Storage Systems

NPTEL Course
Jan 2012
(Lecture 06)

K. Gopinath
Indian Institute of Science

SCSI (Small Computer System Interface)

- Standard interface and communication protocol for computer peripherals
  - Allows device mgmt sw to be indep of storage device
- Based on the Client Server Architecture
- Clients called Initiators issue requests and Severs called Targets respond to initiator requests
- Arch may not be best when 3-party xfers needed
  - eg. backup betw tape and disk but control with sw
- SCSI commands sent in units called Command Descriptor Blocks (CDB)
- SCSI Transport Layer protocols such as iSCSI/TCP/IP, FCP govern the flow of SCSI command and data blocks

SCSI

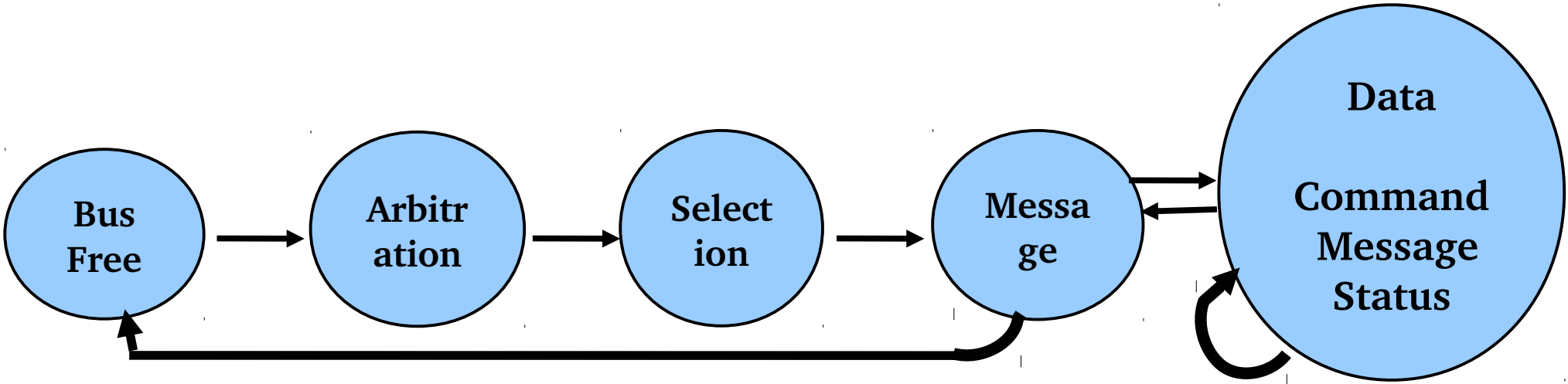
- Early SCSI assumes a bus based architecture
  - not efficient in recovering from packet loss
    - not an issue in bus architecture
    - drivers still based on old SCSI standards and have been retrofitted for a network
  - applications designed to cope with the above
  - pipelining hardly used
- Applications need to commit to stable storage
  - When status OK sent, cannot lose data
- Storage response time
  - Few milliseconds for disks; sub-millisecond for caches
  - Latency budget for interconnect should be less than storage response time

