

## APPROVED MINUTES OF THE 13TH FAIS AD HOC MEETING

The thirteenth meeting of the T11.5 FAIS Ad Hoc group took place at the DoubleTree World Arena Hotel in Colorado Springs, CO, on July 13, 2004.

The following people were present:

ASTUTE NETWORKS	Tanjore Suresh	tsuresh@astutenetworks.com
BROCADE	Ed McClanahan	edmc@brocade.com
CISCO	Claudio DeSanti	cds@cisco.com
CISCO	Maurilio Cometto	maurilio@cisco.com
CISCO	Ronak Desai	rodesai@cisco.com
CNT	David Peterson	dap@cnt.com
DOT HILL	Elizabeth Rodriguez	Elizabeth.Rodriguez@dothill.com
EMC	John Kemeny	kemeny_john@emc.com
EMC	Fred Oliveira	oliveira_fred@emc.com
INCIPIENT	Brian Geisel	bgeisel@incipient.com
IVIVITY	Eddy Quicksall	eddyquicksall@ivivity.com
MCDATA	Gib Cardwell	mike.o'donnell@mcddata.com
MCDATA	Larry Hofer	larry.hofer@mcddata.com
QLOGIC	Craig W. Carlson	craig.carlson@qlogic.com
TROIKA NETWORKS	William Chow	wchow@troikanetworks.com
VERITAS	Roger Cummings	roger.cummings@veritas.com

A total of 16 people from 12 organizations.

### 1. Opening remarks and introductions

The Chairman of FAIS, Claudio DeSanti of Cisco, called the meeting to order and welcomed the attenders, and led a round of introductions.

### 2. Administrivia

## 2.1 Approval of minutes

The minutes of the June 8 FAIS Ad Hoc meeting had been uploaded to www.t11.org by William Chow of Troika Networks as 04-451v0.

CDS had one comment:

- Action items should keep their originally assigned numbers. William will apply this convention starting with this (13th) meeting.

The minutes were approved in the absence of any objections.

## 2.2 Approval of this agenda

Claudio had distributed a draft agenda for this meeting as 04-475v0 via the T11 Web site, and he now asked for changes to that agenda. No modifications were offered.

Claudio moved to approve the agenda. Approved by acclamation(sp?).

## 3. Call for Patents

Claudio stated that amongst rules and policies under which this Working Group operates are the T11, INCITS, and ANSI patent policies. He directed persons wishing to make statements relevant to those policies to do so at the T11.5 or T11 plenary meetings.

## 4. Review of old FAIS action items

Old action items were reviewed. Their status was updated as follows:

AI 1: Ed McClanahan to review the required operation if an i/o crosses a boundary and the second extent is "held". (Opened 7/9/2003)

**Closed by 04-471v0.**

AI 2: Tanjore Suresh to investigate methods of generating WWNs. (Opened 7/9/2003)

**Closed by 04-469v1.**

AI 3: Robert Mulcahy (ER?) to propose support for journaling. (Opened 7/9/2003)

**Replaced by AI 11.**

AI 4: Roger Cummings to provide requirements for statistics gathering in FAIS and present proposals (Opened 9/10/2003)

**Tabled until object model finalized.**

AI 5: Ed McClanahan and William Chow will work on a draft of Chapter 5. (Opened 11/9/2003)

**Open.**

AI 7: Tanjore Suresh will cross-reference names/types of capabilities against existing standards/prior work. (Opened 1/13/2004)

