

T10/04-297 revision 0

Date: September 08, 2004

To: T10 Committee (SCSI)

From: George Penokie (IBM/Tivoli)

Subject: SAM-3 SAS-1.1 FCP-3 - Retry Delay

1 Overview

For a long time systems have complained that they have no way of knowing how long a logical unit will stay busy either because the logical unit is not ready or the queue is full. Many factors come into play in a logical unit when allocating resources among an unknown number of initiators. Also the amount of resources available to the target side of a large subsystem can fluctuate based on the workload. So in many cases it is impossible for a logical unit to guarantee a fixed amount of queue slots for an initiator without having to tie-up resources (and as a result reduce performance) for commands that may never happen.

This proposal defines a reporting mechanism that will give an indication to initiators as to how long the logical unit current resources will be tied-up and a upper limit as to the amount of time should be spent retrying before declaring the logical unit dead.

I want to acknowledge that the idea of for the 'additional sense' field was borrowed from a proposal Rob Elliott is working on.

2 SAS 1.1 Changes

2.0.0.1 RESPONSE information unit

2.0.0.1.1 RESPONSE information unit overview

Table 1 defines the response IU. The RESPONSE frame is sent by an SSP target port to deliver SCSI status (e.g., GOOD or CHECK CONDITION) and sense data, or to deliver SSP-specific status (e.g., illegal frame

format). The maximum size of the RESPONSE frame is the maximum size of any IU in an SSP frame (see 9.2.1).

Table 1 — RESPONSE information unit

Byte\Bit	7	6	5	4	3	2	1	0	
0	Reserved								
7	Reserved								
8	ADDITIONAL STATUS								
9	Reserved								
10	Reserved					DATAPRES			
11	STATUS								
12	Reserved								
15	Reserved								
16	(MSB)	SENSE DATA LENGTH (n bytes)							
19								(LSB)	
20	(MSB)	RESPONSE DATA LENGTH (m bytes)							
23								(LSB)	
24	RESPONSE DATA								
23+m	RESPONSE DATA								
24+m	SENSE DATA								
23+m+n	SENSE DATA								

[The ADDITIONAL STATUS field contains the additional status code \(see SAM-3\).](#)

3 FCP-3 changes

3.1 FCP_RSP IU

3.1.1 Overview and format of FCP_RSP IU

....

Table 2 - FCP_RSP IU Payload

Bit Byte	7	6	5	4	3	2	1	0
0	RESERVED							
7	RESERVED							
8	ADDITIONAL STATUS							
9	ADDITIONAL STATUS							
10	FCP_ BIDI_ RSP	FCP_ BIDI_ READ_ RESID_ UNDER	FCP_ BIDI_ READ_ RESID_ OVER	FCP_ CONF_ REQ	FCP_ RESID_ UNDER	FCP_ RESID_ OVER	FCP_ SNS_ LEN_ VALID	FCP_ RSP_ LEN_ VALID
11	SCSI STATUS CODE							
12	(MSB)	FCP_RESID						(LSB)
15	FCP_RESID						(LSB)	
16	(MSB)	FCP_SNS_LEN (= n)						(LSB)
19	FCP_SNS_LEN (= n)						(LSB)	
20	(MSB)	FCP_RSP_LEN (= m)						(LSB)
23	FCP_RSP_LEN (= m)						(LSB)	
24	(MSB)	FCP_RSP_INFO (m bytes long)						(LSB)
23+m	FCP_RSP_INFO (m bytes long)						(LSB)	
24+m	(MSB)	FCP_SNS_INFO (n bytes long)						(LSB)
23+m+n	FCP_SNS_INFO (n bytes long)						(LSB)	
24+m+n	(MSB)	FCP_BIDIRECTIONAL_READ_RESID						(LSB)
27+m+n	FCP_BIDIRECTIONAL_READ_RESID						(LSB)	

[The ADDITIONAL STATUS field contains the additional status code \(see SAM-3\).](#)

4 SAM-3 changes

5 SCSI command model

5.1 The Execute Command procedure call

An application client requests the processing of a command by invoking the SCSI transport protocol services described in 5.4, the collective operation of which is modeled in the following procedure call:

Service Response = Execute Command (IN (I_T_L_Q Nexus, CDB, Task Attribute, [Data-In Buffer Size], [Data-Out Buffer], [Data-Out Buffer Size], [Command Reference

Number], [Task Priority]), OUT ([Data-In Buffer], [Sense Data], [Sense Data Length], Status, [\[Additional Status\]](#))

Input Arguments:

I_T_L_Q Nexus: The I_T_L_Q nexus identifying the task (see 4.12).

CDB: Command descriptor block (see 5.2).

Task Attribute: A value specifying one of the task attributes defined in 8.6. SCSI transport protocols may or may not provide the ability to specify a different task attribute for each task (see 8.6.1). For a task that processes linked commands, the Task Attribute shall be the value specified for the first command in a series of linked commands. The Task Attribute specified for the second and subsequent commands shall be ignored.

Data-In Buffer Size: The number of bytes available for data transfers to the Data-In Buffer (see 5.4.3).

