

iSCSI Extensions for RDMA

Peter Dunlap
12/6/2007

iSER Overview

- Replaces TCP/IP transport with RDMA operations
 - > Initiator and Target negotiate iSER support during the login phase using TCP/IP
 - > If both Initiator and Target agree to use iSER the connection is switched to “iSER-assisted” mode
 - > Subsequent communication for that session takes place using RDMA operations

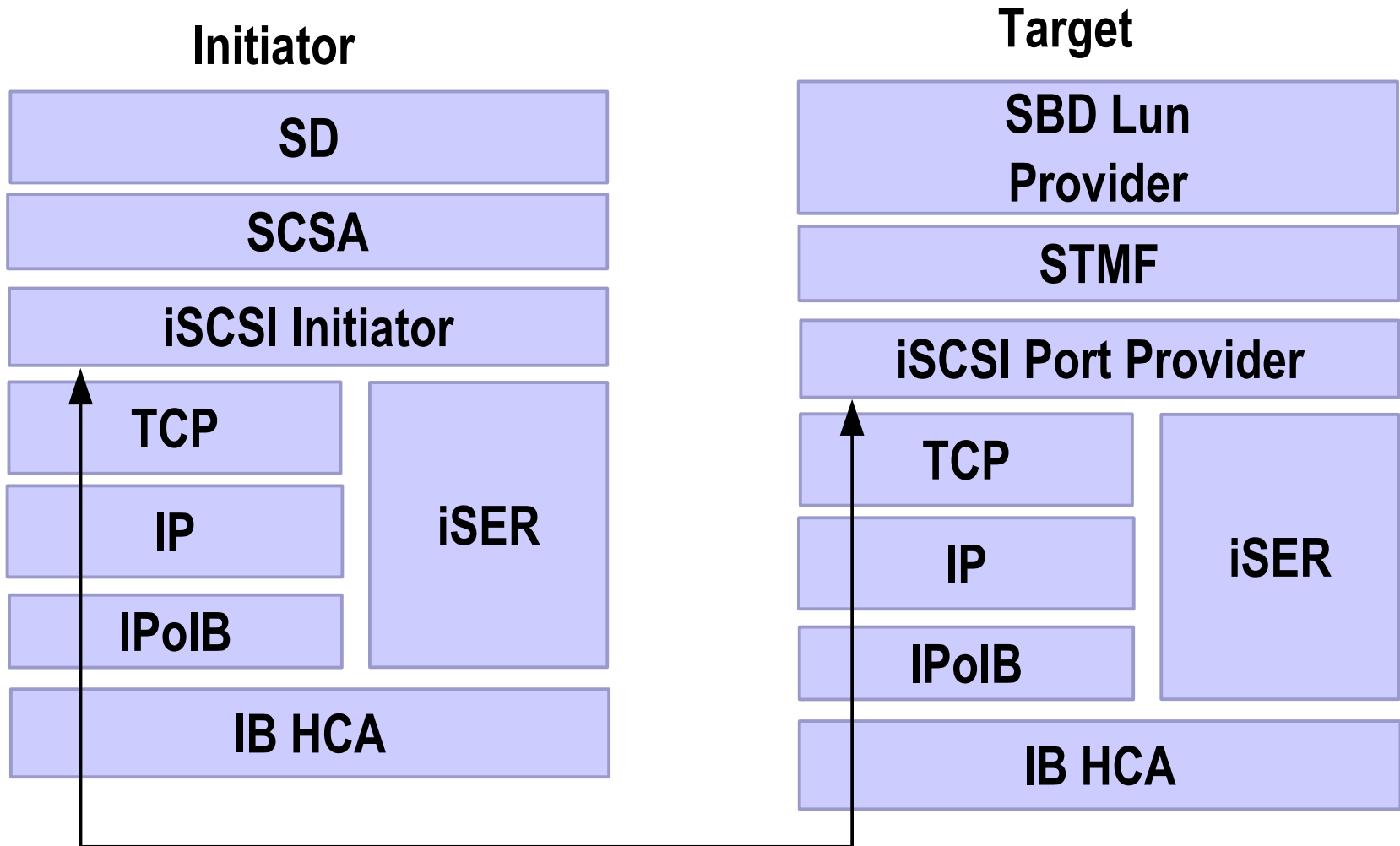
iSER Overview (continued)

