

1/18/05 ICSC ITWG Meeting Minutes

Taking minutes: HP (FW)

Y HP Jim Hamrick (JH)
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 **1/31/05** (no meeting on 1/25/05)

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.

NEW AIs:

5. AI (JH): Update create_sr man page to include more detailed discussion of protection zone issues
6. AI (FN): Document stronger ordering rules for RMDA DTOs related to overlapping buffers
7. 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)
8. 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)
9. AI (FN): Generate functional change list for MM calls as an example for review

Minutes:

0) Additional agenda items

- None

1) Approval of previous minutes

- Minutes for 14. Dec. 2004, email "ICSC ITWG draft minutes for 12/14/04" sent on 15. Dec. 2004 by Jay Rosser.
- Minutes for 21. Dec. 2004, email "ICSC ITWG draft minutes for 12/21/04" sent on 22. Dec. 2004 by Jay Rosser.

Minutes Approved

2) Action item review

- AI (JR): Drop MPA marker control for IT-API consumers from man pages.
 - **Completed**
- 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): Change cm_msg page to reflect the IRD/ORD usage model (see also Minutes of 21. Dec. 2004).
 - Updated man pages (in latest release) to document change to IRD/ORD usage
 - ITAPI 1.0 conn est. event stated IRD/ORD fields were not valid
 - Changed this so that on active side in 2 way con est. IRD/ORD would represent the passive IRD/ORD
 - In passive side, they will have values in them that represent a snapshot of the initial offering from the active side
 - Consumer should note that these values may change
 - Issue is captured in latest set of man pages
 - **Completed**
- 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)
 - o Why a mapping/unmapping service?
 - o See first discussion in Minutes for 23. Nov. 2004
 - **Still pending**
- AI (JR): Produce requirements for RDDP WG informational draft to support IOH functionality.
 - Updated IOH def in man pages in latest (4th) draft, which could serve as basis for RDDP draft
 - 16 byte field
 - Suggest **further discussion** of proposal in draft
- AI (JR): Implement non-peer rejects for graceful closes during MPA Startup (RDMAC vs. Non-Permissive IETF RNIC). See email thread starting on 12/15/2004,

subject "Graceful closes during MPA Startup (RDMAC vs. Non-Permissive IETF RNIC)"

- [Completed \(in 4th draft, 13 Jan 2005\)](#)
- Reject reason code for IETF non-permissive device not being able to interoperate with RDMAC device
- See email thread on graceful closes
- In some cases, local RNIC can actually detect that it is incapable of interacting with the remote device; can detect that remote has DDP version 0 and it has DDP version 1 and can not downgrade
- Creates a new type of failure, for which a new error reason code should be created
- Latest draft does not support this new error reason code
- Two different completion errors that can occur:
 - o Connection Broken
 - o Non-peer reject event
- Suggesting new reason code to cover case above
 - o Is Bad Conn. Params a sufficient reason code?
 - o Name is OK, but description can be improved
- FN will provide feedback to JR off line
- h. AI (JR): To send OpenGroup style guide to FN
- [Completed \(sent by JR 21Dec2004\)](#)
- Additional discussion: Proposal for Completion Error section
 - o Helps with readability
 - o Discussed as example `it_post_rdma_write` man page in draft 2.03 of MM man pages sent by FN ON 12 Jan 2005, Subject: v2.03 of MM pages with full Narrow RMR support
 - o May not be necessary for every call
 - o Particularly useful for link and many of the new post calls as well
 - o Also to RDMA DTOs
 - o Would clarify the sync/async nature of a call
 - o Concerns raised about the amount of work necessary to do for all calls, inconsistency of doing for some but not all calls
 - o Alternative: move this type of information to the end of the description section without creating a specific title/section for it
 - o Completion errors:
 - § Class of remote completion errors that can surface locally
 - § This type of error can be difficult to describe
 - § Also comments on completion errors that are common to multiple calls
 - EG “completion status other than `it_dt_success` will break the connection” is found on DTO man page
 - Should this be moved to the global behavior section?
 - § Note that v2 has a whole range of completion errors that did not exist in v1
 - o Discussed replicating the completion error information; create a global section, but do not remove information from the individual man pages
 - o Discussion of overall formatting

- § Changes in formatting done to make man pages more organized, cleaner than v1 documents
 - § However, the level of change (moving sections of the page around, rewording, etc) makes the change-bar version of the document impractical – too many changes for a consumer familiar with v1 to quickly determine what has changed in substance from the v1 call
 - § Agreed that if this new format is used, the change bar version of the spec for such pages should NOT be provided
- Places where changes have NOT already been made:
 - § CM man pages
 - § UD address resolution (SIDR)
- New section proposed for all man pages would be for completion errors only, NOT also for discussion of sync/async return codes
 - § There is a section of open group documents for Async errors
 - § Could use that heading for this section, but could give consumers the incorrect impression that this section was only for async issues and not also for completion errors
 - § Async errors are discussed on it_post_rdma_write man page
 - § Errors can be expressed as a completion error or async error
 - § If affiliated async errors ARE to be discussed in the same section, then “Completion Errors” may be the wrong title for the section
- Can “Async Errors” also include “Completion Errors”:
 - § Both appear on EVDs
 - § Both are asynchronous
 - § One class of these errors are called completion errors in IB
 - § They are all “Event types”
 - § No specific objections to using the term “Asynchronous errors” for both error types
- **AGREED:**
 - § Use the title “Asynchronous Errors” as the heading for both async and completion errors
 - § Add a line to the global section to emphasize that completion errors are included in async errors
- IF MM man page format were adopted universally, then:
 - § Add “applicability” section to all man pages
 - § Style change in description for all RMDA work requests
 - Information on access priv. on source / destination buffers that was not there before
 - § Add Async Errors section to all man pages
 - § Makes change bar version of the document unusable for those pages where there is a lot of churn
 - Number of pages with a lot of churn may be quite large (50+%)
- What’s different section
 - § Ex: Section for it_post_rdma_write
 - What was completion errors in v1 can be async errors in v2

