Date: May 05, 2004

To: T10 Committee (SCSI)

From: George Penokie (IBM/Tivoli), David Chambliss (IBM Almaden Research Center)

Subject: SAS-1, FCP-3, SRP-2 Per-Command Priority Tagging

1 Overview

The following proposed wording represents changes to FCP-3, SAS 1.1, and SRP-2 to enable the transmission of priority information on a per-command basis.

This proposal standardizes the interface by which device servers can offer differentiated quality of service to different applications associated with the same initiator. Examples of its use would include offering lower priority on IO associated with background destage writes within a storage controller or on IO associated with background applications, so that response time may be reduced for those IO operations that directly affect the responsiveness offered to the end user.

The method defined in this proposal to accomplish this involves changes to the protocol standards to accommodate an extension to the task attribute field to allow different priorities to be assigned to simple task attributes.

2 FCP-3, SAS 1.1, and SRP-2 additions

The changes below will be applied to the FCP-3, SAS 1.1, and SRP-2 standards. Only the section numbers are different.

.

Table 1 — COMMAND information unit

Byte\Bit	7	6	5	4	3	2	1	0
0	(MSB)							
7				LOGICAL UNIT NUMBER				(LSB)
8	Reserved							
9	Reserved	PRIORITY				TASK ATTRIBUTE		
10	Reserved							
11	ADDITIONAL CDB LENGTH (n dwords)					Reserved		
12				CDB				
27								
28				ADDITIONAL CDB BYTES				
27+n×4								

The PRIORITY field specifies the relative scheduling of this task in relation to other tasks already in the task set for processing by the device server (see SAM-3).

The TASK ATTRIBUTE field is defined in table 2.

Table 2 — TASK ATTRIBUTE **field**

Task Attribute Code	Task attribute	Priority Code	Description
000b	SIMPLE	0h	Requests that the task be managed according to the rules for a simple task attribute (see SAM-3).
		1h-Fh	Requests that the task be managed according to the rules for a simple task attribute and priority (see SAM-3).
001b	HEAD OF QUEUE	Reserved	Requests that the task be managed according to the rules for a head of queue task attribute (see SAM-3).
010b	ORDERED	Reserved	Requests that the task be managed according to the rules for an ordered task attribute (see SAM-3).
011b	Reserved		
100b	ACA	Reserved	Requests that the task be managed according to the rules for an automatic contingent allegiance task attribute (see SAM-3).
101b-111b	Reserved		

3 Additions to SPC

3.1 Reservations

3.1.1 Persistent Reservations overview

.

Table 3 — SPC commands that are allowed in the presence of various reservations (part 1 of 3)

Command	Addressed LU has this type of persistent reservation held by another I_T nexus				
	From any I_T nexus		From registered I_T nexus (RR all types)	From I_T nexus not registered	
	Write Excl	Excl Access		Write Excl RR	Excl Access – RR
ACCESS CONTROL IN	Allowed	Allowed	Allowed	Allowed	Allowed
ACCESS CONTROL OUT	Allowed	Allowed	Allowed	Allowed	Allowed
CHANGE ALIASES	Conflict	Conflict	Allowed	Conflict	Conflict
EXTENDED COPY	Conflict	Conflict	Allowed	Conflict	Conflict
INQUIRY	Allowed	Allowed	Allowed	Allowed	Allowed
LOG SELECT	Conflict	Conflict	Allowed	Conflict	Conflict
Key: **LU**=Logical Unit, **Excl**=Exclusive, **RR**=Registrants Only or All Registrants, **<>** Not Equal					
[a] Exceptions to the behavior of the RESERVE and RELEASE commands described in SPC-2 are defined in 5.6.2.					

Table 3 — SPC commands that are allowed in the presence of various reservations (part 2 of 3)

