

# Improving Storage System Resiliency: Stealth Intelligent Change Manager

## AT A GLANCE

This paper takes a look at the remarkable benefits realized by storage system OEMs and end users by using the Stealth Intelligent Change Manager capabilities within back-end applications.

## PRODUCTS

InSpeed SOC 422

## APPLICATIONS

SBODs  
Fully Switched Architectures

## Introduction

Storage Area Network (SAN) systems are fundamentally comprised of two distinct areas: The front-end SAN and the back-end storage. The back-end of the vast majority of SANs today consist of Fibre Channel technologies and there the Fibre Channel Arbitrated Loop (FC-AL) protocol dominates. With FC-AL systems all devices are serially connected; meaning a problem with one device can cause loss of communication to all devices – much as one bad light in a string of holiday lights can cause all lights in that string to go out. This is a primary reason why Fibre Channel and all Enterprise-class storage systems are implemented with channel redundancy – if access is not available to a drive from one path, then access to that drive can be achieved through the redundant path.

Several years ago, Emulex introduced InSpeed® technology to offer system architects switched point-to-point connections to all drives in the back-end of storage systems and alleviate the serial connection problems of FC-AL. InSpeed technology has been so successful that millions of ports of InSpeed have now been successfully deployed throughout the world – making it the de facto standard for back-end switching.

Some persistent problems, however, have been noted in storage systems containing a large number of disks regardless of the technology or protocol used. These system resiliency issues typically are the direct result of changes made to the system infrastructure. These changes can occur when a hard drive is removed, replaced, or becomes rogue, when a link shows intermittent behavior, or any other of a number of events causing a change in the overall system configuration.

InSpeed technology includes patented and patent pending technology known as the Stealth Intelligent Change Manager that takes a major step forward in removing the remaining system resiliency issues. Stealth Intelligent Change Manager was originally designed for applications, such as video streaming and editing, or tape backup systems, where disruptions due to change could not be tolerated. Stealth Intelligent Change Manager is being used today by most tape back-up and video editing vendors due to this special, uninterrupted streaming capability. The technology behind the Stealth Intelligent Change Manager is so beneficial to system stability it also meets the requirements imposed by Microsoft for allowing FC-AL in applications previously reserved for only fabric switches. Microsoft Knowledge Base Article ID: 317162 covers this capability.

Recently, Emulex has advanced the Application Programming Interface (API) of its InSpeed products to extend Stealth Intelligent Change Manager capabilities to back-end embedded applications. This use of Stealth Intelligent Change Manager between the disk drives and initiators is being recognized as a dramatically powerful tool improving resiliency in embedded, back-end disk storage systems.

## System Resiliency

The ability for a large number of disk drives to become and remain available and in communication with initiators through a deterministic, rapid method is an indication of the resiliency of a storage system. The benefits of a resilient storage system include greater system reliability and availability, quicker identification of faulty devices, faster system initialization, fewer opportunities for rogue system behaviors, and fewer disruptions to a system when changes do occur. Multiple conditions may affect the ability of disk drives to recover from a change in the system topology. One such condition is the removal of a disk drive whether intentional or due to hard or intermittent failure. In FC-AL systems, removal of a disk causes a LIP (Loop Initialization Primitive) to occur. A LIP traverses the entire loop and causes all devices in the loop to reset their loop state machines. When all devices are folded into a loop, command timeout errors can occur, which may add seconds to a recovery. Extended recovery greater than a second is often unacceptable in storage systems, which is just one of the reasons SAS expander systems are having problems scaling and are adding auto-configuration capabilities.

**Table 1: Change Effects on a Back-End System using the Stealth Intelligent Change Manager**

CHANGE EVENT	SYSTEM EFFECT
Drive Inserted	Initiators and disk drive updated
Drive Removed	Only initiators updated
Initiator Inserted	Only initiators updated
Initiator Removed	No effects
Drive Generates a LIP	Initiators and one disk drive updated
Initiator Generates a LIP	Only initiators updated
Link Inserted or Removed	Only initiators updated
Trunk Link Inserted or Removed	No effects

Stealth Intelligent Change Manager technology enhances the point-to-point nature of InSpeed by limiting LIP change behavior in a storage system to a very small set of devices. The various possible combinations are shown in Table 1.

The above system effects are based on initiators not communicating with each other, which is the default mode and is quite common. If initiators actively communicate with each other, policies can be changed to include identified initiators whenever a change does occur.

One of the keys to the change effects in Table 1 are that targets typically do not care about other targets or initiators, and therefore do not need to be informed of any changes. Initiators, by comparison, must be informed that a target has either appeared or disappeared so that it knows how and where to direct its data, or to determine if it needs to begin a RAID rebuild cycle.

