

Date: March 13, 2007

To: T10 Committee (SCSI)

From: George Penokie (IBM)

Subject: SAS-2: Expander Notification of Temporary Shutdown

## **1 Overview**

Within the Service Delivery Subsystem in the SAS topology we have intelligent devices that will have to have their software updated. The software is updated by issuing WRITE BUFFER commands to a SES device within the expander device. When this occurs there may be a period of time while the flash is being written when the expander will not permit SMP functions and may limit traffic between it's phys. This period of time may, depending on the implementation, take longer than a minute. After the code is written the expander does a reset sequence on all it's phys before allowing traffic to resume.

This may cause timeouts at every layer of SAS for any operations outstanding at the time of the code download. This is a problem that can be handled within the current SAS and SCSI definitions by the application client/initiator device that issued the WRITE BUFFER command. However, any other application client/initiator device may not be happy, as any outstanding operations could without warning just disappear.

This proposal defines a Broadcast (Expander) that expanders may use to notify all devices that it will be temporarily (as in seconds) be shutting down the SMP functions and may be limiting traffic.

After a SAS expander sends a Broadcast (Expander) and the time-out expires that SAS expander responds to attempts to open a port with an OPEN REJECT (RETRY) if access to that port is blocked as a result of the action that resulted in the Broadcast (Expander). After the action that resulted in the BROADCAST (EXPANDER) is complete the SAS expander initiates a link reset sequence on all phys.

The action a SAS device takes with it receives a Broadcast (Expander) depends on the type of device it is:

- a) Expanders pass the Broadcast(Expander) through;
- b) Target devices ignore the Broadcast (Expander); and
- c) Initiator devices may (this behavior is not intended to be specified in SAS, this is only informational):
  - A) after receiving a Broadcast (Expander) issue a REPORT GENERAL function to determine the source and cause of the Broadcast (Expander). If the source of the broadcast was as a result of an expander limiting action, then the Initiator should terminate any outstanding I\_T\_L\_Q nexus; and
  - B) terminate any I\_T\_L\_Q nexus that receive an OPEN REJECT (RETRY) for a length of time longer than indicated by the WRITE BUFFER commands code update timeout.

This requires a new bit in REPORT GENERAL function that provides information on the source and cause of the broadcast.

## **2 SAS-2 changes**

### **4.1.13 Broadcasts**

Broadcasts are used to notify all phys in the SAS domain about certain events. Broadcasts are transmitted using BROADCAST (see 7.2.5.5) or the SMP ZONED BROADCAST function (see 10.4.3.17).

Table 1 defines the types of Broadcast supported.

**Table 1 — Broadcast types**

Broadcast	Primitive	Description
Broadcast (Change)	yes	Originated by an expander device to notify SAS initiator ports that a SAS domain change has occurred (see 7.11). May also be originated by SAS initiator ports. Ignored by SAS target ports.
Broadcast (Reserved Change 0)	yes	Reserved. SAS ports (i.e, SAS initiator ports and SAS target ports) shall process this Broadcast the same as Broadcast (Change).
Broadcast (Reserved Change 1)	yes	Reserved. SAS ports shall process this Broadcast the same as Broadcast (Change).
Broadcast (SES)	yes	Originated by a logical unit with a peripheral device type set to 0Dh (i.e., enclosure services device) (see SPC-4 and SES-2) accessible through a SAS target port in the SAS domain to notify SAS initiator ports of an asynchronous event.  SSP initiator ports should poll all the logical units in the SAS domain with peripheral device types set to 0Dh to determine the source.  SAS target ports shall ignore this Broadcast.
Broadcast (Expander)	yes	Originated by an expander device to notify SAS initiator ports that an expander event has occurred, including: <ul style="list-style-type: none"> <li>a) a phy event information peak value detector has reached its threshold value; <del>ef</del></li> <li>b) a phy event information peak value detector has been cleared by an SMP CONFIGURE PHY EVENT INFORMATION function (see 10.4.3.26); <u>or</u></li> <li>c) <a href="#">an expander device is going to temporarily have reduced function (see 4.6.8).</a></li> </ul> Expander events do not include SAS domain changes, which are communicated with Broadcast (Change).
Broadcast (Asynchronous Event)	yes	Originated by an SSP target port when an event occurs that causes one or more unit attention conditions to be established for one or more logical units accessible through the SSP target port.  An SSP target port shall only originate one Broadcast (Asynchronous Event) for each event that affects multiple logical units accessible through the SSP target port (e.g., only one Broadcast (Asynchronous Event) is originated when a hard reset occurs).
Broadcast (Reserved 3)	yes	Reserved. SAS ports shall ignore this Broadcast.
Broadcast (Reserved 4)	yes	
Broadcast (Zone Activate)	no	Initiates the zone activate step (see 4.9.6.4).  Devices that are not locked zoning expander devices shall ignore this Broadcast.
<p><sup>a</sup> All Broadcasts are supported by the SMP ZONED BROADCAST function (see 10.4.3.17). Broadcasts labeled “yes” are also transmitted via BROADCAST primitive sequences (see 7.2.5.5).</p>		

