

Date: Oct. 21, 1998
To: T10 Committee
From: Gerry Houlder, Seagate Technology
Subj: Proposed SCSI Device Locks

The additions to describe the Device Locks feature involve:

- (a) Device Locks feature description added to model section,
- (b) Device Locks command description, and
- (c) Device Locks mode page description.

I believe all of these items belong in the SPC standard because this feature can be implemented on many different types of devices.

5.6 Device Locks

Device locks are mechanisms used in distributed environments to facilitate mutual exclusion of shared resources. They can further be used to maintain coherence of data that is cached in several locations. The locks are implemented on storage devices and accessed with the device lock command. This command is independent of the normal functions of the storage devices, so the devices supporting the locks have no awareness of the nature of the resource that is locked. Each lock requires only a small amount of memory, allowing devices to support thousands of these locks with minimal amounts of memory.

The operating system may assign several resources to a particular lock. When a client acquires a particular lock, all of the associated resources are reserved to that client even though it may use only one of the associated resources. When the client updates one resource, the lock's version number is changed. This means all of the resources associated with that lock must be treated as if they have changed. The operating system may use activity history of the resources when deciding which resources should be associated to a common lock.

Device locks are used among multiple initiators that require operations to be protected across initiator failures, which may involve hard resets. Even though different protocols that transport SCSI handle hard resets differently (e.g., parallel uses a reset signal, fibre channel uses log outs) the device lock shall be protected.

Device locks held by failing initiators may become unassigned after a timeout interval or may be pre-empted by another initiator as part of the recovery process. Device locks shall be retained by the device server until released, pre-empted, or cleared by mechanisms specified in this standard. The locks are not preserved across power cycles, however.

yy.yy DEVICE LOCKS command

The DEVICE LOCKS command (see table 1) is used to acquire, release, or obtain information about device locks.

Table 1 – DEVICE LOCKS command

Bit Byte	7	6	5	4	3	2	1	0
0	Operation Code (83h)							
1	Reserved				Action Code			
2	(MSB)							
3	Lock Number							
4								
5	(LSB)							
6	(MSB)							
7	Version Number							
8								
9	(LSB)							
10	(MSB)							
11	Allocation Length							
12								
13	(LSB)							
14	reserved							
15	Control							

The DEVICE LOCKS command shall not be affected by reservations or persistent reservations.

The action code field specifies a device lock action. If the action code field contains an unsupported value, the device server shall return a CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST and additional sense data shall be set to INVALID FIELD IN CDB.

The lock number field contains the address of the desired device lock. A value of FFFFFFFFh indicates all device locks. Some action codes specifically allow this as a valid value, otherwise this is treated as an invalid value. If the lock number field contains an invalid or unsupported value, the device server shall return a CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST and additional sense data shall be set to INVALID FIELD IN CDB.

The version number field is ignored unless specified otherwise in the action code description.

The actual length of the DEVICE LOCKS parameter data is available in a parameter data field. The allocation length field in the CDB indicates how much space has been reserved for the returned parameter data. If the length is not sufficient to contain the entire parameter list, the first portion of the list shall be returned. This shall not be considered an error. If the remainder of the list is required, the application client should send a new DEVICE LOCKS command with an allocation length field large enough to contain all of the data.

yy.yy.1 DEVICE LOCKS action codes

The DEVICE LOCKS action codes are defined in table 2.

Table 2 – DEVICE LOCKS action codes

Code	Name	Description	Data Returned
0h	No Operation	No change, return current lock state	Type 1
1h	Lock shared	Acquire shared lock	Type 1
2h	Lock Exclusive	Acquire exclusive lock	Type 1
3h	Force Lock Exclusive	Acquire exclusive lock, preempting lock if needed	Type 1
4h	Refresh Lock	Reset timer on lock	Type 1
5h	Unlock	Release lock	Type 1
6h	Unlock Increment	Release lock and increment version number	Type 1
7h	Activity On	Turn on activity monitor	Type 1
8h	Activity Off	Turn off activity monitor	Type 1
9h	Report Expired	Report which locks have expired	Type 2
Ah-Fh	Reserved		

There are two formats for data returned by the device locks actions. The returned data column specifies which format for returned data is used for each action code.