- Initial support for InfiniBand only
- It should be straightforward to add support for other physical transports later
- Ethernet and other TCP/IP physical transports can use the iWarp protocol suite to provide RDMA operations
 - > Effectively requires TOE (TCP Offload Engine) that includes iWarp protocol offload

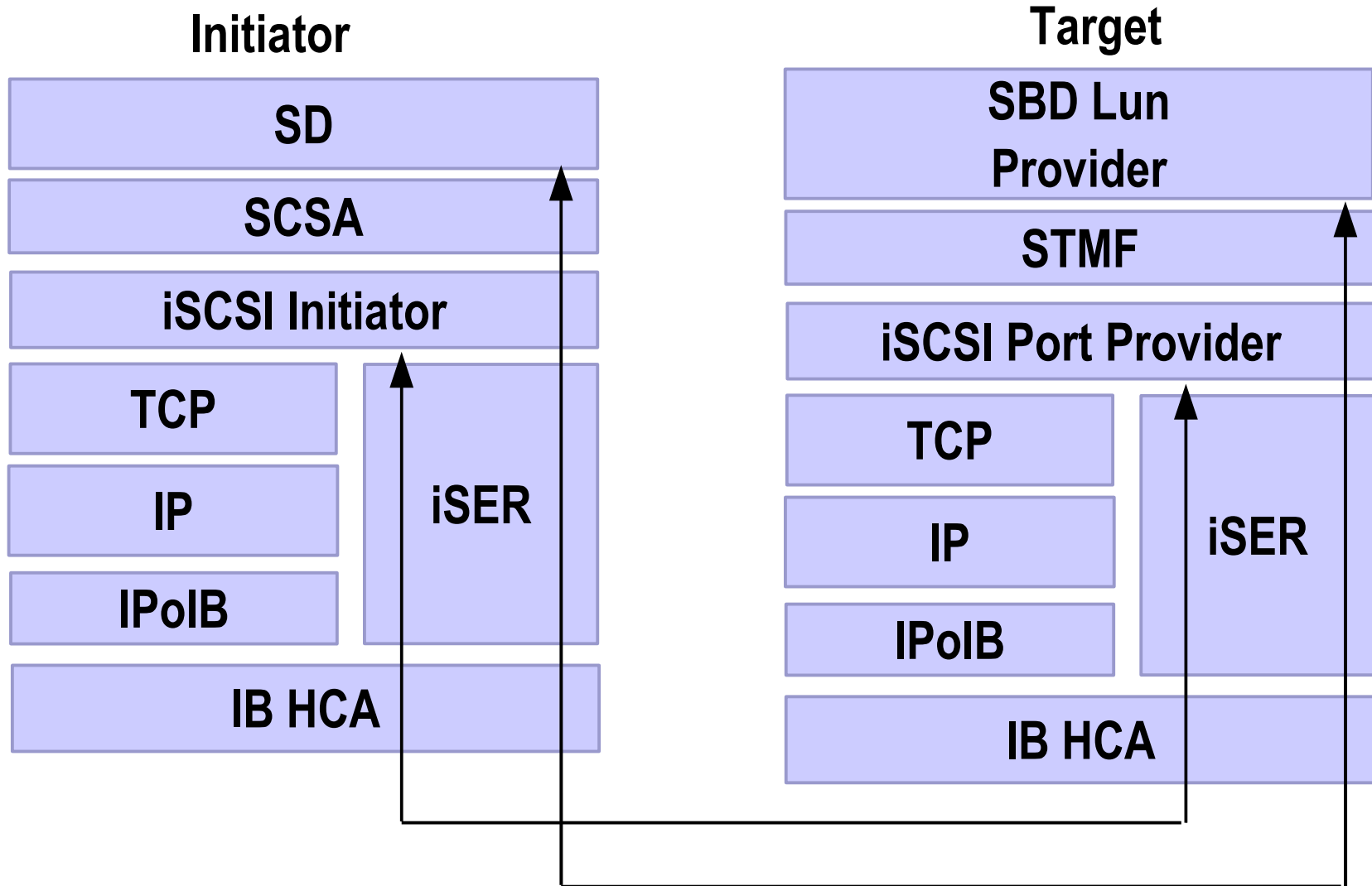
iSER Overview (continued)