SCSI Terms

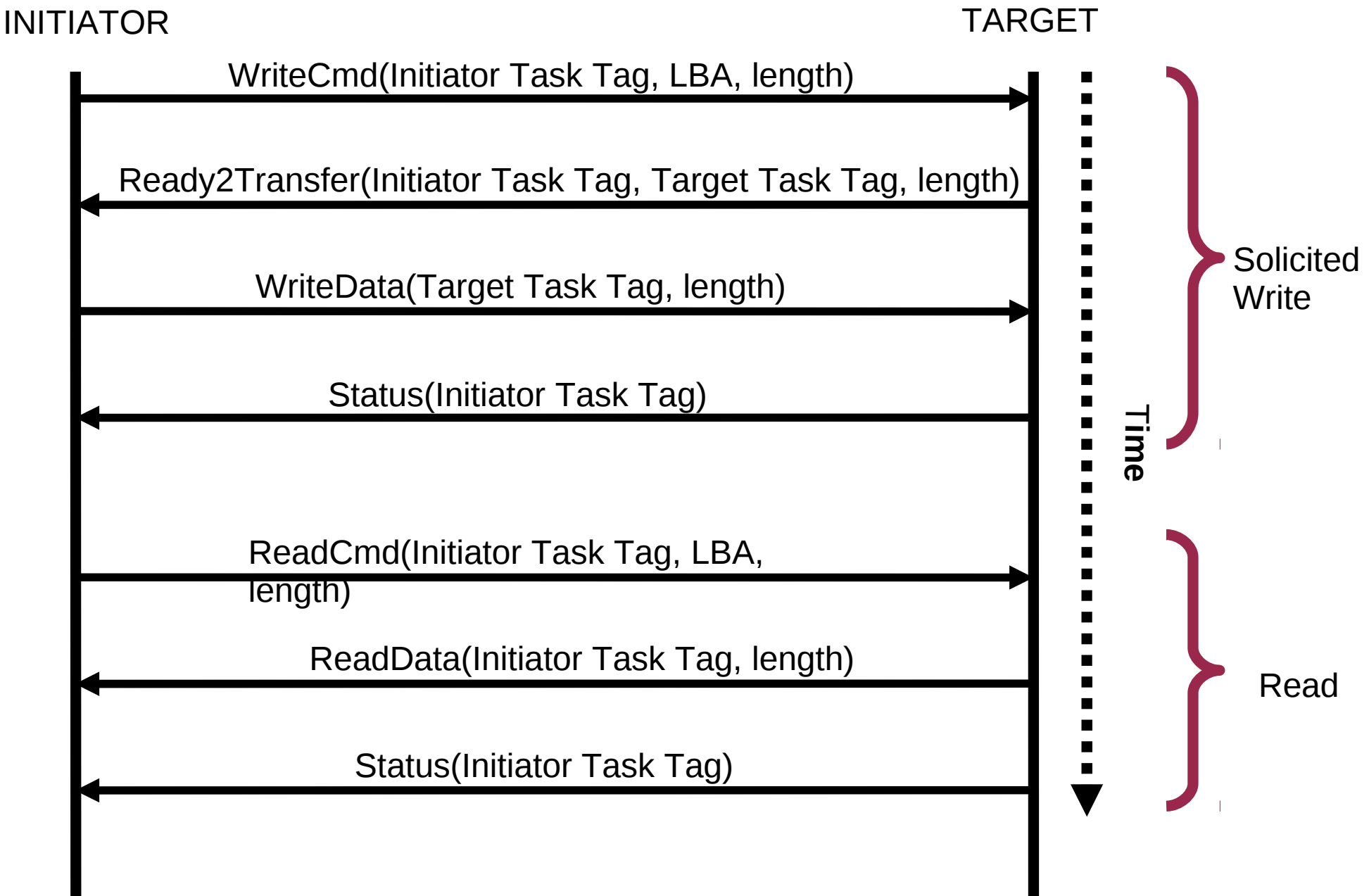
- "target": a collection of logical units, directly addressable on the network
- "initiator": creates and sends SCSI commands to the target
- "task": a linked set of SCSI commands
  - task attr: Untagged, Simple, Ordered, Head of Q, ACA
  - max one command in a task can be outstanding at any given time
- "task tag": used by target to distinguish between tasks
- “LBA”: Logical Block Address
- “LUN”: Logical Unit Number

