GREEN SCSI

Giri Basava,

Consultant Software Engineer, EMC²

Agenda

- Objective
- What is Green Storage?
- What is Green SCSI?
- SAN
  - Who is in control?
  - Host
  - Array
- SCSI Standards
  - Power Conditions
  - Power Condition Control Methods
    - Power Condition Mode Page
    - Start Stop Unit
  - Power Condition Change Notifications

Agenda…

- Standards... work in progress
- Call for action
- Q&A

Objective

- Customers demand reduction in power and cooling costs.
- Modern disk drives support the ability to directly control power consumption, but it is necessary to extend this to storage area networks to fully meet customer demands.
- This presentation will not discuss how a storage array can spin up/down the multiple types of disk drives it may have.
- This presentation reviews existing SCSI power management capabilities and outlines how to use these capabilities to allow operating systems and external storage arrays to cooperatively reduce power consumption.

What is Green Storage?

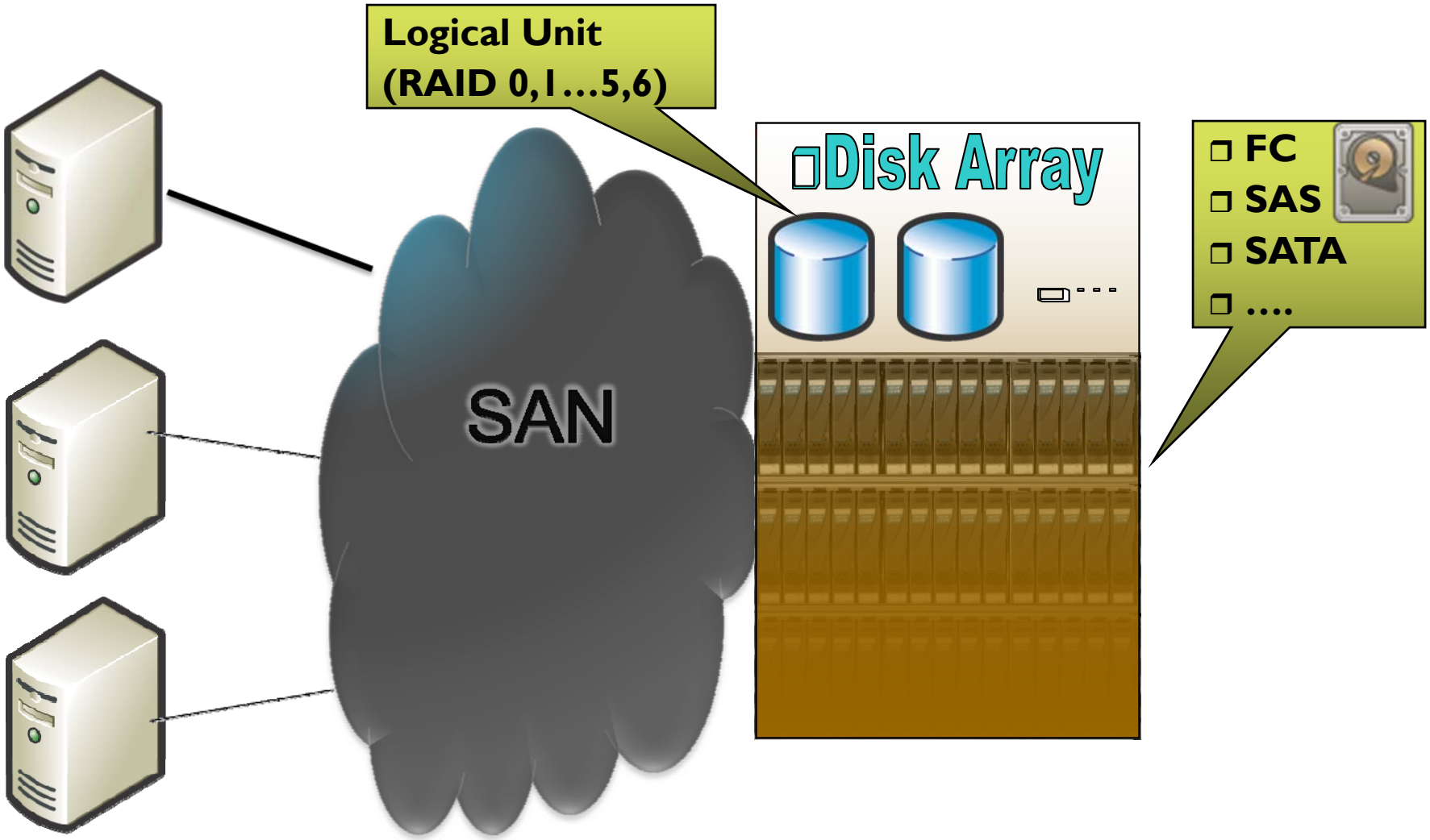
- Green storage means
  - Reducing hazardous materials (by recycling storage systems and components)
  - Reducing power consumption in the storage eco system
    - Disk Arrays
    - Switches
    - Power distribution units
    - Uninterruptible power supplies
    - Fans, Controllers, Hard drives
    - Applications/Software