All of the device lock actions that return type 1 data check the status of the addressed device lock to determine if that lock has expired. If the lock has expired, the state of that lock is set to unlocked and the expired field is set according to the previously expired state. The device lock action that returns type 2 data checks all device locks to determine if they have expired. All expired device locks have their state and expired fields updated before data is returned.

yy.yy.1.1 No Operation

The addressed device lock is unchanged unless it was locked and the timer has expired. The returned data describes the current status of the addressed device lock. The result bit is always set to 1.

yy.yy.1.2 Lock Shared

If the addressed device lock is unlocked, then

- (a) if the expired field is zero, the state field is set to locked shared; else the state field is set to locked exclusive;
- (b) the expired field reflects whether the lock expired from the previous lock holder,
- (c) the number of holders is set to 1,
- (d) the initiator's WWN is put in the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

If the addressed device lock is in locked shared state and another holder can be added, then

- (a) the state field remains set to locked shared,
- (b) the expired field remains unchanged,
- (c) the number of holders is incremented by 1,
- (d) the initiator's WWN is added to the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

If the addressed device lock is in locked exclusive state and initiator's WWN is the holder, then

- (a) the state field is set to locked shared,
- (b) the expired field remains unchanged,
- (c) the number of holders remains set to 1,
- (d) the initiator's WWN remains in the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

Otherwise the lock status is unchanged and the result bit is set to zero.

yy.yy.1.3 Lock Exclusive

If the addressed device lock is unlocked, then

- (a) the state field is set to locked exclusive,
- (b) the expired field reflects whether the lock expired from the previous lock holder,
- (c) the number of holders is set to 1,
- (d) the initiator's WWN is put in the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

If the addressed device lock is in locked shared or locked exclusive state, the number of holders is 1, and the initiator's WWN is the holder, then

- (a) the state field is set to locked exclusive,
- (b) the expired field remains unchanged,
- (c) the number of holders remains at 1,
- (d) the initiator's WWN remains on the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

Otherwise the lock status is unchanged and the result bit is set to zero.

yy.yy.1.4 Force Lock Exclusive

If the addressed device lock is unlocked, then

- (a) the state field is set to locked exclusive,
- (b) the expired field reflects whether the lock expired from the previous lock holder,
- (c) the number of holders is set to 1,
- (d) the initiator's WWN is put in the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

If the addressed device lock is in locked shared or locked exclusive state and the version number in the CDB matches the version number of the lock, then

- (a) the state field is set to locked exclusive,
- (b) the expired field remains unchanged,
- (c) the number of holders is set to 1,
- (d) the WWN list is cleared and the initiator's WWN is put in the list of holders,
- (e) the expire time is reset, and
- (f) the result bit is set to 1.

Otherwise the lock status is unchanged and the result bit is set to zero.

yy.yy.1.5 Touch Lock

If the lock number field is less than FFFFFFFFh, the addressed lock is in locked shared or locked exclusive state, and the initiator's WWN is on the list of holders; then the device resets the expire time on the addressed device lock and sets the result bit to 1. If these conditions aren't met, the device lock status is unchanged and the result bit is set to 0.

If the lock number field is set to FFFFFFFFh, the device resets the expire time on all device locks that are in locked shared or locked exclusive state and include the initiator's WWN as a holder. All other fields are unchanged. If at least one device lock had its expire time reset, the result bit is set to 1, else the result bit is set to 0.

[GAH What is returned data format if lock number field is FFFFFFFFh? Type 1 with all fields except Result field set to zero?]

yy.yy.1.6 Unlock

If the addressed lock is in unlocked state, then device status is unchanged and the result bit is set to 0.

If the addressed lock is in locked shared or locked exclusive state and the initiator's WWN is in the list of holders, then

- (a) the number of holders is decremented by 1,
- (b) the initiator's WWN is removed from the list of holders,
- (c) the expired field is set to 0,
- (d) if the number of holders has been reduced to 0, the state is set to unlocked; else the state is unchanged;
- (e) if the activity bit is set to 1, increment the version number; else the version number is unchanged; and
- (f) the result bit is set to 1.

Otherwise the lock status is unchanged and the result bit is set to zero.

yy.yy.1.7 Unlock Increment

If the addressed lock is in unlocked state, then device status is unchanged and the result bit is set to 0.

If the addressed lock is in locked shared or locked exclusive state and the initiator's WWN is in the list of holders, then

- (a) the number of holders is decremented by 1,
- (b) the initiator's WWN is removed from the list of holders,
- (c) the expired field is set to 0,
- (d) if the number of holders has been reduced to 0, the state is set to unlocked; else the state is unchanged;
- (e) the version number is incremented by 1, and
- (f) the result bit is set to 1.

Otherwise the lock status is unchanged and the result bit is set to zero.

yy.yy.1.8 Activity On

The activity bit for the addressed lock is set to 1. All other device lock status fields remain unchanged. The result bit is set to 1.

yy.yy.1.9 Activity Off

The activity bit for the addressed lock is set to 0 and the version number is incremented by 1. All other device lock status fields remain unchanged. The result bit is set to 1.

yy.yy.1.10 Report Expired

For this action the lock number field is ignored. The device server will scan all supported device locks and record any device lock whose time has expired. If any device lock has expired, the result bit is set to 1; else the result bit is set to 0.