When an expander port receives a Broadcast, the BPP (see 4.6.5) shall forward the Broadcast on at least one phy in each other expander port if zoning is disabled, or forward the Broadcast as described in 4.9.5 if zoning is enabled.

An expander device is not required to queue multiple identical Broadcasts for the same expander port. If a second identical Broadcast is requested before the first Broadcast has been transmitted, the second Broadcast may be ignored.

See 10.4.3.3 for details on counting Broadcast (Change)s originated in an expander device. See 4.11 for details on phy event information.

#### **4.6.8 Expander reduced function**

Broadcast (Expander) shall be transmitted by an expander port to specify that the expander device is going to temporarily have reduced function (e.g., disable SMP access, reduced performance, disable phy to phy communication) for a period of time. The maximum period of time the expander device is going to have reduced function may be determined:

- a) by requesting write buffer command information using a REPORT SUPPORTED OPERATION CODES command (see SPC-4) issued to a enclosure services device within the expander device; and
- b) from the contents of the MAXIMUM REDUCED FUNCTION TIME field (see 10.4.3.3).

After the originating expander sends a Broadcast (Expander) that expander device shall:

- a) set the REDUCED FUNCTION bit to one in the REPORT GENERAL function (see 10.4.3.3);
- b) initialize and start the time to reduced function timer to the value indicated by the INITIAL TIME TO REDUCED FUNCTION field in the REPORT GENERAL response (see 10.4.3.3);
- c) wait for the time to reduced function timer to expire before reducing any expander functions; and
- d) not stop or restart the time to reduce function timer until after the expander device enters the reduced function condition.

If an expander phy receives an OPEN that attempts to connect to a phy or expander SMP target port that is not accessible because of the reduced function, then that expander device shall respond with an OPEN REJECT (RETRY) until the operation that caused the expander device to reduce function is complete.

After the operation that caused the reduced function is complete the expander device shall:

- 1) set the REDUCED FUNCTION bit to zero; and
- 2) originate a Broadcast (Change) or a link reset sequence on all phys.

For the SAS initiator device rules on determining the cause of a Broadcast (Expander) see 7.9.2.

#### **7.9.2 SAS initiator device rules**

After a link reset sequence, or after receiving a Broadcast (Change), a management application client behind an SMP initiator port should perform a discover process (see 4.7).

When a discover process is performed after a link reset sequence, the management application client discovers all the devices in the SAS domain. When a discover process is performed after a Broadcast (Change), the management application client determines which devices have been added to or removed from the SAS domain.

The discover information may be used to select connection rates for connection requests (see 7.8.3).

After receiving a Broadcast (Expander), a management application client behind an SMP initiator port should issue a REPORT GENERAL function (see 10.4.3.3) to all expander devices to determine:

- a) which expander devices, if any, are reducing their function (i.e., the REDUCED FUNCTION bit set to one); and
- b) the amount of time remaining until the reduced function occurs (i.e., the contents of the TIME TO REDUCED FUNCTION field).

### 10.4.3.3 REPORT GENERAL function

The REPORT GENERAL function returns general information about the SAS device (e.g., a SAS device contained in an expander device). This SMP function shall be implemented by all management device servers.

Table 2 defines the request format.

**Table 2 — REPORT GENERAL request**

Byte\Bit	7	6	5	4	3	2	1	0	
0	SMP FRAME TYPE (40h)								
1	FUNCTION (00h)								
2	Reserved								
3	REQUEST LENGTH (00h)								
4	(MSB)	CRC							
7							(LSB)		

The SMP FRAME TYPE field shall be set to 40h.

The FUNCTION field shall be set to 00h.

