To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 18 January 2005
Subject: 04-340r2 SAS-1.1 OPEN_REJECT BAD DESTINATION handling

Revision history
Revision 0 (18 October 2004) First revision
Revision 1 (24 December 2004) Incorporated changes based on November SAS protocol WG meeting - recommend that expanders to send NO DESTINATION instead of requiring that end devices treat BAD DESTINATION as retry-class.
Revision 2 (18 January 2005) Incorporated comments from January SAS protocol WG.

Related documents
sas1r06 - Serial Attached SCSI 1.1 revision 6

Overview
When the physical link to a destination SAS address goes down, the response can be either OPEN_REJECT (NO DESTINATION) or OPEN_REJECT (BAD DESTINATION) depending on whether there is a subtractive port in one of the expanders (see figure).

OPEN_REJECT (NO DESTINATION) is in the retry class of responses, meaning the OPEN will be attempted again until the I_T nexus loss timer expires. OPEN_REJECT (BAD DESTINATION) is in the abandon class, meaning the OPEN is not reattempted.

Since the physical link might become enabled again, the retry behavior is appropriate.
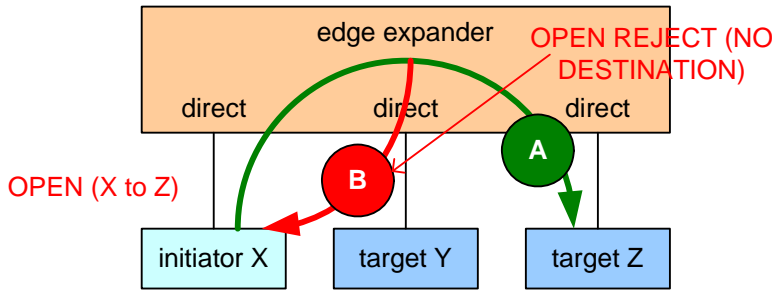
Possible corrections:

a)   Require that expanders return NO DESTINATION instead of BAD DESTINATION for phys using either the subtractive or table routing methods (not just subtractive as currently defined). Phys using the direct method (i.e. attached to an end device) would still return BAD DESTINATION. This means the request must have arrived through a subtractive phy in an attached expander; or
b)   Treat BAD DESTINATION as a retry-class response rather than an abandon-class response.

The latter was suggested in revision 0 of this proposal; the disk drive companies in the November SAS protocol WG protested that approach; although some handle this in firmware, others apparently handle this in hardware.
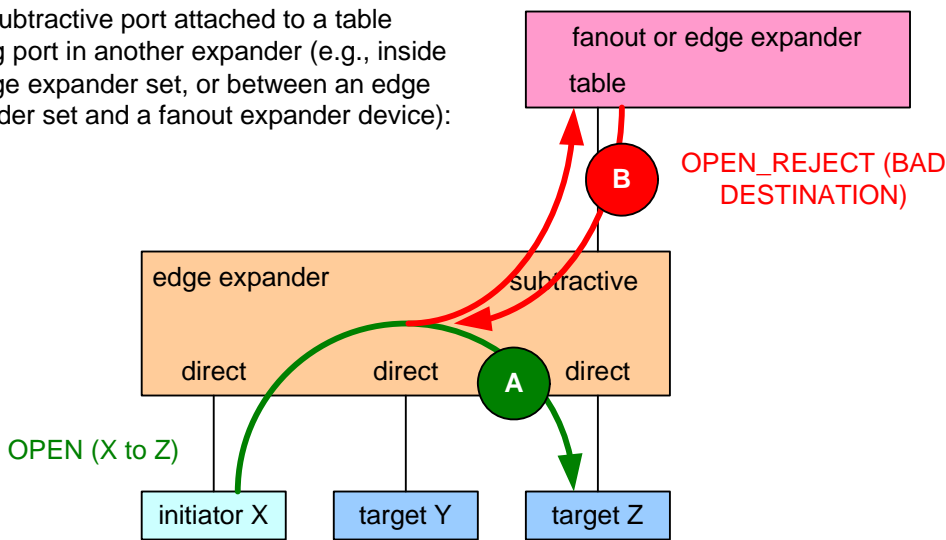
This revision proposes option a). Since this is handled in hardware by all existing expanders, it is only proposed as a "should".
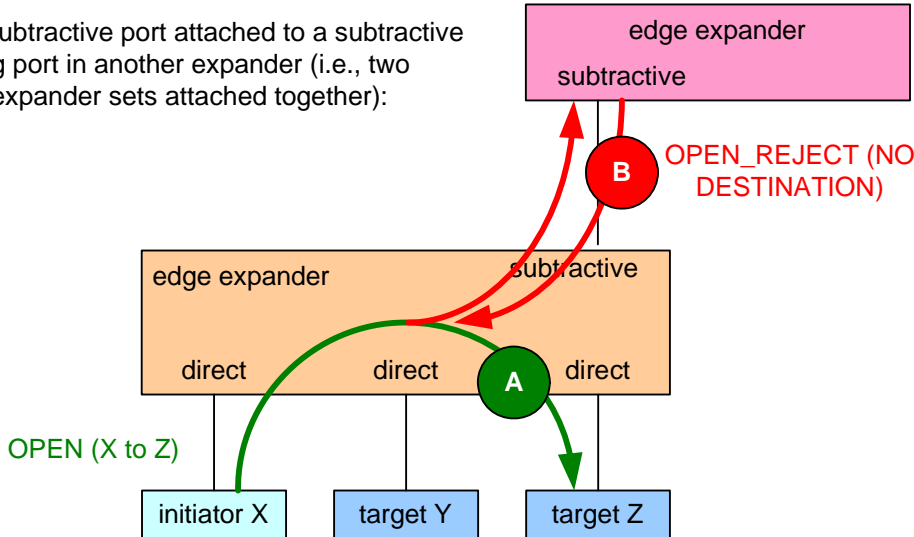
With no subtractive port:

edge expander

OPEN REJECT (NO DESTINATION)

direct   direct   direct

A

B

OPEN (X to Z)

initiator X   target Y   target Z

A) before target physical link goes offline, OPEN routed to target
B) after target physical link goes offline. OPEN_REJECT (NO DESTINATION) from edge expander

With subtractive port attached to a table
routing port in another expander (e.g., inside
an edge expander set, or between an edge
expander set and a fanout expander device):

fanout or edge expander

table

B

OPEN_REJECT (BAD DESTINATION)

edge expander   subtractive

direct   direct   direct

A

OPEN (X to Z)

initiator X   target Y   target Z

A) before target physical link goes offline, OPENs routed to target
B) after target physical link goes offline, OPEN_REJECT (BAD DESTINATION) from top expander

With subtractive port attached to a subtractive
routing port in another expander (i.e., two
edge expander sets attached together):

edge expander

subtractive

B

OPEN_REJECT (NO DESTINATION)

edge expander   subtractive

direct   direct   direct

A

OPEN (X to Z)

initiator X   target Y   target Z

A) before target physical link goes offline, OPENs routed to target
B) after target physical link goes offline, OPEN_REJECT (NO DESTINATION) from other edge expander

<u>**Suggested changes**</u>

4.6.6.3 ECM interface

...

Table 1 describes the confirmations from the ECM to an expander phy. These confirmations are sent in confirmation of a Request Path request.

Table 1 — ECM to expander phy confirmations <u>(part 1 of 2)</u>

Message	Description
Arbitrating (Normal)	Confirmation that the ECM has received the Request Path request.
Arbitrating (Waiting On Partial)	Confirmation that the ECM has determined that: a) there is a destination port capable of routing to the requested destination SAS address; b) at least one phy within the destination port supports the requested connection rate; c) each of the phys within the destination port is returning a Phy Status (Partial Pathway) or Phy Status (Blocked Partial Pathway) response; and d) at least one of the phys within the destination port is returning a Phy Status (Partial Pathway) response.
Arbitrating (Blocked On Partial)	Confirmation that the ECM has determined that: a) there is a destination port capable of routing to the requested destination SAS address; b) at least one phy within the destination port supports the requested connection rate; c) each of the phys within the destination port is returning a Phy Status (Blocked Partial Pathway) response.
Arbitrating (Waiting On Connection)	Confirmation that the ECM has determined that the connection request is blocked due to one of the following reasons: a) the connection request is blocked by an active connection; or b) there are insufficient routing resources within the expander to complete the connection request. A connection request shall be considered block by an active connection when: a) there is a destination port capable of routing to the requested destination SAS address; b) at least one phy within the destination port supports the requested connection rate; c) each of the phys within the destination port is returning a Phy Status (Partial Pathway), Phy Status (Blocked Partial Pathway), or Phy Status (Connection) response; and d) at least one of the phys within the destination port is returning a Phy Status (Connection) response.
Arb Won	Confirmation that an expander phy has won path arbitration.
Arb Lost	Confirmation that an expander phy has lost path arbitration.

