May 15, 2000 T10/00-232 revision 0

To: John Lohmeyer, chairperson, T10
From: Bob Snively
Date: May 12, 2000
Subject: Asymmetrical SCSI behavior

1 Introduction

A significant number of SCSI storage subsystems have the property of asymmetrical accessibility to logical units through various ports. Typically, one port may provide full performance access to a logical unit, while another port, possibly on a different physical controller, provides either lower performance access or supports a subset of the available SCSI commands to the same logical unit. In some cases, the logical unit can be modified to provide full performance access to the limited port if the original full performance port fails. This proposal proposes a set of SCSI tools necessary to properly support such asymmetrical access and failure recovery.

2 Overview

Symmetrical access to logical units is very desirable, since it provides for very rapid recovery from link failures and it provides the infrastructure that supports dynamic load balancing capabilities. Symmetrical access is characteristic of almost all simple disk drives and JBODs. Symmetrical access is managed by the simple SCSI mechanisms already defined by SAM-2, SPC-2 and other documents.

Asymmetrical access is useful, since it can be implemented for very large storage sub-systems with very simple and low-cost storage controller configurations. Asymmetrical access requires additional SCSI mechanisms to indicate which ports are fully accessible and which ports have only partial access to a particular logical unit. Mechanisms are also required to allow controlled transfer of the primary port for a logical unit from one port to another. The proposals in this document provide those mechanisms.

3 Proposed mechanisms for managing asymmetrical access to SCSI logical units

3.1 General considerations

The SCC command set is not appropriate for these devices, since a true SCC device is almost never available to manage the LUNs. A command set based on the SPC-2 command set is appropriate, because it manages each LUN as an individual object behind the set of ports that can access it.

Brocade Communications Systems, Inc.
1901 Guadalupe Parkway ■ San Jose, CA 95131
T 408.487.8000 ■ F 408.487.8101
www.brocade.com

Port state

Each port or symmetrical group of ports must be in one of the following states with respect to the ability to access a particular LUN:

Active:

The port is capable of immediately accessing the logical unit. All commands operate exactly as specified by SPC-2, SBC, and SES today.

Inactive:

The port is capable of performing a limited set of SPC-2 commands. Those commands that operate behave precisely as specified by SPC-2, SBC, and SES today. Those commands that do not operate will provide the specified error indication. Commands that operate include those necessary for:

Diagnosing and testing the logical unit and its paths

Identifying the path

Identifying the logical unit

Determining the operational state of the logical unit

Determining the active/inactive state of the unit

Allowed commands include:

INQUIRY,

LOG SELECT and LOG SENSE

MODE SELECT (6/10) and MODE SENSE (6/10)

REPORT LUNS

RECEIVE DIAGNOSTIC RESULTS and SEND DIAGNOSTIC.

Task management functions to inactive ports will not effect operations or tasks enqueued through active ports.

Unavailable:

The port cannot access the requested logical unit. The only commands that operate normally are INQUIRY, REQUEST SENSE, and, for logical unit 0, REPORT LUNS. Those commands that do not operate provide the specified error indication.

Management function

The following management operation can be performed for each port:

Make port(s) active:

The port or group of ports indicated is made active. The previously active ports not on this list are made inactive. (See D below)

This function can be done explicitly or automatically. See (A) below to determine whether the device requires explicit or implicit change. [Alternatively, implicit or explicit behavior can be specified by adding a few bits to item D below.]

Discovery requirements

The following information must be discoverable by an appropriate mechanism.

Identify asymmetrical access requirement:

A value is provided to indicate that only one port group at a time is allowed normal access to a logical unit. (See A, below)

Identify port group to logical unit:

A value is provided identifying the port or port group through which a command is being passed. This can be optionally associated with a port world-wide name through either the SCSI command set or the Fibre Channel command set. (See B.1, below)

Report active port group:

A value is provided identifying the port group that is active. (See B.2, below)

Report port groupings:

A list of available port groups is provided for the logical unit. It is up to the driver to associate the ports to the logical unit with one of the port groups.

[I propose that this be extrapolated from the exhaustive following of the topology, correlated with the logical unit world-wide name. See C below]

3.2 SUMMARY OF FORMAL PROPOSAL FOR CHANGES TO SPC-2 DOCUMENT

A Section 7.5.1, Standard INQUIRY data

Asymmetric port behavior bit

A new bit will be placed in byte 5 of INQUIRY data to indicate that the logical unit has asymmetric ports and supports asymmetric port behavior. The default of 0 indicates symmetric or unspecified port behavior. The value of 1 indicates that asymmetric port behavior is observed by this logical unit and that the asymmetric port management and discovery tools are supported.

Implicit asymmetric port behavior bit

A new bit will be placed in byte 5 of INQUIRY data to indicate that the logical unit supports asymmetric behavior using only implicit port group activation. The default of zero is explicit port group activation. If the asymmetric port behavior bit is zero, the implicit asymmetric port behavior bit is reserved and shall be zero. (The committee may choose to move this to a mode select page).

B Section 8.4.3, Device Identification Page

The document presently defines a 4-byte “relative port identifier” that, when included, with the association value of 1, indicates the port (relative to some arbitrary internal ordering) through which the command is passed.

The device identification page allows simultaneous presentation of identifiers, including the port WWN, the LUN WWN, the relative port identifier, and any other similar information.

B.1

An additional 4-byte identifier is added to the available identifiers to indicate the “relative port group identifier”, using the same formats and very similar text to that for the relative port identifier. Only 16 bits of the field are allowed to be used for the identifier. The other 2 bytes are reserved.

B.2

An additional bit is placed in the identification descriptor field to indicate that the port is active.

[Note that we could probably take the top byte of the “relative port identifier” and place a “group/port” selection bit and an “active/inactive” selection bit without causing anybody any grief. That could replace B.1 and B.2.]

C Section 5, new section of model

The model needs a new section telling how asymmetrical port groups work and indicating the proper behavior. Proper error indications are defined in this section. The error indications would include:

COMMAND INVALID FOR INACTIVE PORT

COMMAND INVALID FOR UNAVAILABLE PORT

D Port group management

The obvious place to put this control field is the Control Mode Page, described in clause 8.3.4. There are 16 contiguous bits in bytes 10 and 11, allowing 64 Ki groups to be defined and tested. A “set port group” validity bit is placed in byte 4.