- Leverages existing iSCSI infrastructure
 - > Device discovery
 - > Naming
 - > Management
- Supersedes the current SRP protocol
 - > SRP lacks naming and management functionality
 - > No current development on the SRP standard
 - > iSER allows the use of iSCSI mechanisms to discover and manage InfiniBand storage targets

iSCSI/iSER During Login



iSCSI/iSER During Full-Feature Mode



Background

- Solaris has existing iSCSI initiator and target implementations
 - > iSER initiator and target implementations are needed
- iSER project requires modifications to the current iSCSI transports
 - > Connect via TCP/IP
 - > Negotiate RDMAExtensions key
 - > Switch to RDMA assisted mode
- Current code is socket-centric and will require substantial modification to accommodate iSER

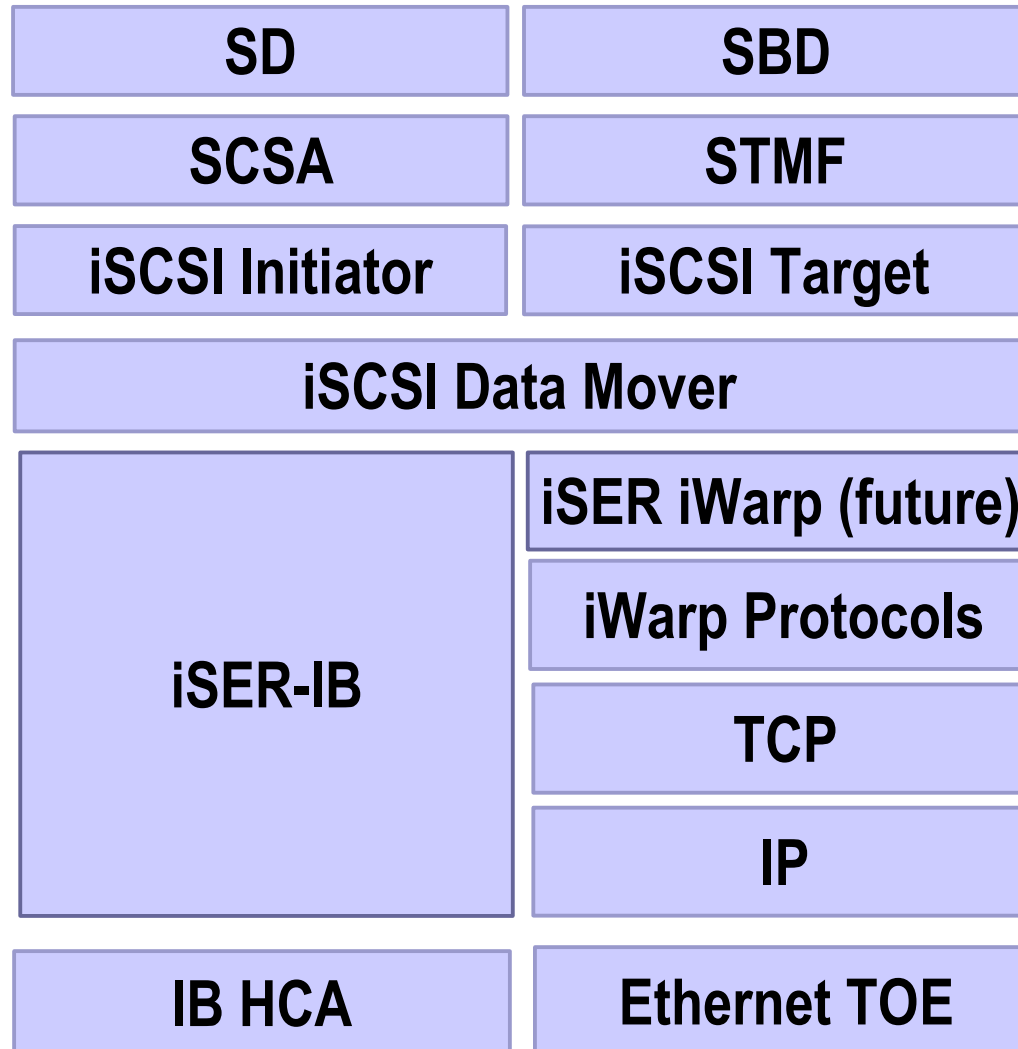
Background

- iSCSI Data Mover Architecture
 - > RFC 5047 (<http://www.ietf.org/rfc/rfc5047.txt>) defines an abstraction layer called “Data Mover”
 - > Separates data movement from the rest of the iSCSI protocol
 - > Doesn't seem originally targeted at TCP/IP transport
 - > Nothing appears to preclude implementing sockets as a sort of “default” transport within the data mover architecture
 - > **No constraints – we are following the concept not the API
- Common Multiprotocol SCSI Target Architecture (COMSTAR)
 - > iSER target will be a Comstar plugin