  - Apply technologies like thin provisioning, deduplication, SSD, MAID
  - Etc.
- Green storage is a combination of technologies, practices and policies that lead to lower and more efficient energy use

What is Green SCSI?

- Green SCSI is the subset of SCSI standards to enable Operating Systems(or applications) and Storage to reduce power consumption cooperatively.

SAN



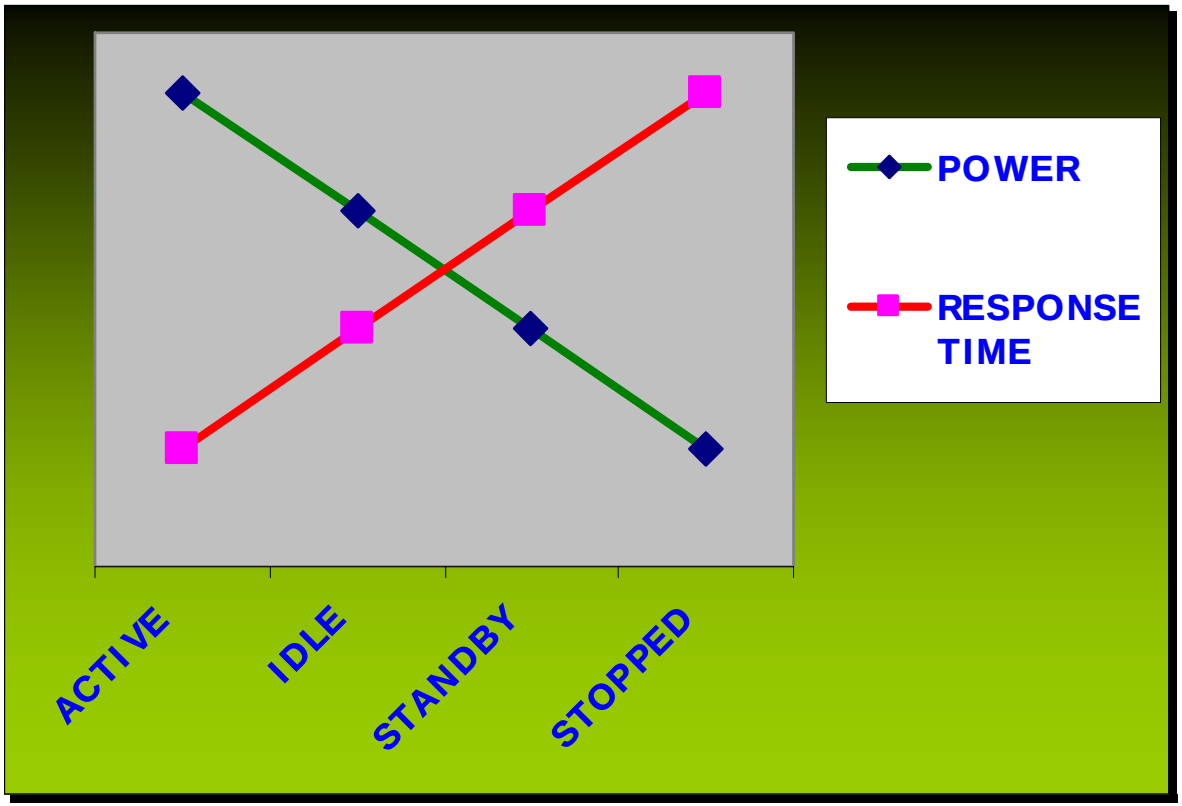
SAN - Who is in control?

