return RESERVATION CONFLICT status.

Should be changed to:

> Superseding reservations is mandatory if the RELEASE(10) command is implemented. An application client that holds a current logical unit reservation may modify that reservation by issuing another RESERVE command to the same logical unit. The superseding RESERVE command shall release the previous reservation state when the new reservation request is granted. The current reservation shall not be modified if the superseding reservation request ~~cannot be~~ is not granted. If the superseding reservation cannot be granted because of conflicts with a previous reservation, other than the reservation being superseded, then the device server shall return RESERVATION CONFLICT status.

This change is required because the term 'cannot' is not in the keyword list and therefore has to be removed.

Response & Resolution

The proposed change will be made as written.

1.4 Clause 8.4.1 (PDF page 226)

Comment

The following paragraph:

> This subclause describes the vital product data page structure and the vital product data pages (see table 167) that are applicable to all SCSI devices. These pages are optionally returned by the INQUIRY command (see 7.3) and contain vendor specific product information about a target or logical unit. The vital product data may include vendor identification, product identification, unit serial numbers, device operating definitions, manufacturing data, field replaceable unit information, and other vendor specific information. This standard defines the structure of the vital product data, but not the contents.

Should be changed to:

> This subclause describes the vital product data page structure and the vital product data pages (see table 167) that are applicable to all SCSI devices. These pages are optionally returned by the INQUIRY command (see 7.3)~~. and contain vendor specific product information about a target or logical unit. The vital product data may include vendor identification, product identification, unit serial numbers, device operating definitions, manufacturing data, field replaceable unit information, and other vendor specific information. This standard defines the structure of the vital product data, but not the contents~~.

This change is required because VPD data pages are now defined that go beyond what is currently defined. By deleting the indicated text the currently defined VPD pages become legal.

Response & Resolution

The proposed change will be made as written.

1.5 Clause 7.3.1 (PDF page 99)

Comment

The following paragraph:

> If the standard INQUIRY data changes for any reason, the device server shall generate a unit attention condition for all initiators (see SAM-2). The device server shall set the additional sense code to INQUIRY DATA HAS CHANGED.

Should be changed to:

> If the standard INQUIRY data changes for any reason, the device server shall generate a unit attention condition for all initiators (see SAM-2). The device server shall set the additional sense code to INQUIRY DATA HAS CHANGED. If INQUIRY VPD data changes for any reason, the device server may generate a unit attention condition for all initiators (see SAM-2). The device server shall set the additional sense code to INQUIRY VPD DATA HAS CHANGED.

This change is required to allow a VPD data change to generate a unit attention.

Response

The juxtaposition of 'shall' and 'may' in the proposed new sentences could led to unacceptable questions about whether the new unit attention is a requirement or a suggestion.

Resolution

The cited text will be changed to:

> If the standard INQUIRY data changes for any reason, the device server shall generate a unit attention condition for all initiators (see SAM-2). The device server shall set the additional sense code to INQUIRY DATA HAS CHANGED. If INQUIRY VPD data changes for any reason, the device server may generate a unit attention condition for all initiators (see SAM-2), setting~~. The device server shall set~~ the additional sense code to INQUIRY VPD DATA HAS CHANGED.

INQUIRY VPD DATA HAS CHANGED is not currently defined as an additional sense code. ASC/ASCQ 3Fh/12h will be assigned to INQUIRY VPD DATA HAS CHANGED. The Additional Sense Code tables in 7.20.6 and C.2 will be updated to reflect the new additional sense code.

1.6 Clause 7.13.6 (PDF page 134)

Comment

The following paragraph:

> The READ BUFFER command shall return the same number of bytes of data as received in the prior echo buffer mode WRITE BUFFER command from the same initiator. If a prior echo buffer mode WRITE BUFFER command was not successfully completed the echo buffer mode READ BUFFER command shall terminate with a CHECK CONDITION status, the sense key shall be set to ILLEGAL REQUEST and the additional sense code to COMMAND SEQUENCE ERROR. If the data in the echo buffer has been overwritten by another

initiator the target shall terminate the echo buffer mode READ BUFFER command with a CHECK CONDITION status, the sense key shall be set to ABORTED COMMAND and the additional sense code to ECHO BUFFER OVERWRITTEN.

Should be changed to:

The READ BUFFER command shall return the same number of bytes of data as received in the prior echo buffer mode WRITE BUFFER command from the same initiator. If the allocation length is insufficient to accommodate the number of bytes of data as received in the prior echo buffer mode WRITE BUFFER command, the data returned shall be truncated as described in 4.3.4.6, and this shall not be considered an error. If a prior echo buffer mode WRITE BUFFER command was not successfully completed the echo buffer mode READ BUFFER command shall terminate with a CHECK CONDITION status, the sense key shall be set to ILLEGAL REQUEST and the additional sense code to COMMAND SEQUENCE ERROR. If the data in the echo buffer has been overwritten by another initiator the target shall terminate the echo buffer mode READ BUFFER command with a CHECK CONDITION status, the sense key shall be set to ABORTED COMMAND and the additional sense code to ECHO BUFFER OVERWRITTEN.

