

DRAFT MINUTES OF THE 11TH FAIS AD HOC MEETING

The eleventh meeting of the T11.5 FAIS Ad Hoc group took place at the Embassy Suites in Monterey, CA, on May 3-4, 2004.

The following people were present:

AAROHI	Parag Bhide	parag@aarohi.net
ASTUTE NETWORKS	Tanjore Suresh	tsuresh@astutenetworks.com
BROCADE	Ed McClanahan	edmc@brocade.com
CERTANCE	Paul Suhler	paul.a.suhler@certance.com
CISCO	Claudio DeSanti	cds@cisco.com
CISCO	Silvano Gai	sgai@cisco.com
CISCO	Maurilio Cometto	maurilio@cisco.com
CISCO	Ronak Desai	rodesai@cisco.com
DOT HILL	Elizabeth Rodriguez	Elizabeth.Rodriguez@dothill.com
EMC	Fred Oliveira	oliveira_fred@emc.com
EMC	David Black	black_david@emc.com
HDS	Shoji Kodama	shoji.kodama@hds.com
IVIVITY	Eddy Quicksall	eddyquicksall@ivivity.com
IVIVITY	Zulfiqar Qazilbash	zqazilbash@ivivity.com
MCDATA	Michael O'Donnell	mike.o'donnell@mcddata.com
TROIKA NETWORKS	William Chow	wchow@troikanetworks.com
VERITAS	Roger Cummings	roger.cummings@veritas.com
VERITAS	Amitara Guha	amitara.guha@veritas.com (?)

A total of 18 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 April 9 FAIS Ad Hoc meeting and April 16 conference call had been uploaded to www.t11.org by William Chow of Troika Networks as 04-330v0 and 04-309v0, respectively.

Claudio DeSanti requested the following changes:

- In section 1, a reference to his company association needs to be changed to Cisco.

Pursuant to these changes, 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 04-341v0 via the T11 Web site, and he now asked for changes to that agenda. No modifications were offered.

Claudio moved to approve the agenda; Ed McClanahan seconded. No objections.

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

The review of old action items was deferred until the next FAIS Ad Hoc meeting.

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)

Open.

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

Open.

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

Open. Notify Brian Geisel.

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: Mike O'Donnell to post updated terminology/glossary document for review at the next meeting. (Opened 10/7/2003)

Closed by 04-338.

- AI 6: Mike O'Donnell to look at wording of HBA API and determine how linkage independency and thread safety/reentrancy is treated. (Opened 10/7/2003)
Closed.
- AI 7: William Chow to post revisions to the requirements document to address specifics of "error" cases - I/O faults, backside I/O errors, timeouts. (Opened 10/7/2003)
Closed.
- AI 8: Silvano Gai will work on a draft of Chapter 6. (Opened 11/9/2003)
Closed by 04-325.
- AI 9: Roger will work with Claudio on a draft of Chapter 4. (Opened 11/9/2003)
Closed by 04-340.
- AI 10: Ed McClanahan and William Chow will work on a draft of Chapter 5. (Opened 11/9/2003)
Open.
- AI 11: Roger Cummings to check on possible restrictions for upcall context. (Opened 12/9/2003)
Open.
- AI 12: Tanjore Suresh will cross-reference names/types of capabilities against existing standards/prior work. (Opened 1/13/2004)
Open.
- AI 13: Roger Cummings/Fred Oliveira/Bob Mulcahy will each provide a list of capabilities that they must be able to query support for. (Opened 1/13/2004)
Open. Notify Brian Geisel.
- AI 14: Ed McClanahan will propose an approach for supporting failover in the event of CPP failure. (Opened 1/13/2004)
Open.
- AI 15: George Penokie will produce a UML conventions section for inclusion in the standard. (Opened 2/3/2004)
Open.
- AI 16: Roger Cummings will provide specific requirements for DPC pass-thru support. (Opened 2/3/2004)
Open.
- AI 17: William Chow will post a new requirements document per the final resolution of the letter ballot comments. (Opened 2/3/2004)
Closed.
- AI 18: Elizabeth Rodriguez will propose support for journaling. (Opened 4/6/2004)
Open.