- Host (Explicit Spin up/down)
  - Operating Systems or Applications initiate the request to spin down the logical units on the arrays.

- Storage Array (Implicit Spin up/down)
  - Everything is spinning, and array spins down the disks based on policies.
    - Works best for applications and Operating Systems that can detect the implicit spin down
  - MAID
    - With MAID, everything is idle -- array spins the disks up when requiried.
    - Works best for MAID aware applications and Operating Systems that can handle the performance hit.
  - Disk Drives Primitives
    - SCSI/SAS
      - Active, Idle, Standby, Stopped
    - SATA
      - Idle, Idle Immediate
      - Sleep
      - Standby, Standby Immediate
    - Etc…
    - Disk Drive manufacturers are continuously innovating to reduce the power consumption. More power states and primitives are being added to the standards.

SCSI Standards – Power Conditions

- Current SCSI standards define the following Power Conditions for a Logical Unit.
  - Active
  - Idle
  - Standby
  - Stopped



SCSI Standards – Power Conditions

- Active - while in the active power condition
  - Array is capable of responding to all of its supported commands including media access requests;
  - A logical unit completes processing of operations in the shortest time when compared to the time required for completion while in the idle, standby or stopped power conditions; and
  - The SCSI target device may consume more power than when the logical unit is in the idle power condition (e.g., a disk drive's spindle motor may be active)

SCSI Standards – Power Conditions

- Idle - while in the idle power condition
  - Array is capable of responding to all of its supported commands including media access requests;
  - A logical unit may take longer to complete processing a command than it would while in the active power condition (e.g., the device may have to activate some circuitry before processing a command); and
  - The power consumed by the SCSI target device should be less than or equal to the power consumed when the logical unit is in the active power condition and may be greater than the power consumed when the logical unit is in the standby power condition.

SCSI Standards – Power Conditions

- Standby - while in the standby power condition
  - Array is not capable of processing media access commands; and
  - The power consumed by the SCSI target device should be less than or equal to the power consumed when the logical unit is in the idle power condition (e.g., a disk drive's spindle motor is stopped).

SCSI Standards – Power Conditions

- Stopped - while in the stopped power condition
  - Array is not capable of processing media access commands; and
  - The power consumed by the SCSI target device should be less than or equal to the power consumed when the logical unit is in active, idle or standby power conditions.
  - The array will terminate each medium access command or TEST UNIT READY command with CHECK CONDITION status with the sense key set to NOT READY and the additional sense code set to LOGICAL UNIT NOT READY, INITIALIZING COMMAND REQUIRED.

SCSI Standards – Power Condition Control Methods

- Power Condition Mode Page
- Start Stop Unit

SCSI Standards - POWER CONDITION MODE PAGE

- **POWER CONDITION MODE PAGE: Use this mode page to activate and set timer based spin down of the Logical Unit.**

Power Condition mode page

Bit / Byte	7	6	5	4	3	2	1	0
0	PS	SPF (0b)	PAGE CODE (1Ah)					
1	PAGE LENGTH (0Ah)							
2	Reserved							
3	Reserved						IDLE	STANDBY
4	(MSB)	IDLE CONDITION TIMER						
7								(LSB)
8	(MSB)	STANDBY CONDITION TIMER						
11								(LSB)

SCSI Standards - POWER CONDITION MODE PAGE ...