### Stealth Intelligent Change Manager

Each port on an InSpeed switch can be set to operate in one of four different modes, as is shown in Table 2. All back-end InSpeed SBOD<sup>®</sup> (Switched Bunch Of Disks) applications to date – excluding streaming – have used the normal FC-AL port operation mode for connection to targets and initiators, and the Switch-to-Switch mode for connecting one SBOD to another. With the introduction of the latest version of the InSpeed API firmware, Version 2.2, the other two modes are being made available to OEMs to help close a substantial portion of the remaining gaps in system resiliency. These

modes control how disruptions are handled and operate. All four modes are described below.

#### Normal Mode

The normal mode of operation of a port defaults to standard FC-AL behavior. This means that a normal mode port receiving a LIP from a device will allow that LIP to pass to the InSpeed switch, and if a LIP comes to a normal mode port from the internal router, it will transmit that LIP. In this mode, an attached device is not isolated when a change occurs in the system.

#### Target Mode

When an InSpeed port is configured as a target, it is typically being used for connection to a disk drive or tape drive. When a target is inserted into a target mode port, the target is allowed to complete FC-AL initialization, but is connected to only those initiators configured to receive notifications. During the initialization process, InSpeed modifies the appropriate initialization frames to reserve the addresses assigned to other devices in the system that are not participating in the initialization. The net effect is the target and all identified initiators detect all devices in the system and believe they have undergone an FC-AL initialization sequence.

#### Initiator Mode

When an InSpeed port is set to the initiator mode, it is typically being used for connection to an initiator. When an initiator is removed, no effect on the remaining system occurs unless other initiators are set by policy to be informed of such a change. When an initiator is inserted, only that port undergoes the initialization sequence. The initiator proceeds through all of the stages of initialization, and is appropriately informed of the existence of all other devices in the system through InSpeed frame modifications. This may include other initiators. The net effect on the system is the new initiator learns of all other devices but with no disruptions to existing streams of data.

Table 2: Stealth Intelligent Change Manager Port Settings

Tx LIP	Rx LIP	Mode
Yes	Yes	FC-AL port operation
Yes	No	Target
No	Yes	Initiator
No	No	Switch-to-switch communication

### Switch-to-Switch Mode

When an InSpeed switch is connected to another InSpeed switch, the ports connecting the two switches are set to the switch-to-switch communication mode. In this mode, the routers of any connected InSpeed switches can be updated to ensure that device lists are kept coherent and that any connected initiators are kept informed of changes on any switches upstream or downstream of the initiator. This coherency occurs even when an initiator is added or removed and no devices are updated with the changes.

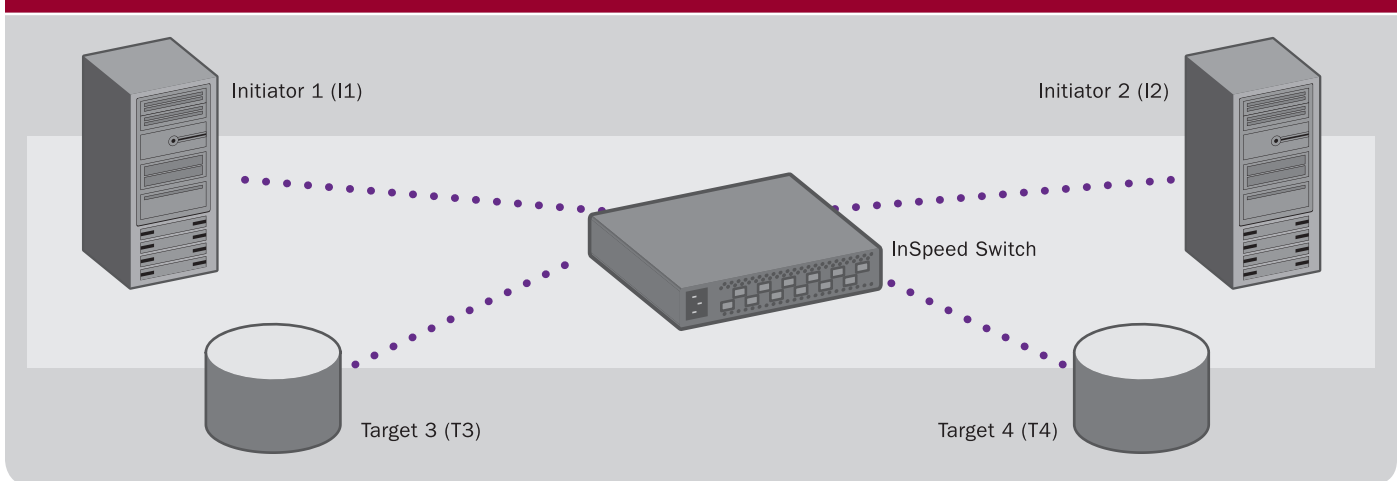
From an external view, all of the actions taken by InSpeed appear completely transparent to all devices in the system. Attached devices believe they are connected to an FC-AL compliant infrastructure. Applied to disk array systems, Stealth Intelligent Change Manager eliminates the interactions that occur between the large number of drives

in a system and ensures initiator drive lists are kept accurate. The end result is the Stealth Intelligent Change Manager provides ultra fast discovery and minimal disruptions to embedded storage systems.

