

## 2/8/05 ICSC ITWG Meeting Minutes

Taking minutes: HP (FW)

Y HP Jim Hamrick (JH) (joined at 10:40 PST)  
Y HP Jay Rosser (JR)  
Y HP Fred Worley (FW)  
Y IBM Fredy Neeser (FN)  
N NetApp Arkady Kanevksy (AK)  
N Sun Matt Pearson (MP)

cascading ascii art attendance diagram. (if you have more than 1 minus visible, you are not eligible to vote.)

	hp	ibm	netapp	sun
	----	-----	-----	----
m-3	+	+	-	-
m-2	+	+	-	-
m-1	+	+	-	-
m-0	+	+	-	-

### Next Meeting:

Tuesday 2/15/05, 10am-12pm PST

### **ACTION summary:**

#### **PENDING AIs:**

1. AI (FN): Get rid of MM-16.4.D2.1 and MM-16.4.D2.2
2. AI (FN): Document that an unlink operation that fails remotely will not directly manifest as an error locally. Also add text to Implementers guide that they need to absorb the Async Error and not manifest it via the IT API.
3. AI (JR) Generate a proposal for a mapping service to generate a PBL (with bus addresses) given a virtual address range and a virtual address space ID(virtual addresses)
4. AI (JR): Lead further discussion to complete: Produce requirements for RDDP WG informational draft to support IOH functionality.
5. AI (JR): Add macro for rdma\_read\_incoming / rdma\_read\_outgoing field names to support IRD / ORD as alternate (additional) names; determine where this changes should be documented (includes msg\_events page)
6. AI (JR/FN): Determine if keeping around socket descriptors for lifetime of converted connection is too resource-consuming
7. AI (FN): Evaluate "detailed status code" idea.

#### **NEW AIs:**

8. AI (ToDo list/JR) Reformat Reference page to use separate columns for reference tag and reference
9. AI (JR/FW, others?) Add this text to the man pages for all blocking calls
10. AI (FN) will keep to-do list for edits that still need to be done; noted edits should be provided to FN for tracking

11. AI (JR) Implement resolution for GN55=HP15 proposed by FN
12. AI (JH) Implement resolution to GN23=SUN11
13. AI (All) Review portion of implementers guide corresponding to owned man pages to see if transport specifics need to be called out
14. AI (ToDo list) Implement terminology changes described in mail from FN on 2/8/05  
Subject: Re: Terminology
15. AI (Add to ToDo list) Add text to Implementers Guide to specify that on IB one must modify the EP after creation (with `it_ep_rc_create`) to ensure that the endpoint attributes match those requested in endpoint creation
16. AI (ToDo list) Document the completion or async errors that can occur associated with RDMA enables being disabled
17. AI (ToDo list) Document how to use the endpoint attributes in connection with RDMA Reads
18. AI (FN) Review `it_rmr_link` to ensure existence of text discussing ramifications of stale STags.
19. AI (ToDo list) Remove proposed fix and reference to `INVALID_SGID/SLID` from latest draft of v2.0 document
20. AI (ToDo list) Update DTO flags

## Minutes:

- **Additional agenda items**
  - Discussion of GN13 and GN62
- **Approval of previous minutes**
  - No minutes to approve (no mtg 2 weeks ago)
- **Action item review**
  - AI (FN): Get rid of MM-16.4.D2.1 and MM-16.4.D2.2  
§ Still pending
  - AI (FN): Document that an unlink operation that fails remotely will not directly manifest as an error locally. Also add text to Implementers guide that they need to absorb the Async Error and not manifest it via the IT API.  
§ Still pending
  - AI (JR) Generate a proposal for a mapping service to generate a PBL (with bus addresses) given a virtual address range and a virtual address space ID(virtual addresses)  
§ Still pending
  - AI (JR): Lead further discussion to complete: Produce requirements for RDDP WG informational draft to support IOH functionality.  
§ Still pending
  - AI (JR): Add macro for `rdma_read_incoming` / `rdma_read_outgoing` field names to support IRD / ORD as alternate (additional) names; determine where this changes should be documented (includes `msg_events` page)  
§ Global edit in man pages to replace the name complete  
§ Add of macro still pending (until header file creation complete)
  - AI (JR): Figure out how to reassemble man pages

