

To: T10 Technical Committee
From: Cris Simpson, Intel Corp. *cris.simpson@intel.com*
Date: 16 January 2002
Subject: Proposed SRP Model Clause

Revision History

Revision 0: First Revision

Related Documents

SRPr10

InfiniBand™ Architecture Specification, Volume 1, Release 1.0

3.1 Definitions

3.1.1 initiator interconnect port: An interconnect layer-resident object that connects the SCSI initiator port to the interconnect through which IUs and data transfer requests and responses are routed.

3.1.2 interconnect nexus: An association between an initiator interconnect port (see 3.1.1) and a target interconnect port (see 3.1.4) such that messages and data transmitted by one interconnect port will be delivered to the other interconnect port.

3.1.3 interconnect port: An interconnect layer-resident object that connects the SCSI port to the interconnect through which IUs and data transfer requests and responses are routed. Interconnect port is synonymous with either a initiator interconnect port (see 3.1.1) or a target interconnect port (see 3.1.4).

3.1.4 target interconnect port: An interconnect layer-resident object that connects the SCSI target port to the interconnect through which IUs and data transfer requests and responses are routed.

4 RDMA Service Model

4.1 Layering

As shown in figure 1, a SCSI protocol, in this case, the SCSI RDMA Protocol, accepts protocol service requests from the SCSI Application Layer above. To fulfill these requests, the SCSI RDMA Protocol instance makes interconnect service requests to the interconnect layer through the interconnect service interface.

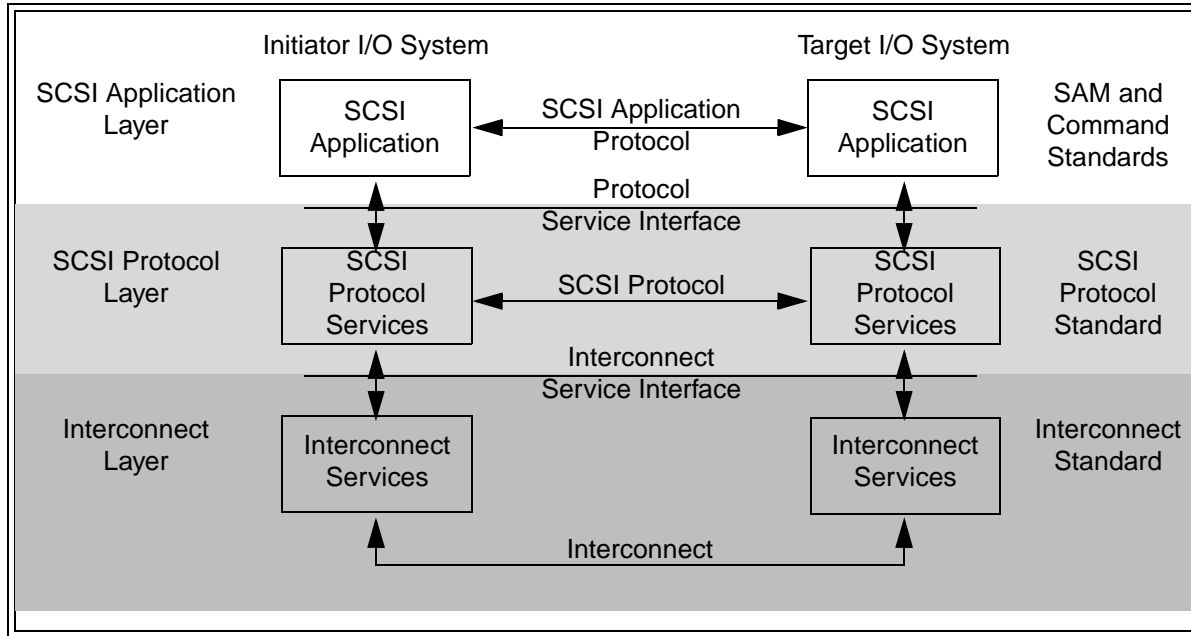


Figure 1 - Protocol service reference model

4.2 Interconnect characteristics

4.2.1 Reliability

The interconnect layer shall deliver each message once, complete and without error, in the order transmitted over an interconnect nexus, and shall deliver an acknowledgement to the requestor. Otherwise, the interconnect layer shall deliver an error notification to the requestor.