19. Old business

There was no old business at this meeting.

20. Scheduled FAIS business

20.1 Requirements for FAIS

04-037v1

Mike O'Donnell was actioned to carry terms from this doc into his newly proposed glossary section, and identify any conflicts b/t them.

BITL/BLU are simply acronym expansions.

William was actioned to post an updated version of the requirements document, with the aforementioned modifications.

20.2 FAIS Execution Model - Brocade

04-322v0

Tanjore Suresh asked why do we need a client context, i.e. isn't it implicit in the downcall?

Eddy Quicksall asked why would a completion callback go back to a different process than that which issued it? Ed responded that this is desirable but not critical.

Fred Oliviera asked what the timeout means. E.g., does it guarantee a return in that timeframe? Ed responded: no, since the timeframe is dependent on time spent in upcalls. The timeout only indicates how long the provider should block any time it waited for new events.

20.3 FAIS Execution Model - Cisco

04-335v0

Ed asked if there is an async version of every sync call. Maurilio responded: yes (at least for the ones which require a timeout).

Ed asked how an async request is identified. If not possible, how do you cancel an async request? Maurilio responded that async requests are not identified and cancellations would be handled sequentially "in the order they were submitted".

Ed asked how many upcalls were done in a FAIS_dispatch call. Maurilio responded: one.

Silvano Gai asked folks where we stand on the differences.

Elizabeth Rodriguez indicated that she thought we had decided on async for all functions except for the registration. William Chow proposed that we simply define an operation as either sync or async (i.e., the provider does not need to support both for any operation). There were no objections.

Ed proposed the notion of "global event objects". Ronak Desai suggested that their APP_ID supports this. Ed said the APP_ID does not allow different contexts/threads to support different events for the same object. Tanjore said it's simplest for the provider/API to just dispatch all upcalls to a single context. Silvano suggested taking a straw poll on: "for the objects owned by a FAIS_client, all events to the owning context, i.e. not on a per-request basis".

Claudio suggested limiting any further discussion to one hour.

Silvano asked for a straw poll for the case: “the provider is aware of multiple client contexts and demultiplexes events to the appropriate one”. Vote: y=7, n=1, a=2.

Silvano then asked what were the differences b/t the 2 proposals. He identified one issue is on where the dispatch loop is implemented. Silvano asked for a straw poll for the case: an exception should not be made to use select(). Vote: y=7, n=2, a=1.

Ed and Maurilio were actioned to collaborate on a new proposal.

20.4 Chapter 6

04-325v0

1.1.1

William asked if “xmap entry” derives from “xmap”. Silvano responded: no, that several arrows in figure 1 appeared wrong. Silvano will verify arrows/inheritance issues against previously presented diagrams.

Fred asked about layering xmaps. He was actioned to propose another model.

David Black suggests that it is “typical” to have an I_T nexus object (e.g., useful for LUN mapping/masking). William asked whether this just means adding one new object to the model for the I_T nexus. David responded: yes. David and Silvano were actioned to propose an updated model to support this addition.

Eddy suggested replacing WWN with “port identifier” or whatever SAM-3 says (i.e., to include the iSCSI ISID). William suggested using “SCSI port name”. Eddy prefers “SCSI port identifier”. William suggested that the name must be persistent. Eddy and William were actioned to suggest a replacement for “WWN”.

Ed suggested adding separate object IDs for client and provider. Ed was actioned to propose wording for adding a provider-defined object identifier.

1.1.2

Parag Bhide asked about the “V2P” term. Silvano said he would replace this with “DPC mapping tables”.

Mike asked about “path weight”. Silvano suggested that it allows some paths to be defined as passive-until-active-fails, i.e. so provider can dynamically pick any valid path with the lowest weight. Ed and David noted that such failover could induce trespass mechanisms in the physical storage. Parag proposed removing support for provider-handled failover (i.e., the provider does not need to perform the trespass). Silvano said he would remove support to implicit failover.

Ed suggested that path weighting is not necessarily a useful load-balancing mechanism. David suggested leaving this open-ended, e.g. allow “path weight” to be flexible in its meaning. Silvano said he would generalize the name.

