To: X3T10 Committee (SCSI)

From: George Penokie (IBM)

Subject: More Simplification of SACL Configuration

1 SCSI Implementation

This implementation of the simplified configuration of a storage subsystem will use the CREATE/MODIFY STORAGY ARRAY CONFIGURATION service action to deliver the configuration parameters to the storage subsystem.

1.1 CREATY/MODIFY STORAGE ARRAY CONFIGURATION service action

The CREATE/MODIFY STORAGE ARRAY CONFIGURATION service action (see table 1) requests the creation of a new volume set and redundancy group, or the modification of an existing volume set and redundancy group. If the create operation fails to complete successfully the command shall be terminated with a CHECK CONDITION status. The sense key shall be set to HARDWARE ERROR, and the additional sense code set to CREATION OF LOGICAL UNIT FAILED. If the modification operation fails to complete successfully the command shall be terminated with a CHECK CONDITION status. The sense key shall be set to HARDWARE ERROR, and the additional sense code set to MODIFICATION OF LOGICAL UNIT FAILED.

Table 1 - CREATE/MODIFY STORAGE ARRAY CONFIGURATION service action

Bit Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (BFh)							
1	RESERVED			SERVICE ACTION (08h)				
2	REDUNDANCY TYPE IDENTIFIER							
3	BUSFAIL	RESERVED						
4	(MSB)	LUN_V						
5								(LSB)
6	(MSB)	LIST LENGTH						
7								
8								
9								(LSB)
10	CREATE/MODIFY		CONFIGURE		RESERVED			IMMED
11	CONTROL							

The REDUNDANCY GROUP IDENTIFIER field indicates the type of protection that shall be used within the redundancy group being created or modified. See table 2 for the format of the REDUNDANCY GROUP IDENTIFIER field.

Table 2 - **REDUNDANCY GROUP IDENTIFIERS**

Codes	Description
00h	No redundancy
01h	Copy redundancy
02h	XOR redundancy
03h	P+Q redundancy
04h	P+S redundancy
05h	S redundancy
06h-7Fh	Reserved
80h-FFh	Vendor specific

A bus fail (BUSFAIL) bit of zero indicates that the target shall be configured such that a single bus failure may cause the application client to loose access to user data within the volume set being created or modified. A BUSFAIL bit of one indicates that the target shall be configured so a single bus failure shall not cause the application client to loose access to any user data within the volume set being created or modified.

The LUN_V field specifies the address of the volume set that shall be created or modified.

An immediate (IMMED) bit of zero indicates that status shall be returned after the create/modify storage array operation has completed. An IMMED bit of one indicates that the storage array shall return status as soon as the command descriptor block has been validated, and the entire SIMPLE CREATE/MODIFIY VOLUME SET parameters list has been transferred.

The CONFIGURE field is defined in table 3.

TABLE 3 - CONFIGURE

Codes	Description
00b	Any unassigned p_extent(s) within the target that received the CREATE/MODIFY STORAGE ARRAY CONFIGURATION service action may be used to configure the selected volume set and redundancy group to the requested capacity. Any CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTORS (table 6) shall be ignored.
01b	The target shall use the CREATE/MODIFY STORAGE ARRAY CONFIGURATION parameter list (table 5) to determine the configuration of the volume set and redundancy group.
10b	All unassigned p_extents within the target that received the SIMPLIFY CREATE/MODIFY VOLUME SET service action shall be configured into a volume set and a redundancy group. The VOLUME SET CAPACITY field (table 5) and any CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTORS (table 6) shall be ignored.
11b	Reserved

The CREATE/MODIFY field is defined in table 4.

TABLE 4 - CREATE/MODIFY

Codes	Description
00b	The target shall create a volume set and a redundancy group and shall assign to the created volume set the logical unit number contained in the LUN_V field. The target shall assign to the created redundancy group a logical unit number per the addressing rules (xxx). If the addressed volume set already exists within the target the target shall modify the existing volume set and its associated redundancy group as requested in the CREATE/MODIFY STORAGE ARRAY CONFIGURATION service action.
01b	The target shall create a volume set and a redundancy group, and shall assign to the created volume set and redundancy group logical unit numbers per the addressing rules (xxx). The LUN_V field shall be ignored.
10b	The target shall modify the volume set addressed in the LUN_V field and its associated redundancy group. If the addressed volume set does not exist the target shall terminate the command with a CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST, and the additional sense code set to LOGICAL UNIT NOT CONFIGURED.
11b	Reserved

The CREATE/MODIFY STORAGE ARRAY CONFIGURATION parameter list (table 5) contains user data mapping information and a list of CREATE/MODIFY PS_EXTENT DESCRIPTORS that are used to create or modify the addressed volume set and its associated redundancy group.