Table 1 — ECM to expander phy confirmations (part 2 of 2)

Message	Description
Arb Reject (No Destination)	Confirmation that the ECM did not find an operational expander phy capable of routing to the requested destination SAS address, or has determined that the requested destination SAS address maps back to the requesting port (see 7.12.5.2 and 7.12.5.3).
Arb Reject (Bad Destination)	Confirmation that the ECM has determined that the requested destination SAS address maps back to the requesting port and, the requesting port is using the direct routing method or the table routing method, and the ECM has not chosen to return Arb Reject (No Destination) (see 7.12.5.2 and 7.12.5.3).
Arb Reject (Bad Connection Rate)	Confirmation that the ECM has determined that there is a destination port capable of routing to the requested destination SAS address but no phys within the destination port are configured to support the requested connection rate.
Arb Reject (Pathway Blocked)	Confirmation that the ECM has determined that the requesting expander phy shall back off according to SAS pathway recovery rules.

...

7.2 Primitives

7.2.5 Primitives not specific to types of connections

7.2.5.11 OPEN_REJECT

OPEN_REJECT indicates that a connection request has been rejected and indicates the reason for the rejection. The result of some OPEN_REJECTs is to abandon (i.e., not retry) the connection request and the result of other OPEN_REJECTs is to retry the connection request.

All of the OPEN_REJECT versions defined in table 2 shall result in the originating device abandoning the connection request.

Table 2 — OPEN_REJECT abandon primitives

Primitive	Originator	Description
OPEN_REJECT (BAD DESTINATION)	Expander phy	~~An expander device receives a request in which the destination SAS address equals the source SAS address, or a~~ A connection request arrives through an expander phy using the direct routing or table routing method and the expander device determines the connection request would have to be routed to the same expander port as the expander port through which the connection request arrived (e.g., the destination SAS address equals the source SAS address), and the expander device has not chosen to return OPEN_REJECT (NO DESTINATION) (see 7.12.2.2 and 7.12.2.3).
OPEN_REJECT (CONNECTION RATE NOT SUPPORTED)	Any phy	The requested connection rate is not supported on some physical link on the pathway between the source phy and destination phy. When a SAS initiator phy is directly attached to a SAS target phy, the requested connection rate is not supported by the destination phy. The connection request may be modified and reattempted as described in 7.12.2.2.
OPEN_REJECT (PROTOCOL NOT SUPPORTED)	Destination phy	Device with destination SAS address exists but the destination device does not support the requested initiator/target role, protocol, initiator connection tag, or features (i.e., the values in the INITIATOR PORT bit, the PROTOCOL field, the INITIATOR CONNECTION TAG field, and/or the FEATURES field in the OPEN address frame are not supported).
OPEN_REJECT (RESERVED ABANDON 0)	Unknown	Reserved. Process the same as OPEN_REJECT (WRONG DESTINATION).
OPEN_REJECT (RESERVED ABANDON 1)	Unknown	Reserved. Process the same as OPEN_REJECT (WRONG DESTINATION).
OPEN_REJECT (RESERVED ABANDON 2)	Unknown	Reserved. Process the same as OPEN_REJECT (WRONG DESTINATION).
OPEN_REJECT (RESERVED ABANDON 3)	Unknown	Reserved. Process the same as OPEN_REJECT (WRONG DESTINATION).
OPEN_REJECT (STP RESOURCES BUSY)	Destination phy	STP target port with destination SAS address exists but the STP target port has an affiliation with another STP initiator port or all of the available task file registers have been allocated to other STP initiator ports (see 7.17.4). Process the same as OPEN_REJECT (WRONG DESTINATION) for non-STP connection requests.
OPEN_REJECT (WRONG DESTINATION)	Destination phy	The destination SAS address does not match the SAS address of the SAS port to which the connection request was delivered.