1.1.3

Ed asked about (c): position of VDEV in parent. Silvano responded that it really should be reversed (i.e., position of child VDEV).

Ed asked whether much of this really belongs in a separate “programmer’s guide” section. Silvano responded that we can first agree on the text, and then the editor can move them to the appropriate section.

William asked about objects in figure 2. Silvano said he will rename instances using names in object model.

1.1.4

William asked about whether we can unify block ranges with xmap entries, since there is significant overlap. Silvano suggested tabling this discussion until Fred presented his “layered xmap” proposal.

Ed indicated that FAULT is new. William proposed making this be an optional state. DB suggested limiting optional features since clients are typically coded to the lowest common subset of supported features.

Ed asked why QUIESCE is an attribute but can only be modified thru quiesce-specific calls. David noted that the primary issue relates to who remembers the post-quiesce state. Ed indicated that the client will typically handle the quiesce state as a transient condition, whereas the other attributes are long-lived. William indicated that it is easier to implement a provider if QUIESCE is a simple state. David suggested that this precludes multiple, possibly unrelated, clients from quiescing/unquiescing independently.

William asked where the sense/status is defined for REJECT. Ronak responded: per-xmap. Ed asked whether anyone even uses this. Silvano said he would remove it.

1.1.5

Eddy suggests replacing “port WWN” with “target port name”, and “node WWN” with “target device name”. Roger noted that “device name” isn’t equivalent to a WWNN. William asked if we even need WWNN. Ed responds that it is needed in fabric/port logins. William noted that the WWNs can be provided by the client (e.g., via upcall to client upon a login request). Ed noted that given that approach, both WWPN and WWNN could be removed.

William asked why “FLU list”. Silvano indicated that it was an error and will be removed.

Ed indicated that we must keep the I_T_L nexus object since we need initiator-specific access permissions. Ronak suggested that this could be identified via a tuple of object IDs, instead of a new object.

[day 2]

Silvano suggested that we just create new FAIS terms for multi-transport names and addresses. Ed proposed deriving transport-specific object instances which contain transport-specific

attributes which are not necessarily shared with port objects for other transports. David suggested using a tagged union. Ed presented alternatives via pseudo-code. Eddy suggested using a “real” inheritance (i.e. using type-specific pointers).

Mike asked how one determines what transports to support. William responded that it can be either implicit (e.g. based on which DPC to associate the FAIS_SCSI_port) or explicit (type specified in the descriptor used to create it). Amitara Guha noted that a DPC may support multiple transports.

Silvano suggested that it is simpler to write the specification using generic terms. David suggested using a generic term just for the transport-specific stuff and keep it in the port object. Then the object model is not affected as other objects only reference the port objects. William asked whether there are separate objects for port and connections; Ed responded: yes, the latter has references to the former. Silvano indicated that he will add a new FAIS_Transport_Endpoint. He said he will also merge his model with Ed’s so that his objects will then have references to the FAIS_Transport_Endpoint. Silvano will move FC-specific references into a separate transport-specific section; Tanjore was actioned to provide similar text in this new section for iSCSI.

Silvano noted that we can skip the Xmap-related sections since it is subject to change pending resolution of the xmap-layering issue. Ed indicated that he would flesh out the description of each of the objects in his section, which overlaps with this doc’s section 1.1.

1.2.1

William asked why was it necessary to have a separate COMMIT operation. Maurilio responded that the client may have mapping updates to multiple providers that may fail separately, so it may be necessary to abort if SYNC didn’t work on all providers. Ed suggested that the COMMIT operation is useful if it was used to determine when to download a collection of mapping updates (instead of a download per update) and also, if it was associated with quiescing. He suggested that w/o quiesce, it isn’t very useful. Fred agreed because he would quiesce (a block range) before updating maps and “commit” by unquiescing. He said he would use this approach by doing a “range-lock” b/t the sync and commit. Ed proposed adding a range-lock to the sync operation, so that there is a single message to each DPC instead of two. William responded that the window b/t sync and quiesce is less relevant than that b/t quiesce and unquiesce. Ed noted that another benefit of sync is the ability to undo. Ed asked whether people are more concerned about the I/O quiesce timeframe or the number of operations the client needs to handle. Ed and Fred chose the latter; all others chose the former.

