SCSI Enclosure Services (SES)

Rajendra Divecha
Patni Americas Inc.

Topics

- What is SCSI Enclosure Services (SES)?
- Commands for Enclosure Services
- Diagnostic pages
- SES model
- Implementation ideas
- Possible improvements

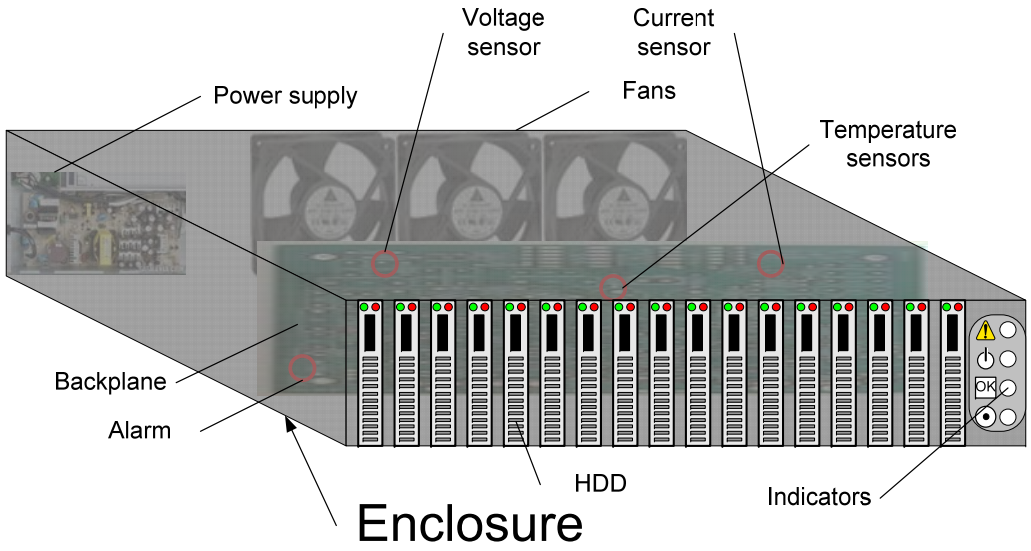
(This presentation is based off of SES 2 r20.)

What is

SCSI Enclosure Services (SES)?

What is SCSI Enclosure Services (SES)?

- Enclosure is a box/unit that houses one or more SCSI devices, specifically end devices (ex. drives) and/or expander devices.
- Besides SCSI devices, it has supporting devices which are non-SCSI.



- Examples of these non-SCSI devices are, but not limited to, power supplies, fans, displays, indicator lights, locks, temperature sensors, current sensors and voltage sensors.

What is SCSI Enclosure Services (SES)?

- Like SCSI devices, these non-SCSI devices are also controlled and managed. This was, and can be, done in a proprietary way. However, you lose on interoperability.
  SES provides a standard way of managing these devices.
- SES belongs to the SCSI-3 family and is one of the many command sets.
- Various versions of SES are 1.0, 1.1 and 2.0.
- The physical connection between the host and the Enclosure can be any interface defined in the SCSI Standards Architecture (http://www.t10.org/scsi-3.htm), ex. Serial Attached SCSI (SAS), Fibre Channel (FC) etc.
- Host has distinct "logical" connections to end devices (drives) in an enclosure, and in most cases to the enclosure itself, through one or more "physical" connector(s). (Example on another slide.)
- The physical interfaces and connections are outside the scope of SES and won't be discussed in this presentation.

What is SCSI Enclosure Services (SES)?

Brief introduction of a few entities before we continue:

- **Application client**: Initiates SCSI commands and task management.
- **Device Server**: Processes SCSI commands.
- **Task manager**: Controls the sequencing of SCSI commands and processes SCSI task management functions.
- **SCSI/SAS initiator port**: Acts as a connection between the application client and the service delivery subsystem through which requests and confirmations are routed.
- **SCSI initiator device**: Container of the application client and the initiator port.
- **SCSI/SAS target port**: Acts as a connection between device server & task manager and the service delivery subsystem through which requests, indications, responses and confirmations are routed.
- **SCSI target device**: Container of the device server, task manager and the target port.

Legends: Text color – Software, Text color – Software object, Text color – Hardware, **Text color** – Theoretical entity

What is SCSI Enclosure Services (SES)?

