To: X3T9.2 Committee (SCSI)

From: George Penokie (IBM)

Subject: SCSI Disk Array Model

1. Denotes changes from rev 5 to rev 6
   + - Denotes changes from rev 6 to rev 7

1.0 Glossary

1. Check Data - Information contained within a LBA set which allows lost or destroyed user data to be reconstructed.

1. Check Data Mapping - The distribution of the check data within a LBA set.

1. Data Mapping - The distribution of user data within a LBA set.

1. Disk Array Conversion Layer (DACL) - Converts input logical unit identifiers to output logical unit identifiers and may convert input LBAs to output LBAs.

1. Data LBA Range - The addressable continuum of logical blocks which are input to the Disk Array Conversion Layer. Check Data is not part of this addressable continuum of logical blocks.

1. LBA Set - A LBA range consists of a Data LBA Range and the Check Data (Check data may be null). Sets shall not intersect and shall be independent from one another. Sets may span more than one device. A single device may contain more than one LBA set or may contain the entire LBA set.
1. LBA Set Mapping - The distribution of user data and check data (if any) within a LBA set. User data and check data shall be contained within non-intersecting ranges of logical blocks.

1. SCSI Disk Array (SDA) - A device which processes SCSI command descriptor blocks and performs the services of a disk array conversion layer.

2.0 Model for SCSI disk arrays

The following is a model which defines SCSI disk arrays. It is not intended to define any requirements nor is it intended to alter any requirements defined elsewhere in this standard.

3.0 SCSI Disk Array Requests

3.1 Maintenance Services

The Maintenance Services provides individual field replaceable unit (FRU) addressing information. This information provides a way to bypass the DACLs manipulation of the logical unit identifier and LBA.

A SCSI disk array returns the following parameters for all addressable field replaceable units under its control in the confirmation of a request for maintenance service:

- Logical Unit Identifier
- FRU Type Identifier

Any time a DACL encounters a Logical Unit Identifier which is not its own, but addresses a device under its control the DACL shall send the task, without modification, to the addressed device.

Implementors Note: This provides the means for a system to corrupt data.

3.2 Configuration Services

A SCSI disk array configuration is initialized and controlled by requests for configuration services. Configuration services are used to define the following parameters of a LBA set:

- Number of blocks
- Number of bytes per block
- Transfer Rate
- I/Os per second
- Level of integrity
- Availability
- Locality

- Method of reporting informational exception conditions

3.2.1 Number of blocks
The number of blocks parameter shall be used by the DACL
to determine the number LBAs within the Data LBA Range.

3.2.2 Number of bytes per block

The number of bytes per block shall be used by the DACL to
determine the size of the blocks within the Data LBA Range.

3.2.3 Transfer Rate

[Ed Note: Need to define a scale for Transfer Rate. GOP]

3.2.4 I/Os per second

[Ed Note: Need to define a scale for I/Os per sec. GOP]

3.2.5 Level of integrity

[Ed Note: Need to define a scale for level of integrity.
GOP]

3.2.6 Availability

[Ed Note: Need to define a scale for Availability and
a definition. GOP]

3.2.7 Locality

[Ed Note: Need to define a scale for Locality and a
definition. GOP]

3.2.8 Reporting informational exception conditions

The higher level configures the DACL to report informational
exception conditions using one of the following methods:

- Periodic Check Condition with a Recoverd Error Sense Key,
- Check Condition with Unit Attention Sense Key, or
- No Check Condition; system will poll to determine current
state of DACL.

3.3 SCSI Disk Array Services

The disk array conversion layer performs the following
services:

- Regenerate the check data within a LBA set,
- Reconstruct the user data within the LBA set using
  check data information or duplicate data,
  + - Verify user data within a LBA set,
+ - Enable/Disable a LBA range,
+ - Enable/Disable writes, and
+ - Enable/Disable generation of check data within a LBA set.

4.0 Disk Array Conversion Layer Functions
A disk array conversion layer initiates several functions when a media access service is requested. The type of media access and the mapping in effect determine which functions are used. The following are the functions available within a disk array conversion layer:

- Translation of input Logical Unit Identifiers to output Logical Unit Identifiers.
- Translation of input LBAs to output LBAs.
- Reading data from and writing data to locations based on the mapping in effect for the addressed LBA set.
- Calculates and updates the check data (if any).
- Reconstruction of user data within the LBA set using check data information or duplicate data.
- Regeneration of the check data within a LBA set.
- Determines when a LBA range should be disabled and/or replaced.
- Returns a confirmation to the higher level as to the success or failure of a service request and, in the case of a failure, gives possible corrective actions.

5.0 Exception Conditions

Exception conditions inform a higher level that:

- A change occurred to the DACL's configuration,
- A change occurred in the operating state of the DACL,
- A repair action is requested,
- A repair action is required to restore the LBA sets level of integrity,
- A repair action is required to restore the LBA sets performance,
- A repair action is required to restore the LBA sets availability, or
- An error occurred.

Which exception conditions are returned in confirmations and how often they are returned is based on configuration requests received from a higher level.

6.0 System Layering Model

A system is typically composed of many protocol conversion layers, and these layers may exist in hardware or software. Each of these layers has input(s) and output(s). The next layer accessed is determined by the preceding layers output.

These protocol conversion layers include, but are not limited to: transport modules, host adapter drivers, SIMs, HBAs, bridge controllers, and disk drives. Each of these layers will be represented by a simple block which has an input and output.

All requests to or from a DACL contain LUIs but not all requests contain LBAs.

Note: The logical unit identifier is fully defined in SAM.
6.1 Generic Layer

Generic Layers do not modify the logical unit identifier or the LBA.

