

Fibre Channel Tape Profile Meeting
Tucson, Arizona, 16 December 1998
Stewart Wyatt, Secretary

T11/98-619v0

1. Introductions: Chairman Dale Lafollette (STK) called the meeting called to order at 12:35 PM and had the participants introduce themselves.

2. Approval of this Agenda: T11/98-592v0

The agenda was approved with the addition of a new item, 6.1 FCP-2: Bob Snively

3. Approval of 11/3 Minutes: T11/98-567v0

Minutes were approved.

4. Review of Old Action Items:

Secretary Stewart Wyatt, HP, reviewed the action items.

#1 Stewart Wyatt will get a resolution on HP SCSI command comments. Resolved in HP letter ballot comments.

#2 Erich Oetting will propose to the T10 working group that the Write Exclusive Persistent Reservations be made illegal for SSC devices. Erich was not present, deferred to next T10 meeting.

#3 Erich Oetting will propose to the T10 working group that the communications devices command set be made obsolete. Erich was not present, deferred to next T10 meeting.

#4 Dave Peterson will attempt to document EOD behavior more precisely in SSC. Dave has posted a new version of the SSC, deferred to the next T10 meeting.

#5 Dave Peterson will enter a letter ballot to correct the I5 IU specified in FC_Tape. Completed.

#6 Bob Snively will take a recommendation to the T10 working group to obsolete the Flag bit in the Control byte. Completed.

#7 Bob Snively will enter a letter ballot comment to make FC-Tape compatible with the changes in FCP-2 regarding handling of CRN's. (8.2.1.1 and Annex D.) Completed.

#8 Bob Snively will enter a letter ballot comment to remove annexes form FC-Tape for items documented in FCP-2. Completed.

5. Letter Ballot Comment Review:

Editor Dave Peterson, STK, had brought a list of letter ballot comments that he wanted to have the

group review. The other comments he felt he could handle on his own.

#1. Bob Snively (SUN) Comment ID 47. The REC does not provide enough information to create a properly formed SRR.

Bob Snively, SUN, referred to figure 10 in Annex B. The REC accept does not provide relative offset for resumption of a transfer with the Enable Modify Data Pointers (EMDP) bit set. Matt Wakeley, HP, responded that this figure was just requesting the retransmission of the transfer ready and would not exhibit a problem.

Neil Wanamaker, Crossroads, suggested that Figure 17 in Annex B was a better example as it includes an error on a data transfer. The REC indicates the number of bytes transferred but with the EMDP bit set the sequences are not required to be sent in-order so it is not always clear where the recovery should begin. Bill Martin, Gadzooks, proposed that on a read failure the SRR could request the RO at the beginning of the transfer. As the problem occurs when reads operations are streamed, Matt Wakeley suggested precluding streaming reads, which Bob observed would seriously limits tape performance. Dale suggests retransmitting the whole record again, which he clarified was the entire command, typically 256K bytes. Write transfers do not have the problem since the transfer ready limits the recovery to only the last sequence.

Two possible solutions were considered: Specify the REC and SRR more completely or give the host the option that redo the entire transfer. Neil Wanamaker noted that reads are less common than writes so reducing the performance of the read may be acceptable. Dal Allan, ENDL, noted that recovery is a high anxiety moment - reducing the performance of the recovery may not be good.

Rob Basham, IBM, interjected a case he was concerned about which he called the fixed block problem. In this case recovery becomes more difficult when it starts in the middle of a physical block. Dale replied that the solution is to not release the data buffer until you are certain that the host got it all so that you do not have to recover the data from the tape.

The conclusion was to recover streamed reads from a relative offset of zero if there is any ambiguity, meaning that the entire command is re-transferred. If there is no ambiguity, recovery can begin from some later position in the transfer.

Bob Snively, who is the editor of the FCP-2, noted that some of these issues belong in the FCP-2. He would like the issues to be resolved by this group and placed in annexes in the FC-Tape document. He will put them in the FCP-2 and they can be removed from tape document before it is published.