- **Service delivery subsystem**: Part of SCSI I/O system that transmits information between source (SCSI/SAS initiator port or ATA host) and destination (SCSI/SAS target port or ATA device).
- Phy: Basic interface that connects one device to the other. Phys can be physical or virtual. A port can be formed of one or more phys.
- **Logical unit (LU)**: Externally addressable entity within a SCSI target device.
- **Enclosure services process (ESP)**: A process that implements SES.
- **Element**: A (non-SCSI) device in SES context.
- **Enclosure domain**: A boundary enclosing all the "ESP managed" devices.

Legends: Text color – Software, Text color – Software object, Text color – Hardware, **Text color** – Theoretical entity

What is SCSI Enclosure Services (SES)?

Example of a standalone SAS enclosure:

- Host has a Host Bus Adaptor (HBA) that has a Mini-SAS connector.
- Enclosure also has a Mini-SAS connector (call A).
- Host connects to the Enclosure using a Mini-SAS cable.
- The Enclosure has a SAS expander (a switch) and 20 SAS/SATA drives. Drives connect to the expander using other connectors (call B).
- Connectors have one or more Phys. Assume that the expander's route table is configured such that phys in A can talk to phys in B.
- Host discovers expander and the drives while sending and receiving information through one physical Mini-SAS connector on the enclosure (i.e. A).
- Expander exposes a Logical Unit (LU) with device type as Enclosure Services that has an Enclosure Services Process (ESP) as its Device Server.
- Host connects to the drives and the Enclosure Service using their respective LUs.

What is SCSI Enclosure Services (SES)?

- Other examples might be where an enclosure has two or more connectors that connects to either one or multiple hosts.

- SES uses two SCSI commands to send and receive diagnostic pages.
  - SEND DIAGNOSTIC – to control SES elements
  - RECEIVE DIAGNOSTIC RESULTS – to get enclosure configuration & status of SES elements

  (These are defined in SCSI Primary Commands (SPC) standard.)
- SES standard defines control and status type diagnostic pages that carry control and status information respectively.
- SES has the concept of subenclosures (secondary enclosures). Multiple enclosures are managed by one ESP.
- SES also allows more than one ESP to manage a single enclosure through more than one target port.

Commands for Enclosure Services

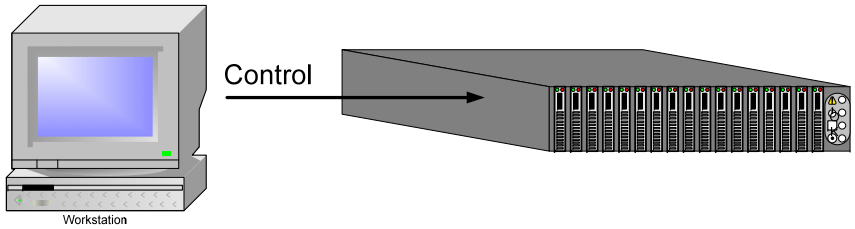
Commands for Enclosure Services

- SES standard defines the commands for the enclosure services peripheral device.
- Mandatory commands for Enclosure Services device:
  - INQUIRY
  - REPORT LUNS
  - REQUEST SENSE
  - TEST UNIT READY
  - SEND DIAGNOSTIC and
  - RECEIVE DIAGNOSTIC RESULTS

(These are defined in SCSI Primary Commands (SPC) standard.)

Commands for Enclosure Services

- Application client uses SEND DIAGNOSTIC command with PF bit set to one to transmit control information (diagnostic page) to the ESP with.
  - The ESP may ignore or override the control information to ensure proper state.
  - ESP ignore features not yet implemented.
  - ESP should ignore instructions that might adversely affect the enclosure.

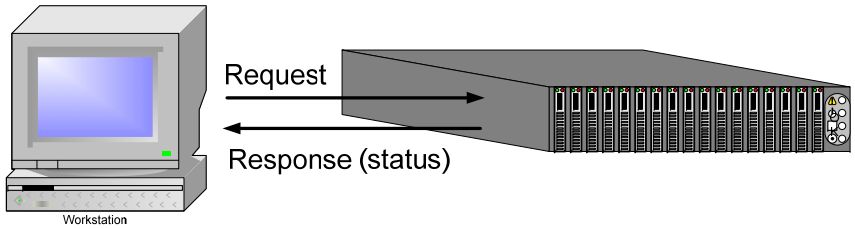


Application client is not necessarily a desktop application. It can be any software in any initiator device.

It is possible that the Application client might read the status of the enclosure after sending a control page and find certain unexpected status. One reason for this is that the ESP has “all” the rights to ignore or override control information from the application client.