Power Condition mode page

Bit / Byte	7	6	5	4	3	2	1	0
0	PS	SPF (0b)	PAGE CODE (1Ah)					
1	PAGE LENGTH (0Ah)							
2	Reserved							
3	Reserved						IDLE	STANDBY
4	(MSB)	IDLE CONDITION TIMER						
7								(LSB)
8	(MSB)	STANDBY CONDITION TIMER						
11								(LSB)

- Power condition mode page allows the host to set the following timers to transition the logical unit in to the desired power state.
  - IDLE condition timer
  - STANDBY condition timer
- These timers are number of milliseconds in increments of 100ms.
- Enable Bits: IDLE & STANDBY – when set to 1 the respective timers are enabled.

SCSI Standards - POWER CONDITION MODE PAGE …

Power Condition mode page

Bit / Byte	7	6	5	4	3	2	1	0
0	PS	SPF (0b)	PAGE CODE (1Ah)					
1	PAGE LENGTH (0Ah)							
2	Reserved							
3	Reserved						IDLE	STANDBY
4	(MSB)	IDLE CONDITION TIMER						
7								(LSB)
8	(MSB)	STANDBY CONDITION TIMER						
11								(LSB)

- It is possible to move the logical unit into the desired state immediately by setting the timers to '0'.
- When the array receives a command while in a power condition based on a setting in the power condition mode page, the logical unit shall transition to the power condition that allows the command to be processed.
- Using this mode page to set the power states, leaves the control of the power state with the array.

SCSI Standards - START STOP UNIT

- **START STOP UNIT: Host can use SSU to bring a logical unit in to a specific power condition.**

START STOP UNIT command

Byte\Bit	7	6	5	4	3	2	1	0
0	OPERATION CODE (1Bh)							
1	Reserved							IMMED
2	Reserved							
3								
4	POWER CONDITION				Reserved		LOEJ	START
5	CONTROL							

SCSI Standards - START STOP UNIT…

START STOP UNIT command

Byte\Bit	7	6	5	4	3	2	1	0
0	OPERATION CODE (1Bh)							
1	Reserved							IMMED
2	Reserved							
3								
4	POWER CONDITION				Reserved		LOEJ	START
5	CONTROL							

- Start Stop Unit has precedence over Power Condition Mode Page.
- LU can be moved into STOP and START states.

SCSI Standards - START STOP UNIT…

POWER CONDITION **field**

Code	Name	Description
0h	START_VALID	Process the START and LOEJ bits.
1h	ACTIVE	Place the device into the active power condition.
2h	IDLE	Place the device into the idle power condition.
3h	STANDBY	Place the device into the standby power condition.
7h	LU_CONTROL	Transfer control of power conditions to the logical unit.
Ah	FORCE_IDLE_0	Force the idle condition timer to zero.
Bh	FORCE_STANDBY_0	Force the standby condition timer to zero.

- The following power conditions can also be set with this command.
  - ACTIVE
  - IDLE
  - STANDBY
  - LU_CONTROL – Transfer control of power settings to the array (i.e array will enable the idle condition timer if it is active and disable the standby condition timer if it is active).
  - FORCE_IDLE_0 – Reset idle timer.
  - FORCE_STANDBY_0 – Reset Standby Timer.

SCSI Standards - START STOP UNIT…

START STOP UNIT command

Byte\Bit	7	6	5	4	3	2	1	0
0	OPERATION CODE (1Bh)							
1	Reserved							IMMED
2	Reserved							
3								
4	POWER CONDITION				Reserved		LOEJ	START
5	CONTROL							

- When ACTIVE, IDLE & STANDBY are used by the host, the array should not change the power conditions on it's own until another START STOP or LU Reset. (i.e. it should disable the idle and standby timers)
- When LU_CONTROL is sent by the client, array gets the control of power condition states (i.e. array will enable the idle condition timer if it is active and disable the standby condition timer if it is active).
- When FORCE_IDLE_0 or FORCE_STANDBY_0 is received, array gets the control of the power condition and moves the LU in to IDLE or STANDBY state.