SCSI Bus Phases

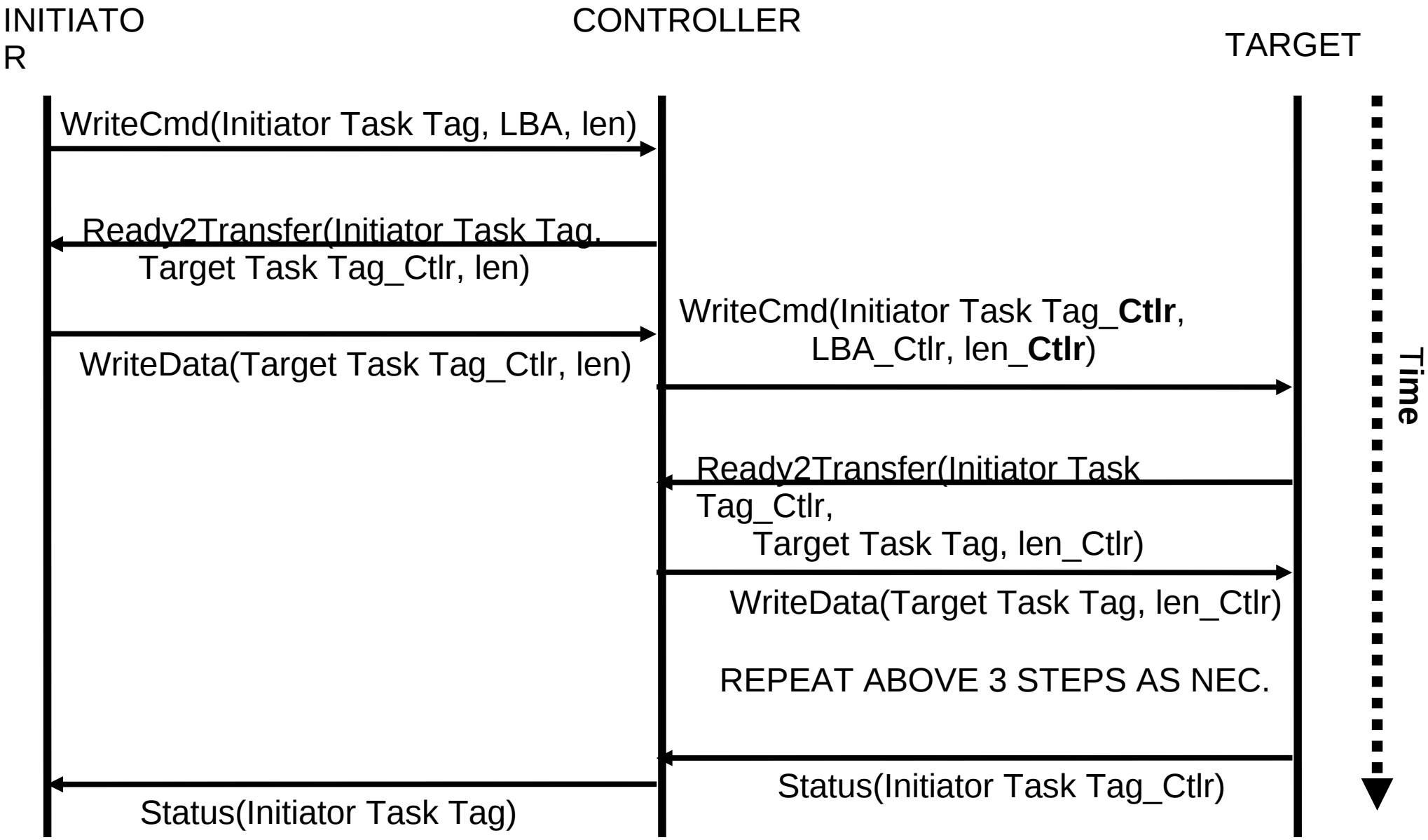
- Bus Free
- Arbitration
- Selection / Reselection: after this, target in control
  - target decides when to recv msg, cmd, data from initiator
  - or send status to initiator
- Message: used by target to send/recv protocol info
  - eg: identify, cmd complete, msg parity error, abort, tagged Q cmds, disconnect/reconnect, data xfer req, kill IOP
- Data, Command, Message, Status