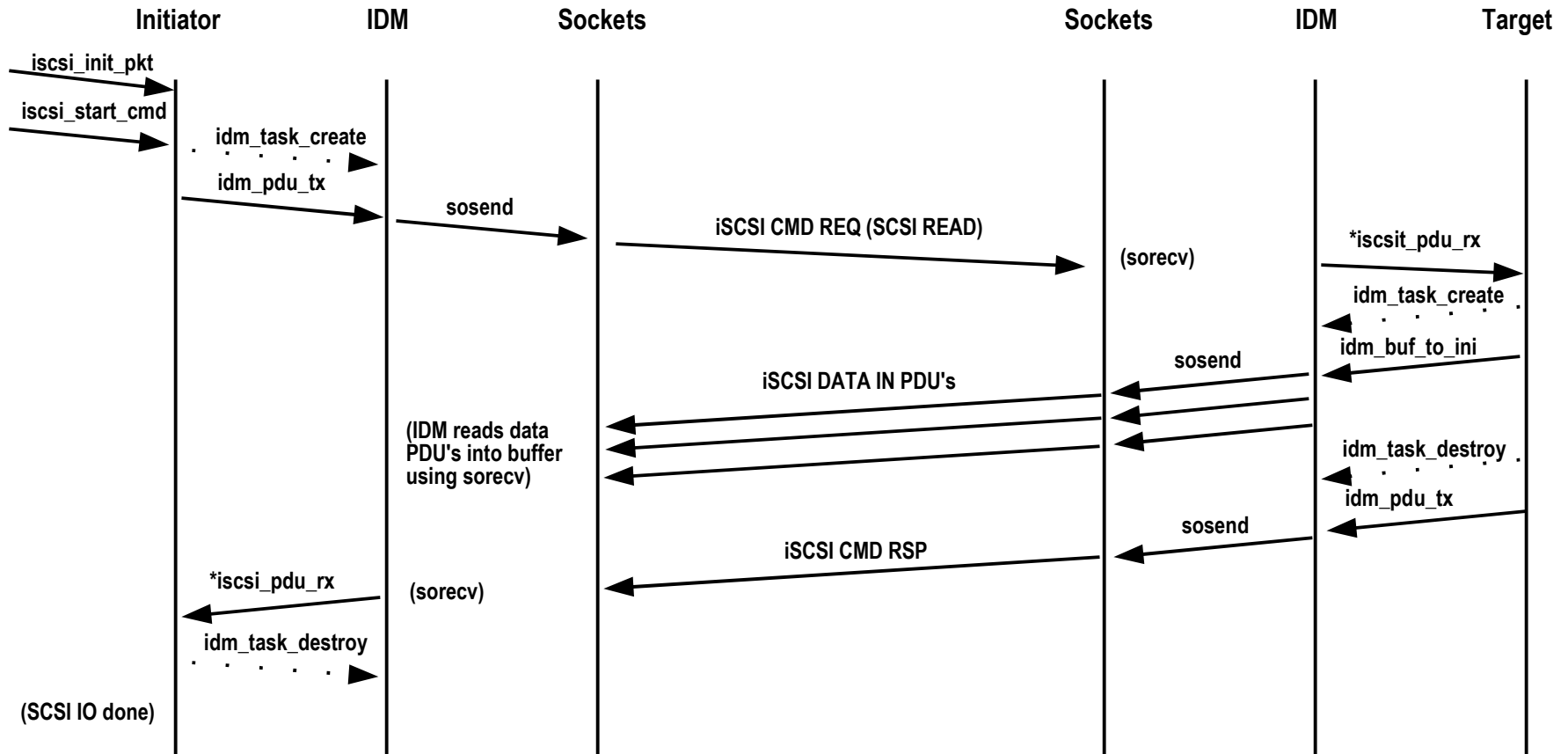
Solution: “Data Mover” module

- IDM: iSCSI Data Mover
- Create unified connection kernel module that will handle connection-related tasks between initiator and target. IDM will be responsible for:
 - > Connection establishment using the appropriate kernel socket services
 - > Manage connection state machine
 - > Transition from standard mode to iSER mode
 - > Generate useful events to client (initiator/target)
 - > “Ready to login”
 - > “Full Feature Phase”
 - > “Connection lost”