The current text for Read Data from echo buffer (1010b) as written implies the number of bytes returned is the same as the number of bytes received in the previous Write echo buffer command and makes no mention of allocation length. A strict reading could interpret this to mean the allocation length is ignored for this mode. The above change clears this up.

Response & Resolution

The proposed change will be made as written.

2. Congruent Software Comment

Comment — Subclause 7.2.6 (PDF page 68)

In clause 7.2.6, "Target descriptors", there is no information for a target descriptor for IEEE 1394. I believe that this omission is inappropriate since Serial Bus Protocol 2 (SBP-2) is an NCITS-developed transport protocol suitable for SCSI. It is, in fact, elsewhere mentioned in the SPC-2 draft standard (see table 165, table C.5).

This omission is relatively simple to rectify. I have not included proposed text with this EMail since I wish to present it in the same format used by draft standard SPC-2. I believe that I will have the proposed text available before the close of the public comment period on May 21. But in the event that I fail to send you the text by the deadline, I wish this public comment to reflect my ability and willingness to provide the supplemental text to T10 in a timely fashion.

Because of the long development cycle for many T10 projects, I think it is very desirable to remedy this omission in SPC-2 before its approval as a standard. I believe that the remedy is not controversial and is therefore unlikely to significantly extend the approval process for SPC-2.

Response & Resolution

Work on definition of new target descriptors is continuing in SPC-3, see T10 document T10/398r3 already approved by T10 for inclusion in SPC-3. Also, it is possible that as yet incomplete work exemplified by T10/00-425r3 may obviate the need for a specific IEEE 1394 target descriptor. For these reasons and because there is no specific proposal for an IEEE 1394 target descriptor, the urgency for making changes in SPC-2 has not been demonstrated.

No changes will be made.

3. Veritas Comment

Comment — Subclause 5.5.3.6.1 (PDF page 50)

Introduction

VERITAS has decided to take the serious step of submitting a Public Review comment on this draft standard because of experience gained in testing prototype equipment that is compatible with the Persistent Reservation requirements in the SPC-2 draft along with our clustering applications.

Background

One of the major uses for the Persistent Reservations defined for the first time in SPC-2 is expected to be in clusters. VERITAS has a cluster application that can support up to 32 nodes, and which assumes a Storage Area Network shared by all of those nodes is used to provide the connection to storage. VERITAS there intends to use Persistent Reservations to provide controlled access from a subset of the nodes in each cluster to each storage device, via a Write Exclusive - Registrants Only Persistent Reservation. When the cluster is created, each node will register with a set of storage devices as directed by the management logic, and the reservation will be initiated.

Problem

When one of the nodes in the cluster is to be taken offline for maintenance, the shutdown procedure will cause the node to remove its registration from its set of storage devices. If this node was the one that originally established the reservation, then the following text from subclause 5.5.3.6.1 (page 32) applies:

> When a reservation key has been removed, no information shall be reported for that unregistered initiator in subsequent READ KEYS service action(s) until the initiator is registered again (see 5.5.3.4). Any persistent reservation associated with that unregistered initiator shall be released. If that released persistent reservation was of the type Write Exclusive - Registrants Only or Exclusive Access - Registrants Only the device server shall establish a unit attention condition or all registered initiators other than the initiator that issued the PERSISTENT RESERVE OUT command with PREEMPT or PREEMPT AND ABORT service action. The sense key shall be set to UNIT ATTENTION and the additional sense data shall be set to RESERVATIONS RELEASED.

VERITAS has a number of problems with both the definition and the text in this paragraph, as follows:

1) This definition has a significant impact on the remaining nodes in the cluster, each of which now have to execute the following process:
   a. Respond to the resulting Unit Attention condition;
   b. Clear the resulting Auto Contingent Allegiance condition;
   c. Interact with the other nodes to determine which should reissue the reservation;
   d. Reissue the Write Exclusive - Registrants Only Persistent Reservation if selected;
   e. Communicate with all other nodes that the reservation has been re-established;
   f. Reissue all of the commands that completed with ACA Active status.
2) The end result of the process in 1) above is only to recreate a reservation that is almost identical to the reservation that existed before the node was taken offline!
3) The process in 1) above is a significant impediment to scalability. We are attempting to expand our clustering products to support hundreds of nodes, and tens of thousands of disk drives. For each node/disk pair the process in 1) above has to be performed at a time when those nodes are under full load since a configuration change is being processed and cluster resources are reduced. Note that this happens even in the case of a graceful shutdown of a node.
4) There is a time period between steps a. and d. in the process described in 1) above where an unregistered node can access the device and cause data corruption.

5) The text in the first paragraph under Problem above that states "If this node was the one that originally established the reservation" is actually an assumption on the part of VERITAS. The text in the quoted paragraph states *"Any persistent reservation associated with that unregistered initiator"*. The verb "associated" is only used elsewhere in the Persistent Reservation section in terms of keys associated with a reservation. Elsewhere (e.g. in 5.5.3.3.3 b)), the text describes *"the initiator that holds the persistent reservation"*. Is the assumption correct?