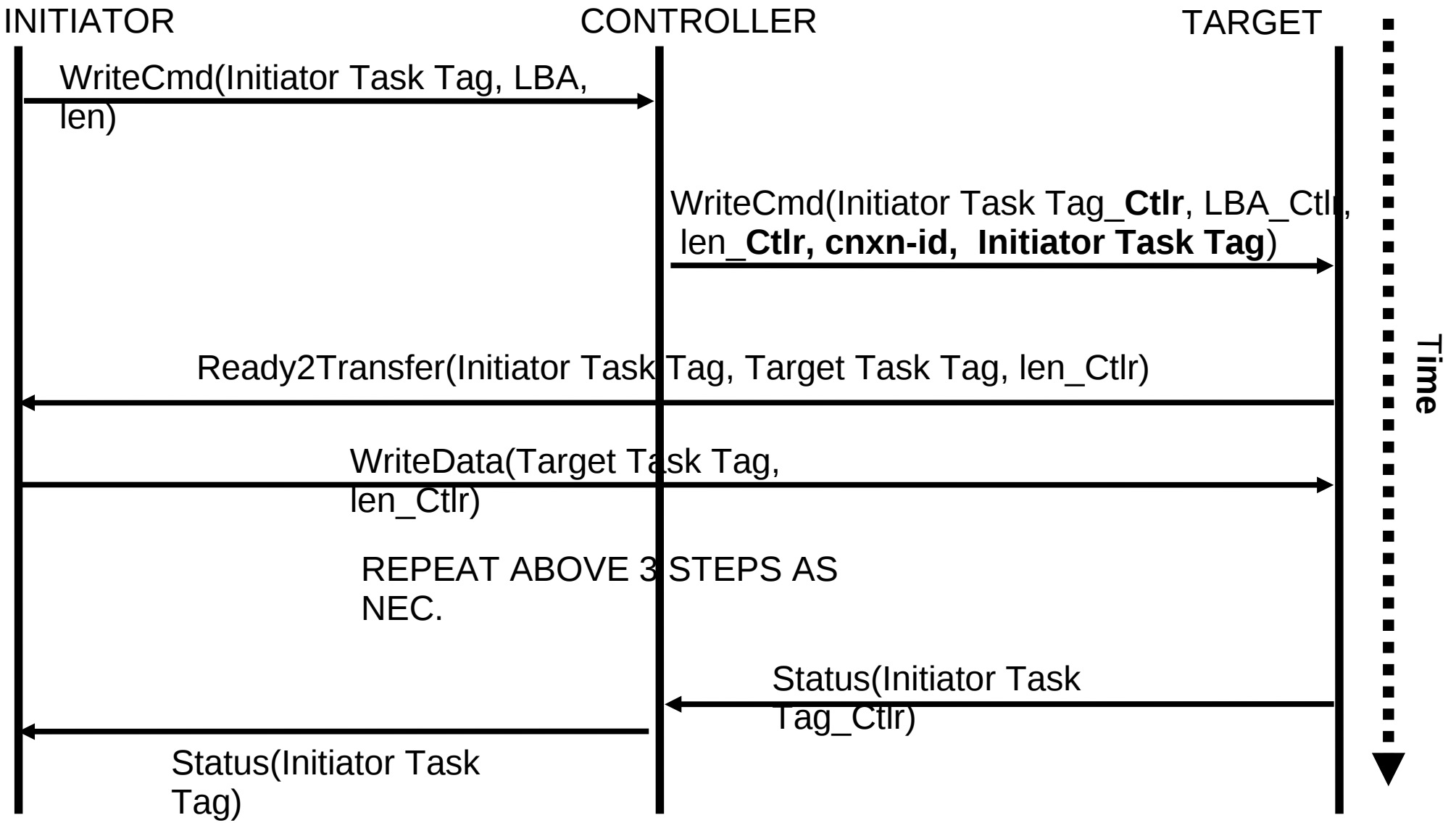
SCSI Protocol