Commands for Enclosure Services

- Application client uses RECEIVE DIAGNOSTIC RESULTS command with PCV bit set to one and PAGE CODE fields set to the page being requested to obtain various status information.
- In the response…
  - The information shall indicate the actual state of the enclosure.
  - This state can be a combination of control from one or more application client and the ESP itself.



Here “actual” does not mean the settings that the ESP “tried to” set earlier. Those settings might not have been actually set in the device due to some hardware problem. Thus, rely on the latest reading from the devices and report the same as the actual state.

Diagnostic Pages

Diagnostic Pages

- Two types of Diagnostic pages
  - Control and
  - Status
- Control pages send control information like enable/disable, identify, on/off, set indicators etc for Elements managed by the ESP.
- Status pages give configuration of the Enclosure and the status of the Elements.
- There are 14 standard SES diagnostic pages and provision for 16 vendor specific SES diagnostic pages.
- List of standard diagnostic pages on the next slide…

Diagnostic Pages

Code	Diagnostic Page	Type
00h	Supported Diagnostic Pages	Status
	SES diagnostic pages…	
01h	Configuration	Status
02h	Enclosure Control	Control
	Enclosure Status	Status
03h	Help Text	Status
04h	String Out	Control
	String In	Status
05h	Threshold Out	Control
	Threshold In	Status
07h	Element Descriptor	Status
08h	Short Enclosure Status	Status
09h	Enclosure Busy	Status
0Ah	Additional Element Status	Status

Code	Diagnostic Page	Type
0Bh	Subenclosure Help Text	Status
0Ch	Subenclosure String Out	Control
	Subenclosure String In	Status
0Dh	Supported SES Diagnostic Pages	Status
0Eh	Download Microcode Control	Control
	Download Microcode Status	Status
0Fh	Subenclosure Nickname Control	Control
	Subenclosure Nickname Status	Status

Some diagnostic pages, ex. 02h, come in pairs. When the page code is in the SEND DIAGNOSTIC command, it is a control page. When the page code is in the RECEIVE DIAGNOCTIC RESULTS command, it is a status page.

Diagnostic Pages

- Page 00h – Supported Diagnostic Pages is mandatory.
- Simple enclosures support only page 08h – Short Enclosure Status.
- For regular enclosures, page 01h – Configuration and 02h – Enclosure Control/Status pages are mandatory.
- Application client starts by querying the following pages in the order…
  - Page 00h – Supported Diagnostic Pages - to figure out if the enclosure is simple or regular.
  - Page 01h – Configuration - if the enclosure supports this page, it gives a complete idea of how many enclosures there are, how many ESP each enclosure has and how many elements/devices are supported by each.

Diagnostic Pages

Configuration diagnostic page

- The order of elements in this page defines the order of elements in other diagnostic pages, namely…
  - Enclosure Control/Status
  - Threshold Out/In
  - Element Descriptor and
  - Additional Element Status
- This page gives the Number Of Secondary Enclosures, Enclosure Descriptor list, Type Descriptor Header list and Type Descriptor Text list in the order.
- First Enclosure descriptor represents the primary enclosure, others define the secondary enclosures if any.
- Enclosure descriptor gives the Subenclosure Id, Logical Id (unique), Vendor and Product information, Number Of ESP, Number Of Type Descriptor Headers and other Vendor Specific information.

Diagnostic Pages

- Type Descriptor Header describes an element type.

Byte	Field
0	ELEMENT TYPE
1	NUMBER OF POSSIBLE ELEMENTS
2	SUBENCLOSURE IDENTIFIER
3	TYPE DESCRIPTOR TEXT LENGTH

- Devices can be controlled either individually or as a group.
  - If Number Of Possible Elements field in a Type Descriptor Header is 0, ESP is controlling all the elements as a group (Overall control).
  - If Number Of Possible Elements field in a Type Descriptor Header is not 0, ESP is controlling elements individually (Individual control).
- Type Descriptor Text provides an ASCII name for the element type that the application client can use for display.
- Type Descriptor Texts is ordered identical to Type Descriptor Headers.

