

ICSC ITWG Meeting Minutes 3/29/2005

3/29/05 ICSC ITWG Meeting Minutes

Taking minutes: HP (FW)

Y HP Jim Hamrick (JH) (joined at 10:34 PST)
Y HP Jay Rosser (JR)
Y HP Fred Worley (FW)
Y IBM Fredy Neeser (FN)
N NetApp
N Sun

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 4/5/05, 10am-12pm PDT

ACTION summary:

PENDING AIs:

1. AI (JR): Do some experimentation to fix how it_post_rdma_read_to_rmr appears in the "legal blackline" comparison document.
2. AI (JR): Request status update from Martin with expected delivery date of the document back to the WG
3. AI (JH): Request more justification from the requestor of 0 as a valid watermark; specifically how this feature would be used
4. AI (FN): Update MM Detailed Requirements document to reflect v2.0 (e.g. Absolute vs. Relative Addressing)

NEW AIs:

5. AI (FN): Draft improved text for Application Usage section of it_socket_convert regarding use of sd after Socket Conversion.

Agenda bashing, approve minutes:

- [No minutes to approve](#): No minutes for 15. March 2005 (Meeting was cancelled)

Action item review:

- AI (JR): do some experimentation to fix how it_post_rdma_read_to_rmr appears in the "legal blackline" comparison document.

ICSC ITWG Meeting Minutes
3/29/2005

o Pending

- AI (JR): Propose errata on 2.0 that having successfully completed socket conversion calling close on the SD has no effect on the RDMA connection
 - o See also related email thread "Erratum on 2.0: "close()" shall have no effect on a converted socket"
 - o Change made stating "calling close has no effect on the endpoint"
 - § This position is likely not satisfactory
 - § Needs to be resolved
 - o Review of email sent by FN on 3/29/95, Subject "TCP keepalive option; calling close during/after socket conversion"
 - § SDP implementation uses `it_socket_convert` on the socket used by the SDP implementation for setting up the RDMA connection (`s2`)
 - § Some RNICs may not allow operations on a socket after conversion
 - § Could mandate that socket options are only allowed prior to socket conversion
 - However, could create conflict with SDP – consumer could change socket at any time in this model
 - § RNICPI has not addressed this issue to date
 - State that `llp_stream_handle` is modifyable
 - Does not say that `llp_stream_handle` may change
 - Some vendors have stated that socket will not be operable after the transition to RTS state
 - § If the LLP handle that is an output of the `modifyQP` RNICPI API call allows some means to change keepalive (which is implementation dependent) then an implementation could abstract this
 - Concern raised in RNICPI WG that after the RTS, socket may not be eligible to have keepalive turned on for it
 - In some implementations, `fd` may be gone after conversion
 - Note that this implies an OS implementation, which is outside the scope of the RNICPI
 - When you convert, options that were available through the LLP may now only be available through the QP interface
 - Summary: Some vendors do not guarantee that the consumer can perform operations on the LLP handle that change the behavior of the underlying stack
 - § Keepalive in SDP implementation?
 - IB SDP specifies that the SDP implementation should emulate keepalive
 - Implementation of keepalive is optional in iWARP SDP spec
 - o Section 14.3 on keepalive messages
 - § Keepalive is optional
 - § If implemented, must do so by turning on the TCP keepalive timer

ICSC ITWG Meeting Minutes 3/29/2005

- An SDP implementation sitting on top of RNICPI (directly) would have a difficult time providing keep alive support
- Unless the ITAPI provides support directly for keepalive, SDP implementations that want to support this feature would need to use a mixed API model
- Alternatives:
 - Could provide endpoint attributes to modify LLP keep alive attribute
 - Could agree not to support keep alive
 - What if different methods are used?
 - If one side of an SDP implementation did keepalive using TCP keepalive capability and the other side used 0 length writes (as in IB SDP spec) they would probably still interoperate
 - This would violate the text of the spec
- Further discussion:
 - A keepalive solution that is strongly RNIC dependent would be a bad solution
 - ULPs will need to support a range of RNICs
 - Note that keepalive may be a vital mechanism for a given application
 - Will there be a portable implementation underneath the ITAPI that will support keepalive?
 - If ITAPI implementations need to depend on the presence or non-presence of this feature then the implementation will be difficult
 - Vendor feedback that some implementations will not support keepalive directly
 - Note that the text in `it_socket_convert` leaves room for an implementation flexibility to support keepalive (or not) in a number of ways
 - Does the spec terminology cover the cases that we are considering, or do we need to make modifications for this issue before closure of the 2.0 spec?
 - § Could add text stating that programming model may not be portable between RNIC vendors
 - § Note that there is already text to this effect (support is optional)
 - § Man page gives an implementation the leeway to implement support for keepalive via the Berkeley Sockets API but does not mandate it
 - **AGREED that current wording is sufficient**