The REQUEST LENGTH field shall be set to 00h.

The CRC field is defined in 10.4.3.1.

Table 3 defines the response format.

**Table 3 — REPORT GENERAL response (part 1 of 2)**

Byte\Bit	7	6	5	4	3	2	1	0	
0	SMP FRAME TYPE (41h)								
1	FUNCTION (00h)								
2	FUNCTION RESULT								
3	RESPONSE LENGTH ( <del>00</del> 36h)								
4	(MSB)	EXPANDER CHANGE COUNT							
5							(LSB)		
6	(MSB)	EXPANDER ROUTE INDEXES							
7							(LSB)		
8	Reserved								
9	NUMBER OF PHYS								
10	TABLE TO TABLE SUPPORTE D	Reserved			ZONE ADDRESS RESOLVED SUPPORTED	CONFIGURES OTHERS	CONFIGURING	EXTERNALLY CONFIGURABLE ROUTE TABLE	

Table 3 — REPORT GENERAL response (part 2 of 2)

Byte\Bit	7	6	5	4	3	2	1	0
11	Reserved							
12	ENCLOSURE LOGICAL IDENTIFIER							
19	ENCLOSURE LOGICAL IDENTIFIER							
20	Reserved							
29	Reserved							
30	(MSB)	STP BUS INACTIVITY TIME LIMIT						(LSB)
31	STP BUS INACTIVITY TIME LIMIT							
32	(MSB)	STP MAXIMUM CONNECT TIME LIMIT						(LSB)
33	STP MAXIMUM CONNECT TIME LIMIT							
34	(MSB)	STP SMP I_T NEXUS LOSS TIME						(LSB)
35	STP SMP I_T NEXUS LOSS TIME							
36	Reserved			ZONE LOCKED	PHYSICAL PRESENCE SUPPORTED	PHYSICAL PRESENCE ASSERTED	ZONING SUPPORTED	ZONING ENABLED
37	Reserved							
38	(MSB)	MAXIMUM NUMBER OF ROUTED SAS ADDRESSES						(LSB)
39	MAXIMUM NUMBER OF ROUTED SAS ADDRESSES							
40	ACTIVE ZONE MANAGER SAS ADDRESS							
47	ACTIVE ZONE MANAGER SAS ADDRESS							
48	(MSB)	ZONE LOCK INACTIVITY TIME LIMIT						(LSB)
49	ZONE LOCK INACTIVITY TIME LIMIT							
50	Reserved							
51	Reserved							
52	<a href="#">REDUCED FUNCTION</a>	Reserved						
53	<a href="#">TIME TO REDUCED FUNCTION</a>							
54	<a href="#">INITIAL TIME TO REDUCED FUNCTION</a>							
55	<a href="#">MAXIMUM REDUCED FUNCTION TIME</a>							
56	(MSB)	CRC						(LSB)
57	CRC							

....

The REDUCED FUNCTION bit set to one indicates that the expander device is scheduled to reduce its function in the time indicated in the TIME TO REDUCED FUNCTION field or that the expander device is currently operating with reduced function. A REDUCED FUNCTION bit of zero indicates the expander device is not scheduled to reduce any functions and that the contents of the TIME TO REDUCED FUNCTION field shall be ignored.

If the REDUCED FUNCTION bit set to one, then the TIME TO REDUCED FUNCTION field contains the time, in 100 ms increments, remaining until the expander device is scheduled to reduce any functions. The expander device starts the time to reduced function timer after originating a Broadcast (Expander) (see 4.6.8).

The INITIAL TIME TO REDUCED FUNCTION field contains the minimum period of time, in 100 ms increments, that an expander device waits from originating a Broadcast (Expander) to reducing any expander function (see 4.6.8). The expander device should set the default value for the INITIAL TIME TO REDUCED FUNCTION field to at least 2 000 ms (i.e., 14h).

The MAXIMUM REDUCED FUNCTION TIME field contains the maximum time, in 1 s increments, the expander device issues OPEN REJECT (RETRY) in response to an OPEN that attempts to connect to a phy or expander SMP target port that is not accessible as a result of the expander device reducing its function. This timer starts after the time to reduced function expires.

#### **10.4.3.15 CONFIGURE GENERAL function**

The CONFIGURE GENERAL function requests actions by the device containing the management device server. This SMP function may be implemented by any management device server. In zoning expander devices, if zoning is enabled then this function shall only be processed from SMP initiator ports that have access to zone group 2 (see 4.9.3.2).