Diagnostic Pages

Standard Element Types

Code	Name
01h	Device Slot
02h	Power Supply
03h	Cooling
04h	Temperature Sensor
05h	Door Lock
06h	Audible Alarm
07h	Enclosure Services Controller Electronics
08h	SCC Controller Electronics
09h	Nonvolatile Cache
0Ah	Invalid Operation Reason
0Bh	Uninterruptible Power Supply (UPS)
0Ch	Display
0Dh	Key Pad Entry
0Eh	Enclosure

Code	Name
0Fh	SCSI Port/Transceiver
10h	Language
11h	Communication Port
12h	Voltage Sensor
13h	Current Sensor
14h	SCSI Target Port
15h	SCSI Initiator Port
16h	Simple Subenclosure
17h	Array Device Slot
18h	SAS Expander
19h	SAS Connector

UPS is not inside the enclosure but is inside the enclosure services domain.

Diagnostic Pages

- Besides the 25 standard element types, there is a provision for 128 vendor-specific element types.
- Temperature Sensor, Audible Alarm, Voltage Sensor and Current Sensor are the ones that support the **`DISABLE`** bit.
- Temperature Sensor, UPS, Voltage Sensor and Current Sensor are the ones that supports threshold values.
- Device Slot and Array Device Slot type descriptor headers appear before any other type descriptor headers regardless of their subenclosure identifier.

Order element type headers on 1) element type and 2) subenclosure id for consistency. More than one type descriptor headers can have the same element type.

Disk drives also have temperature reporting capability and one might be tempted to add support for disk drive temperature sensors. However, the Temperature Sensor element type is not meant for this purpose. It represents a temperature sensing device only.

Diagnostic Pages

Element Descriptor diagnostic page

- This page provides the ASCII name for all the elements that the application client can use for display.

Help Text and Subenclosure Help Text diagnostic pagess

- Provides text that describes the present state of the enclosure and indicates what corrective action can be taken.

String Out/In and Subenclosure String Out/In diagnostic pages

- Provides a mechanism to pass any vendor specific binary data between the application client and the ESP. It may have indicator states, switch states, text, graphics or even control/status information for elements not defined in the configuration diagnostic page.

Diagnostic Pages

Threshold Out/In diagnostic pages

- Helps establish threshold values for the element types that support it.
- Supported thresholds are high critical, high warning, low warning and low critical.
- There is always an overall threshold control/status descriptor for every element type in the Threshold Out/In page.
- Overall & Individual descriptors are identical.

These values are compared with the reading of an element and used to indicate the status of the associated element and the overall enclosure condition.

Standard says that this page has threshold descriptor for “each” type descriptor header in the Configuration diagnostic page. However, “each” in this context means only those type descriptor headers that define an element type that supports threshold values. Thus, you do not have threshold descriptors for Array Device Slot or Cooling or Enclosure etc. in the Threshold Out/In diagnostic page. This is obvious by the fact that the threshold descriptor do not have a SELECT bit.

Diagnostic Pages

Enclosure Busy diagnostic page

- This page can be returned in place of a requested page to indicate that the ESP is busy.
- One can also return a vendor specific value in the given field to indicate specific reason.

Short Enclosure Status diagnostic page

- Only simple enclosures support this page.
- Returns a vendor specific status of the enclosure.

Enclosure Control/Status diagnostic pages

- Used to send and receive control and status information respectively.
- Information is passed using control or status descriptor for various element types.

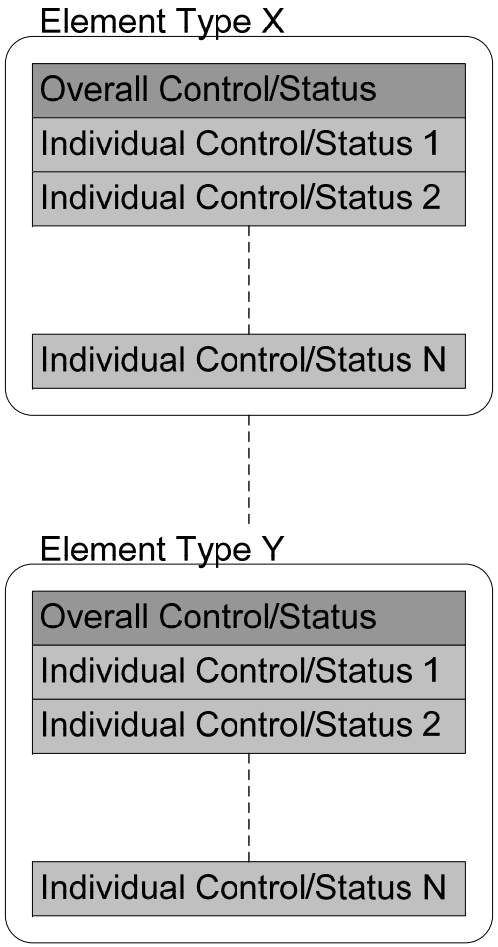
cont…

Diagnostic Pages

- Device Slot and Array Device Slot element types represents the disk drive slot in the enclosure. The descriptors are similar. Array Device Slot has provision for a few more indicators that are useful for disk drives participating in RAID configuration.