**Open.**

- AI 8: Roger Cummings/Fred Oliveira/Brian Geisel will each provide a list of capabilities that they must be able to query support for. (Opened 1/13/2004)  
**Open.**
- AI 9: Ed McClanahan and John Kemeny will propose an approach for supporting failover in the event of CPP failure. (Opened 1/13/2004)  
**Open.**
- AI 10: George Penokie will produce a UML conventions section for inclusion in the standard. (Opened 2/3/2004)  
**Open.**
- AI 11: Roger Cummings will provide specific requirements for DPC pass-thru support. (Opened 2/3/2004)  
**Closed by 04-116v1.**
- AI 12: Elizabeth Rodriguez will propose support for journaling. (Opened 4/6/2004)  
**Open.**
- AI 13: Mike O'Donnell will update the glossary per any overlap with the terms defined in the requirements document. (Opened 5/8/2004)  
**Open.**
- AI 15: Ed McClanahan and Maurilio Cometo will collaborate on a new proposal for the execution model. (Opened 5/8/2004)  
**Open.**
- AI 18: David Black and Silvano Gai will propose modifications to the object model to directly support the I\_T nexus. (Opened 5/8/2004)  
**Closed, overcome by events.**
- AI 19: Ed McClanahan will propose chapter 6 text for separate client and provider-specified object handles. (Opened 5/8/2004)  
**Open.**
- AI 20: Tanjore Suresh will propose chapter 6 text for iSCSI-specific port attributes. (Opened 5/8/2004)  
**Open.**
- AI 21: Ed McClanahan and Silvano Gai will unify their object models. (Opened 5/8/2004)  
**Closed, overcome by events.**
- AI 22: Ed McClanahan will propose a draft for chapter 1 and 2. (Opened 5/8/2004)  
**Open.**
- AI 23: Mike O'Donnell will propose a draft for chapter 3. (Opened 5/8/2004)  
**Open.**
- AI 24: Ed McClanahan will incorporate the operational model (04-340) to the draft specification as chapter 1. (Opened 5/8/2004)  
**Open.**
- AI 25: Fred Oliveira will update EMC's LU layout proposal per feedback from June meeting. (Opened 6/8/2004)  
**Closed by 04-401v0.**
- AI 26: Fred Oliveira will update EMC's FAIS API proposal per feedback from June meeting. (Opened 6/8/2004)  
**Closed by 04-402v0.**

## **5. Old business**

There was no old business at this meeting.

## **6. Scheduled FAIS business**

### **6.1 Proposed Object Model Modifications (04-467v1, 04-470v1) William Chow**

Maurilio Cometo suggested that you can allocate COW space up-front, so there may not be a lot of Concat entries. William responded that the issue isn't the number of Concat or Xmap entries, but that they are separate and thus requires separate extent-capable data structures in the provider.

Maurilio wants to keep the ability to fork reads from writes. William suggests using the "multi-path" object, which is a generalization of the BITLSet that can be layered over other Vdevs while avoiding the requirement for extent-capable resources. MC asked when would layouts differ b/t the read and write paths. WC responded that one example is when reads can be fastpath'ed directly to storage but writes cannot be, such as when write data needs to be journaled by the Client or an external device.

Maurilio remarked that he thought the current object model was already accepted. William responded that he didn't believe there has been any agreement on a final object model. He pointed out that there have been changes proposed to it in each of the last several meetings. Claudio suggested that to make forward progress, proposals need to take the form of text that can be included in the draft standard. William said that he would present modifications to the EMC-proposed API and Cisco-proposed chapter 6 text.

### **6.2 Xmap Boundary Crossing (04-471v0)**

**Ed McClanahan**

William asked about handling for layered Xmaps. Ed McClanahan responded that each Xmap is independently processed under the same rules.

William asked how to handle I/O atomicity in the face of partial transfers. Ed responded that he didn't know of any targets that guaranteed I/O atomicity. Eddy Quicksall said there is a mode page setting; Tanjore Suresh said he believed this was the "restricted reordering" setting.

Ronak Desai asked why not just pass handles and leave I/O state in the DPC (for I/O takeover by Client)? William added that if the Client only cares to handle SCSI-level data and response, then transport-level information is not needed. Ed agreed that transport info is not necessarily needed if this is case. Action for Ronak to write up API requirements for Client-controlled I/O (hand-off and I/O management).

William asked why force quiesce? Ed said he believes that all clients will do it anyways. William says that this lets you do multiple operations (including changing state) before waiting to drain for

each one. John says that it would be convenient to update the state on multiple regions, and then drain after updating all. Ed responded that he can make it an option to an “xmap update” call.

Elizabeth Rodriguez said she believes that proposal #2 can leave data in an difficult to predict state. She wants the capability for the client to choose whether or not to enable “check all regions”; she wants the provider/platform to support both. Ed suggested thinking about it offline.

### **6.3 Proposed FAIS Object Model (04-401v1)**

**Fred Oliviera**

Fred Oliviera said the only change was replacing LUNMap with individual FITLs; no ITs. Ed asked if Fred didn't see any operations on an IT. William suggested that logout is useful (e.g., freeing up a flow if a Client knows it doesn't need a connection). He also said this is separate from the concept of eliminating the LUNMap; e.g., useful on the backend for BITLs. Tanjore said that an IT makes it easier to do “LUN masking”; Fred responded that the FITL already supports this directly.