Suggested Changes

VERITAS would like to amend the definition of the Registrants Only reservations such that the reservation survives as long as there are registrations in place for that Logical Unit. Thus even if the initiator that holds the reservation unregisters, the reservation survives. This will convert the Registrant Only reservations to a true peer basis, and will mean that there is NO processing required on the remaining cluster nodes after a node unregisters. There will also be no time period during that process during which a competing reservation can be established or data can be corrupted, and no Unit Attention conditions have to be created.

Suggested text for the paragraph quoted from subclause 5.5.3.6.1 (page 32) above is:

> When a reservation key has been removed, no information shall be reported for that unregistered initiator in subsequent READ KEYS service action(s) until the initiator is registered again (see 5.5.3.4). Any Registrants Only persistent reservation held by that unregistered initiator shall be released only when no other registrations exist for that Logical Unit, otherwise the reservation shall not be effected. Any non Registrants Only persistent reservation held by that unregistered initiator types held by that initiator shall be released.

Also, the following additional paragraph is suggested following the one above:

> When a Registrants Only persistent reservation continues after the initiator that originally created the reservation unregisters, the device server shall report the reservation key of any of the initiators with remaining registrations in a PERSISTENT RESERVE IN command with a READ RESERVATION service action.

Response & Resolution

The changes requested substantially alter the behavior intended by T10 when defining the Persistent Reservations feature. The intent of T10 is that releasing a reservation should be part of a controlled software shutdown process during which cooperating members of a cluster should use Persistent Reservation commands to gracefully transfer reservations ownership to remaining cluster members. The behavior described in SPC-2 is consistent with that intent.

Furthermore, changing the definition in this way at this late date will materially harm those who worked with T10 during the process of defining Persistent Reservations and as such would be a disservice to the industry.

If the new behavior proposed is desired, it should be proposed for SPC-3 as a new Persistent Reservation Type (e.g., Write Exclusive - All Registrants).

No changes will be made.

4. Comments from the ANSI Editor

4.1 Subclause 3.5 (PDF page 29)

Comment

Regarding:

> [Result =] Procedure Name (IN ([input-1] [,input-2] …]), OUT ([output-1] [,output-2] …))

After the ellipsis, there is a closing bracket that has no corresponding opening bracket. Should it be deleted? If not, where should the opening bracket be placed?

Response & Resolution

The "...])" should be "...)". The identified square bracket will be deleted.

4.2 Subclause 7.2.1 (PDF page 62)

Comment

On page 44, in table 15, the bit that follows bit n+1 is labeled n+1+l. Should the "l" be changed to a "1"?

Response & Resolution

The "l" (letter l) should not be changed to "1" (number one). However, you have noted a lack of clarity since l and 1 look similar.

The "l" (letter l) will be changed to "s" (letter s). This may require a change in the column width, but I think that is acceptable.

4.3 Subclause 7.10.4.1 (PDF page 124)

Comment

On page 106, in 7.10.4.1, is it OK to change the cross-reference from "table 38" to "table 69" to match the table shown in the subclause?

Response & Resolution

The incorrect cross reference with be fixed as proposed.

4.4 Subclause 7.20.2 (PDF page 155)

Comment

On page 137, the parts of the second level of the list in the middle of the page were changed from "a)" and "b)" to "1)" and "2)" to avoid confusion with the main parts of the list. Changes OK?

Response

It is unacceptable to change lettered lists to numbered lists at any level in the list hierarchy. Numbered lists indicate a specific order of processing requirement whereas lettered lists indicate no ordering requirements.

Resolution

Throughout SPC-2 second level lettered lists will be changed from lower case letters to upper case letters to avoid confusion with main parts of the list.

4.5 Subclause 7.26.1 (PDF page 182)

Comment

On page 164, in 7.26.1, is it OK to change the cross-reference from "table 77" to "table 119" to match the table shown in the subclause?

Response & Resolution

The incorrect cross reference with be fixed as proposed.

4.6 Subclause 8.3.2 (PDF page 207)

Comment

On page 189, in 8.3.2, is it OK to change the cross-reference from "table 91" to "table 146" to match the table shown in the subclause?

Response & Resolution

The incorrect cross reference with be fixed as proposed.

4.7 Subclause 10.2 (PDF page 236)

Comment

On page 218, in 10.2, is it OK to change the cross-reference from "table 121" to "table 184" to match the table shown in the subclause?

Response & Resolution

The incorrect cross reference with be fixed as proposed.

5. Non-Substantive Comments From Other Sources

5.1 Subclause 7.20.2 (PDF page 155)

Comment & Resolution

In the last sentence of the first list item a), "two bytes" will be changed to "four bytes". The sentence is describing the INFORMATION field and the size of the INFORMATION field is four bytes not two bytes.

5.2 Subclause 8.4.4 (PDF page 230)

Comment & Resolution

In Table 174, "Vendor ID (see annex C)" will be changed to "Vendor ID (see annex D)". Vendor IDs are listed in Annex D, not Annex C.