[JH joins]

ICSC ITWG Meeting Minutes 3/29/2005

- Noting, however, that this is not a portable technique; potential issue for consumers
 - But without support in RNICPI there is little alternative
 - An endpoint attribute would still need an IA attribute to validate the endpoint attribute
- Note also that the SDP implementation may not provide keepalive support
- Resolved for time being due to the current lack of support in RNIC-PI; however, FN requests improved guidance for Consumers in the Application Usage section of it_socket_convert (see below).
 - AI (FN): Draft improved text for Application Usage section of it_socket_convert regarding use of sd after Socket Conversion.
- AI (FN): Investigate how RNICPI will address the keep alive option and whether they will add a transport dependent mechanism to deal with this or simply expose through LLP handle (the latter being implementation dependent)
 - Completed
- AI (JR): Request status update from Martin with expected delivery date of the document back to the WG
 - Pending
- AI (FN): Update MM Detailed Requirements document to reflect v2.0 (e.g. Absolute vs. Relative Addressing)
 - Pending
- AI (FN): Generate formal proposal for alternate name for PBLs
 - Completed
- AI (JR): Add text in it_evd_events: "Indeterminate behavior is constrained to the evd concerned; overflow on a CQ associated with a particular EVD must not effect any other EVD or endpoint associated with other EVDs"
 - Completed
- AI (FN): Add mapping service requirements to memory management detailed requirements
 - Completed
- AI (FN): Change the terminology and associated type names in the memory management requirements from "Physical Buffer List" to "I/O Buffer List"
 - Completed

ICSC ITWG Meeting Minutes 3/29/2005

- AI (JH): Request more justification from the requestor of 0 as a valid watermark; specifically how this feature would be used
 - Pending
- AI (FN): Make a best effort attempt to document all resolved defects in Appendix D
 - Completed by JR
- **Fixing errata against v2.0 draft**
 - See latest spreadsheet: itapi_errata_sorted_resolutions_032805-2.xls
 - Any errata against v2.0 draft that should be fixed before publication?
- **Publication Process for IT-API v2.0**
 - ICSC Status Update
 - Other TODOs?
- **Work towards IT-API v2.1**
 - Callback discussion for IT-API 2.1. See email thread started by Jim Hamrick on 03/15/2005
 - What is allowed to be done from within a callback function?
 - New MM Requirements:
 - Mapping service for IT-API v2.1. See JR's revised proposal from 03/11/2005
 - Support for lists of Work Requests (WR seems to be a preferred term now over DTO, as it includes MM operations)
 - IB Atomics
 - UD Multicast support (e.g. for MPI implementation on top of IT-API)
- **Next steps**
- **Any other business**
 - § Documentation of close behavior for it_socket_convert
 - Paragraph in it_socket_convert man page that talks about keepalive
 - Only other legal option is calling close
 - Missing from the spec is description of what happens on call to close
 - JR added statement that calling close has no effect on the endpoint
 - What is the purpose of calling close?
 - § Free up OS resource, e.g.
 - § Request to add statement supporting reasons for calling close
 - § Proposed text: “A consumer wishing to indicate that a socket descriptor is no longer needed may use close()”

ICSC ITWG Meeting Minutes
3/29/2005

- § Alternative “or they pass it to close() to close the fd if no further operations on the SD are needed”
- What happens if calling close during or after socket conversion
 - § Current text says can not call prior to conversion (will fail conversion), after conversion can only do close or set_sock_opt
- § List of fixed errata
 - Resolved by JR
 - Exception: GN44 and GN45 not noted as resolved
 - § No functional changes section for the header file
 - Both are errata in the header file
 - § 45 was problem with original perl scripts used to generate header files

Meeting adjourned, 11:12am PST