Data-Out Buffer: A buffer containing command specific information to be sent to the logical unit (e.g., data or parameter lists needed to process the command). The buffer size is indicated by the Data-Out Buffer Size argument. The content of the Data-Out Buffer shall not change during the lifetime of the command (see 5.5) as viewed by the application client.

Data-Out Buffer Size: The number of bytes available for data transfers from the Data-Out Buffer (see 5.4.3).

Command Reference Number (CRN): When this argument is used, all sequential commands of an I_T_L nexus shall include a CRN argument that is incremented by one. The CRN shall be set to one for each I_T_L nexus involving the SCSI port after the SCSI port receives a hard reset or detects I_T nexus loss. The CRN shall be set to one after it reaches the maximum CRN value supported by the protocol. The CRN value zero shall be reserved for use as defined by the SCSI transport protocol. It is not an error for the application client to provide this argument when CRN is not supported by the SCSI transport protocol or logical unit.

Task Priority: The priority assigned to the task (see 8.7).

Output Arguments:

Data-In Buffer: A buffer to contain command specific information returned by the logical unit by the time of command completion. The **Execute Command** procedure call shall not return a status of GOOD, CONDITION MET, INTERMEDIATE, or INTERMEDIATE-CONDITION MET unless the buffer contents are valid. The application client shall not assume that the buffer contents are valid unless the command completes with a status of GOOD, CONDITION MET, INTERMEDIATE, or INTERMEDIATE-CONDITION MET. While some valid data may be present for other values of status, the application client should rely on additional information from the logical unit (e.g., sense data, to determine the state of the buffer contents). If the command ends with a service response of SERVICE DELIVERY OR TARGET FAILURE, the application client shall consider this argument to be undefined.

Sense Data: A buffer containing sense data returned in the same I_T_L_Q nexus transaction (see 3.1.46) as a CHECK CONDITION status (see 5.9.6). The buffer length is indicated by the Sense Data Length argument. If the command ends with a service response of SERVICE DELIVERY OR TARGET FAILURE, the application client shall consider this argument to be undefined.

Sense Data Length: The length in bytes of the Sense Data.

Status: A one-byte field containing command completion status (see 5.3). If the command ends with a service response of SERVICE DELIVERY OR TARGET FAILURE, the application client shall consider this argument to be undefined.

Additional Status: [Additional information about the indicated status code \(see 5.3.3\).](#)

Service Response assumes one of the following values:

TASK COMPLETE: A logical unit response indicating that the task has ended. The Status argument shall have one of the values specified in 5.3 other than INTERMEDIATE or INTERMEDIATE-CONDITION MET.

LINKED COMMAND COMPLETE: Logical unit responses indicating that the task has not ended and that a linked command has completed successfully. As specified in 5.3, the Status argument shall have a value of INTERMEDIATE or INTERMEDIATE-CONDITION MET.

SERVICE DELIVERY OR TARGET FAILURE: The command has been ended due to a service delivery failure (see 3.1.113) or SCSI target device malfunction. All output parameters are invalid.

5.2 Command descriptor block (CDB)

....

5.3 Status

5.3.1 Status codes

The status codes are specified in table 3. Status shall be sent from the device server to the application client whenever a command ends with a service response of TASK COMPLETE or LINKED COMMAND COMPLETE.

Table 3 — Status codes

Status Code	Status	Task Ended	Service Response
00h	GOOD	Yes	TASK COMPLETE
02h	CHECK CONDITION	Yes	TASK COMPLETE
04h	CONDITION MET	Yes	TASK COMPLETE
08h	BUSY	Yes	TASK COMPLETE
10h	INTERMEDIATE	No	LINKED COMMAND COMPLETE
14h	INTERMEDIATE-CONDITION MET	No	LINKED COMMAND COMPLETE
18h	RESERVATION CONFLICT	Yes	TASK COMPLETE
22h	Obsolete		
28h	TASK SET FULL	Yes	TASK COMPLETE
30h	ACA ACTIVE	Yes	TASK COMPLETE
40h	TASK ABORTED	Yes	TASK COMPLETE
All other codes	Reserved		

Definitions for each status code are as follows:

GOOD. This status indicates that the device server has successfully completed the task.

CHECK CONDITION. This status indicates that sense data has been delivered in the buffer defined by the Sense Data argument to the **Execute Command** procedure call (see 5.9.6). Additional actions that are required when CHECK CONDITION status is returned are described in 5.9.1.

CONDITION MET. This status shall be returned whenever the requested operation specified by an unlinked command is satisfied (see the PRE-FETCH commands in the SBC standard).

BUSY. This status indicates that the logical unit is busy. This status shall be returned whenever a logical unit is temporarily unable to accept a command. The recommended application client recovery action is to issue the command again at a later time. If the UA_INTLCK_CTRL field in the Control mode page contains 11b (see SPC-3), termination of a command with BUSY status shall cause an unit attention condition to be established for the SCSI initiator port that sent the command with an additional sense code of PREVIOUS BUSY STATUS unless a PREVIOUS BUSY STATUS unit attention condition already exists.