Since Array Device Slot descriptor is a superset of the Device Slot descriptor, it is better to implement only the former element type. In case you use parallel SCSI, you will need the Device Slot also as its status descriptor has the SLOT ADDRESS field.

- There is always an overall control/status descriptor for every element type in the Enclosure Control/Status page.
- Overall & Individual descriptors are identical.



Diagnostic Pages

- Control/Status descriptor format…

Byte	Field
0	COMMON CONTROL/STATUS
1 – 3	ELEMENT TYPE SPECIFIC CONTROL/STATUS

- Common Control format...

Byte\Bit	7	6	5	4	3	2	1	0
0	SELECT	PRDFAIL	DISABLE	RST SWAP	RESERVED			

Control bits are considered by the ESP only if the SELECT bit is set to one.

- Common Status format...

Byte\Bit	7	6	5	4	3	2	1	0
0	RESERVED	PRDFAIL	DISABLED	SWAP	ELEMENT STATUS CODE			

Diagnostic Pages

Element type specific control descriptor samples:

- Array Device Slot control descriptor

Byte\Bit	7	6	5	4	3	2	1	0
0	COMMON CONTROL							
1	RQST OK	RQST RSVD DEVICE	RQST HOT SPARE	RQST CONS CHECK	RQST IN CRIT ARRAY	RQST IN FAILED ARRAY	RQST REBUILD/ REMAP	RQST R/R ABORT
2	RQST ACTIVE	DO NOT REMOVE	RESERVED	RQST MISSING	RQST INSERT	RQST REMOVE	RQST IDENT	RESERVED
3	RESERVED		RQST FAULT	DEVICE OFF	ENABLE BYP A	ENABLE BYP B	RESERVED	

  - Most bits instruct to turn on/off their respective visual indicators.
  - **DO NOT REMOVE**, **RQST INSERT**, **RQST REMOVE** bits might control a mechanical lock or a visual indicator.
  - **DEVICE OFF** bit set to one is an instruction to turn (or keep) OFF the device. Zero is an instruction to turn (or keep) ON the device.
  - **ENABLE BYP A/B** bits are instruction to bypass a port.

Diagnostic Pages

- Cooling control descriptor

Byte\Bit	7	6	5	4	3	2	1	0
0	COMMON CONTROL							
1	RQST IDENT	RESERVED						
2	RESERVED							
3	RESERVED	RQST FAIL	RQST ON	RESERVED		REQUESTED SPEED CODE		

  - **RQST IDENT** and **RQST FAIL** are for visual indications.
  - **RQST ON** bit set to one is an instruction to turn (or keep) ON the cooling device (ex. a fan). Zero is an instruction to turn (or keep) OFF the device.
  - **REQUESTED SPEED CODE** field is an instruction to set the fan speed or specifically the rate of cooling. There are seven possible rate of cooling.

Map actual speeds to the SES speed codes.

Diagnostic Pages

Element type specific status descriptor samples:

- Array Device Slot status descriptor

Byte\Bit	7	6	5	4	3	2	1	0
0	COMMON STATUS							
1	OK	RSVD DEVICE	HOT SPARE	CONS CHECK	IN CRIT ARRAY	IN FAILED ARRAY	REBUILD/R EMAP	R/R ABORT
2	APP CLIENT BYPASSED A	DO NOT REMOVE	ENCLOSURE BYPASSED A	ENCLOSURE BYPASSED B	READY TO INSERT	RMV	IDENT	REPORT
3	APP CLIENT BYPASSED B	FAULT SENSED	FAULT REQSTD	DEVICE OFF	BYPASSED A	BYPASSED B	DEVICE BYPASSED A	DEVICE BYPASSED B

- Most bits reflect the (actual) state of the indicators. This state is a combination of the result of handling the request from application client(s) and the action that the ESP took on its own.