### Specifics of Stealth Intelligent Change Manager

Stealth Intelligent Change Manager operates on the initialization frames of the FC-AL protocol. In particular, the LIFA (Loop Initialization Fabric Assigned) and LIPA (Loop Initialization Previous Assigned) frames are manipulated by InSpeed. To give an example of what happens during a power-up initialization cycle between four devices, the sequences involving several different change events are shown in the following tables. Figure 1 shows the configuration for the example sequences. The initiators in the figure are typically RAID controllers and the drives connected to the InSpeed switch are typically an SBOD.

Figure 1: Configuration for Examples in the Following Tables



## Power-Up Baseline Example

Table 3 shows the contents of one byte of the initialization frames, assuming two initiators (I1 and I2 – typically RAID controllers) and two targets (T3 and T4 – typically connected through an SBOD) as shown in Figure 1. The initiators during power-up typically take an address during the LISA (Loop Initialization Soft Assigned) frame, and the targets typically take their address during the LIHA (Loop Initialization Hard Assigned) frame. This sequence establishes the baseline upon which the Stealth Intelligent Change Manager operates.

**Table 3: Power-Up Sequence - Stealth Intelligent Change Manager or FC-AL**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x00	(2) 0x00	(3) 0x00	(4) 0x00
LIPA	(5) 0x00	(6) 0x00	(7) 0x00	(8) 0x00
LIHA	(9) 0x00	(10) 0x00	(11) 0x01	(12) 0x03
LISA	(13) 0x83	(14) 0xC3	(15) 0xC3	(16) 0xC3
Result	0xC3 - Baseline			

## Initiator Change Examples

When Initiator I2 is removed under FC-AL, the three remaining devices take their address during the LIPA frame, since they already have their addresses. All devices are affected, and the sequence takes a relatively large number of steps as shown in Table 4.

**Table 4: Change Sequence - FC-AL: Remove Initiator 2 (I2)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x00		(2) 0x00	(3) 0x00
LIPA	(4) 0x80		(5) 0x81	(6) 0x83
LIHA	(7) 0x83		(8) 0x83	(9) 0x83
LISA	(10) 0x83		(11) 0x83	(12) 0x83
Result	0x83 - All devices affected			

Table 5 shows that when Initiator I2 is removed when Stealth Intelligent Change Manager is enabled, there is no effect on the system at all. This behavior is based on the typical default policy of not informing other initiators of changes through the loss or addition of a different initiator.

**Table 5: Change Sequence - Stealth Intelligent Change Manager: Remove Initiator 2 (I2)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA				
LIPA				
LIHA				
LISA				
Result	No effect on the system			

When an initiator is re-inserted under FC-AL, as is shown in Table 6, previous devices take their address under the LIPA frame and the new initiator takes its address under the LISA frame. All devices are affected by the insertion.

**Table 6: Change Sequence - FC-AL: Re-Insert Initiator 2 (I2)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x00	(2) 0x00	(3) 0x00	(4) 0x00
LIPA	(5) 0x80	(6) 0x80	(7) 0x81	(8) 0x83
LIHA	(9) 0x83	(10) 0x83	(11) 0x83	(12) 0x83
LISA	(13) 0x83	(14) 0xC3	(15) 0xC3	(16) 0xC3
Result	0xC3 - All devices affected			

Under Stealth Intelligent Change Manager, a re-inserted initiator affects only that inserted initiator, as shown in Table 7. Again, if the policy is set to ensure existing initiators are informed when other initiators are added or removed, Initiator I1 would have also been included in the cycle.

**Table 7: Change Sequence - Stealth Intelligent Change Manager: Re-Insert Initiator 2 (I2)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA		(1) 0x83		
LIPA		(2) 0x83		
LIHA		(3) 0x83		
LISA		(4) 0xC3		
Result	0xC3 - No effect on other devices			

### Target Change Examples

Removing and replacing a target under FC-AL is the same as when removing or replacing an Initiator. Table 8 shows the removal cycle, where all remaining devices take their addresses during the LIPA frame.