Ronak said that it is useful to support a default LUNMap that is used for all hosts that aren't explicitly created. He suggested that FITLs may be created using such a LUNMap, so that a Provider can offload this from the Client, reducing API calls. Ed asked if a FITL must be supported thru the API (e.g. a Client must know about all FITLs to support UA). Ronak said that there does exist FITLs, even if they are a result of the default LUNMap. Fred said that a default LUNMap would be a wildcard for the I but specific to an FT; the FT and FITL can exist before the initiator logs in. William suggested that if a default LUNMap was supported, it should be optional and preferably handled thru API calls separate from those used for Client-created FITLs.

William asked why there isn't a “BI” if there exists an FT. Fred responded that he isn't sure if there is a need for such a thing. William said that these are dynamically created like FTs. He also suggested that a BIT is even more important to the backend than a FIT is to the front, since the Provider must initiate these connections. Fred agreed that this might be useful.

### **6.4 2-Step V2P Download to DPC (not posted)**

**Ronak Desai**

Ronak reiterated the reasons for 2-step download:

- batching multiple updates into a single DPC message
- quiescing can be paired with the batch

Ronak said download/activate can be implicit and immediate in each mapping update, and no-op'ed when called specifically called. Clients would determine when quiesce operation needs to be performed by checking with the provider (thru some other mechanism). William responded that this results in different Client logic for different providers. He suggested that a “begin batch” operation would solve this problem.

Ed asked what activate really does. He said that activate wouldn't even be performed if the update was immediate.

William asked how the TRAN\_NO was assigned/allocated. Ronak said that it is allocated by the Client and *told* to the Provider. William said this is not ideal as related transactions consume resources in the Provider and also need to be linked. It would be preferable for the TRAN\_NO to be a provider-provided handle that directly referenced the provider's data structures. Ed suggested that the TRAN\_NO arg could be an in/out that could then be used by the Provider to allocate on the first update. William responded that this prevents the Provider from occasionally rejecting the creation of a transaction (e.g. if it's a limited resource), in which case the Client could then "fall back" to quiescing around immediately activated updates.

Ed said that it would be important for there to be atomicity guarantees for activate. He said that the activate call must implicitly quiesce for its duration to ensure atomicity. Ronak responded that the Client will quiesce around its activate call. This also allows quiesce to protect activate calls across multiple providers (e.g., if it might fail more likely so than the subsequent unquiesce). William said that it would be simpler for a separate, standalone quiesce/unquiesce primitive, and *not* bundle it into mapping update calls.

Ronak said that he is okay with adding a "begin transaction" operation. He wasn't sure whether they could support this as async. Ed said he still doesn't see any value with adding download/activate calls unless they piggybacked with quiesce/unquiesce. William asked how you would determine *what* to quiesce. Ed said this could be inferred from the updates associated with the transaction. Ronak responded that it would be too complex for the Provider to figure this out; he said he could accept appending full quiesce parameters to the download call.

## 6.5 EMC FAIS API v2 (04-479v0)

John Kemeny

William asked which category Client-initiated I/Os would fall into. Action for William to propose something.

Eddy suggested putting common async params into a separate "header" structure that is contained within operation-specific callback structures. He added that this makes it easier to modify the common header without editing all operation-specific structures. Ed suggested that it would be useful to add pointers for linked-list management.

William asked what "vol ID" is? John Kemeny said this is a Vdev handle. Ed asked why there is a restriction for it to only reference an Xmap (in FAIS\_VOL\_QUIESCERANGE\_S). John said he thought this was agreed to in the last meeting. Ed responded that there was no agreement for this restriction. John said we could just use this as the starting point and people can propose modifications.

John proposes using this document as a starting point; Ed so moves. William said that there hasn't been sufficient time for review; Ed noted that it was posted only on last Friday (7/9). John said this could be deferred until the next meeting.

## 6.6 FAIS Proposals Under Consideration

04-480v0

Claudio posted a list of proposals he suggested that people use as the baseline for subsequent proposals. He suggested that future proposals should include modifications to these documents to ensure further progress.

## **6.7 FAIS\_PortName**

**04-426v0**

Eddy asked whether the FAIS\_PortName is the “SCSI port name”. Ed responded yes.

Eddy said that a SCSI port name is equal to the FC WWN of the port, but this isn’t the same for iSCSI. Ed said the “name” are simply whatever uniquely identifies the I or T of an I\_T nexus. Eddy said he thought that this only needed to include SAM-only concepts, and not transport-level information. Ed responded no, that it needed to include transport-level information.