Table 5 - CREATE/MODIFY STORAGE ARRAY CONFIGURATION parameter list

Bit Byte	7	6	5	4	3	2	1	0
0	(MSB)							
1				CAPACITY				
2								
3								(LSB)
4	RESERVED							
5	RESERVED							
6	(MSB)			BYTES PER BLOCK				
7								(LSB)
8	(MSB)			NORMAL USER DATA TRANSFER SIZE				
9								(LSB)
10	REBUILD/RECALCULATE RATE							
11	PERCENTAGE OF SEQUENTIAL TRANSFERS							
	CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTORS(S) (if any)							
12				CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTOR 0				
15								
	. . .							
n-3				CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTOR x				
n								

The CAPACITY field contains the size to configure the volume set and the redundance group in logical blocks. If the CREATE/MODIFY field is 10b the new size of the volume set being modified shall be set to the value in the CAPACITY field and the new size of the redundancy group shall be set to the value in the CAPACITY field.

The BYTES PER BLOCK field contains the size, in bytes, of the logical blocks in the CAPACITY field and the NORMAL USER DATA TRANSFER SIZE field. A value of zero in the BYTES PER BLOCK field shall indicate the number of bytes per logical block is 512.

The NORMAL USER DATA TRANSFER SIZE field contains the number of logical blocks the application client normally requests transferred on each user data transfer. The target shall treat the NORMAL USER DATA TRANSFER SIZE field as an advisory parameter. A NORMAL USER DATA TRANSFER SIZE field of zero indicates the application client has no information on the size of user data transfers.

The REBUILD/RECALCULATE RATE field contains the length of time the target should take to do a rebuild

operation or a recalculate operation. A value of one in the REBUILD/RECALCULATE RATE field indicates the target should use the longest rebuild or recalculate time. A value of 255 in the REBUILD/RECALCULATE RATE field indicates the target should use the shortest rebuild time. The target shall treat the REBUILD/ RECALCULATE RATE field as an advisory parameter. A REBUILD/RECALCULATE RATE field of zero indicates the application client has no information on the time to do rebuilds or recalculates. If the REDUNDANCY GROUP IDENTIFIER field contains a zero (i.e., no redundancy) then the target shall ignore the REBUILD/RECALCULATE RATE field.

NOTE 1 - The effect of different rebuild/recalculate times is to increase and decrease the performance of a target. Lower values increase performance but at a cost of being exposed to data loss for a longer time. Higher values decrease performance but keep the exposure to data loss at a minimum.

The PERCENTAGE OF SEQUENTIAL TRANSFERS field contains the percent of times a application client accesses sequential logical blocks on consecutive user data transfers. The target shall treat the PERCENTAGE OF SEQUENTIAL TRANSFERS field as an advisory parameter. A PERCENTAGE OF SEQUENTIAL TRANSFERS field of zero indicates the application client has no information on the sequentially of user data transfers. If a value greater then or equal to 100 is contained in the PERCENT OF SEQUENTIAL TRANSFERS field then the percentage of sequential transfers is set to 100% by the target.

The CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTOR contains information the target shall use to control the user data mapping within peripheral devices. See table 6 for the format of the CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTOR.

Table 6 - Data format of CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTOR

Bit Byte	7	6	5	4	3	2	1	0
0	LUN_P							
1								
3	RESERVED							
4	PERCENT OF USER DATA							

The LUN_P field defines the address of the peripheral device to place user data.

All fields within the CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTOR shall be bounded by the addressed peripheral device. It is not an error for a group of peripheral device(s) that define a volume set and redundancy group to contain different parameters within the CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTORS.

The PERCENT OF USER DATA field contains the percentage of the volume sets requested capacity that would placed on the selected peripheral device (e.g., a value of 20 in the PERCENT OF USER DATA field would cause the target to configure 20% of the requested capacity onto the addressed peripheral device). If a value greater then or equal to 100 is contained in the PERCENT OF USER DATA field the target shall place 100% of the requested capacity on the addressed peripheral device. If the requested capacity will not fit on the addressed peripheral device the target shall terminate the command with a CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST, and the additional sense code set to PARAMETER VALUE INVALID.

If the sum of all the PERCENT OF USER DATA fields for all the CREATE/MODIFY PERIPHERAL DEVICE DESCRIPTORs is not 100 then the target shall terminate the command with a CHECK CONDITION status.

The sense key shall be set to ILLEGAL REQUEST, and the additional sense code set to PARAMETER VALUE INVALID.

