

To: Membership of X3T10  
From: Gerry Houlder - Seagate  
Date: September 23, 1994  
Subject: Minutes of the XOR Study Group Meeting on September 12, 1994

X3T10/94-199R1

Attendees:

Gerry Houlder	Seagate
Jay Elrod	Seagate
Paul Hodges	IBM
Bill Hutchinson	HP
Thai Nguyen	Storage Technology <a href="mailto:thai_nguyen@stortek.com">thai_nguyen@stortek.com</a>
James McGrath	Quantum
Edward Fong	Amdahl
Larry Lamers	Adaptec
John Lohmeyer	NCR

Gerry Houlder acted as chairman for the meeting. The issues discussed are summarized below.

(1) XDWRITE command descriptor block layout - In response to criticism that the 16 byte CDB didn't follow accepted SCSI guidelines, Gerry Houlder proposed a slight rearrangement of fields to follow even byte boundaries and the general structure given in SBC. This proposal included a 3 byte secondary address field and a LongID bit to allow for an 8 byte address in the data phase. Discussion on item (2) subsequently reduced the secondary address field to one byte and eliminated need for LongID bit.

(2) Use mode page to store/define redundancy group addresses. The group preferred using a small secondary address field that contains an index into an internal table of 8 byte addresses. Gerry Houlder will draft a proposal for this. A key problem to be solved is that different areas of a target may be in different redundancy groups, and different addresses and or numbers of devices may apply to each group. A mode page that works like the notch page (page Ch) will be drafted that allows a definition of redundancy group ranges as well as the number of group members and address of each group member. The XDWRITE, REGENERATE, and REBUILD commands will need a different (and simpler) structure to make use of the addresses from the mode page.

(3) Transfer of error handling - This was a discussion of document X3T10/94-184, which was sent on the SCSI reflector earlier. The group preferred using 3<sup>rd</sup> party reservation technique for error recovery (item C in the document) because it doesn't require defining any new constructs. The XOR command document will add this procedure.

(4) Other error handling issues - The group discussed what the "primary target" should do if the "secondary target" returns Reservation Conflict or ACA Active status. The preferred response was to return Check status to the initiator and return Command Aborted sense key with new ASC for Command Blocked. The initiator should assume that the parity drive has not been updated and the data drive may be partially updated (i.e., is in an unpredictable state). Its error recovery action should include restoring both the data and parity drives.

Action of Busy or Queue Full statuses wasn't discussed, but the primary target should retry the secondary command a reasonable number of times before resorting to the Check status w/command blocked response.

(5) Multi-controller data validation problem - Paul Hodges (IBM) posed the problem of one initiator doing an update write (XDWRITE with an XPWRITE to another drive) while another initiator is doing a regenerate on the same LBAs. The regenerate operation could read new data from the data drive (because XDWRITE is done or a cache hit on new data occurs) and get old data from the parity drive (because XPWRITE hasn't happened yet).

Our conclusion is that this problem is not unique to XOR command architectures and can only be solved by having RAID controllers co-operate with each other on such activities. We didn't identify any particular implementation rules that should be added to the XOR commands.