**Table 8: Change Sequence - FC-AL: Remove Target 3 (T3)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x00	(2) 0x00		(3) 0x00
LIPA	(4) 0x80	(5) 0xC0		(6) 0xC2
LIHA	(7) 0xC2	(8) 0xC2		(9) 0xC2
LISA	(10) 0xC2	(11) 0xC2		(12) 0xC2
Result	0xC2 - All devices affected			

When a target is removed under Stealth Intelligent Change Manager, two key behaviors must occur. First, the initiators must be informed of the removal of the target, and second, any remaining target must not be disturbed. This behavior is shown in Table 9.

**Table 9: Change Sequence - Stealth Intelligent Change Manager: Remove Target 3 (T3)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x02	(2) 0x02		
LIPA	(3) 0x82	(4) 0xC2		
LIHA	(5) 0xC2	(6) 0xC2		
LISA	(7) 0xC2	(8) 0xC2		
Result	0xC2 - Only Initiators affected			

A target re-insertion under FC-AL, as shown in Table 10, affects all devices. All devices already in the system take their address during the LIPA frame, and the newly inserted target takes its address during the LIHA frame.

**Table 10: Change Sequence - FC-AL: Re-Insert Target 3 (T3)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x00	(2) 0x00	(3) 0x00	(4) 0x00
LIPA	(5) 0x80	(6) 0xC0	(7) 0xC0	(8) 0xC2
LIHA	(9) 0xC2	(10) 0xC2	(11) 0xC3	(12) 0xC3
LISA	(13) 0xC3	(14) 0xC3	(15) 0xC3	(16) 0xC3
Result	0xC3 - All devices affected			

The re-insertion of a target under Stealth Intelligent Change Manager causes that target and all initiators to be included in the initialization cycle, as is shown in Table 11. In a system with a large number of drives, this will always result in (1 Target) + (# of Initiators) being initialized, as compared to (# of Devices). In a single initiator system with 120 drives, that is the difference between 121 devices initializing under FC-AL versus 2 devices initializing under Stealth Intelligent Change Manager.

**Table 11: Change Sequence - Stealth Intelligent Change Manager: Re-Insert Target 3 (T3)**

Frame/Device	I1 (0x80)	I2 (0x40)	T3 (0x01)	T4 (0x02)
LIFA	(1) 0x02	(2) 0x02	(3) 0x02	
LIPA	(4) 0x82	(5) 0xC2	(6) 0xC2	
LIHA	(7) 0xC2	(8) 0xC2	(9) 0xC3	
LISA	(10) 0xC3	(11) 0xC3	(12) 0xC3	
Result	0xC3 - Only initiators and new target affected			

There are many other scenarios in multiple topologies where Stealth Intelligent Change Manager has been extensively tested, including multiple switch configurations where up to eight SBODs are serially connected to each other. The fourth mode of operation of Stealth Intelligent Change Manager, the Switch-to-switch mode, ensures the routers of each switch keep up to date with changes and also manage those changes within their own realm.

## Directed LIPs

Some InSpeed back-end implementations use selective reset LIPs (also called “directed LIPs”) for resetting a disk drive that has been deemed to be unresponsive. There are two types of directed LIPs. The first is sent directly from an initiator to a specific target, which is a request for that target to perform a vendor specific reset. The second type is sent from an initiator to all targets with all targets requested to undergo a vendor specific reset.

Under Stealth Intelligent Change Manager, because LIPs are carefully managed, directed LIPs are handled through separate mechanisms.

## Summary

Stealth Intelligent Change Manager technology is common today in streaming applications like tape backup and video editing and streaming. With Version 2.2 of the InSpeed API, this same technology can now be applied to embedded storage systems to introduce advanced system resiliency.

Now disruptions in the back-end where a large number of drives are connected to one or more initiators are limited to just an affected disk drive and any initiators. Limiting the number of devices involved in system change events to just two or three devices provides several benefits:

- ▶ Significant reduction in the time required to resolve system changes, since the smallest possible set of devices participate.
- ▶ Other devices not involved in the change notification process potentially continue sending traffic.
- ▶ Selective resets can be performed that now truly only affect targeted devices. Other targets are not even aware a reset has occurred.
- ▶ Defective devices can potentially be identified more quickly. Often these devices cause system changes to occur, and with Stealth Intelligent Change Manager, the number of devices involved is drastically reduced, making it much easier to find the device causing the system change.

This document refers to various companies and products by their trade names. In most, if not all cases, their respective companies claim these designations as trademarks or registered trademarks. This information is provided for reference only. Although this information is believed to be accurate and reliable at the time of publication, Emulex assumes no responsibility for errors or omissions. Emulex reserves the right to make changes or corrections without notice. This report is the property of Emulex Corporation and may not be duplicated without permission from the Company.