Agenda item 6.1

For the convenience of the editor, there was a break in the letter ballot review while Bob Snively reviewed the FCP-2 status.

Bob provided an overview of the changes in FCP-2 from FCP. The first topic was mode pages.

The first one Bob mentioned was the Disconnect-reconnect page. The next was the Fibre Channel Logical Unit Control page (18). Interesting bits on this page are the Maximum Burst Size field (which defines the maximum sequence length) and EMDP bit. The next mode page is the Fibre Channel Logical Unit Control page which contains the EPDC which enables the precise delivery function (command CRNs). The last one mentioned was the Fibre Channel Port Control Page (19) which contains port configuration bits.

In reviewing this page, Tak Seto, Quantum, noted that these definitions were inconsistent with previous documentation. Bob noted that these functions had been defined in several different documents and that the FCP-2 would be the final authority.

Pak noted that DTOLI requires that FC-AL-2 behavior be required and would obsolete FC-AL devices. He asked if FCP-2 should require FC-AL2 compliance. This was discussed. It was noted that many of the bits required FC-AL2 to be implemented anyway. This secretary was of the opinion that the discussion conclusion was that AL2 compliance was required. Jeff Williams, Western Digital, and Jim Coomes, Seagate, agreed that these features needed to be reviewed and discussed. Having Bob collect them into the FCP-2 is providing a forum for the review to occur.

Jim Coomes noted that an additional bit had been proposed for the Port Control page. He called it Enable auto response and it had been located in byte 4, bit 0. This bit is necessary because some initiators cannot accept the response in the same latency as the data on a read operation.

George Penoki, IBM, expressed concern that Bob Snively, The FCP-2 editor was making technical changes to the FCP-2 without obtaining approval by submitting a formal request to the appropriate groups. Bob replied that some of the tape development will be moved into FCP-2 without a new proposal, as he felt this requirement was met by the tape discussions. He considered this presentation to be the formal proposal for several of the features discussed.

Other changes to the FCP-2 include adding Confirmed delivery, RR_TOV, an extended CDB proposal, eliminating operation associators, eliminating sequences with mixed command and data as well as mixed data and response. The concept of confirmation is being dropped.

After Bob concluded the agenda was returned to agenda item 5, reviewing the letter ballots comments:

#2 Matt Wakeley (HP) Comment titled: Public behavior not documented.

Stewart Wyatt noted that there are several places in the profile that presume private behavior that need to be corrected. The biggest area of concern with public behavior was the subject of authentication after a loop initialization. Currently the profile refers to the FLA for public initialization and copies material from the PLDA for private loop initialization. Jeff Stai, Brocade, said he had heard some hallway talk about a suggestion to generate a generic SCSI and Fabric behavior. Bob Snively thought that would belong in FCP-2. Stewart Wyatt felt that a common initialization process needed to be defined that demonstrated the process of discovering that the fabric was not present and proceeding to private loop initialization. Stewart was also interested in how hosts on remote loops were informed of initialization changes. Jeff said that was done through state change

notification. Stewart noted that this was prohibited in this profile which he thought was an indication that the process had not been thought through thoroughly.

In conclusion editor Dave Peterson took an action item to provide an annex covering the whole process by enlisting the help of a few experts.

#3 Matt Wakeley (HP) Comment titled: The Tape Profile should not prohibit establish commands.

Dal Allan groaned, noting that this is the same argument that has been rehashed many times. Dal argued that if people would only read the document definition of prohibit this would not be a problem. George Penoki agreed with Dal but sided with the letter ballot comment that he felt that prohibited is too strong of a term. Bob Snively noted that there are two parts to compliance. All required commands must be implemented and all other commands must never be required. There was some discussion of compliance testing.