Cont…

Diagnostic Pages

- **BYPASSED A/B** bits shows if the port is bypassed or not.
- **APP CLIENT BYPASSED A/B** bits shows if the port was bypassed due to a request from the application client or not.
- **ENCLOSURE BYPASSED A/B** bits shows if the port was bypassed due to a request from the ESP or not.
- **DEVICE BYPASSED A/B** bits shows if the port was bypassed by the Device itself or not.
- **DO NOT REMOVE, READY TO INSERT and RMV** bit shows if the mechanical lock is prepared or not or the visual indicator is set or not.
- **REPORT** bit: Most status descriptors have this field. A one indicates that the device described by the status descriptor, containing the field, is one of the devices through which the diagnostic page was transmitted.
- **FAULT SENSED** bit indicates if the ESP sensed a fault condition or not. The Fault indicator may be ON.
- **FAULT REQSTD** bit reflects the value of **RQST FAULT** bit in the last control descriptor received by the ESP.

Diagnostic Pages

- Cooling status descriptor

Byte\Bit	7	6	5	4	3	2	1	0
0	COMMON STATUS							
1	IDENT	RESERVED				(MSB)		
2	ACTUAL FAN SPEED							(LSB)
3	HOT SWAP	FAIL	RQSTED ON	OFF	RESERVED	ACTUAL SPEED CODE		

- **IDENT** and **FAIL** are for visual indications.
- **RQSTED ON** bit reflects the value of **RQST ON** bit in the last control descriptor received by the ESP. Alternatively, it also reflects any manual operation that might have been performed.
- **OFF** bit indicates if the device is OFF or not.
- **ACTUAL FAN SPEED** is the actual speed in RPM divide by 10. Application client has to multiply the value by 10 to get the speed in RPM.
- **ACTUAL SPEED CODE** bit shows the actual rate of cooling. There are seven possible rates.

Your mapping of actual speeds to the SES speed code is vendor specific. Thus, the value in the ACTUAL FAN SPEED and the ACTUAL SPEED CODE should be matching. You cannot have the ACTUAL SPEED CODE showing what the ESP was trying to do instead of what the current actual speed rate is.

Diagnostic Pages

Additional Element Status diagnostic page

- This page provides additional information for…
  - Device Slot
  - Array Device Slot
  - SAS Expander
  - SCSI Initiator Port
  - SCSI Target Port and
  - Enclosure Services Controller Electronics
- Additional information might include protocol identifier, number of ports/phys, port/phy descriptors, port/phy identifier etc.

Diagnostic Pages

Download Microcode Control/Status diagnostic pages

- This page allows one to download a vendor-specific microcode, ex. a firmware image or a manufacturing page.
- The destination of the microcode is specified by the BUFFER ID field. The buffer ids are vendor-specific.
- Microcode can be sent in parts if it does not completely fit in one page.

Subenclosure Nickname Control/Status diagnostic pages

- This page sets or gets the nick name for the enclosure(s).
- Nick name strings are controlled by Language element if supported. Else, they are ASCII strings.
- Nick names can be persistently stored.

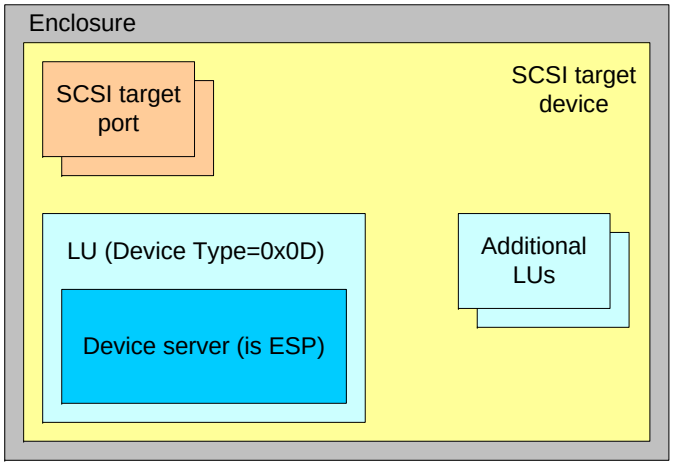
SES Model

SES Model

- ESP can be either **Standalone** or **Attached**.

Standalone ESP

- A standalone ESP's LU has a device type 0x0D (enclosure services) and sets the ENCSERV bit to one in the Standard INQUIRY data.
- ESP is the Device Server for this LU.