The returned data will be of type 2 format. If the result bit is 0, the data length shall also be zero. If the result bit is 1, the entire device lock bit map shall be returned and the data length is set to the length of the bit map.

yy.yy.2 DEVICE LOCKS parameter data

There are two different formats used for parameter data provided in response to a DEVICE LOCKS command. The format used by each action code is shown in Table 2. In general, the type 1 response is used to return all parameter data for a specific device lock and type 2 response is used to return only the expired status for all of the device locks.

yy.yy.2.1 Type 1 parameter data

The type 1 parameter data format is shown in table 3.

Table 3 – Type 1 DEVICE LOCKS parameter data

Bit Byte	7	6	5	4	3	2	1	0
0	(MSB) _____							
1	Version Number							
2	_____							
3	_____							(LSB)
4	Result	Act	reserved		Expired		State	
5	Number of Holders							
6	(MSB) _____							
7	Holder List Length (n-7)							(LSB)
	List of Holders							
8	(MSB) _____							
15	World Wide Name (first)							(LSB)

n-7	(MSB) _____							
n	World Wide Name (last)							(LSB)

The version number field indicates whether a resource represented by the lock may have been updated. The value is set to 0 after a power cycle. It is also incremented as defined in the DEVICE LOCKS action code descriptions (see clause yy.yy.1).

A result bit of 1 indicates that the requested action was carried out successfully. A result bit of 0 indicates the requested action could not be performed.

An Act (Activity) bit of 1 indicates that the version number is incremented by more actions than just the unlock increment action code. An Act bit of 0 indicates that only the unlock increment action code causes the version number to increment.

The expired field indicates whether the previous holder of the addressed device lock unlocked it within the lock timeout interval or whether the timer expired. The values are described in table 4.

Table 4 – Expired field values

Code	Description
0	Not expired
1	Expired from locked shared
2	Expired from locked exclusive
3	Reserved

The state field indicates the current state of the addressed device lock. The values are described in table 5.

Table 5 – State field values

Code	Description
0	Unlocked
1	Locked shared
2	Locked exclusive
3	Reserved

The number of holders field indicates the number of initiators that hold the addressed lock.

The holder list length indicates the length in bytes for the list of holders. Each list entry is the 8 byte World Wide Name (WWN) for each of the holders. The holder list length must be a multiple of 8. The holder list length may be 0 in some cases even though the number of holders is non-zero.

yy.yy.2.2 Type 2 parameter data

The type 2 parameter data format is shown in table 6.

Table 6 – Type 2 DEVICE LOCKS parameter data

Bit	7	6	5	4	3	2	1	0
Byte								
0	Result	Reserved						
1	Reserved							
2	(MSB)	Data Length (n-3)						(LSB)
3								
Bitmap of expired device locks								
4	L7	L6	L5	L4	L3	L2	L1	L0
5	L15	L14	L13	L12	L11	L10	L9	L8
6	L23	L22	L21	L20	L19	L18	L17	L16
:								
n	LM							
Where $M = (n - 3) * 8 - 1$								

A result bit of 1 indicates that at least one expired lock was found. A result bit of 0 indicates that no expired locks were found.

The data length field indicated the number of bytes in the expired device lock bit map. When the result bit is 0, the data length field may also be set to zero. When the result bit is 1, the data

length shall reflect the number of bytes required to return a bit map of all device locks supported by the device server. Each 1 bit in the bit map indicates that the timer expired for the device lock associated with that bit.

zz.zz Device locks page

The device locks page (see table 100) provides control and reporting of information about the device locks feature. If the device locks mode page is supported, the DEVICE LOCKS command (see clause yy.yy) shall be supported.

Table 100 – Device locks page

Bit Byte	7	6	5	4	3	2	1	0
0	PS	Resvd	Page code (xxh)					
1	Page length (0Ah)							
2	Reserved							
3	Maximum Clients per Lock							
4	(MSB)	Number of Locks						
5								
6								
7								(LSB)
8	(MSB)	Lock Timeout Interval						
9								
10								
11								(LSB)

[GAH Need to get a reserved page code assignment for this mode page.]

The maximum clients per lock field indicates the maximum number of clients that can simultaneously hold a lock in the lock shared state.

The number of locks field indicates the number of locks that exist on this device.

The lock timeout interval field indicates the number of milliseconds after which an untouched lock will timeout. A value of zero or a value of FFFFFFFFh indicate an infinite timeout interval. If the device cannot implement the exact time interval that is requested, the device may round the value up to the nearest implemented value.