SCSI Standards - START STOP UNIT…

START STOP UNIT command

Byte\Bit	7	6	5	4	3	2	1	0
0	OPERATION CODE (1Bh)							
1	Reserved							IMMED
2	Reserved							
3								
4	POWER CONDITION				Reserved		LOEJ	START
5	CONTROL							

- IMMED
  - If this bit is set to ‘0’, array will return status only after the requested operation is completed.
  - If this bit is set to ‘1’, array will return status as soon as the CDB is validated.

SCSI Standards - POWER CONDITION CHANGE NOTIFICATIONS

- There will be no notification to the host that a logical unit has transitioned from one power condition to another.
- A host may determine the current power condition of a logical unit by issuing a REQUEST SENSE command.

SCSI Standards - REQUEST SENSE

- If the logical unit is in a power condition other than the active or stopped, array should respond with one of the following.

Description	Sense/ASCQ
LOW POWER CONDITION ON	5Eh/ 00h
IDLE CONDITION ACTIVATED BY TIMER	5Eh/ 01h
STANDBY CONDITION ACTIVATED BY TIMER	5Eh/ 02h
IDLE CONDITION ACTIVATED BY COMMAND	5Eh/ 03h
STANDBY CONDITION ACTIVATED BY COMMAND	5Eh/ 04h

SCSI Standards – Work In Progress

- Proposal - http://www.t10.org/ftp/t10/document.08/08-126r1.pdf
  - MANAGE POWER OUT
  - MANAGE POWER IN
  - Power Management VPD Page
  - New Power conditions
    - Sleep – Consumes the least amount of power possible, short of actually removing power from the device, requires a hard reset to wake up.
    - Low-Rpm Idle - The disk keeps spinning at a reduced rate, providing a shorter recovery time than the full spin-down required by the standby power condition. Not as much power savings as standby.
    - Park/Retract Heads - Reduce friction on the spinning disk during the idle power condition. Very fast recovery time. Not as much power savings as low-rpm idle.

SCSI Standards – Work In Progress

Table 2 — MANAGE POWER OUT command

Byte\Bit	7	6	5	4	3	2	1	0
0	OPERATION CODE (A4h)							
1	Reserved				SERVICE ACTION (nnh)			
2	Reserved							
5								
6	(MSB)			PARAMETER LIST LENGTH				
9								(LSB)
10	Reserved							
11	CONTROL							

SCSI Standards – Work In Progress

- Proposal - http://www.t10.org/ftp/t10/document.08/08-184r3.pdf
- This proposal add the following low power modes : idle2 and idle3.
- Also adds a recovery time parameter for each of the low power modes.

SCSI Standards – Work In Progress

Table 319 -- Power Condition Mode Page

Byte \ Bit	7	6	5	4	3	2	1	0
0	PS	SPF(0b)	Page Code (1Ah)					
1	Page Length							
2	Reserved							
3	Reserved			STANDBY2	IDLE3	IDLE2	IDLE	STANDBY
4	IDLE CONDITION TIMER							
5								
6								
7								
8	STANDBY CONDITION TIMER							
9								
10								
11								
12	IDLE2 CONDITION TIMER							
13								
14								
15								
16	IDLE3 CONDITION TIMER							
17								
18								
19								
20	STANDBY2 CONDITION TIMER							
21								
22								
23								
24	IDLE CONDITION RECOVERY TIME							
25								
26	STANDBY CONDITION RECOVERY TIME							
27								
28	IDLE2 CONDITION RECOVERY TIME							
29								
30	IDLE3 CONDITION RECOVERY TIME							
31								
32	STANDBY2 CONDITION RECOVERY TIME							
33								
34	RESERVED							
39								

SCSI Standards – Work In Progress

- Proposal – Link Layer Power Management
  - http://www.t10.org/ftp/t10/document.08/08-249r0.pdf
  - http://www.t10.org/ftp/t10/document.08/08-206r0.pdf
  - The SATA interface already defines a method to invoke lower power transceiver modes.
  - This proposal extends the similar functionality to SAS.

Call for Action

- Think what can you do to contribute to “GREEN”
- Review the standards, proposals and pass on your feedback to your(company) representative to T10.
- Let’s go “GREEN”…

Q&A

- Questions?