- § Merge is complete
- § Most change bars have disappeared in merged document
- § Discussed re-merging the document to retain change bars for review
- § JR distributed a document created by comparing the merged document with the original 1.0 document
  1. This mechanism re-created change bars
  2. Seems very usable; sufficient to re-create tracking of changes
  3. Discussed distribution of this document as informational
- § Merged document does not preserve separate change colors for errata vs new functionality
  1. Discussed that this may not be critical if Functional Changes document is published
  2. MM man pages have comments for errata changes that are preserved by the merge
- § AGREED that separate colors for errata fixes and new functionality are not required
- AI (FN): Generate standard text for blocking man pages
  - § Complete (sent to reflector by FN on 2/7/04, Subject: AI: Boiler-plate text for API calls that may block)
  - § AI (JR/FW, others?) Add this text to the man pages for all blocking calls
  - § AI (FN) will keep to-do list for edits that still need to be done; noted edits should be provided to FN for tracking
    1. Send either placeholder (e.g. “man page x should be changed”) or suggested wording (e.g. “change ‘Logical’ to ‘Bitwise’ on line 13 of page y”) to FN, CC: JR
- AI (JR): Create global references page with list of agreed tags
  - § Complete
  - § Reference page (page xiii in merged doc sent 2/7/04 by JR, Subject: Second compilation attempt (PDF)) followed Open Group template
  - § FN suggests 2 column format
    1. AI (ToDo list/JR) Reformat Reference page to use separate columns for reference tag and reference
- AI (All): Expunge RETURN VALUE section from data structure pages
  - § Complete
- AI (All): Place resolved Errata in Functional Changes doc
  - § Complete
- AI (JR): Add list of IT-API 1.0 Errata as Appendix to specification
  - § Complete

[JH joins 10:40am PST]

- AI (JR/FN): Determine if keeping around socket descriptors for lifetime of converted connection is too resource-consuming
  - § Still pending

- § Discussed that since this is an optional feature, the resource consumption issue may be less important
  - § Could specify, e.g., a flag to the socket\_convert call to keep the socket alive – would allow resources to be dedicated only for those sockets that need to be kept alive
  - § JR/FN will continue research
- AI (FN): Evaluate "detailed status code" idea.
  - § **Still pending**
- **Editorial issues**
  - Email from Jim Hamrick, subject "Resolution for GN54=HP14", sent 1/31/05 5:24PM PT
    - § No discussion
  - Email from Jim Hamrick, subject "Resolution for GN55=HP15", sent 1/31/05 5:24PM PT
    - § No discussion
  - Email thread started by Jay Rosser, subject "Proposal to resolve Erratum GN62=HP19", sent 1/31/05 12:59PM PT
    - § JR accepts FN's revised language as more accurate
    - § **AI (JR) Implement resolution for GN55=HP15 proposed by FN**
  - Email thread started by Jim Hamrick, subject "Propose we reject GN23=SUN11", sent 1/31/05 2:52PM PT
    - § Add to To Do list for it\_convnet\_netaddr
    - § **AI (JH) Implement resolution to GN23=SUN11**
  - Compiled specification drafts sent via private email due to size
    - § Implementer's Guides: should there be one global one or one for each transport?
      1. In current IG, if text does not specify that it is particular to a given transport, is it applicable to all transports?
        - § Should be the case
        - § Do we need to confirm this?
      2. E.g. it\_addr\_handle\_create talks about IB-specifics
        - § May be inconvenient to have implementers guide that has things mixed up between transports
      3. **AI (All) Review portion of implementers guide corresponding to owned man pages to see if transport specifics need to be called out**
        - § Discussed purpose of implementers guide; does it need to be updated for completeness?
          - § **Agreed that implementers guide need not have extensive additional work to improve completeness; adequate to update transport specifics**
- **Misc issues**
  - Email thread started by Fredy Neeser, subject "Scheduling of IT-API meetings", sent 2/1/05 12:43PM PT
    - § No disagreement, no discussion