Note: There are types of layers other than generic and disk array conversion layers, however, these are not covered in this model.

6.2 Disk Array Conversion Layer (DACL)

The disk array conversion layer is capable of extensive manipulation on the logical unit identifier and the LBA, based upon a consistent algorithm which follows the defined configuration. It is possible that a single LUI(INPUT) or LBA(INPUT) may be converted to multiple different LUI(OUTPUT)s and LBA(OUTPUT)s. Disk array conversion layers will be shown with 'DACL' in the block.

+ The LUI(OUTPUT)s may not be SCSI devices.

Disk array conversion layers provide the functions listed in section 4.0.

7.0 Examples of SCSI Disk Array Variations

Typically a system diagram will be composed of many layers combined into a tree. For example, a driver may connect to multiple HBAs, which in turn may connect to multiple Targets, etc. The example below shows a system which consists of:
- One initiator has two SCSI targets attached on a single SCSI bus and this SCSI bus is not expandable.
- One initiator has two target attached on a single SCSI bus and this SCSI bus is expandable. One of the targets is a DACL.
- The DACL initiator has three SCSI buses with targets attached and is capable of driving more SCSI buses. Two of the SCSI buses contain two targets each and these SCSI buses are not expandable. One of the SCSI buses contain one target and this SCSI bus is expandable.

In the following diagrams, only a single branch of this tree will be shown.
7.1 Branch of Generic Layers

This diagram shows that all layers pass the logical unit identifier and LBA directly through.

```
Operating system
    |    |    |    |
    | LUI(X) + LBA(X)    |
    v   |
---------------------------------  |
    |  |
    | Software (driver)   |
    |
---------------------------------  |
    |  |
    | HBA (Initiator)     |
    |
---------------------------------  |
    |  |
    | Drive (Target)      |
```
7.2 Software Array with a Branch of SCSI Disks

This diagram shows software performing SCSI Disk Array functions. These functions convert the input logical unit identifier (LUI(X)) and the input LBA (LBA(X)) to the output logical unit identifier (LUI(Y)) and output LBA (LBA(Y)). All other layers pass the logical unit identifier and LBA through.

```
Operating system
    \   /
    |  |
    V  V
----------
    |  |
    | DACL |
    ----------
    |  |
    |  |
    |  |
    |  |
    V  V
----------
    |  |
    | Software (driver) |
    |  |
    |  |
    |  |
    |  |
    V  V
----------
    |  |
    | HBA (Initiator) |
    |  |
    |  |
    |  |
    |  |
    V
----------
    |  |
    | Drive (Target) |
    |  |
    |  |
    |  |
    |  |
    V
----------
```
7.3 Branch with HBA SCSI Disk Array

This diagram shows a HBA performing SCSI Disk Array functions. These functions convert the input logical unit identifier (LUI(X)) and the input LBA (LBA(X)) to the output logical unit identifier (LUI(Y)) and output LBA (LBA(Y)). All other layers pass the logical unit identifier and LBA through.

```
Operating system
   / LUI(X) + LBA(X)
  /                  
/                     Software (driver)

/ LUI(X) + LBA(X)
/                  
/                     HBA (Initiator)

/ DACL
/                  
/                     Drive (Target)
```

Date: Jul, 06 1993
7.4 Branch with Bridge Controller SCSI Disk Array

This diagram shows a bridge controller performing SCSI Disk Array functions. These functions convert the input logical unit identifier (LUI(X)) and the input LBA (LBA(X)) to the output logical unit identifier (LUI(Y)) and output LBA (LBA(Y)). All other layers pass the logical unit identifier and LBA through.

```
                                 \ Operating system
                                 /                     \ Software (driver)
                                / \                     / \ HBA (Initiator)
                               /   \                   /   \ Bridge Controller
                              /     \                 /     \ (Target/Initiator)
                             /       \               /       \ Drive (Target)
                            /         \             /         \                  
                           /           \           /           \                      
                          LUI(X) + LBA(X)         LUI(X) + LBA(X)           LUI(Y) + LBA(Y)
                           \               /       \               /       \               /       \ 
                            \         /         \         /         \         /         \ 
                             \     /           \     /           \     /           \ 
                              \   /             \   /             \   /             \ 
                               \ /               \ /               \ /               \ / 
                                \                     \                     \                     
                                 \                               \                               
                                 \                             \                             
                                 \                           \                           
                                 \                       \                       
                                 \                  \                  
                               \          \          
                                \    \    
                                 \  \  
                                  \\  
                                   \ 
                                    
```

M
7.5 Branch with Multiple Layers of SCSI Disk Arrays

This diagram shows one example with two disk array controller layers. This example shows a software layer performing SCSI disk array functions and a bridge controller performing SCSI disk array functions.

The software functions convert the input logical unit identifier (LUI(X)) and the input LBA (LBA(X)) to the output logical unit identifier (LUI(Y)) and output LBA (LBA(Y)). The bridge controller functions then convert the input logical unit identifier (LUI(Y)) and the input LBA (LBA(Y)) to the output logical unit identifier (LUI(Z)) and output LBA (LBA(Z)). All other layers pass the logical unit identifier and LBA through.

```
        Operating system
          | LUI(X) + LBA(X)
        |----------------|
        | Software (driver)
        |----------------|
        | LUI(Y) + LBA(Y)
        |----------------|
        | HBA (Initiator)
        |----------------|
        | LUI(Y) + LBA(Y)
        |----------------|
        | Bridge Controller
        | (Target/Initiator)
        |----------------|
        | LUI(Z) + LBA(Z)
        |----------------|
        | Drive (Target)
```