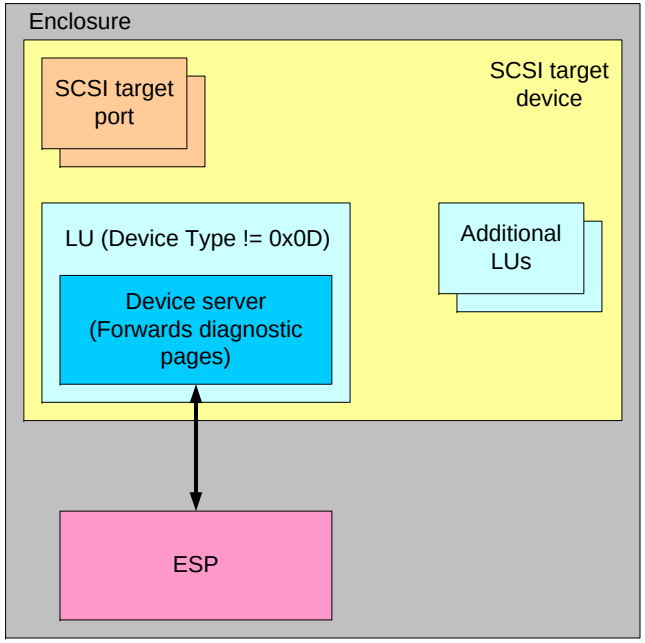


SES Model

Attached ESP

- An attached ESP is indirectly accessible through a LU which has a device type other than 0x0D.
- The interface between the Device server and the attached ESP is outside the scope of SES.
- The Device server sets the ENCSERV bit to one in the Standard INQUIRY data.

[cont…]



SES Model

- LU does not communicate hard reset, LU reset or I_T nexus loss to an attached ESP. Power On is the only supported device condition.
- Device server does not validate the CDB and returns a GOOD status. ESP reports CDB errors using **INVOP** bit in the Enclosure Status or Threshold In diagnostic page. ESP may also use Invalid Operation Reason element when returning one of the above mentioned status page with INVOP bit set to one.

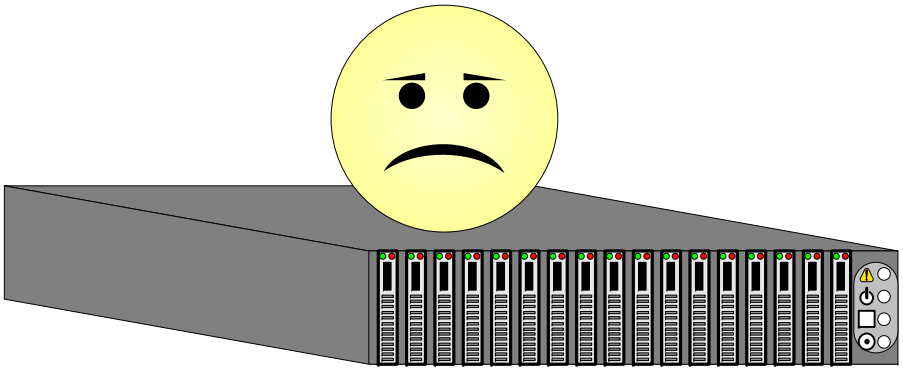
It is a good idea to support Invalid Operation Reason element for an attached ESP.

- Device server directs diagnostic pages 01h-2Fh to the attached ESP and reports all 00h-2Fh in the Supported Diagnostic Pages diagnostic page.

SES Model

Simple enclsoure.

- This may be defined as an enclosure with limited functionality.
- They respond with Short Enclosure status diagnostic page, no matter which SES diagnostic page is requested.
- SEND DIAGNOSTIC command is terminated with a check condition. All one can get is status.