Table 4 defines the request format.

**Table 4 — CONFIGURE GENERAL request**

Byte\Bit	7	6	5	4	3	2	1	0
0	SMP FRAME TYPE (40h)							
1	FUNCTION (80h)							
2	Reserved							
3	REQUEST LENGTH (034h)							
4	(MSB)	EXPECTED EXPANDER CHANGE COUNT						(LSB)
5								
6	Reserved							
7								
8	Reserved				<a href="#">UPDATE INITIAL TIME TO REDUCED FUNCTION</a>	UPDATE STP SMP I_T NEXUS LOSS TIME	UPDATE STP MAXIMUM CONNECT TIME LIMIT	UPDATE STP BUS INACTIVITY TIME LIMIT
9	Reserved							
10	(MSB)	STP BUS INACTIVITY TIME LIMIT						(LSB)
11								
12	(MSB)	STP MAXIMUM CONNECT TIME LIMIT						(LSB)
13								
14	(MSB)	STP SMP I_T NEXUS LOSS TIME						(LSB)
15								
<a href="#">16</a>	<a href="#">INITIAL TIME TO REDUCED FUNCTION</a>							
<a href="#">17</a>	<a href="#">Reserved</a>							
<a href="#">19</a>								
<del>16</del> <a href="#">20</a>	(MSB)	CRC						(LSB)
<del>19</del> <a href="#">23</a>								

...

[An UPDATE INITIAL TIME TO REDUCED FUNCTION bit set to one specifies that the INITIAL TIME TO REDUCED FUNCTION field in the SMP REPORT GENERAL response \(see 10.4.3.3\) shall be set to the contents of the INITIAL TIME TO REDUCED FUNCTION field. An UPDATE INITIAL TIME TO REDUCED FUNCTION bit set to zero specifies that the INITIAL TIME TO REDUCED FUNCTION field shall be ignored.](#)

...

The INITIAL TIME TO REDUCED FUNCTION field specifies the minimum period of time, in 100 ms increments, that an expander device shall wait from originating a Broadcast (Expander) to reducing any expander function (see 4.6.8). The INITIAL TIME TO REDUCED FUNCTION field in the SMP REPORT GENERAL response (see 10.4.3.3) shall be set to this value.

### 3 SPC-4 changes

Add the following section to the WRITE BUFFERS command. This change assumes that 05-284 is accepted.

#### 6.36.1 WRITE BUFFER commands command processing times descriptor

---



---

Editor's Note 1: All new in this section

---



---



---



---

Editor's Note 2: The editor will place this new section in the REPORT SUPPORTED OP CODES command section with refernaces to the WRITE BUFFER command section.

---



---

The command processing times descriptor (see table 5) reported by the REPORT SUPPORTED OPERATION CODES command (x.x) for a WRITE BUFFER command indicates timeout information specific to the WRITE BUFFER command.

**Table 5 — Command Processing Times descriptor**

Byte/Bit	7	6	5	4	3	2	1	0
0	(MSB)	DESCRIPTOR LENGTH (0Ah)						(LSB)
1								
2		Reserved						
3		MAXIMUM REDUCED FUNCTION TIME						
4	(MSB)	NOMINAL COMMAND PROCESSING TIME						(LSB)
7								
8	(MSB)	RECOMMENDED COMMAND TIMEOUT						(LSB)
11								

The MAXIMUM REDUCED FUNCTION TIME field contains the maximum time, 1 s increments, access to the SCSI device is limited or not possible through any SCSI ports associated with the logical unit that receives the WRITE BUFFER command. A value of zero in the MAXIMUM REDUCED FUNCTION TIME field indicates that there is no maximum time specified. The MAXIMUM REDUCED FUNCTION TIME field shall only apply when the following modes are specified:

- a) Download microcode mode (04h);
- b) Download microcode and save mode (05h);
- c) Download microcode with offsets mode (06h);
- d) Download microcode with offsets and save mode (07h);
- e) Download microcode with offsets and defer activation mode (0Eh) only if the microcode is activated by an event other than an activate deferred microcode mode; and



**T10/07-008 revision 8**

- f) Activate deferred microcode mode (0Fh).

The NOMINAL COMMAND PROCESSING TIME field and RECOMMENDED COMMAND TIMEOUT field are defined in x.x.