William asked whether the sync operation is feasible from the client’s perspective.. Fred suggests that it is easy for a client to do, since it fits into his existing algorithms; so if providers want it, then he wouldn’t have a problem using it. Silvano asked whether the sync/commit is acceptable. Ed responded that this is aggressive for FAIS v1. William noted that the requirement to support undo imposes complexity on the provider. Silvano asked for a straw poll for transaction model; no objections were noted. David suggested that we distinguish batch from undo, since the former is a provider optimization. Ed says this is what Xpath does, i.e. it doesn’t support undo. William asked what about operations that might need to be undone within a transaction is, e.g. does creating a BITL cause a login? William offers to support it if the batch operation is optional; this requires

that the client must know whether to quiesce before doing a mapping update (i.e. if a batch-create operation fails). Silvano will investigate how to support this proposal.

20.5 FAIS Managed Objects

04-012v2

Several people suggested a 1-1 relationship b/t DPC and provider (i.e. no DPC sharing b/t providers), but Ed expressed a desire for sharing b/t multiple providers (e.g. multiple instances of the same provider).

Ed asked whether or not to keep WWNs. Claudio indicated that we agreed to keep it; Ed said that is still open pending discussion on UML. Ed supported exposing (but not setting) FCID. Ed suggested that some environments want to have a user-configurable domain area ID. Maurilio suggested that that could be handled by other APIs outside of FAIS. Ed indicated that there are “millions of hosts which expect a fixed FCID for a given WWPN”, so a FAIS_client running in a switch will want to set this.

Claudio asked what the ProviderWWN is. Ed responds that it is the GUID, e.g. for static binding of the client. However, he noted that there needs to be a way for the DPC to uniquely identify multiple provider instances. William suggests that could be internal and proprietary to the provider.

Claudio asked what the ChassisWWN is for. Ed responded that it provides info on how various DPCs are related. However, he noted that he hasn't yet heard a strong argument for why a DPC needs to be exposed to a client. Maurilio responded that it allows a client to specifically use a DPC “close to” a particular FAIS_scsi_port.

Mike asked whether we decided on settability of FCID. Ed asked whether the configurability is optional, i.e. provider can handle it by default. Claudio suggested that this is a platform-specific “setup phase”. Fred responded that an application needs to handle it, but it's not clear which API to use for this.

Claudio suggested using FC-FS terms, i.e. “node name” and “host name” instead of “ChassisWWN” and “FC WWPN”. Ed responded that these terms do not support iSCSI. William asked what “chassis” means, i.e. does it only establish a physical relationship b/t managed objects.

Claudio suggested renaming the objects to match the terms agreed upon for the requirements doc. Ed indicated that he would add new terms: FAIS_InitiatorPort and FAIS_TargetPort.

Eddy asked to remove “FC”, e.g. FC_Port_Name. Ed responded that he's provided the FC-specific cases and someone needs to provide iSCSI-specific versions.

Ed asked again about configurability of WWNs. Silvano responded that he supports this. He suggested taking a straw poll; no objections were noted. Ed now asks about configurability of FCID. Mike suggested that the API could allow a provider to reject a request to configure the FCID; however, he expressed that he is philosophically opposed to it. Silvano asked if there are any objections; none were noted.

Ed asked about MapOfInitToLUN and relationship to David's proposal for an I_T nexus object. Silvano asked what one can't do with an I_T_L object that you can with one for I_T. David responded that there is no technical difference; the difference is only that it is "typical" for existing s/w to operate on an I_T. William asked whether a new object is really necessary, or whether we could simply provider operations around an <I,T> tuple. David responded that if there are operations around an I_T, then it is natural to have an object for it.

Maurilio asked about the need for a "fabric ID" for a FAIS_SCSI_Port, since a DPC can be in multiple fabrics. William noted that this needs to be optional, i.e. provider can reject client attempts to set it. Ed indicated that he would make it requestable/configurable.