- **S-RQ issues**

- Email thread started by Fredy Neeser, subject "S-RQ and multiple Protection Zones", sent 2/1/05 1:00PM PT
  - § Resolved by later thread, "it\_srq\_create: Different PZ for S-RQ and EP", sent 2/7/05 1:23AM PT; see below
- Email from Fredy Neeser, subject "Fw: Re: S-RQ implementation options", sent 2/2/05 5:02AM PT
  - § Pending
- Email thread started by Fredy Neeser, subject "it\_srq\_create: Different PZ for S-RQ and EP", sent 2/7/05 1:23AM PT
  - § Question of what to call send queues and receive queues in the spec
  - § JR preference is to use the terms Send Queue and Receive Queue as they match the terms in the spec; propose to call the set of such things Request Queues
    1. Why not "Work Queues" for the set?
      - § Request Queue terminology already occurs in the v1.0 spec (usage includes send and receive queues)
      - § FN: The term Work Queue is used more often in the spec than Request Queue - would create less churn to standardize on the term Work Queue
      - § Verbs identify Work Queue as either Send or Receive Queue; however, Shared Receive Queue is also a Work Queue
      - § FN has proposed wording to define Work Queue that includes all 3 queue types
      - § Resizable Work Queue refers to either Send Queue or Receive Queue but not Shared Receive Queue
        - § FN suggests "ep\_work\_queues\_resizeable" to refer to such queues
    2. In v1.0, there was discussion about not using terms that were defined in the IB spec
      - § Concern raised that IBTA may object to use of terms defined in IB spec
      - § Expect now that there is very little risk that IBTA would object to use of terminology common with the IB spec
      - § Note that this terminology is common to the IB and IETF iWARP terminology
  - § **RESOLVED:** Adopt terminology as described in mail from FN on 2/8/05 Subject: Re: Terminology, specifically:
    1. Eliminate use of the term Request Queue
    2. Use the term Work Queue to refer to either Send, Receive or Shared Receive queues
    3. Rename the IA attribute "resizable\_work\_queue" to "ep\_work\_queue\_resizeable"

§ AI (ToDo list) Implement terminology changes described in mail from FN on 2/8/05 Subject: Re: Terminology

• **MM issues**

- Email thread started by Fredy Neeser, subject "Man page check-out and files to be merged", sent 2/1/05 9:36AM PT
  - § (become Terminology discussion - Send/Receive Queues, etc)
  - § See above
- Email thread started by Jim Hamrick, subject "Review of it\_post\_recv", sent 2/1/05 3:00PM PT
  - § Text incorporated by FN in latest draft
- Email thread started by Jay Rosser, subject "missing flag in it\_ia\_info\_t", sent 2/2/05 10:26AM PT
  - § First point: Creation of QP
    1. Can modify attributes later on but can not specify at create time
    2. What should the default be
      - § Suggest both on
      - § Would need to document defaults
    3. Does this need to be fixed in it\_ep\_rc\_create?
      - § If attributes are already specified, then this is a non-issue
      - § Does spec mandate that they be specified?
      - § Should Implementers Guide specify that on IB one must modify the EP after creation to ensure that the endpoint attributes are as specified in the ep\_attr structure?
        - § Sounds correct
        - § AI (Add to ToDo list) Add text to Implementers Guide to specify that on IB one must modify the EP after creation (with it\_ep\_rc\_create) to ensure that the endpoint attributes match those requested in endpoint creation
  - § **Second point: Additional checks to verify access**
    1. AI (ToDo list) Document the completion or async errors that can occur associated with RDMA enables being disabled
  - § **Third point: If you use an endpoint to do an RDMA Read, you have to enable the endpoint for RDMA Write as well**
    1. Boolean already says this, extended desc. In post\_rdma\_read also mentions this – is this sufficient?
    2. Yes, however endpoint itself also has to have remote write enabled
    3. AI (ToDo list) Document how to use the endpoint attributes in connection with RDMA Reads
- Email thread started by Fredy Neeser, subject "it\_post\_rdma\_read\_to\_rmr", sent 2/3/05 7:43AM PT
  - § Conclusion reached in mail on reflector

