

APPROVED MINUTES OF THE 15TH FAIS AD HOC MEETING

The fifteenth meeting of the T11.5 FAIS Ad Hoc group took place at the Crowne Plaza Hotel in Nashua, NH, on September 14, 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
EMC	Anshul Chadda	chadda_anshul@emc.com(?)
EMC	John Kemeny	kemeny_john@emc.com
EMC	Fred Oliveira	oliveira_fred@emc.com
INCIPIENT	Brian Geisel	bgeisel@incipient.com
MCDATA	Larry Hofer	larry.hofer@mcddata.com
TROIKA NETWORKS	William Chow	wchow@troikanetworks.com
VERITAS	Amitava Guha	amitava.guha@veritas.com

A total of 11 people from 8 organizations.

1. Opening remarks and introductions

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

2. Administrivia

2.1 Approval of minutes

The minutes of the August 3 FAIS Ad Hoc meeting had been uploaded to www.t11.org by William Chow of Troika Networks as 04-598v0.

The following corrections were noted:

- The minutes for the July 13 meeting were not approved, as this action was actually deferred.

- John Kemeny noted that a motion was incorrectly attributed to him in the 7/13 minutes.
- Claudio requested that the presenter be noted in the section headings.
- Claudio indicated that the next meeting time (for Sep) was incorrect.
- Claudio indicated that a new patent statement is forthcoming.

Pursuant to these changes, the minutes were approved by acclamation.

2.2 Approval of this agenda

Claudio had distributed a draft agenda for this meeting as 04-612v0 via the T11 Web site, and he now asked for changes to that agenda.

Brian Geisel moved to approve the agenda; Ed McClanahan seconded. The agenda was approved by acclamation.

3. Call for Patents

Claudio stated that this Working Group operates under the INCITS procedures, which are available on the web at <http://www.incits.org/>. Amongst these procedures are the ANSI patent policies. He directed persons to forward any patent related issue to the T11.5 and T11 Plenary meetings.

4. Review of old FAIS action items

Old action items were reviewed. Their status 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. Reassigned to Amitava Guha.
- 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. A related section from SAM-4 is expected soon.
- AI 12: Elizabeth Rodriguez will propose support for journaling. (Opened 4/6/2004)

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 and Eddy Quicksall 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: Larry Hofer 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. (Opened 7/13/2004)

Open.

AI 28: William Chow will propose modifications to the EMC- proposed API (04-402v1) for his object model modifications and Client-initiated I/Os. (Opened 7/13/2004)

Open.

AI 29: Roger Cummings will propose object model modifications to support pass-thru of non-read/write I/Os. (Opened 7/13/2004)

Open. Tabled until object is finalized.

AI 30: Larry Hofer will merge the conventions and glossary sections into a single proposal. (Opened 8/3/2004)

Open.

5. Old business

There was no old business at this meeting.

6. Scheduled FAIS business

6.1 Proposed FAIS Object Model (04-401v2)

John Kemeny

John points out that we need initial drafts from which to work with, instead of starting from scratch at each meeting. Larry Hofer would like to see a target date for functional cutoff, i.e. enter revision-only mode. Ed suggested that the primary impasse has been debate on the object model. He suggests making this the short-term focus.

John points out that this is simplified over this last (aug) object model proposal. He also noted that does not include William's proposal for Xmap modifications. EM asked whether there was any conclusion about LUNMap vs FITL. John responds that there seemed to be consensus that FITL is preferred.

John noted that the MirroredVDEV now has separate policies for read (includes weights) and write. Ed notes that it might be useful for a MirroredVDEV to support non-broadcast writes that are serialized by the Provider. Anshul Chadda asked whether this could be handled by a serialization-capable BITLSet. EM responded that this adds complexity to the BITLSet that is already present in the MirroredVDEV. Tanjore Suresh asked why this needs to be visible to FAIS. William responded that this is necessary when the backend devices do not provide the same serialization guarantees that the virtual one provides (e.g. if former has QAM=1 but virtual LU has QAM=0).

Tanjore said he would like to explicitly see whether each attribute is settable by. E.g., bitmap of "static" or "settable" by provider or client. Tanjore asked why statistics gathering is not one of the operations of an object. William suggested generalizing "return the attributes of the object" so that it can be inclusive of statistics.

Anshul asked about the need for the FT object. Ed responded that he needs this to associate requests to a specific Client that created the FAISSCSIPort.

William asked about the FITL and whether there should be a "state" (e.g. fault, allow) attribute. Anshul said that "aborting I/Os" is not a visible state. WC responded: yes, but the others (allow/fault) are.

William asked about the need for a FLU. He suggested this should not morph into a VDEV since it would otherwise require a new VDEV to be supported in the mapping tree. Tanjore asked about reservations. William suggested removing any wording about any specific SCSI commands. Ed suggested saying the FITL is for initiator-specific operations and the FLU is for operations that apply to all initiators. Tanjore asked why there isn't a LUN in the FLU. Ed responded that this is handled by the FITL.