- Etc
- § Could create document by doing this on a call-by-call basis
- § Text does not need to be verbose; those seeking details can read new pages
- § More human-intensive than the change bar method
- § **ACTION (FN): Generate functional change list for MM calls as an example for review**
- § Can review with that example whether:
 - A) summary is sufficient
 - B) with summary, change bars are also required
 - C) whether effort to add summary would be overwhelming

3) CM issues

Discussion of fourth draft of CM man pages

- Sent by JR, 13Jan2005, Subject: Fourth draft of CM man pages
- Discussion
 - o Names
 - § rdma_read_incoming / rdma_read_outgoing
 - Names are somewhat cumbersome
 - Would be nice if spec had IRD and ORD instead
 - Should we implement a macro to support the names IRD/ORD, and use the new IRD/ORD in the man pages
 - **ACTION (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)**

Delivery date discussion:

- Have slipped delivery dates several times
- Should determine when we expect to release v2 spec
- There are very few remaining function questions
- Editorial questions remain
 - o Expect resolution on editorial questions quickly
- Implementation guide may require significant additional work
- Most amount of work remaining is to add the new sections proposed in today's meeting, generate the list of functional changes (per call) and compile the document

Goal: [Complete v2.0 by 11 Feb 2005](#)

4) MM issues

- LMR/RMR addressing modes and terminology
 - o Issue of term "virtual address" for kernel consumers
 - o Kernel addresses may have a different type of address (e.g. logical address for Linux kernel)
 - o Terminology from RDMA verbs may be confusing
 - o Suggest usage of "absolute" vs "relative" addressing
 - o Works for virtual addressing, logical address space for Linux kernel, etc
 - o Currently, just saying "address" in most places

- If you search for “virtual”, there are very few hits; some care apparently taken not to use the term
- Issue is mostly for MM man pages
- Also issue for discussion of RMR / LMR in man pages
 - § Need to clearly document how addressing modes can be combined for RMRs and LMRs
 - § Combinations are described in a table in the verbs
 - § Can not document that RMRs can not be bound to 0 based LMRs, since 0 based addressing for LMRs does not exist
 - Can completely document without discussing 0-based LMRs
 - Can do by adding discussion of addressing to LMRs page
 - § AGREED to make naming consistent as proposed by FN
- AGREED to use the terms “absolute” and “relative” for previous “virtual” and “zero-based” terminology
- ACTION (FN): Update man pages to use the terms “absolute” and “relative” in place “virtual” and “zero-based” terminology.
- Ordering rule for RDMA Write preceding a Send
 - Discussion on reflector concluded that there is ordering for the completions that occur locally
 - Ex: If RDMA writes and sends are posted in a row, they will complete locally in order; however, they may not complete remotely in order
 - If RMDA write is posed before a send, then when Send completes remotely through matching receive, then peer knows that all data from the previous RMDA write have been placed; ordering for completions of these operations is guaranteed
 - There is NOT necessarily ordering of the placement of data itself
 - If the RMDA write and the Send happen to target the same remote buffer (or overlapping remote buffers), then the content of that overlapping buffer would not be defined (because placement can be out of order).
 - FN feels current text is misleading (by omission) for RMDA writes preceding a Send
 - § Text is OK for RDMA writes in sequence that target overlapping buffers (content of overlapping buffers is undefined) – this was not documented before
 - § Suggest that additional text to clarify the lack of ordering between RDMA Writes and Sends (for data)
 - Need to highlight this change as a function change between v1 and v2
 - § Discussion that in IB, this may not be an issue (recollection is that IB is strongly ordered w.r.t. data for RMDA write and Send)
 - § Consumers should be made aware of this change in behavior of v2
 - § ACTION (FN): Document stronger ordering rules for RMDA DTOs related to overlapping buffers
- Zero-based addressing support for LMRs: Add to v2.0 for consistency with RMRs, or defer to v2.1?
 - See discussion above
 - AGREED to add 0-based addressing for LMRs to v2.0 for consistency

5) S-RQ issues

- Different Protection Zones for S-RQ and Endpoints?
- two sorts of protection zones of concern:
 - o Local resources
 - o Remote resources
- Discussion on reflector that it would be nice to have two different protection zones – one for local, one for remote operations
- What we have instead is local protections associated with SRQ, not the endpoint; remote protections associated with the endpoint
- Suggestion that we revisit the use model for different protection zones in create_sr man page; section there may not reflect today's discussion
- **ACTION (JH): Update create_sr man page to include more detailed discussion of protection zone issues**

6) Any other business

- FN will be unable to attend next meeting
 - o Discussion of changes to document format will be via email
 - o No meeting next week
 - o Next meeting 31 Jan 2005
- Definition of capitalized terms in document
 - o Some inconsistency in definition of capitalized terms
 - o Suggest that all defined proper nouns be capitalized consistency
 - o **AGREED:**
 - § Either define a term (capitalized nouns) in the definitions man page or, if no definition exists, then the term should not be capitalize
 - § Verbs/conjugated versions of terms not be capitalized
 - o Either define a term (capitalized nouns) in the definitions man page or, if no definition exists, then the term should not be capitalized

7) Next steps

- Focus on man page generation
- Occupy spare time in telecons with errata review, next round of detailed requirements to be prioritized.

Meeting adjourned, 12:03pm PST