- § If you have to do an RDMA Read followed by an invalidate operation and wanted not to have to worry about ordering issues w.r.t. that, you have to fence the invalidate operation and thus stall the send queue
  - 1. JR agrees
- § This type of fencing seems useful; without would need to use a read fence which would block your send queue
- § Flag will be introduced in v2.1
- § AI (JR/ToDo list) Incorporate the man page for `it_post_rdma_read_to_rmr` into the latest v2.0 draft
- Email from Fredy Neeser, subject "Resolution of remaining MM Errata", sent 2/4/05 3:19PM PT
  - § 3 errata in mail
  - § GN21=SUN9
    - 1. JR resolved in 2.0 spec in section 3.9 ("Output Parameters")
      - § Specify global in nature, but passing bad pointers into these routines could have severe consequences to the consumer
      - § In 1.0, no mention of side effects of bad parameters
      - § Calling out in 2.0 that there is no attempt by Implementation to detect bad pointers/parameters provided by Consumer
      - § Rationale: This is a programming error that should be avoided
      - § If there were an error code would have to check all routines for invalid pointer
      - § [Agreement that this resolution is acceptable](#)
  - § GN46=SUN12: `it_lmr_free` (and `it_lmr_modify`) - what happens to the RMR context?
    - 1. Proposed fix does not take into account that stale STags could exist
    - 2. See proposed text for `it_LMR Modify`
      - § New context can be retrieved with `it_lmr_query`
      - § Problem is with revocation of remote access privileges
      - § It is not possible to revoke
      - § `lmr_free` man page says nothing about LMR context
      - § For consumers that are trying to be careful this information could change their programming model
      - § Would be helpful to look again at state STag issue to see if this needs to be addressed globally
      - § LMR could end up using the same (a stale) STag
      - § [AI \(FN\) Review `it\_rmr\_link` to ensure existence of text discussing ramifications of stale STags.](#)
- § GN31=OG11
  - 1. Segments language is confusing
  - 2. Consumer does get back ownership of the array, but not the local buffer identified by the array

- 3. Agreement that proposed resolution is correct and should be adopted (already in the spec in current v2.0 draft)
        - Email thread started by Fredy Neeser, subject "Ballot on Detailed MM Requirements for RDMA Read to RMR", sent 2/4/05 6:34PM PT
          - § Adopted by email vote
- **CM issues**
  - Some outstanding unresolved Errata discussions on reflector remain
- **Additional agenda items**
  - GN13 discussion
    - § Proposal to deprecate INVALID\_SGID and INVALID\_SLID to have same value as INVALID\_SOURCE\_PATH
    - § Rationale is that INVALID\_SGID/SLID is misleading – there could be other things wrong than the SGID/SLID; INVALID\_SOURCE\_PATH is more generic
    - § Side effect is that code written to distinguish between those errors would not be run any more
    - § FN suggests we simply eliminate INVALID\_SGID and INVALID\_SLID and force a compilation error
    - § **AI (ToDo list) Remove proposed fix and reference to INVALID\_SGID/SLID from latest draft of v2.0 document**
  - DTO flag discussion
    - § Discussion of lack of extended description for iWARP
    - § **AI (ToDo list) Update DTO flags**
- **Next steps**
  - Focus on IT-API 2.0 specification generation, next round of detailed requirements to be prioritized.
- **Any other business**

Meeting adjourned, 12:27pm PST