All of the OPEN_REJECT versions defined in table 3 shall result in the originating device retrying the connection request.

Table 3 — OPEN_REJECT retry primitives

Primitive	Originator	Description
OPEN_REJECT (NO DESTINATION) [c]	Expander phy	Either: a) No such destination device; b) ~~a connection request arrives through an expander phy using the subtractive routing method and~~ the expander device determines the connection request would have to be routed to the same expander port as the expander port through which the connection request arrived (e.g., the destination SAS address equals the source SAS address), and it has not chosen to return OPEN_REJECT (NO DESTINATION) (see 7.12.2.2 and 7.12.2.3); or c) the SAS address is valid for an STP target port in an STP/SATA bridge, but the initial Register - Device to Host FIS has not been successfully received (see 10.4.3.7).
OPEN_REJECT (PATHWAY BLOCKED) [b]	Expander phy	An expander device determined the pathway was blocked by higher priority connection requests.
OPEN_REJECT (RESERVED CONTINUE 0) [a]	Unknown	Reserved. Process the same as OPEN_REJECT (RETRY).
OPEN_REJECT (RESERVED CONTINUE 1) [a]	Unknown	Reserved. Process the same as OPEN_REJECT (RETRY).
OPEN_REJECT (RESERVED INITIALIZE 0) [c]	Unknown	Reserved. Process the same as OPEN_REJECT (NO DESTINATION).
OPEN_REJECT (RESERVED INITIALIZE 1) [c]	Unknown	Reserved. Process the same as OPEN_REJECT (NO DESTINATION).
OPEN_REJECT (RESERVED STOP 0) [b]	Unknown	Reserved. Process the same as OPEN_REJECT (PATHWAY BLOCKED)
OPEN_REJECT (RESERVED STOP 1) [b]	Unknown	Reserved. Process the same as OPEN_REJECT (PATHWAY BLOCKED).
OPEN_REJECT (RETRY) [a]	Destination phy	Device with destination SAS address exists but is not able to accept connections.

[a] If the I_T Nexus Loss timer (see 8.2.2) is already running, it is stopped.
[b] If the I_T Nexus Loss timer is already running, it continues running. Stop retrying the connection request if the I_T Nexus Loss timer expires.
[c] If the I_T Nexus Loss timer is already running, it continues running; if it is not already running, it is initialized and started. Stop retrying the connection request if the I_T Nexus Loss timer expires.

When a destination device detects more than one reason to transmit an OPEN_REJECT, the device shall transmit only one OPEN_REJECT and shall select the primitive using the following priority:

1) OPEN_REJECT (WRONG DESTINATION) (highest priority selection);
2) OPEN_REJECT (PROTOCOL NOT SUPPORTED);
3) OPEN_REJECT (CONNECTION RATE NOT SUPPORTED);
4) OPEN_REJECT (STP RESOURCES BUSY); or
5) OPEN_REJECT (RETRY) (lowest priority selection).

When an expander device detects more than one reason to transmit an OPEN_REJECT, the expander shall transmit only one OPEN_REJECT primitive and shall select that primitive using the following priority:

1) OPEN_REJECT (BAD DESTINATION) or OPEN_REJECT (NO DESTINATION) (highest priority selection);
2) OPEN_REJECT (CONNECTION RATE NOT SUPPORTED); or
3) OPEN_REJECT (STP RESOURCES BUSY) or OPEN_REJECT (PATHWAY BLOCKED) (lowest priority selection).

See 7.12 for details on connection requests.

7.12 Connections

7.12.2 Opening a connection

7.12.2.2 Connection responses

Table 4 lists the responses to an OPEN address frame being transmitted.