1.2 REPORT STORAGE ARRAY CONFIGURATION service action

The REPORT STORAGE ARRAY CONFIGURATION service action (see table 1) requests that information regarding the SCSI storage array's configuration be sent to the application client. If this service action requests information on a volume set that has more than one associated redundancy group the target shall terminate the command with a CHECK CONDITION status. The sese key shall be set to ILLEGAL REQUEST, and the additional sense code set to INVALID FIELD IN CDB.

Table 7 - REPORT STORAGE ARRAY CONFIGURATION service action

Bit Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (BEh)							
1	RESERVED			SERVICE ACTION (02h)				
2	RESERVED							
3	RESERVED							
4	(MSB)	LUN_V						
5								(LSB)
6	(MSB)	ALLOCATION LENGTH						
7								
8								
9								(LSB)
10	RESERVED							RPTSEL
11	CONTROL							

The LUN_V field specifies the address of the volume set for which information shall be reported per table 8. If the requested logical unit has not been configured the command shall be terminated with a CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST, and the additional sense code set to LOGICAL UNIT NOT CONFIGURED.

A report selected bit (RPTSEL) of zero indicates the target shall report on all the volume set(s) within the target that have a single associated redundancy group. The LUN_V field shall be ignored when the RPTSEL bit is zero. A RPTSEL bit of one indicates the target shall only report information on the volume set addressed by the LUN_V field.

The REPORT CONFIGURATION parameter list is defined in table 8.

Table 8 - REPORT CONFIGURATION parameter list

Bit Byte	7	6	5	4	3	2	1	0
0	RESERVED							
1	REDUNDANCY GROUP IDENTIFIER							
2	BUSFAIL	RESERVED						
3	RESERVED	STATE OF THE VOLUME SET						
4	(MSB)	CAPACITY						
5								
6								
7								(LSB)
8	RESERVED							
9	RESERVED							
10	(MSB)	BYTES PER BLOCK						
11								(LSB)
12	REBUILD/RECALCULATE RATE							
13	PERCENTAGE OF SEQUENTIAL TRANSFERS							
14	(MSB)	REPORT CONFIGURATION DESCRIPTOR LIST LENGTH						
15								(LSB)
	VOLUME SET PERIPHERAL DEVICE DESCRIPTOR(S)							
16	VOLUME SET PERIPHERAL DEVICE DESCRIPTOR (First)							
19								
	. . .							
n-3	VOLUME SET PERIPHERAL DEVICE DESCRIPTOR (Last)							
n								

The REDUNDANCY GROUP IDENTIFIER field (see table 2) indicates the type of protection being used within the redundancy group associated with the addressed volume set. For a description of the redundancy group methods see xxx.

A bus fail (BUSFAIL) bit of zero indicates that the target is configured such that a single bus failure causes the application client to loose access to user data within the addressed volume set. A BUSFAIL bit of one indicates that the target is configured such that single bus failure does not cause the application client to

loose access to any user data within the addressed volume set.

The VOLUME SET STATE field is defined in xxx.

The CAPACITY field indicates the size of the addressed volume set in logical blocks.

The BYTES PER BLOCK field indicates the size, in bytes, of the logical blocks in the CAPACITY field and the NORMAL USER DATA TRANSFER SIZE field. A value of zero in the BYTES PER BLOCK field indicates the number of bytes per logical block is 512.

The NORMAL USER DATA TRANSFER SIZE field indicates the number of logical blocks the target expects during a normal user data transfer. A NORMAL USER DATA TRANSFER SIZE field of zero indicates the target has no expectations on the size of user data transfers.

The REBUILD/RECALCULATE RATE field indicates the length of time the target takes to do a rebuild operation or a recalculate operation. A REBUILD/RECALCULATE RATE field of zero shall indicate the target has no information on how long it will take to do rebuilds or that the associated redundancy group is configured as no redundancy.

The PERCENTAGE OF SEQUENTIAL TRANSFERS field indicates the percentage of times the target expects accesses to sequential logical blocks on consecutive user data transfers. A PERCENTAGE OF SEQUENTIAL TRANSFERS field of zero indicates the target has no expectations on the sequentially of user data transfers.

The VOLUME SET PERIPHERAL DEVICE DESCRIPTOR contains a list of peripheral devices associated with the addressed volume set. See table 9 for the format of the VOLUME SET PERIPHERAL DEVICE DESCRIPTOR field.

TABLE 9 - VOLUME SET PERIPHERAL DEVICE DESCRIPTOR

Bit Byte	7	6	5	4	3	2	1	0
0	LUN_P							
1								
3	RESERVED							
4	PERCENT OF USER DATA							

The LUN_P field indicates the address of a peripheral device associated with the addressed volume set.