Silvano asked why the FAIS_PlatformPort is even needed. William responded that it is necessary for N-ports. However, he noted that it can be an attribute of the FAIS_SCSI_Port objects. Silvano responded that this would be preferable so that the FAIS_SCSI_Ports can be directly associated with the Remote_SCSI_Ports.

Ed said he would remove LU mapping references, as they are handled in Silvano's model. They were actioned to collaborate on cross-referencing or unifying them.

Action for Tanjore to provide iSCSI-specific attributes for FAIS_SCSI_Ports.

20.6 FAIS Operational Model

04-340v0

William asked about whether the scope of FAIS is covered. Roger responded that it should, but the content can be moved to other sections as necessary.

1.1

Silvano asked about the term "information routing". David suggested replacing it with "command/data forwarding".

Claudio moved to incorporate the proposal as chapter 1. Ed seconded. No objections.

An action was noted for Ed to produce drafts for chapters 1 and 2. An action was noted for Mike to provide a draft for chapter 3.

20.7 FAIS Glossary

04-338v0

Ed said "block virtualization" does not cover non-block functions, e.g. control path pass-thru. William asked whether we can remove it since we have "LBA remapping". Mike offered to remove it; no objections.

William asked whether definition of "concatenation" is too broad. Mike offered to strike 2nd sentence.

Control Path:

William noted that this should be CPP (not CP). Mike indicated that he needed to sync definition with requirements document.

Fabric-based storage virtualization:

This was not used anywhere, so Mike will remove.

Fast mirror resync:

It is not clear if needed, but will save it for later.

FastPath:

William suggested not adding this as a new term. Mike will merge this definition into that of the DPC.

20.8 Rationale for Hierarchical Xmaps

???

Fred said this allows services to operate at any desired layer, e.g. copy service operate on the logical volume under a striped VDEV. Otherwise, the copy service would need to use the strip-specific Xmap entries, instead of the logical subdevice. Also, a single Xmap requires the client to collapse its logical layers into a single Xmap for the relevant operation (e.g. COW) but then restore the Xmap based on a similar collapse.

Silvano asked how one handles faults at a lower layer Xmap? Fred responded that the fault goes to the owner of the lower layer Xmap. Maurilio asked why not use FITL-BITL connections? He prefers this because it allows him to work only with FC frames, i.e. the fault is generated based on the ITL in the command frame header.

Fred was actioned to post his current proposal and address the above concerns in an updated proposal.

21. Unscheduled FAIS business

Review of chapter assignments:

Action for Fred to propose an initial draft for chapter 8 (function descriptions). Goal is post document 1-2 weeks prior to June meeting.

22. Review of 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)

Open.

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

- Open.**
- AI 3: Robert Mulcahy to propose support for journaling. (Opened 7/9/2003)
- Open.**
- 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 6: Roger Cummings to check on possible restrictions for upcall context. (Opened 12/9/2003)
- 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/Bob Mulcahy 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 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)
- 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.
- New.**
- AI 14: William Chow will post an updated requirements document per feedback from May meeting.
- New.**
- AI 15: Ed McClanahan and Maurilio Cometo will collaborate on a new proposal for the execution model.
- New.**
- AI 16: Silvano Gai will update the chapter 6 draft per feedback from May meeting.
- New.**
- AI 17: Fred Oliviera will submit his proposal for xmap layering.
- New.**
- AI 18: David Black and Silvano Gai will propose modifications to the object model to directly support the I_T nexus.
- New.**
- AI 19: Ed McClanahan will propose chapter 6 text for separate client and provider-specified object handles.
- New.**

AI 20: Tanjore Suresh will propose chapter 6 text for iSCSI-specific port attributes.

New.

AI 21: Ed McClanahan and Silvano Gai will unify their object models.

New.

AI 22: Ed McClanahan will propose a draft for chapter 1 and 2.

New.

AI 23: Mike O'Donnell will propose a draft for chapter 3.

New.

AI 24: Ed McClanahan will incorporate the operational model (04-340) into the draft specification as chapter 1.

New.

25. Next meeting schedule

The next meeting of the FAIS group will be at the T11 Plenary week in Chicago, Il on June 8-9.

26. Adjournment

A motion to adjourn the meeting was passed unanimously.