Table 4 — Connection responses

Response	Description
AIP	Arbitration in progress. When an expander device is trying to open a connection to the selected destination port, it returns an AIP to the source phy. The source phy shall reinitialize and restart its Open Timeout timer when it receives an AIP. AIP is sent by an expander device while it is internally arbitrating for access to an expander port.
OPEN_ACCEPT	Connection request accepted. This is sent by the destination phy.
OPEN_REJECT	Connection request rejected. This is sent in response by the destination phy or by an expander device. The different versions are described in x.x. See 4.5 for I_T nexus loss handling.
OPEN address frame	If AIP has been previously detected, this indicates an overriding connection request. If AIP has not yet been detected, this indicates two connection requests crossing on the physical link. Arbitration fairness determines which one wins (see 7.12.3).
BREAK	The destination port or expander port may reply with BREAK indicating the connection is not being established.
Open Timeout timer expires	The source phy shall abort the connection request by transmitting BREAK (see 7.12.6). See 4.5 for I_T nexus loss handling.

...

7.12.4 Arbitration and resource management in an expander device

7.12.4.1 Arbitration overview

The ECM shall arbitrate and assign or deny path resources for Request Path requests from each expander phy.

Arbitration includes adherence to the SAS arbitration fairness algorithm and path recovery. Path recovery is used to avoid potential deadlock scenarios within the SAS topology by deterministically choosing which partial pathway(s) to tear down to allow at least one connection to complete.

The ECM responds to connection requests by returning an Arb Won, Arb Lost, or Arb Reject confirmation to the requesting expander phy.

Several of the Request Path arguments are used for arbitration. The Arbitration Wait Time, Source SAS Address, and Connection Rate arguments are filled in from the received OPEN address frame and are used to by the ECM to compare Request Path requests. The Retry Priority Status argument is used to prevent the

Arbitration Wait Time argument from being considered during an arbitration which occurs after a Backoff Retry response is sent by an expander phy (see 7.15.4).

An expander phy shall set the Retry Priority Status argument to IGNORE AWT when it requests a path after:

a) it has forwarded an OPEN address frame to the physical link;
b) an OPEN address frame is received with higher arbitration priority (see 7.12.3); and
c) the destination SAS address and connection rate of the received OPEN address frame are not equal to the source SAS address and connection rate of the transmitted OPEN address frame (see 7.15.4 and 7.15.9).

Otherwise, the expander phy shall set the Retry Priority Status argument to NORMAL.

If two or more Request Path requests contend and all of the Request Path requests include a Retry Priority Status argument set to NORMAL, the ECM shall select the winner by comparing the OPEN address frame contents described in table 5.

Table 5 — Arbitration priority for contending Request Path requests in the ECM when all requests have Retry Priority Status arguments of NORMAL

Bits 83-68 (83 is MSB)	Bits 67-4	Bits 3-0 (0 is LSB)
ARBITRATION WAIT TIME field value	SOURCE SAS ADDRESS field value	CONNECTION RATE field value

If two or more Request Path requests contend and one or more of the Request Path requests include a Retry Priority Status argument set to IGNORE AWT, the ECM shall select the winner from the set of Request Path requests with Retry Priority Status arguments of IGNORE AWT by comparing the OPEN address frame contents described in table 6.

Table 6 — Arbitration priority for contending Request Path requests in the ECM among requests with Retry Priority Status arguments of IGNORE AWT

Bits 67-4 (67 is MSB)	Bits 3-0 (0 is LSB)
SOURCE SAS ADDRESS field value	CONNECTION RATE field value

The ECM shall generate the Arb Reject confirmation when any of the following conditions are met and all the Arb Won conditions are not met:

a) Arb Reject (No Destination) or Arb Reject (Bad Destination) if the connection request does not map to an expander phy that is not part of the same expander port as the requesting expander phy (i.e., there is no direct routing or table routing match and there is no subtractive phy);
b) Arb Reject (Bad Connection Rate) if the connection request does not map to any expander phy that supports the connection rate; or
c) Arb Reject (Pathway Blocked) if the connection request maps to expander phys that all contain blocked partial pathways (i.e., are all returning Phy Status (Blocked Partial Pathway)) and pathway recovery rules require this connection request to release path resources.

The ECM shall generate the Arb Lost confirmation when all of the following conditions are met:

a) the connection request maps to an expander phy that:
   A) supports the connection rate; and
   B) is not reporting a Phy Status (Partial Pathway), Phy Status (Blocked Partial Pathway), or Phy Status (Connection) response unless that expander phy is arbitrating for the expander phy making this connection request;
b) there are sufficient routing resources to complete the connection request; and
c) the destination expander phy of this connection request has received a higher priority OPEN address frame with this expander phy as its destination (i.e., when two expander phys both receive an OPEN address frame destined for each other, the ECM shall provide the Arb Lost confirmation to the expander phy that received the lowest priority OPEN address frame).