[The ADDITIONAL SENSE field may provide the SCSI initiator port with more information on when the command should be retransmitted \(see table 4\).](#)

INTERMEDIATE. This status or INTERMEDIATE-CONDITION MET shall be returned for each successfully completed command in a series of linked commands (except the last command), unless the command is terminated with CHECK CONDITION, RESERVATION CONFLICT, TASK SET FULL, or BUSY status. If INTERMEDIATE or INTERMEDIATE-CONDITION MET status is not returned, the series of linked commands is terminated and the task is ended. This status is the equivalent of GOOD status for linked commands.

INTERMEDIATE-CONDITION MET. This status is returned whenever the requested operation specified by a linked command is satisfied (see the PRE-FETCH commands in the SBC standard), unless the command is terminated with CHECK CONDITION, RESERVATION CONFLICT, TASK SET FULL, or BUSY status. If

INTERMEDIATE or INTERMEDIATE-CONDITION MET status is not returned, the series of linked commands is terminated and the task is ended.

RESERVATION CONFLICT. This status shall be returned whenever a SCSI initiator port attempts to access a logical unit or an element of a logical unit in a way that conflicts with an existing reservation. (See the PERSISTENT RESERVE OUT command and PERSISTENT RESERVE IN command in SPC-3).

If the UA_INTLCK_CTRL field in the Control mode page contains 11b (see SPC-3), termination of a command with RESERVATION CONFLICT status shall cause an unit attention condition to be established for the SCSI initiator port that sent the command with an additional sense code of PREVIOUS RESERVATION CONFLICT STATUS unless a PREVIOUS RESERVATION CONFLICT STATUS unit attention condition already exists.

TASK SET FULL. This status shall be implemented by all logical units.

When the logical unit has at least one task in the task set for a SCSI initiator port and a lack of task set resources prevents accepting a received task from that SCSI initiator port into the task set, TASK SET FULL shall be returned. When the logical unit has no task in the task set for a SCSI initiator port and a lack of task set resources prevents accepting a received task from that SCSI initiator port into the task set, BUSY should be returned.

The logical unit should allow at least one command in the task set for each supported SCSI initiator port that has identified itself to the SCSI target port in a SCSI transport protocol specific manner or by the successful transmission of a command.

If the UA_INTLCK_CTRL field in the Control mode page contains 11b (see SPC-3), termination of a command with TASK SET FULL status shall cause an unit attention condition to be established for the SCSI initiator port that sent the command with an additional sense code of PREVIOUS TASK SET FULL STATUS unless a PREVIOUS TASK SET FULL STATUS unit attention condition already exists.

[The ADDITIONAL SENSE field may provide the SCSI initiator port with more information on when the command should be retransmitted \(see table 4\).](#)

ACA ACTIVE. This status shall be returned when an ACA exists within a task set and a SCSI initiator port issues a command for that task set when at least one of the following is true:

- a) There is a task with the ACA attribute (see 8.6.5) in the task set;
- b) The SCSI initiator port issuing the command did not cause the ACA condition; or
- c) The task created to process the command did not have the ACA attribute and the NACA bit was set to one in the CDB CONTROL byte of the faulting command (see 5.9.1).

The SCSI initiator port may reissue the command after the ACA condition has been cleared.

TASK ABORTED. This status shall be returned when a task is aborted by another SCSI initiator port and the Control mode page TAS bit is set to one (see 5.7.3).

5.3.2 Status precedence

...

5.3.3 Additional status codes

The additional status codes are specified in table 4 and provide additional information about the reason for the status code.

Table 4 — Additional status

Status Code	Bits	Additional Status Code	Description
Busy	15-0	0000h	See 5.3.1
		0001h - FFFEh	The number of milliseconds the initiator port should wait before sending another command to the logical unit.
		FFFFh	The target port is not able to accept the command because the addressed logical unit is not responding.
Task Set Full	15-0	0000h	See 5.3.1
		0001h - FFFEh	The initiator port should wait before sending another command to the logical unit until: a) the number of seconds indicated in the ADDITIONAL STATUS CODE field occur; or b) a command addressed to that logical unit completes.
		FFFFh	The logical unit is not able to accept the command because it is servicing too many other initiator ports. The initiator port should wait before sending another command to the logical unit until a command addressed to that logical unit completes.

5.4 SCSI transport protocol services in support of Execute Command

5.4.2 Execute Command request/confirmation SCSI transport protocol services

...

SCSI Transport Protocol Service Response (from device server):

Send Command Complete (IN (I_T_L_Q Nexus, [Sense Data], [Sense Data Length], Status, Service Response, [\[Additional Status\]](#))) Input Arguments:

Additional Status: [The additional status code for the command \(see 5.3\).](#)

...

SCSI Transport Protocol Service Confirmation:

Command Complete Received (IN (I_T_L_Q Nexus, [Data-In Buffer], [Sense Data], [Sense Data Length], Status, Service Response, [\[Additional Status\]](#)))

Input Arguments:

Additional Status: [The additional status code for the command \(see 5.3\).](#)