To: T10 Technical Committee
From: Rob Elliott, Compaq Computer Corporation (Robert.Elliott@compaq.com)
Date: 3 December 2000
Subject: Mode pages for SRP

**Revision History**
Revision 0: first revision 21 November 2000
Revision 1: reflects input from 29 November 2000 CAP meeting

**Related Documents**
SRP revision 1 – SCSI over RDMA protocol
FCP-2 revision 4b – SCSI over Fibre Channel protocol
SPI-4 revision 1 – SCSI over Parallel Interface protocol
T11/FC-FS revision 0.2 – Fibre Channel Framing and Signaling

**Overview**
Two problems may arise when bridging FCP initiators to SRP targets, requiring SRP targets to implement portions of the Disconnect-reconnect mode page.

**Ordering of RDMAs**
Fibre Channel devices are allowed to place requirements on the order of a data stream with two bits in the N_Port Common Service Parameter byte 1:

Continuously Increasing Relative Offset (bit 31)
Random Relative Offset (bit 30).

Continuously Increasing (or no Relative Offset supported) means the frames within a single sequence must be sent in order; Random means the frames within a sequence may be set in any order. This relative offset is at the FC-2 layer.

Fibre Channel FCP devices support a conceptually similar but different concept. The Enable Modify Data Pointers bit in the Disconnect-reconnect mode page indicates whether the target allows sequences within an exchange to use random relative offsets in the application client buffer or whether it requires them to be continuously increasing. This relative offset is at the FC-4 layer. For a 4 KB write command, for example, an FCP target requesting data in 1 KB blocks can only request data in this order when EMDP is disabled:

0, 1K, 2K, 3K

When EMDP is enabled, any order is allowed:

1K, 0, 3K, 2K
0, 2K, 3K, 1K
0, 1K, 2K, 3K
etc.

SRP revision 1 allows the target to issue RDMAs in any order. On InfiniBand (a possible transport for SRP), data within an RDMA may be transferred in any order (e.g. an HCA is not required to deliver data to memory in order), but each RDMA is sent in order relative to other RDMAs. Within an RDMA corresponds to FC-2 ordering and between RDMAs corresponds to the FC-4 ordering described above.

If SRP is bridged to FCP, and an SRP target is communicating to an FCP initiator that doesn't support random FC-4 layer relative offsets, the bridge will have to buffer potentially the entire exchange. In a pure FCP environment, the initiator can clear EMDP in the mode page and not have to worry about the order.

One solution is requiring SRP devices to always use continuously increasing addresses. This may limit performance. The recommended solution is to require SRP targets to support the EMDP bit.

In the other direction, SRP initiators will always accept random addressing, so will never need to turn off EMDP in an FCP target.

A bridge can avoid problems with the FC-2 reordering by limiting the size of RDMAs it issues to be under the size of its temporary buffers.

**Maximum RDMA burst size**

Fibre Channel FCP initiators can limit the burst size used by targets with another field in the Disconnect-reconnect mode page. If a bridge from an FCP initiator to an SRP target is capable of bursts larger than an initiator supports, the bridge won't know to break up read data being returned to the initiator. The target can fix this problem by supporting the Maximum Burst Length field.

In the other direction, SRP initiators that have length restrictions can set the Maximum Burst Length in the FCP target. They can also use a SG list (see 00-410r0) to break up transfers. Bridges should not have length restrictions that matter – they can break up the transfers into separate SRP RDMAs or FC sequences as needed.

**Suggested changes**

Add the disconnect-reconnect mode page to SRP with appropriate fields. Text is taken from FCP-2 revision 4b and SPI-4 revision 1. Placeholders are also added for the protocol-specific mode pages, although no fields are currently defined for them.

## 10 SCSI mode parameters

### 10.1 SCSI mode parameter overview and codes

This subclause describes the block descriptors and the pages used with MODE SELECT and MODE SENSE commands that influence, control and report the behavior of the SCSI over RDMA interface. All mode parameters not defined in this standard shall influence the behavior of the SCSI devices as specified in the appropriate command set document. The mode pages are addressed to the device server of a logical unit. The mode pages associated with SRP are listed in table xx.