The ECM shall generate the Arb Won confirmation when all of the following conditions are met:

    a) the connection request maps to an expander phy that:
        A) supports the connection rate; and
        B) is not reporting a Phy Status (Partial Pathway), Phy Status (Blocked Partial Pathway), or Phy Status (Connection) response, unless that expander phy is arbitrating for the expander phy making this connection request;
    b) there are sufficient routing resources to complete the connection request;
    c) no higher priority connection requests are present with this expander phy as the destination; and
    d) the connection request is chosen as the highest priority connection request in the expander device mapping to the specified destination expander phy.

7.12.5 Expander devices and connection requests

7.12.5.1 All expander devices

Before an expander device transmits AIP, it may have transmitted an OPEN address frame on the same physical link. Arbitration fairness dictates which OPEN address frame wins (see 7.12.3).

After an expander device transmits an AIP, it shall not transmit an OPEN address frame unless it has higher arbitration priority than the incoming connection request.

Expander devices shall transmit no more than three consecutive AIPs without transmitting an idle dword. Expander devices may transmit three consecutive AIPs to provide better tolerance of errors. Expander devices shall transmit at least one AIP every 128 dwords while transmitting AIP (NORMAL), AIP (WAITING ON PARTIAL), or AIP (WAITING ON CONNECTION).

> NOTE 1 - Future versions of this standard may require that expander devices transmit three consecutive AIPs.

Expander devices shall transmit an AIP (e.g., an AIP (NORMAL)) within 128 dwords of receiving an OPEN address frame.

7.12.5.2 Edge expander devices

When an edge expander device receives a connection request, it shall compare the destination SAS address to the SAS addresses of the devices to which each of its phys is attached. For all phys which have table routing attributes (see 4.6.7.1) and are attached to edge expander devices, it shall compare the destination SAS address to all the enabled routed SAS addresses in the expander route table.

If it finds a match in one or more phys, then the expander device shall arbitrate for access to one of the matching phys and forward the connection request.

If it does not find a match, but at least one phy has the subtractive routing attribute and is attached to an expander device (e.g., either an edge expander device or a fanout expander device)(i.e., is using the subtractive routing method), and the request did not come from that expander device, the connection request shall be forwarded to the expander device through any of the subtractive routing phys.

If it does not find a match and no phy using the subtractive routing ~~phy~~method is available, the edge expander device shall reply with OPEN_REJECT (NO DESTINATION).

If the destination phy is in the same expander port as the source phy and ~~the source phys are~~ is using the subtractive routing method, the edge expander device shall reply with OPEN_REJECT (NO DESTINATION). ~~If the source phys are not using subtractive routing, the edge expander device shall reply with OPEN_REJECT (BAD DESTINATION).~~

If the destination phy is in the same expander port as the source phy and is using the direct routing method or the table routing method, the edge expander device shall reply with either OPEN_REJECT (NO DESTINATION) or OPEN_REJECT (BAD DESTINATION); it should reply with OPEN_REJECT (NO DESTINATION).

> NOTE 2 Edge expander devices compliant with previous versions of this standard were required to return OPEN_REJECT (BAD DESTINATION).

7.12.5.3 Fanout expander devices

When a fanout expander device receives a connection request, it shall compare the destination SAS address to the SAS addresses of the devices to which each of its phys is attached. For all phys that are attached to edge expander devices, the fanout expander device shall compare the destination SAS addresses to all the enabled SAS addresses in the expander route table.

If the fanout expander device finds a match in one or more phys, it shall arbitrate for access to one of the matching phys and forward the connection request.

If the fanout expander device does not find a match, it shall reply with OPEN_REJECT (NO DESTINATION). If the destination phy is in the same expander port as the source phy, it shall reply with either OPEN_REJECT (NO DESTINATION) or shall reply with OPEN_REJECT (BAD DESTINATION); it should reply with OPEN_REJECT (NO DESTINATION).

> NOTE 3 Fanout expander devices compliant with previous versions of this standard always returned OPEN_REJECT (BAD DESTINATION).