The PERCENT OF USER DATA field indicates the percentage of the addressed volume set's capacity that is on the selected peripheral device (e.g., a value of 20 in the PERCENT OF USER DATA field would indicate 20% of the addressed volume set's capacity is contained on the addressed peripheral device).

Annex A Example of a SCSI storage array configuration using a CREATE/MODIFY ARRAY CONFIGURATION service action

On receipt of a CREATE/MODIFY STORAGE ARRAY CONFIGURATION service action the target will examine the parameters and configure a volume set and a redundancy group using those parameters.

The contents of the NORMAL USER DATA TRANSFER SIZE field and the PERCENTAGE OF SEQUENTIAL TRANSFERS field in combination are used by the target to determine the user data stripe depth mapping.

See table 10 for an example of user data stripe depths the target in this example selects based on the contents of the NORMAL USER DATA TRANSFER SIZE field and PERCENTAGE OF SEQUENTIAL TRANSFERS field.

Table 10 - User data stripe depth mapping selection

		PERCENT OF SEQUENTIAL TRANSFERS			
		1% to 25%	25% to 50% or 0%	51% to 75%	greater than 75%
NORMAL USER DATA TRANSFER SIZE in logical blocks	1 to 128	1	2		3
	129 to 512 or 0	2		4	
	greater than 512	3	4		5

Notes:
1) user data stripe depth = 4 x normal user data transfer size
2) user data stripe depth = 2 x normal user data transfer size
3) user data stripe depth = normal user data transfer size
4) user data stripe depth = normal user data transfer size / (number of disk drives/2)
5) user data stripe depth = normal user data transfer size / number of disk drives

The contents of the REBUILD/RECALCULATE RATE field is used by the target to determine the how long the a rebuild or recalculate operation will take to complete.

See table 4 for an example of how the target in this example selects the rebuild and recalculate rates using the rebuild/recalculate rate.

.

TABLE 11 - Rebuild rate selection

Codes	Description
00h	Rebuild or recalculate at least one stripe for every read or write request from an application client. (default)
01h	Suspend rebuild and recalculate operations during all read/write requests from any application clients.
02h-1Fh	Rebuild or recalculate at least 1/8th of a stripe for every read or write request from an application client.
20h-3Fh	Rebuild or recalculate at least 1/4th of a stripe for every read or write request from an application client.
40h-5Fh	Rebuild or recalculate at least 1/2th of a stripe for every read or write request from an application client.
60h-7Fh	Rebuild or recalculate at least one stripe for every read or write request from an application client.
80h-9Fh	Rebuild or recalculate at least two stripes for every read or write request from an application client.
A0h-BFh	Rebuild or recalculate at least four stripes for every read or write request from an application client.
C0h-DFh	Rebuild or recalculate at least eight stripes for every read or write request from an application client.
E0h-FEh	Rebuild or recalculate at least 16 stripes for every read or write request from an application client.
FFh	Do not accept any read/write requests from an application client until the rebuild or recalculate operation is complete.

Note:
If the redundancy group is configured as copy redundancy or S redundancy the target in this example will rebuild in 1 MByte stripes.

The contents of the NORMAL USER DATA TRANSFER SIZE field and the PERCENTAGE OF SEQUENTIAL TRANSFERS field in combination are used by the target to determine the amount of read ahead information to transfer from the disk drives.

See table 12 for an example of amount of read ahead the target in this example selects based on the contents of the NORMAL USER DATA TRANSFER SIZE field and PERCENTAGE OF SEQUENTIAL TRANSFERS field.