Bob reiterated his goal that he doesn't want to have to consult individual manuals to write a driver. He just wants to use the compliant command list. Dave Peterson felt that limiting commands exceeded the scope of the profile. The goal is to leverage existing SCSI solutions to Fibre Channel. Rob Basham noted that the command set on his Fibre Channel and SCSI drives will be the same. Rob's drives will support every required command. He sees the list as a systems house issue and is irrelevant to tape drives in the short term. However it may become relevant in the future.

Dave questioned the relevance of this activity. Dal and Bob disagreed that a common command set is necessary for interoperability which is within the scope of the project. Dal Allan noted how difficult it is to remove commands. There is substantial cost savings by reducing the number of commands supported because it reduces the testing required for proving compliance. Dave Peterson agreed but said this effort belongs in the SSC. Rob Basham felt it was a secondary issue that should not delay the profile. Matt Wakeley felt that this issue is not transport specific and should apply to any transport therefore it is not relevant to this group. Dal agreed that Matt would be correct if this is a standard, but this is an interoperability profile so it is an appropriate activity.

Rob Basham proposed the profile will only contain commands and parameters that are required for compliance to the profile. This proposal passed on a letter ballot vote 16 to 3.

Matt Wakeley asked for a vote on a proposal to remove the tables that contain the compliant commands, the tables 26 and 27, entirely. This vote failed 2 to 12.

The editor was instructed to remove all of the commands that are prohibited from the table and update statements about interoperability.

#4. Matt Wakeley (HP) Comment (untitled) on alerting an initiator to a CRN sequencing problem.

The problem occurs if the target is unable to get the initiator to respond to a missing CRN. The conclusion was after RR_TOV to explicitly log out the initiator.

Stewart Wyatt questioned the assumption that the target would retain a command received out-of-order. Dale LaFollette replied that the target was expected to execute the commands in order though they might not be received in-order.

#5 Charles Binford (LSI) CommentID 6. Class 3 and Link Control.

The note referred to in this comment suggests that an ACK should be rejected in Class 3. A discussion followed about appropriate behavior and confusion in other documents. Matt Wakeley proposed removing the comment from profile and defer to FC-PH. This was accepted.

#6 Charles Binford (LSI) CommentID 8

Charles requested that the profile requesting text for clause 5.13 which is on sequence and exchange management. Previous versions of the profile included text copied from the PLDA. The group had decided to remove the text and refer to FC-PH. The group voted 9 to 0 to reject the comment (Replacing text previously removed.) Note that Charles Binford was not present to defend his comment.

#7 Charles Binford (LSI) CommentID 9 LPE/LPB should be I not P.

Stewart Wyatt noted that a footnote promotes this feature while the table prohibits it. Bob Snively and Dale LaFollette argued that feature is not useful and that the footnote should be removed. Matt Wakeley was concerned about loops with non-tape devices which may need this feature as he felt it put hosts in a no-win situation. A host may need to use LPEfx which applies to all devices on the loop and may be useful in some environments.

The question was put to a vote: Shall initiators be allowed to initiate LPE/LPB? (Changes P to A) 11 yes 2 no. A second question was voted on: Should the footnote be eliminated altogether? 5 voted yes and 8 voted to keep it. A third vote was should the note be changed? Approved 13 to 0. The note should be modified to say: Allowed so that the initiator may manipulate non-tape devices.

#8 Robert Snively (SUN) CommentID 26. Incorrect value for FCP_TOV

Bob argued that the only reasonable use for FCP_TOV is as a polling interval for REC. The group agreed and the editor was instructed to remove item (a) "time FCP level information unit reply Sequences" from clause 7.6. The group also agree to change timer name to REC_TOV to emphasize its function.

#9 Bob Snively (SUN) CommentID 24. Problem with E_D_TOV definitions.