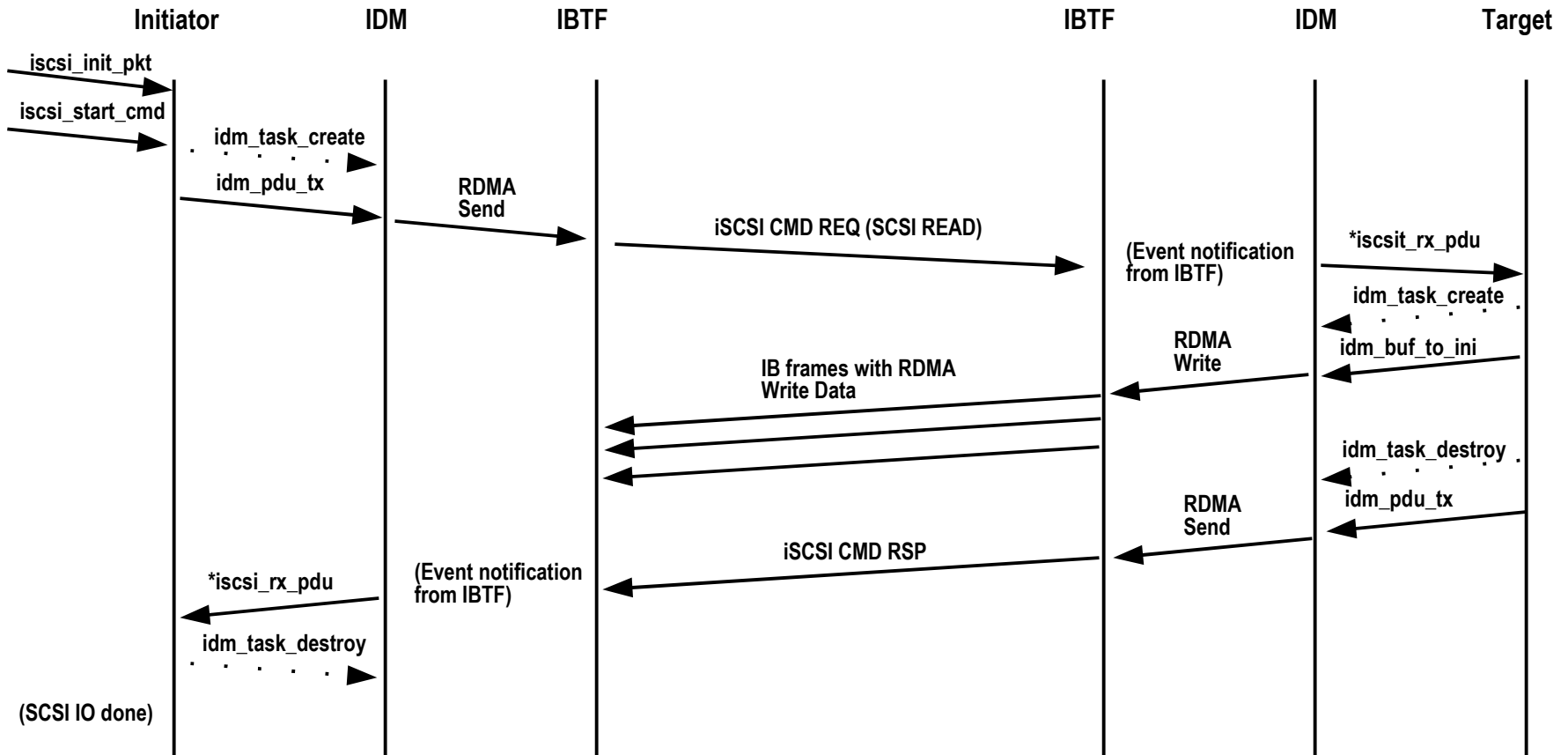
iSCSI Data Mover



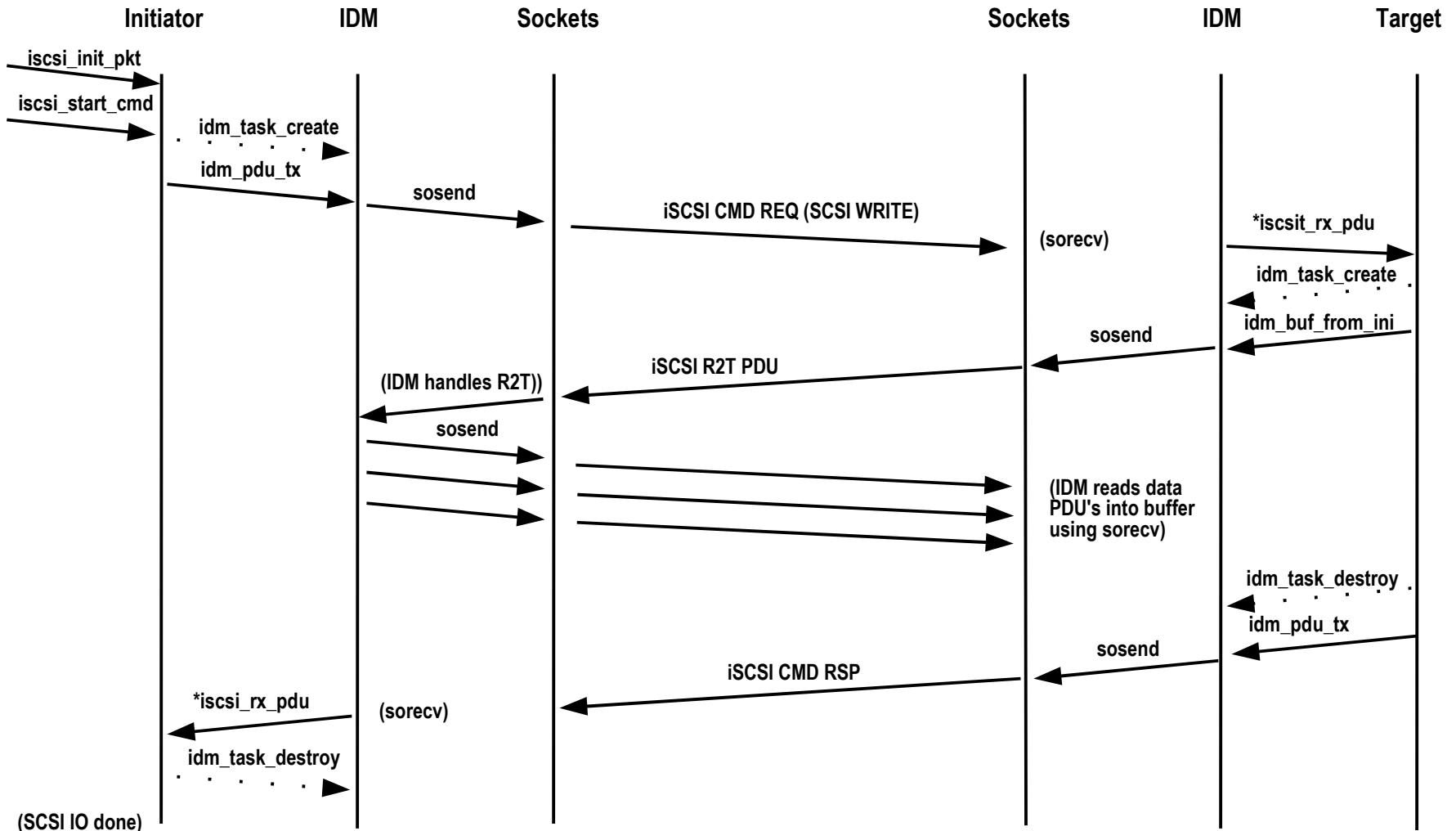
Data flow, SCSI socket read



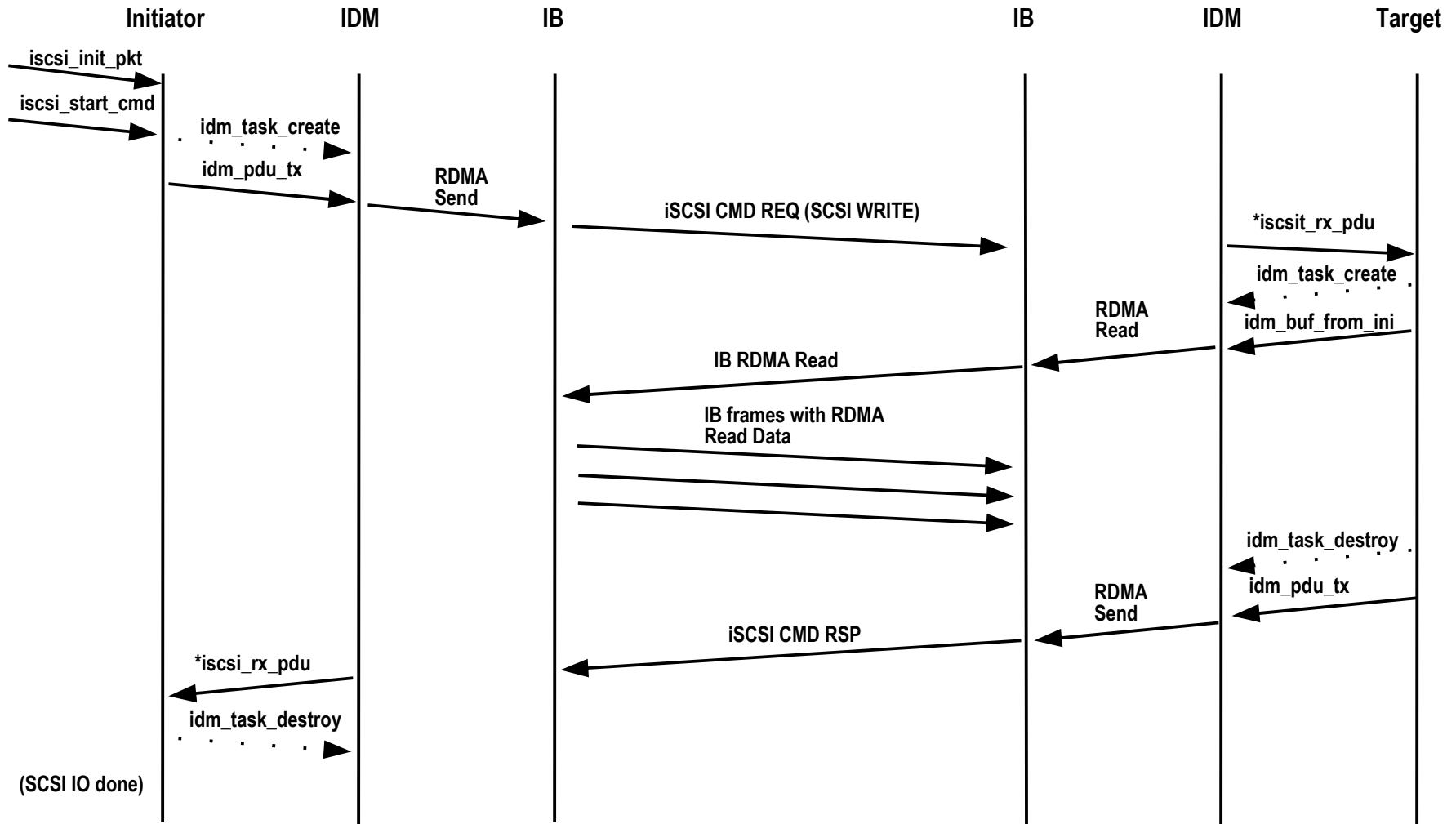
Data flow, SCSI iSER Read



Data flow, SCSI socket write



Data flow, SCSI iSER Write



Development Tasks

- iSCSI Target (Comstar plugin)
 - > Prototype (TCP/IP) planned for Jan 2008
- iSCSI Data Mover
 - > TCP/IP support under development and will be used by the iSCSI target prototype
 - > IDM iSER transport development has started
 - > iSCSI Target plugin + IDM iSER transport = iSER Target

Development tasks

- iSCSI Initiator
 - > Modifications to integrate with IDM have started
 - > iSCSI initiator w/ IDM support + IDM iSER transport = iSER initiator
- iSER Initiator and Target should be feature complete in the mid-April/early-May timeframe
 - > A working prototype should be available earlier
 - > Nevada integration Q1FY09