Table 12 - Read ahead selection

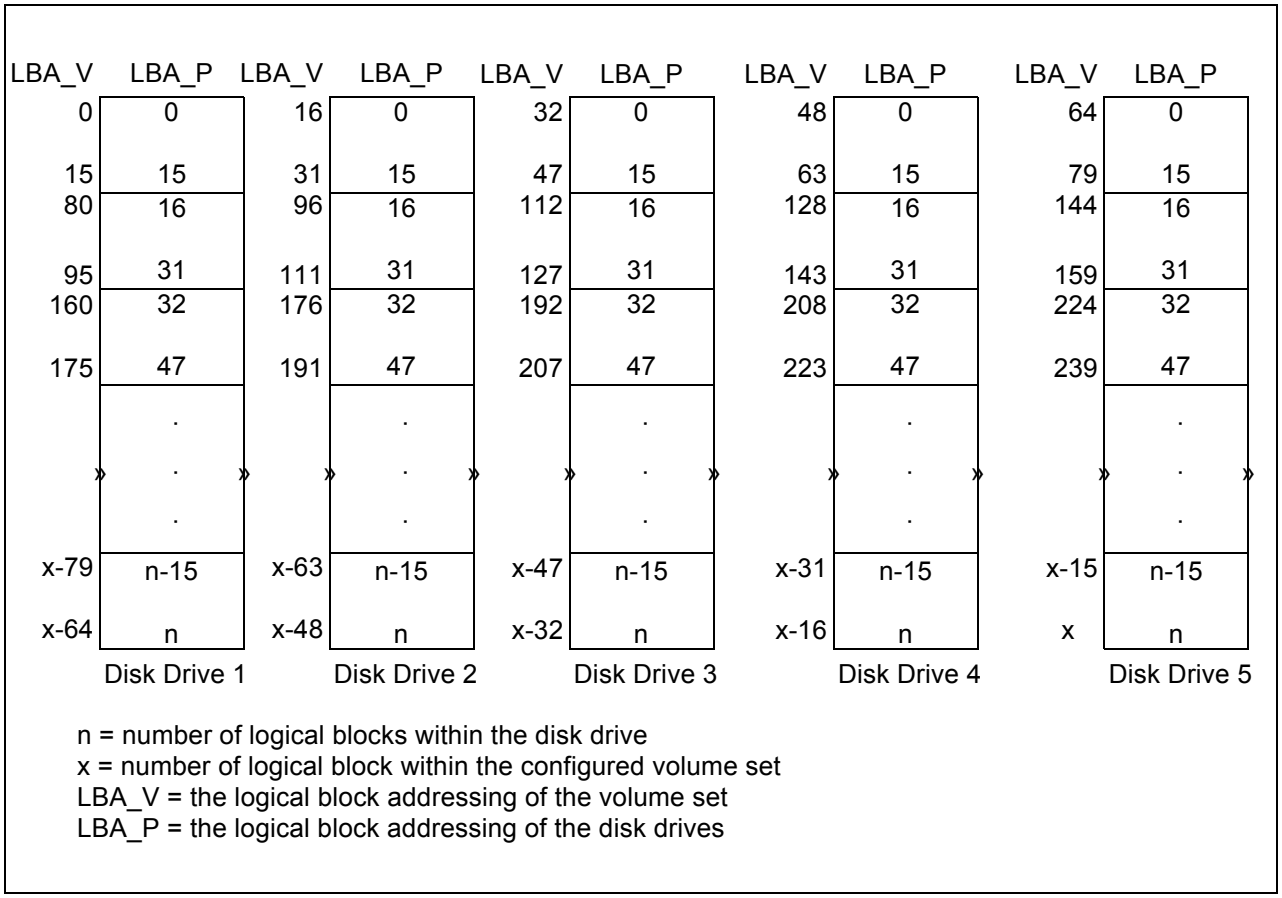
		PERCENT OF SEQUENTIAL TRANSFERS			
		1% to 25%	25% to 50% or 0%	51% to 75%	greater than 75%
NORMAL USER DATA TRANSFER SIZE in logical blocks	1 to 128	1	2		3
	129 to 512 or 0	1	4		2
	greater than 512	1	5		6

Notes:
1) number of logical blocks to read ahead = no read ahead
2) number of logical blocks to read ahead = 2 x normal user data transfer size
3) number of logical blocks to read ahead = 4 x normal user data transfer size
4) number of logical blocks to read ahead = normal user data transfer size
5) number of logical blocks to read ahead = normal user data transfer size / 2
6) number of logical blocks to read ahead = normal user data transfer size up to max cache size

2 Example

A SCSI storage array is connected to an application client that runs applications that normally transfer 4 Kbytes of data at a time with sequential user data requests occurring 60% of the time. In this example the application client issues a CREATE/MODIFY STORAGE ARRAY CONFIGURATION service action to the SCSI storage array with the NORMAL USER DATA TRANSFER SIZE field set to 8 blocks (each block is 512 bytes) and the PERCENTAGE OF SEQUENTIAL TRANSFERS field set to 60.

The SCSI storage array will create a volume set with a user data mapping as shown below:



If, for example, the application client sends a read request for a 4 Kbyte transfer starting with LBA_V 112 then the storage subsystem would read LBA_Ps 16 through 23 from disk drive 3 and transfer that information to the application client. The SCSI storage array would also read ahead LBA_Ps 24 through 31 from disk drive 3 and LBA_Ps 16 through 23 from disk drive 4. The read ahead information is placed into the SCSI storage arrays' cache in anticipation that the next read from the application client will request that information.