Command	Addressed LU has this type of persistent reservation held by another I_T nexus				
	From any I_T nexus		From registered I_T nexus (RR all types)	From I_T nexus not registered	
	Write Excl	Excl Access		Write Excl RR	Excl Access – RR
LOG SENSE	Allowed	Allowed	Allowed	Allowed	Allowed
MODE SELECT(6)/ MODE SELECT(10)	Conflict	Conflict	Allowed	Conflict	Conflict
MODE SENSE(6)/ MODE SENSE(10)	Conflict	Conflict	Allowed	Conflict	Conflict
PERSISTENT RESERVE IN	Allowed	Allowed	Allowed	Allowed	Allowed
PERSISTENT RESERVE OUT	see table 34				
PREVENT ALLOW MEDIUM REMOVAL (Prevent=0)	Allowed	Allowed	Allowed	Allowed	Allowed
PREVENT ALLOW MEDIUM REMOVAL (Prevent<>0)	Conflict	Conflict	Allowed	Conflict	Conflict
READ ATTRIBUTE	Conflict	Conflict	Allowed	Conflict	Conflict
READ BUFFER	Conflict	Conflict	Allowed	Conflict	Conflict
READ MEDIA SERIAL NUMBER	Allowed	Allowed	Allowed	Allowed	Allowed
RECEIVE COPY RESULTS	Conflict	Conflict	Allowed	Conflict	Conflict
RECEIVE DIAGNOSTIC RESULTS	Conflict	Conflict	Allowed	Conflict	Conflict
RELEASE(6)/ RELEASE(10)	As defined in SPC-2 [a]				
REPORT ALIASES	Allowed	Allowed	Allowed	Allowed	Allowed
REPORT DEVICE IDENTIFIER	Allowed	Allowed	Allowed	Allowed	Allowed
REPORT PRIORITY	Allowed	Allowed	Allowed	Allowed	Allowed
REPORT LUNS	Allowed	Allowed	Allowed	Allowed	Allowed
REPORT SUPPORTED OPERATION CODES	Conflict	Conflict	Allowed	Conflict	Conflict
REPORT SUPPORTED TASK MANAGEMENT FUNCTIONS	Conflict	Conflict	Allowed	Conflict	Conflict
REPORT TARGET PORT GROUPS	Allowed	Allowed	Allowed	Allowed	Allowed
REQUEST SENSE	Allowed	Allowed	Allowed	Allowed	Allowed
RESERVE(6)/ RESERVE(10)	As defined in SPC-2 [a]				
SEND DIAGNOSTIC	Conflict	Conflict	Allowed	Conflict	Conflict
SET DEVICE IDENTIFIER	Conflict	Conflict	Allowed	Conflict	Conflict
SET PRIORITY	Conflict	Conflict	Allowed	Conflict	Conflict
Key: **LU**=Logical Unit, **Excl**=Exclusive, **RR**=Registrants Only or All Registrants, **<>** Not Equal					
[a] Exceptions to the behavior of the RESERVE and RELEASE commands described in SPC-2 are defined in 5.6.2.					

Table 3 — SPC commands that are allowed in the presence of various reservations (part 3 of 3)

Command	Addressed LU has this type of persistent reservation held by another I_T nexus				
	From any I_T nexus		From registered I_T nexus (RR all types)	From I_T nexus not registered	
	Write Excl	Excl Access		Write Excl RR	Excl Access – RR
SET TARGET PORT GROUPS	Conflict	Conflict	Allowed	Conflict	Conflict
TEST UNIT READY	Allowed	Allowed	Allowed	Allowed	Allowed
WRITE ATTRIBUTE	Conflict	Conflict	Allowed	Conflict	Conflict
WRITE BUFFER	Conflict	Conflict	Allowed	Conflict	Conflict
Key: **LU**=Logical Unit, **Excl**=Exclusive, **RR**=Registrants Only or All Registrants, **<>** Not Equal					
^a Exceptions to the behavior of the RESERVE and RELEASE commands described in SPC-2 are defined in 5.6.2.					

3.2 REPORT PRIORITY command (new section)

The REPORT PRIORITY command (see table 4) requests the priority that has been assigned to one or more I_T_L nexus.

The REPORT PRIORITY command is a service action of the MAINTENANCE IN command. Additional MAINTENANCE IN service actions are defined in SCC-2 and in this standard. The MAINTENANCE IN service actions defined in SCC-2 apply only to logical units that return a device type of 0Ch or the SCCS bit equal to one in their standard INQUIRY data (see 6.4.2).

Table 4 — REPORT PRIORITY command

Bit Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (A3h)							
1	Reserved			SERVICE ACTION (xxh)				
2	REPORT PRIORITY		Reserved					
3	Reserved							
5								
6	(MSB)							
	ALLOCATION LENGTH (4h or larger)							
9								(LSB)
10	Reserved							
11	CONTROL							

The REPORT PRIORITY field (see table 5) specifies the I_T nexus and the location of the priority value to be reported.

