

# ITWG Meeting Minutes – 09/07/2004

## List of those present at the meeting

Company	Name
Broadcomm	<b>Uri E. (UE)</b> (Was on the phone for a few minutes before dropping off.)
HP	<b>Jim Hamrick (JH)</b>
	<b>Jay Rosser (JR)</b>
	<b>Fred Worley (FW)</b>
IBM	<b>Fredy Neeser (FN)</b>
Network Appliance	Arkady Kanevsky (AK)
Sun	<b>Matthew Pearson (MP)</b>

(People whose names are in **bold** letters were present)

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

```
      hp      ibm      netapp      sun      broadcomm
-----
m-3      +      +      -      +      - <- enforcement starts
m-2      +      +      -      +      -
m-1      +      +      -      +      -
m-0      +      +      -      +      +
```

## Next Meeting

Tuesday 9/14/04

Action items appear in **red** embedded in the document below.

## Discussion

### Matt leaving

MP: Working being done to try to find someone from Sun to replace me, but that someone hasn't been identified so far. Can't guarantee my ability to participate in the future, but will try to at least lurk on the reflector and send questions. Won't be able to attend weekly meetings after next week's meeting.

## Minutes

JR: Minutes to approve from August 24. Any objection? None noted. Minutes approved.

## Guidance on IT-API CM requirements to the RNIC-PI Working Group

FN: Want to add agenda item to convey IT-API CM requirements to RNIC-PI working group. Think we already have pretty good understanding of what we need, and they are looking for our input.

JR: Would feel better about offering input to RNIC-PI working group if we had voted in our detailed requirements in this area.

FN: Don't think we have to wait until we've voted all the CM detailed requirements in. Can create a draft document explaining what we need from them. As soon as we have voted the requirements, we can then immediately send the document to them. Want to avoid a 2-week lag in generating a document for the RNIC-PI working group.

JR: Sounds like a good idea.

FW: How far are we from being able to vote on detailed requirements?

JR: Outstanding issue holding up the vote. Issue is how to deal with iWARP requirement that RDMA responder has to wait for message from RDMA initiator before moving to connected state.

FN: Another issue is how to deal with ULP where MPA REQ message goes in the same direction as the last streaming mode message. Will be messy if we have to support this.

JR: Jim, do you have any further input?

JH: Worried that since the IT-API CEP protocol already resembles this sort of "unidirectional" ULP that we may encounter others as well. Not comfortable architecting the CM interfaces so that this sort of ULP can't be supported by the IT-API.

JR: If we made the Consumer send the MPA REQ/REP frames explicitly, then they could implement such a ULP. Wouldn't be an easy interface to program to, though. Jim, would that address your concern?

JH: Yes. It doesn't have to be easy to code such a ULP on top of the IT-API, it just has to be possible. Would like to see details of what you are proposing, though. **AI: Jay Rosser to publish details of proposal for having Consumer send MPA REQ/REP frames explicitly to reflector.**

## Supporting an RTR state for iWARP

JR: Email from Caitlin regarding notion of using some mechanism other than recommended mechanism in iWARP verbs to trigger transition to connected state.

FN: Verbs have no adequate mechanism to communicate to Consumer that DDP message has been received. Currently only way to detect this is to have a Receive DTO complete. Verbs in particular have no mechanism to detect a zero-length RDMA write. On passive side, if we modify the QP to RTS, QP goes to RTS even though no DDP message has been received and therefore no message can be sent from RTS state, which is counterintuitive. Think trying to synthesize RTR state would be a good thing.

JR: There was much discussion in the RDMAC of the state diagram that was necessary and sufficient to establish a connection. Result was a simpler state diagram for iWARP CM than exists for IB CM. Concerned about derailing existing iWARP vendor HW development efforts if we try to change this.

FN: Should note that there are some conflicting specifications. Difficulty with current verbs is that we are imposing a ULP constraint on passive side that could be avoided if we had an RTR state.

FW: See any solution to this problem, Fredy?

FN: If we had an optional RTR state that RNICs could support, then it would be easier for IT-API to detect arrival of first DDP message. Could have an async event to tell the implementation that the first DDP message had arrived.

FW: How would that impact the Consumer?

FN: Wouldn't be surfaced; would be internal to API implementation.

JR: How to do a zero-length RDMA write? Caitlin assumes well-known STag?

FN: Yes, but I'm not sure it can be implemented.

JR: There will be implementations that can't support RTR, so IT-API use model will have to accommodate those implementations.

FN: Don't see why this needs to be exposed.

JR: Needs to be exposed because for some RNICs Consumer has to post a Send to complete connection.

FN: Currently RNIC developers can't send out anything even though they are in the RTS state because of constraint on not having received a message. Confusing to have an RTS state where you cannot/should not send.

JR: iWARP architects I've polled don't seem to think this is a real problem.