William asked about the applicability of operations to the VDEV. John confirmed that these operations are inherited by all VDEVs. Ed suggested that this could be generalized into a super-class that applies to all objects, i.e. FLU, FITL, etc.

William asked about the value of the ParentOfChild abstract class since they are not common to ALL objects. He suggested eliminating it and putting the appropriate attributes in the specific VDEV children. It was agreed that ParentOfChild should be removed.

William asked about "StripeSizeBlks". Ed suggested renaming to ContiguousColumnBlocks.

William suggested removing VDEV from all names, or add it to Xmap. No specific resolution was agreed to, but consensus was that they will be made consistent.

Anshul asked about whether "fault_one" should be added. It was agreed this should be added. Tanjore asked who manages the I/O. William responded that a fault is still owned by the Provider. Ed suggested adding "control". Anshul suggested adding these attributes to the FITL. Ed suggested that the FITL should also support all of the same attributes. He also suggesting all attributes applied to the FLU. Ed asked about restricting quiesce to an Xmap that is on the top of the mapping tree. Maurilio confirmed that the only concern was whether quiesce could be applied to other VDEVs, not whether an Xmap is top or not.

Ed asked if VDEV children have object Ids (i.e. what are their lifetimes). William suggested an object ID is unnecessary. Maurilio said a "column #" is necessary for the striping algorithm to order the strips. William said that a column can be identified by the Client via the "column #". Ed recapped that the column has: column #, StartChildLBA, object ID of next VDEV. Ed suggested having a "mirror #" as a convenient way to identify a specific mirror (instead of using some subset of its attributes).

Ed asked whether we should rename BITLSet to BLU. He said this makes it clearer since everyone knows what a "logical unit" is, and we already have a corresponding object for the front-end (FLU). Larry noted that the association b/t BITL and BITLSet (1-*) is reversed.

William asked about the object ID for the BITL. He noted that this is not listed in the diagram. EM suggests that a BITL should have an object ID. It was agreed that every diagram object has an ID except for the children contained within the VDEVs. Ed noted that an "object ID" does not need to be added to the specialized VDEVs since they already inherit one from the VDEV abstract class.

Larry moved that we incorporate this proposal (pursuant to the just-discussed corrections) into the draft. Ed seconded. The motion was approved by acclamation.

John was actioned to post an updated document, and send it to Ed in framemaker format. The target date for completion of this action is the end of the month (Sep). Ed was actioned to post a draft chapter incorporating this updated document. The target date for completion of this action item is before the October meeting.

6.2 Proposed Xmap Modifications (04-467v2)

William Chow

William indicated that the primary issue was that extent-based mapping objects can be very expensive. He noted that one portion of his previous proposal was the addition of a "child LBA" the Xmap entry, which was just accepted in the discussion during the previous presentation on the object model. He suggested that this potentially obviates the need for a ConcatVdev.

William said read/write separation is important, but it doesn't need to be per extent (must pay cost on every extent). He would prefer one Vdev that does reads and writes. Maurilio said they have customers that require the layout mentioned on page 4 of the proposal. Ed suggested that one could use Concat Vdevs in the example on page 4 instead of 2 Mirrored Vdevs; this way, you would only need one extent in your Xmap. William responded that the cost for some providers

does not vary by the number of extents. Maurilio said that this is implementation specific, so why shouldn't we leave this up to the provider? William said that we want to give the FAIS client the right primitives, and there is some disagreement about what those primitives are. He said he doesn't see a need in any application for read/write separation on an extent basis as described in the page 4 example. Maurilio responded that the page 4 example is actually incorrect and that it is the first entry where it is desirable to have separate read/write paths; this allows a new mirror side to be updated w/o being exposed for reads.

William suggested that we could support a "single Vdev per extent" mode in the Xmap, where only one Vdev is specified (e.g. read Vdev) and it would apply to both (i.e. also to writes). Maurilio said they would accept this. Ed suggested that another useful "mode" would be to support a single Vdev that applied to all extents. The consensus was to support these 3 modes in an Xmap.

William said that there are several options in supporting read/write redirection w/o extents. One option is to support a non-extent mode in the Xmap, which now supports multiple modes already. Another is to have a Vdev, separate from an Xmap, for this; this Vdev could either be a new Vdev or it can be a generalization of the BITLSet which already supports path selection logic. Ed said the basic proposal is to allow multipathing to appear at any stage in the hierarchy. William asked whether folks were interested in have a Vdev that could point to a set of other Vdevs? He noted that one use case is to support a write-only Vdev such as an external journaling/replication target. William suggested that the multipathing Vdev could be supported with a BITLSet by adding a mode that indicated whether it may be layered over other Vdevs, or only at the bottom. Maurilio stated they would prefer to have another VDEV that is multipathing rather than modify BITLSet. John agreed. William responded that he would propose a new VDEV.