William suggested that a port object should support both initiator and target modes simultaneously. Ed acks that his current proposal does require separate instances in this case, but he said he will consider modifications to address it.

## **6.8 Method for Generating FAIS\_NAMES**

**04-469v1**

Claudio asked whether each Client *instance* needs its own OUI. Tanjore proposed that the “local” component of the name would be standardized into 3 subsections: platform, blade, and port #. Claudio said that it doesn’t seem desirable to standardize this. Ronak said that his naming convention already has existing restrictions that would not be compatible with the proposal. Claudio said that this could be suggested, but not mandated. William asked if this means he wants it to be an informative annex. Claudio responded no, that he’s not ready to suggest that. Elizabeth said that it may not be possible for a Client to support this since this requires a new OUI, which a Client may not be able to acquire since IEEE requires existing allocations must be sufficiently used up.

Roger asked whether having the Client defining the names arbitrarily will result in breaking s/w that currently deciphers WWNs. No one indicated that this would be an issue for them.

Ronak asked whether Client developers wanted to control this or let the provider define it. William responded that this isn’t completely up to the Client, since the provider may not be able to support Client-specified names. Fred responded that the only requirement is that the name be persistent and globally unique.

## **6.9 Additional DPC Functionality**

**04-116v1**

Roger said this is useful for SAN routing from one FITL to a specific BITL. Fred said this would only be a performance issue since the Client could do the pass-thru.

Ronak said he would add an attribute to the Xmap. William said that this isn’t necessary on the per-region level; e.g., he suggested that this could be added to the FITL. Action for Roger to defer the specific location until the object model is finalized.

## 7. **Unscheduled FAIS business**

There was no unscheduled business.

## 8. **Review of FAIS action items**

New action items were reviewed. The list was updated as follows:

- AI 4: Roger Cummings to provide requirements for statistics gathering in FAIS and present proposals (Opened 9/10/2003)  
**Tabled until object model finalized.**
- AI 5: Ed McClanahan and William Chow will work on a draft of Chapter 5. (Opened 11/9/2003)  
**Open.**
- AI 7: Tanjore Suresh will cross-reference names/types of capabilities against existing standards/prior work. (Opened 1/13/2004)  
**Open.**
- AI 8: Roger Cummings/Fred Oliveira/Brian Geisel will each provide a list of capabilities that they must be able to query support for. (Opened 1/13/2004)  
**Open.**
- AI 9: Ed McClanahan and John Kemeny will propose an approach for supporting failover in the event of CPP failure. (Opened 1/13/2004)  
**Open.**
- AI 10: George Penokie will produce a UML conventions section for inclusion in the standard. (Opened 2/3/2004)  
**Open.**
- AI 12: Elizabeth Rodriguez will propose support for journaling. (Opened 4/6/2004)  
**Open.**
- AI 13: Mike O'Donnell will update the glossary per any overlap with the terms defined in the requirements document. (Opened 5/8/2004)  
**Open.**
- AI 15: Ed McClanahan and Maurilio Cometo will collaborate on a new proposal for the execution model. (Opened 5/8/2004)  
**Open.**
- AI 19: Ed McClanahan will propose chapter 6 text for separate client and provider-specified object handles. (Opened 5/8/2004)  
**Open.**
- AI 20: Tanjore Suresh will propose chapter 6 text for iSCSI-specific port attributes. (Opened 5/8/2004)  
**Open.**
- AI 22: Ed McClanahan will propose a draft for chapter 1 and 2. (Opened 5/8/2004)  
**Open.**
- AI 23: Mike O'Donnell will propose a draft for chapter 3. (Opened 5/8/2004)  
**Open.**

AI 24: Ed McClanahan will incorporate the operational model (04-340) to the draft specification as chapter 1. (Opened 5/8/2004)

**Open.**

AI 27: Ronak Desai will write up API requirements for Client takeover and handling of a faulted I/O.

**New.**

AI 28: William Chow will propose modifications to the EMC- proposed API (04-402v1) for his object model modifications and Client-initiated I/Os.

**New.**

AI 29: Roger Cummings will proposed object model modifications for support pass-thru of non-read/write I/Os.

**New. Tabled until object is finalized.**

## **9. Next meeting schedule**

The next meeting of the FAIS group will be at the T11 Plenary week in Keystone, CO on tues evening 6-9 (8/3) and weds 9-12 (8/4).

## **10. Adjournment**

A motion to adjourn the meeting (by Roger, seconded by Brian Geisel) was passed unanimously.