Table 5 — REPORT PRIORITY **field**

Code	Description
00b	Only the priority for the I_T_L nexus associated with this command shall be reported in the REPORT PRIORITY parameter data. (see table 6).
01b	The priority for each I_T_L nexus that is not set to the initial priority shall be reported in the REPORT PRIORITY parameter data.
10b	Reserved
11b	Reserved

The ALLOCATION LENGTH field specifies the number of bytes that have been allocated for the returned parameter data. An allocation length that is not sufficient to contain the entire parameter list shall not be considered an error. If the complete list is required, the application client should send a new REPORT PRIORITY command with an allocation length large enough to contain the entire list.

The format of the parameter data returned by the REPORT PRIORITY command is shown in table 6.

Table 6 — REPORT PRIORITY parameter data

Bit Byte	7	6	5	4	3	2	1	0
0				PRIORITY PARAMETER DATA LENGTH (m - 4)				
3								
				Priority descriptor list				
4				Priority descriptor (first)				
n								
				⋮				
				Priority descriptor (last)				
m								

The PRIORITY PARAMETER DATA LENGTH field specifies the number of bytes of parameter data that follow.

The priority descriptor format is defined in table 7.

.

Table 7 — Priority descriptor

Bit Byte	7	6	5	4	3	2	1	0
0	Reserved				CURRENT PRIORITY			
1	Reserved							
2	RELATIVE TARGET PORT IDENTIFIER							
3								
4	Reserved							
5	Reserved							
6	TRANSPORTID DESCRIPTOR LENGTH (n - 7)							
7								
8	TransportID							
n								

The CURRENT PRIORITY field contains the priority (see SAM-3) assigned to the I_T_L nexus defined by this descriptor. If the REPORT PRIORITY field in this command is set to zero and the priority for the I_T_L nexus associated with this command is set to the initial priority then the CURRENT PRIORITY field shall be set to zero.

The RELATIVE TARGET PORT IDENTIFIER field (see 7.6.4.6) contains the relative target port identifier of the SCSI target port to which the priority applies.

The TRANSPORTID DESCRIPTOR LENGTH field indicates the number of bytes of the TransortID descriptor that follow.

The format of a TransportID descriptor is specified in 7.5.4.

3.3 SET PRIORITY command (new section)

The SET PRIORITY command (see table 8) requests that a priority be set to the specified value. The priority set by this command shall remain in effect until one of the following occurs:

   a) another SET PRIORITY command is received;
   b) hard reset;
   c) logical unit reset; or
   d) power off.

The priority set by this command shall not be affected by an I_T nexus loss.

The SET PRIORITY command is a service action of the MAINTENANCE OUT command. Additional MAINTE-NANCE OUT service actions are defined in SCC-2 and in this standard. The MAINTENANCE OUT service actions defined only in SCC-2 apply only to logical units that return a device type of 0Ch or the sccs bit equal to one in their standard INQUIRY data.

Table 8 — SET PRIORITY command

Bit / Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (A4h)							
1	Reserved			SERVICE ACTION (xxh)				
2	I_T NEXUS TO SET		Reserved					
3	Reserved							
4	Reserved							
5	Reserved							
6	(MSB)							
7	PARAMETER LIST LENGTH							
8								
9								(LSB)
10	Reserved							
11	CONTROL							

The I_T NEXUS TO SET field (see table 9) specifies the I_T nexus and the location of the priority value to be assigned to that I_T nexus.

Table 9 — I_T NEXUS TO SET field

Code	Description
00b	The priority for the I_T nexus associated with this command shall be set to the value contained in the SET PRIORITY field of the priority parameter list (see table 10).
01b	The priority for the I_T nexus identified in the I_T nexus parameter list (see table 10) shall be set to the value contained in the SET PRIORITY field. The device server shall terminate the command with CHECK CONDITION status if the parameter list length results in the truncation of the I_T nexus parameter list. The sense key shall be set to ILLEGAL REQUEST, and the additional sense code shall be set to PARAMETER LIST LENGTH ERROR. On successful completion of a SET PRIORITY command a unit attention shall be generated for the initiator port identified in the I_T nexus parameter list. When reporting the unit attention condition the additional sense code shall be set to PRIORITY CHANGED.
10b	The priority value specified in the INITIAL PRIORITY field of the Control Extension mode page (see 3.3.1) shall be used for all I_T_L nexus regardless of any prior priority. On successful completion of a SET PRIORITY command a unit attention shall be generated for all otherI_T_L nexus. When reporting the unit attention condition the additional sense code shall be set to PRIORITY CHANGED.
11b	Reserved

Editor's Note 1: The PRIORITY CHANGED is a new ASCQ. I would suggest it be coded as 0Fh 12h.