The interconnect layer shall deliver the data transferred by an RDMA Read or RDMA Write without error, and shall deliver an acknowledgement to the requestor. Otherwise, the interconnect layer shall deliver an error notification to the requestor.

The interconnect layer may utilize various techniques, including retries before returning an error notification.

If error recovery is not possible, the interconnect layer shall destroy the interconnect nexus.

4.2.2 Ordering

There is no requirement for ordering on different interconnect port nexi. All ordering references in this subclause are with respect to a single interconnect nexus.

Interconnect ports shall process RDMA Write and Send requests in the order received. Interconnect ports may process RDMA Write and Send requests before RDMA Read requests, but shall not process an RDMA Read request before previously received RDMA requests.

4.2.3 Memory Access

The interconnect layer shall perform no RDMA Read or RDMA Write operation for which the memory descriptor is invalid.

5 Interconnect Services

5.1 Introduction

This clause defines the services that **shall** be provided by the interconnect service interface of an interconnect supporting the SCSI RDMA Protocol. Some services are only meaningful and only required for initiators or for targets, and are thus indicated. If there is such an indication, it is only required as indicated.

Interconnect protocol services support the SCSI protocol services defined in SAM-2. As an example, to fulfill the SCSI protocol request **Send Command Complete**, an SRP target issues the **Send Message** interconnect service request to send the information unit constructed from the **Send Command Complete** parameters.

The message and data transfer services are based on the 1:1 nature of an interconnect nexus, so that by specifying the local interconnect port in a message or data interconnect service request, the remote port is implicitly specified.

Acknowledged

EDITOR'S NOTE 1 - The services are described here as synchronous (i.e. blocking). One of the key features of the interconnect class for which SRP is intended is asynchronous completions. Since this is an abstract services model, not an API, is the synchronous model sufficiently general, or should the services be expressed in terms of separate Post_Request() and Completion pairs?

EDITOR'S NOTE 2 - This list does not include service responses, which would be needed if we stay with the synchronous model. It appears that we don't need anything finer than 'GOOD' and 'INTERCONNECT FAILURE', although we may want 'MEMORY ACCESS FAILURE', not that there would be anything useful to be done about it.

5.2 Message and Data Transfer Services

5.2.1 Send Message(IN (Local Interconnect Port, IU))

Requests that the interconnect layer send (deliver?) the specified information unit to the remote interconnect port associated with the local interconnect port.

Input Arguments:

Local Interconnect Port: The local interconnect port of an interconnect nexus.

IU: Information Unit

5.2.2 Receive Message(IN (Local Interconnect Port, Data-In Buffer), OUT(IU Length))

Requests that the interconnect layer accept an information unit delivered to the local interconnect port and place it in the specified buffer.

Input Arguments:

Local Interconnect Port: The local interconnect port of an interconnect nexus.

Data-In Buffer: A buffer into which the received information unit is to be placed. The buffer length **shall** be no less than the MAXIMUM INITIATOR TO TARGET IU LENGTH or MAXIMUM TARGET TO INITIATOR IU LENGTH, as appropriate.

Output Argument:

IU Length: The length, in bytes, of the received information unit.

5.2.3 RDMA Write(IN (Local Interconnect Port, Buffer, Memory Descriptor))

Requests that the interconnect layer transfer the data in the local buffer to the remote memory location described by the memory descriptor.

Input Arguments:

Local Interconnect Port: The local interconnect port of an interconnect nexus.

Buffer: The local buffer from which the data is to be taken

Memory Descriptor: An SRP memory descriptor as defined in 5.4.1 .