William noted that if BITLSets are always on the backend, then one can't get through the FAIS hierarchy w/o crossing a BITLSet and BITL. He asked whether the BITL needs an object ID? No one responded affirmatively.

William noted that if Vdevs were restricted to supporting only Reads/Writes, this creates a problem for Roger's use case where the device wants to do SAN routing or non-Read/Write pass-through.

Larry said it was unclear how those are handled by the current model - we have a conceptual agreement, but there is no proposal on how you handle pass-through (etc.) commands.

William asked for a vote on his proposal. Claudio said that the API and object model have changed. William said he will make a proposal for the next meeting for the concerns Claudio raised. He noted that there are several options: there could be additional attributes of a BITLSet, or there could be a completely different object that is just like a BITL Set, but different. He asked: what will people accept? Claudio responded that he'd prefer a separate object. William asked him whether Claudio could you confirm this with his technical folks, as Maurilio had already left? Claudio responded: yes.

William said it would be important for the API not to have the overlap of two objects that are virtually identical except for minor differences, such as what they were allowed to layer over. Will-

iam asked what people thought about a new mode for the BITLSet. John said that modes are preferable in terms of configuration. He said that if you don't configure it then you're back to the original mode; if you do have it then you don't have 2 different APIs. William said he would first offer a proposal for a BITLSet with multiple modes; if that is not acceptable then he will propose a totally new VDEV.

7. **Unscheduled FAIS business**

Claudio proposed target dates:

- Feb 2005: cut-off for new functionality
- June 2005: letter ballot

Larry suggested having an editorial review before the letter ballot, so that the latter would be better focused on technical issues. Claudio agreed and suggested April as the target for this.

Claudio was actioned to notify (to the reflector) the remaining schedule of the WG upon completion of the initial draft of the spec.

8. **Review of FAIS action items**

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

- AI 4: Amitava Guha to provide requirements for statistics gathering in FAIS and present proposals (Opened 9/10/2003)
Tabled until object model finalized. Target: Nov meeting.
- AI 5: Ed McClanahan and William Chow will work on a draft of Chapter 5. (Opened 11/9/2003)
Open. Target: Nov meeting.
- AI 7: Tanjore Suresh will cross-reference names/types of capabilities against existing standards/prior work. (Opened 1/13/2004)
Open. Target: the next meeting after completion of action item #8.
- AI 8: Roger Cummings/John Kemeny/Brian Geisel will each provide a list of capabilities that they must be able to query support for. (Opened 1/13/2004)
Open. Target: Nov meeting.
- AI 9: Ed McClanahan will propose an approach for supporting failover in the event of CPP failure. (Opened 1/13/2004)
Open. Target: Dec meeting.
- AI 10: Larry Hofer will produce a UML conventions section for inclusion in the standard. (Opened 2/3/2004)
Open. A related section from SAM-4 is expected soon.
- AI 12: Elizabeth Rodriguez will propose support for journaling. (Opened 4/6/2004)
Open.
- AI 20: Tanjore Suresh and Eddy Quicksall will propose chapter 6 text for iSCSI-specific port attributes. (Opened 5/8/2004)

- Open. Target: Nov meeting.**
- AI 22: Ed McClanahan will propose a draft for chapter 1 and 2. (Opened 5/8/2004)
- Open. Target: Oct meeting.**
- AI 23: Larry Hofer will propose a draft for chapter 3. (Opened 5/8/2004)
- Open. Target: Feb meeting.**
- AI 24: Ed McClanahan will incorporate the operational model (04-340) to the draft specification as chapter 1. (Opened 5/8/2004)
- Open. Target: Oct meeting.**
- AI 27: Ronak Desai will write up API requirements for Client takeover and handling of a faulted I/O. (Opened 7/13/2004)
- Open. Target: Oct meeting.**
- AI 28: William Chow will propose modifications to the EMC- proposed API (04-402v1) for his object model modications and Client-initiated I/Os. (Opened 7/13/2004)
- Open. Target: Oct meeting.**
- AI 29: Roger Cummings will propose object model modifications to support pass-thru of non-read/write I/Os. (Opened 7/13/2004)
- Open. Targeted until object is finalized.**
- AI 30: Larry Hofer will merge the conventions and glossary sections into a single proposal. (Opened 8/3/2004)
- Open. Target: Oct meeting.**
- AI 31: Claudio will notify the membership (via the reflector) of the remaining schedule of the working group upon completion of the initial draft of the spec.
- New.**
- AI 32: Ed McClanahan will incorporate the object model (04-401v3) into the draft specification.
- New.**

9. Next meeting schedule

The next meeting of the FAIS group will be at the T11 Plenary week in Austin, TX on weds 9am-1pm (9/14).

Claudio suggested not hosting an interim meeting at the T10 plenary in Nov, as it conflicts with IETF. There was general agreement. He suggested that he could host a meeting in San Jose if deemed necessary; options were: 10/26-27, 11/15-16 (last resort: 11/1-2).

10. Adjournment

A motion to adjourn the meeting was passed unanimously.