The PARAMETER LIST LENGTH field specifies the length in bytes of the priority parameter list that shall be contained in the Data-Out Buffer. A parameter list length of zero indicates that the Data-Out Buffer shall be empty. This condition shall not be considered as an error.

.

Table 10 — SET PRIORITY commands priority parameter list

Bit Byte	7	6	5	4	3	2	1	0
0	Reserved				SET PRIORITY			
1	Reserved							
2	RELATIVE TARGET PORT IDENTIFIER							
3								
4	Reserved							
5	Reserved							
6	TRANSPORTID DESCRIPTOR LENGTH (n-7)							
7								
8	TransportID							
n								

The SET PRIORITY field indicates the priority to be assigned. The value in the SET PRIORITY field shall be returned in subsequent REPORT PRIORITY commands. A value of zero specifies the I_T_L nexus indicated in this priority parameter list shall be set to the initial priority specified in the INITIAL PRIORITY field (see x.x.x).

The command shall be terminated with a CHECK CONDITION status and the sense key set to ILLEGAL REQUEST if the value in the PARAMETER LIST LENGTH field results in the truncation of the I_T nexus parameter list.

The RELATIVE TARGET PORT IDENTIFIER field (see 7.6.4.6) contains the relative target port identifier of the SCSI target port to which the priority applies.

The TRANSPORTID DESCRIPTOR LENGTH field indicates the number of bytes of the TransortID descriptor that follow. If the I_T NEXUS TO SET field is set to zero then this field shall be set to zero

The format of a TransportID descriptor is specified in 7.5.4.

3.3.1 Control Extension mode page

The Control Extension mode page (see table 223) is a subpage of the Control mode page (see 7.4.6) provides controls over SCSI features that are applicable to all device types. The mode page policy (see 6.7) for this subpage shall be shared. If a field in this mode subpage is changed while there is a task already in the task set, it is vendor specific whether the old or new value of the field applies to that task.

Table 11 — Control Extension mode page

Bit Byte	7	6	5	4	3	2	1	0
0	PS	SPF (1b)	PAGE CODE (0Ah)					
1	SUBPAGE CODE (01h)							
2	(MSB)		PAGE LENGTH (1Ch)					
3								(LSB)
4	Reserved							IALUAE
5	Reserved				INITIAL PRIORITY			
6	Reserved							
31								

The PS bit, SPF bit, PAGE CODE field, SUBPAGE CODE field, and PAGE LENGTH field are described in 7.4.5.

An implicit asymmetric logical unit access enabled (IALUAE) bit set to one specifies that implicit asymmetric logical unit access state changes (see 5.8.6) are allowed. An IALUAE bit set to zero specifies that implicit asymmetric logical unit access state changes be disallowed and indicates that implicit asymmetric logical unit access state changes are disallowed or not supported.

The INITIAL PRIORITY field specifies the priority to be used on tasks received in any I_T_L nexus where the priority has not been modified by a SET PRIORITY command (see x.x.x) since the last power on or hard reset. If a MODE SELECT command specifies an initial priority value that is different than the current initial priority then the device server shall set any priorities that have not be set with a SET PRIORITY command to a value different than the new initiator priority value to the new priority. The device server shall generate a unit attention for any I_T_L nexus that receives a new priority. When reporting the unit attention condition the additional sense code shall be set to PRIORITY CHANGED.

4 SPC-3 additions

4.0.1 Extended INQUIRY Data VPD page

The Extended INQUIRY Data VPD page (see table 12) provides the application client with a means to obtain information about the logical unit.

Table 12 — Extended INQUIRY Data VPD page

Bit Byte	7	6	5	4	3	2	1	0
0	PERIPHERAL QUALIFIER			PERIPHERAL DEVICE TYPE				
1	PAGE CODE (86h)							
2	Reserved							
3	PAGE LENGTH (3Ch)							
4	Reserved					GRD_CHK	APTG_CHK	RFTG_CHK
5	Reserved				PRIOR_SUP	HEADSUP	ORDSUP	SIMPSUP
6	Reserved							
63								

A priority supported (PRIOR_SUP) bit set to one shall indicate that priority (see SAM-3) is supported by the logical unit. A PRIOR_SUP bit set to zero shall indicate that priority is not supported.