5.2.4 RDMA Read(IN (Local Interconnect Port, Buffer, Memory Descriptor))

Requests that the interconnect layer transfer the data from the remote memory location described by the memory descriptor into the local buffer.

Input Arguments:

Local Interconnect Port: The local interconnect port of an interconnect nexus.

Buffer: The local buffer into which the data is to be placed.

Memory Descriptor: An SRP memory descriptor as defined in 5.4.1 .

5.3 Interconnect Nexus Management Services

5.3.1 Connect(IN(Local Interconnect Port, Remote Interconnect Port, Connection Data), OUT([Acceptance Data | Rejection Data]))

Requests that the interconnect layer establish an interconnect nexus, based on the connection data, associating the local interconnect port with the remote interconnect port. (Initiator only)

Input Arguments:

Local Interconnect Port: The local interconnect port of the interconnect nexus to be established.

Remote Interconnect Port: The remote interconnect port of the interconnect nexus to be established.

Connection Data: Data needed to establish the interconnection port nexus and I_T nexus. **Shall** include an SRP_LOGIN_REQ information unit, and may include interconnect-specific information.

Output arguments:

Acceptance Data: Data returned by the target as a result of establishing the interconnect and I_T nexi. **Shall** include an SRP_LOGIN_RSP information unit, and may include interconnect-specific data.

Rejection Data: Data returned by the target as a result of rejecting the I_T nexus establishment request. **Shall** include an SRP_LOGIN_REJ information unit, and may include interconnect-specific data.

EDITOR'S NOTE 3 - The presumption is that If the interconnect nexus establishment request would not succeed, the target need not (**shall** not?) consider the SRP_LOGIN_REQ. In this case, the CONNECT would receive a service response failure.

5.3.2 Listen(IN (Local Interconnect Port), OUT(Remote Interconnect Port, Connection Data))

Requests that the interconnect layer accept a request to establish an interconnect nexus with the specified local interconnect port, returning the connection data present in the connection establishment request. (Target only)

Input Arguments:

Local Interconnect Port: The local interconnect port of the interconnect nexus to be established.

Output Arguments:

Remote Interconnect Port: The remote interconnect port of the interconnect nexus to be established.

Connection Data: Data needed to establish the interconnection port nexus and I_T nexus. **Shall** include an SRP_LOGIN_REQ information unit, and may include interconnect-specific information.

5.3.3 Accept(IN(Local Interconnect Port, Remote Interconnect Port, Acceptance Data))

Requests that that the interconnect layer establish the interconnect nexus defined by the local and remote interconnect port identifiers, and return the acceptance data to the remote interconnect port. (Target only)

Input Arguments:

Local Interconnect Port: The local interconnect port of the interconnect nexus to be established.

Remote Interconnect Port: The remote interconnect port of the interconnect nexus to be established.

Acceptance Data: Data returned by the target as a result of establishing the interconnect and I_T nexi. **Shall** include an SRP_LOGIN_RSP information unit, and may include interconnect-specific data.

5.3.4 Reject(IN(Remote Interconnect Port, Rejection Data))

Requests that the interconnect layer not establish an interconnect nexus with the remote interconnect port, and deliver the rejection data to the remote interconnect port. (Target only).

Input Arguments:

Remote Interconnect Port: The remote interconnect port of the interconnect nexus to be established.

Rejection Data: Data returned by the target as a result of rejecting the I_T nexus establishment request. **Shall** include an SRP_LOGIN_REJ information unit, and may include interconnect-specific data.

5.3.5 Disconnect(IN (Local Interconnect Port))

Requests that the interconnect layer destroy the interconnect nexus associated with the local interconnect port, thereby disassociating the local and remote interconnect ports. Note that an interconnect error may cause the nexus to be destroyed without notice. In such a case, the failure may not be detected until the next request is attempted over the nexus.

Annex B: SRP for the InfiniBand Architecture

B.5 SCSI architecture mapping

Interconnect port identifiers consist of a channel adapter identifier (which may be implicit) and a queue pair number.