**Table xx - Mode page codes for SRP**

| Page code | Description | Clause |
|---|---|---|
| 02h | Disconnect-reconnect page | 10.2 |
| 18h | Logical unit control page (SRP version) | 10.3 |
| 19h | Port control page (SRP version) | 10.4 |

### 10.2 Disconnect-Reconnect mode page

#### 10.2.1 Overview and format of Disconnect-Reconnect mode page

The disconnect-reconnect page (see table xx) provides the application client the means to tune the performance of the service delivery subsystem. The following subclause defines the fields in the disconnect-reconnect mode page of the MODE SENSE or MODE SELECT command that are used by SRP targets.

The application client passes the fields used to control the SRP interface to a device server by means of a MODE SELECT command. The device server then communicates the field values to the target. The field values are communicated from the device server to the target in a vendor specific manner.

SRP SCSI devices shall only use disconnect-reconnect page parameter fields defined below. If any other fields within the disconnect-reconnect page of the MODE SELECT command contain a non-zero value, the device server shall return CHECK CONDITION status for that MODE SELECT command. The sense key shall be set to ILLEGAL REQUEST and the additional sense code set to ILLEGAL FIELD IN PARAMETER LIST.

**Table xx. Disconnect-reconnect mode page**

| Byte Bit | 7 | 6 | 5 | 4 | 3 | 2 | 2 | 0 |
|---|---|---|---|---|---|---|---|---|
| 0 | PS | RSVD | PAGE CODE (02H) | | | | | |
| 1 | PAGE LENGTH (0EH) | | | | | | | |
| 2 | BUFFER FULL RATIO | | | | | | | |
| 3 | BUFFER EMPTY RATIO | | | | | | | |
| 4 | BUS INACTIVITY LIMIT | | | | | | | |
| 5 | | | | | | | | |
| 6 | PHYSICAL DISCONNECT TIME LIMIT | | | | | | | |
| 7 | | | | | | | | |
| 8 | CONNECT TIME LIMIT | | | | | | | |
| 9 | | | | | | | | |
| 10 | MAXIMUM BURST SIZE | | | | | | | |
| 11 | | | | | | | | |
| 12 | EMDP | FAIR ARBITRATION | | | DIMM | DTDC | | |
| 13 | RSVD | | | | | | | |
| 14 | FIRST BURST SIZE | | | | | | | |
| 15 | | | | | | | | |

The BUFFER FULL RATIO field, BUFFER EMPTY RATIO field, BUS INACTIVITY LIMIT field, DISCONNECT TIME LIMIT field, CONNECT TIME LIMIT field, and FAIR ARBITRATION are reserved for SRP devices.

The MAXIMUM BURST SIZE field indicates the maximum size of an RDMA that the device server shall perform. This value is expressed in increments of 512 bytes (e.g., a value of 1 means 512 bytes, two means 1024 bytes, etc.). The device server may round this value down as defined in SPC-2. A value of 0 indicates there is no limit on the amount of data transferred per data transfer operation. This value shall be implemented by all SRP devices. The application client and device server may use the value of this parameter to adjust internal maximum buffering requirements. A router between an SRP device and another protocol device (e.g. FCP) may intercept and adjust this value to reflect its own maximum buffering capabilities.

The ENABLE MODIFY DATA POINTERS (EMDP) bit indicates whether or not the target may use the random buffer access capability to order RDMAs for a single SCSI command. If the EMDP bit is set to 0, the target shall generate continuously increasing RDMA addresses for a single SCSI command. If the EMDP bit is set to 1, the target may issue RDMAs for a single SCSI command in any order. The EMDP bit does not affect the order of frames within an RDMA. The EMDP function shall be implemented by all SRP devices.

The FAIR ARBITRATION FIELD, DISCONNECT IMMEDIATE (DIMM) bit, DATA TRANSFER DISCONNECT CONTROL (DTDC) field, and FIRST BURST SIZE field are reserved for SRP devices.

### 10.3 Logical Unit Control mode page

The Logical Unit Control mode page contains those parameters that select logical unit operation options. The implementation of any parameter and its associated functions is optional. The page follows the MODE SENSE and MODE SELECT rules specified by the SPC-2 standard.

This page is not currently defined for SRP SCSI devices.

### 10.4 Port Control mode page

The Port Control mode page contains those parameters that select target port operation options. The page shall not be implemented by logical units other than LUN 0. The implementation of any parameter and its associated functions is optional. The page follows the MODE SENSE and MODE SELECT command rules specified by SPC-2.

This page is not currently defined for SRP SCSI devices.