JR: Caitlin was trying to construct a scenario that would show that an RDMAC verbs implementation wouldn't behave properly without an RTR state. I think that it will behave properly, by tearing down the connection.

FW: Any disagreement that we have to support what iWARP hardware vendors are developing today? (None noted.)

FN: Goal would be that you don't have to post a Receive message on the passive side to avoid having your connection torn down. Could wait instead until you get into RTS state.

FW: Is this new feature a detailed requirement, or a performance enhancement?

FN: Think it is a detailed requirement, because actions expected from the ULP depend upon whether or not this feature is present.

FW: What interface would the API have to tell the Consumer whether or not they are required to have a Receive posted in order to complete their connection?

FN: An IA attribute would be required that the Consumer could query.

JR: Also need a new async event, and a well-known STag to complete the mechanism.

JH: But the event and the well-known STag aren't Consumer-visible, right?

JR: Right.

FW: From perspective of API, if we want to allow for possibility of RTR state on RNICs in the future, one IA attribute doesn't seem like such a heavy burden.

FN: Jay, is well-known STag even possible?

JR: Don't know.

FN: Don't understand how a well-known STag would be created on different RNICs.

JH, FW: One bit to resolve this issue seems like a reasonable way to make forward progress.

FN: Sounds good to me. **AI: Jay Rosser will add new IA attribute to detailed requirements describing bit that tells Consumer whether or not they have to post a Receive operation before they can successfully complete a connection.**

JR: Can add other parts of the mechanism to Implementor's Guide, don't have to put these other parts of the potential mechanism into the detailed requirements. Do need to describe these new mechanisms to the RNIC-PI working group, though.

Blocking/non-blocking behavior of the it\_ep\_connect call.

FN: Another issue is that statement that it\_ep\_connect is non-blocking doesn't appear on the it\_ep\_connect man page. Would an implementation of it\_ep\_connect that blocked an application work?

JH: Potentially not. If the Consumer codes with an expectation that the call to it\_ep\_connect doesn't block pending remote interaction when in fact it does, Consumer code can break.

JR: Global behavior section covers this.

FN: When can we be blocked in an API call, and what does blocked mean?

JH: Blocked means waiting for interaction to complete with a remote peer. Nearly all calls in the IT-API are non-blocking by this definition.

FN: **AI to Fredy to create additional text for Global Behaviors section to clarify what an asynchronous call is in the IT-API.**

Broken high level requirement CE 7.2.1

JR: How to deal with requirement CE 7.2.1? Voted that we wouldn't allow defeating the parsing of MPA REQ/REP frames by an IETF RNIC. Later discovered that we'd prefer to have liberty to do this in the IT-API.

FN: Maybe we could add some text that this feature is for interworking with legacy devices, and that it isn't recommended to turn off MPA REQ/REP parsing when not necessary for interoperability.

JR: Propose we vote now to overturn CE 7.2.1. **Vote proceeds: Sun, IBM and HP vote to remove high level requirement CE 7.2.1.**

Marker control

FN: Why don't we request a marker-control attribute from RNIC-PI? If attribute is true, can set both transmit and receive markers for the QP.

JR: That would be consistent with our existing set of detailed requirements. If you have an RNIC that allows marker use to be controlled, both directions should be controlled.

FN: 7.0.0.3 seems to be redundant. 7.0.0.1 is the requirement. Do you agree, Jay?

FW: It's redundant, but from an implementation perspective it might be clearer to the Consumers to get the information in this manner.

FN: Why are we requesting a new feature from the RNIC-PI when we know that feature must be there?

FW: Fredy, would it be sufficient to strike 7.0.0.3?

FN: No. Should change requirements so that we have a single marker control flag, not separate flags for transmit markers and receive markers.

FW: 7.0.0.2 and 7.0.0.4 for instance should be combined?

FN: Yes.

JR: Can reduce to a 7.0.0.2 that doesn't talk about direction. **AI: FN to send JR a note summarizing what needs to be changed here.**

JR: Anything else in CM detailed requirements needing to be changed? Suggest that anybody who cares look over document again...

#### Memory management requirements

JR: **AI: Matt to send email to the reflector describing the modifications/additions to the IT-API that will be required if the existing MM detailed requirements are ratified.**

MP: Three existing calls want to have extra parameters added to them: `it_lmr_create`, `it_rmr_create`, `it_rmr_bind`. **AI: To everybody, see if there is any problem with using polymorphic functions to implement the new functionality, or if new function names are going to be required instead.**

JR: Matt's going to be disappearing after next week's meeting. How can we get CM detailed requirements voted upon before he leaves and we are unable to vote anything further?

FN: We should try to get the existing issues dealt with by tomorrow.

JR: Sounds good to me. If we can do this, are folks comfortable voting on it by next meeting? (No objections noted.)

Meeting ends at 12:02pm PST.