Matt noted some errors in Bob's comments which Bob accepted. E_D_TOV definition is flawed since it doesn't consider delays in the loop. (This needs to be fixed but is a project for another group.) Matt Wakeley pointed out that a fabric attached initiator is entirely unaware of the presence of a public loop that a target is attached to. Bob proposed that the timer be large enough that the loop issues would be covered. The comment was accepted with Matt's corrections and a note

that this solution needs to be implemented until the underlying problem are addressed by another group.

Bob was concerned about implementing multiple E_D_TOV timers to time every outstanding sequence. A long discussion followed to convince Bob.

#10 Bob Snively (SUN) CommentID 25 Incorrect definition of RR_TOV.

Bob proposed setting RR_TOV to a constant value of 16 seconds. Rob Basham said that RR_TOV has to be greater than E_D_TOV. Rob likes to have the timers defined with respect to each other so that the order in which they expire is always the same. This goal can also be achieved if they are all defined by constant values. The proposed value was questioned as being too short for some observed exchange authentications to be completed as observed in the interoperability tests. A question was raised about having multiple timers running when a loop initialization occurs. Bill Martin noted that is unavoidable since a fabric connected initiator will have an E_D_TOV time-out running and will be unaware that the loop is initializing. A long discussion followed about how the counters relate to each other. Then Bob Snively led a discussion to calculate the worst case value for RR_TOV.

Stewart Wyatt pointed out that increasing RR_TOV requires a target to keep more status assuming that status cannot be discarded until after RR_TOV expires. Rob Basham gave an example of receiving many inquiries from different initiators where most of them don't send a subsequent command to confirm delivery of the previous status. Matt Wakeley corrected the assumption that RR_TOV was the required time to retain the status. The correct value is 3 times REC_TOV. Rob Basham told the group that they needed to get used to the idea that targets could not keep large amounts of data around for very long. The group decided to separate the RR_TOV issue from holding status and consider only exchange authentication for this discussion.

Dave Baldwin suggests a constant of 5 minutes since he has seen much longer periods than 16 seconds. Jim Coomes thought that 5 minutes was a reasonable value from his experience with disk drives. Note that RR_TOV is a tunable parameter in Mode page 19. Dave's proposal was accepted.

#11 Bob Snively (SUN) CommentID 30 Note (Clause 8.2.1) is incomplete, data overlay

The issue is how a host determines that all of the data was transferred with data overlay, since it cannot rely on counting the bytes transferred and comparing the sum to the FCP_DL when data overlay is allowed. Bob proposed that the note be changed to state that the method to determine that all of the data is transferred is beyond the scope of this profile.

Bob led the group on a tutorial of how the host might make this determination. One approach is for the initiator to depend on the target to inform it that all of the data has been transferred when error recovery occurs with the EMDP bit set. Bob thought the host could do it but that it would be so difficult that it would rely on the target. Rob Basham brought up the issue of read with SILI bit set. In this case the target cannot inform the host if the transfer length is different than what was requested. Bob corrected him by pointing out that the FCP-DL always applies and provides the

expected length. If the target transfers a different amount it reports it as a residual in the FCP_RSP. In this profile the host is required to make the expected data transfer exactly equal to the FCP_DL field.

#12 Charles Binford (LSI) CommentID12 Limit data overlay to error recovery.

It appeared that Charles wanted to limit data overlay to error recovery. Bob Snively felt that if it was allowed at all the target could invoke it at its discretion.

6. T10 New Business:

Note the discussion of the FCP-2 reported earlier.

7. T11 New Business:

None

8. Review New Action Items:

The only new action item was for the editor to create an authentication annex

9. Adjournment:

The chairman noted that there will be no meeting in January. The editor reported that he had reviewed the letter ballot comments and had covered all of the comments in the meeting today that he believed he needed guidance from the group on. He wanted to update the profile with the results of today's discussion and with the remaining letter ballot comments. He felt like he could complete this by February's meeting. A request was made for him to post the results two weeks before the February meeting to allow the group time to review the new draft before the meeting.

The meeting was adjourned about 6 PM.