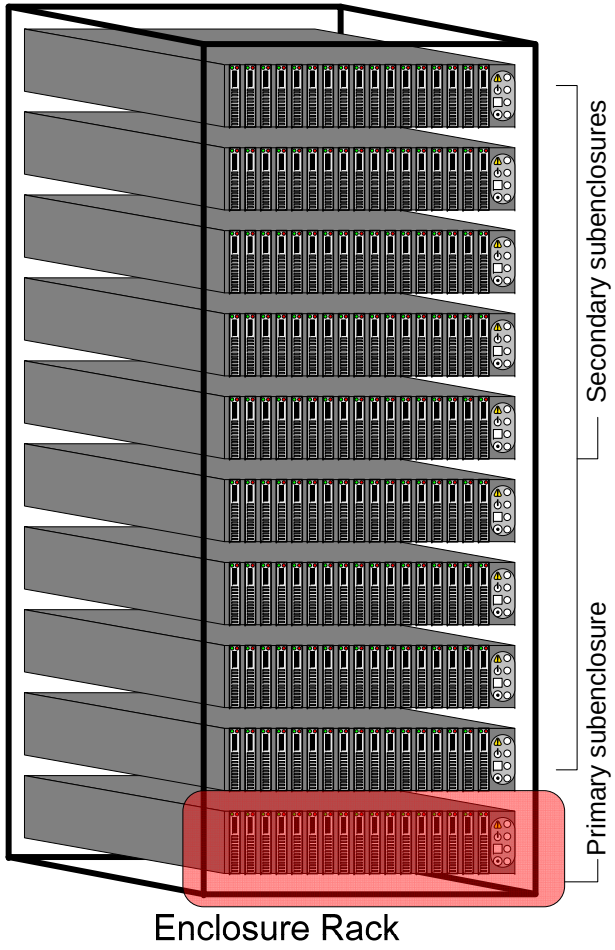


SES Model

Subenclosures

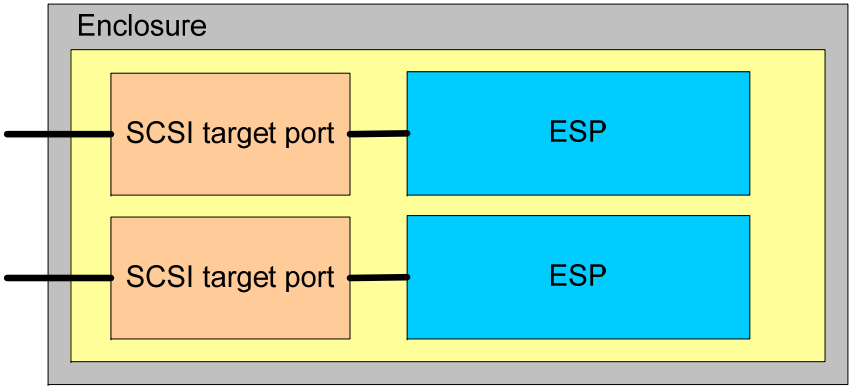
- SES standard allows more than one enclosures to be serviced by a single ESP.
- Primary subenclosure is the one whose ESP communicates with the application client.
- Other subenclosures contribute to the contents of the diagnostic pages transmitted by the primary subenclosure.
- Subenclosures are identified using a subenclosure identifier in the Configuration diagnostic page.
- The primary subenclosure assigns an identifier to each subenclosure. Relationship between a subenclosure identifier and a subenclosure location is vendor specific.
- Due to addition and removal of subenclosures, the subenclosure identifier of a physical subenclosure might change.

[cont…]



SES Model

- **GENERATION CODE**, a common field in most of the diagnostic pages, shall be changed as a result of adding or removing subenclosures.
- Standalone ESP establishes a unit attention condition when there is a change in the generation code.
- Device servers with an attached ESP do not establish a unit attention condition.
- Interface between the primary subenclosure and the secondary subencloures may be either SCSI or vendor specific.
- SES standard also allows for one enclosure to be serviced by more than one ESP.



Implementation Ideas

Implementation Ideas

- ESP needs an SES database where it can store the status of every device it supports. This database will also be used to cache control information from the application client.
- The SES database can be as simple as an array of structures of various element type's control/status descriptors.
- It would help to design a generic module for the ESP with options to "configure" it such that it is easily portable from one hardware to another.
- One might want to make most of the things configurable through configuration files or manufacturing pages to be read at power on state to initialize the SES database.
- Configurable things might include lists of supported pages, list of supported elements, their vital parameters for initialization and read/write operations, thresholds, fan speed - speed code mapping etc.
- In most cases one might want to implement redundant ESP to ensure availability. Thus, implement generic interfaces to share state and data between multiple ESP.
- Try sg tools (http://sg.torque.net/sg/) for testing.

Possible Improvements

Possible Improvements

- The standard might provide a mechanism to associate one element to one or more other elements. For example…
  - Device Slot, Array Device Slot or Power Supply might be associated to Temperature Sensor or Cooling elements.
  - Temperature Sensor and Cooling elements might be associated to each other.
  - UPS might be associated to Power Supply.
  - This association will help define the configuration in more detail and the dependencies clearly identified. For ex. warning in temperature sensor 0 might lead to higher fan speed of cooling elements 0-3, rest untouched.
- A mechanism to identify redundant elements. This helps ESP to identify the correct condition of certain devices. For ex. there might be four power supplies. Power supply element 0 & 1 are redundant and so are 2 & 3. Failure of one from each set is a warning condition where as failure of one set is critical.

Thanks

...

Q & A