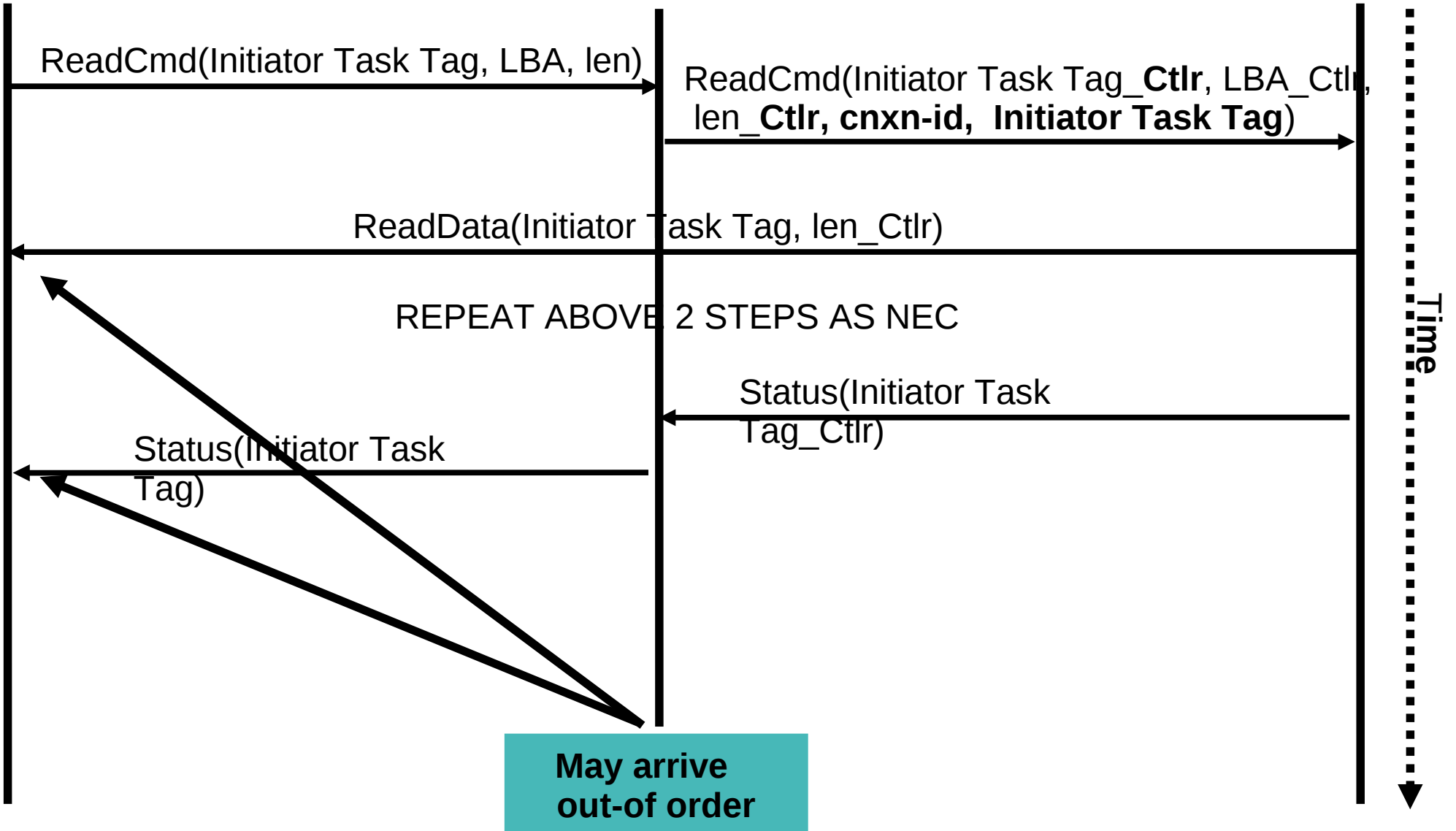
Write with Controller + SAR



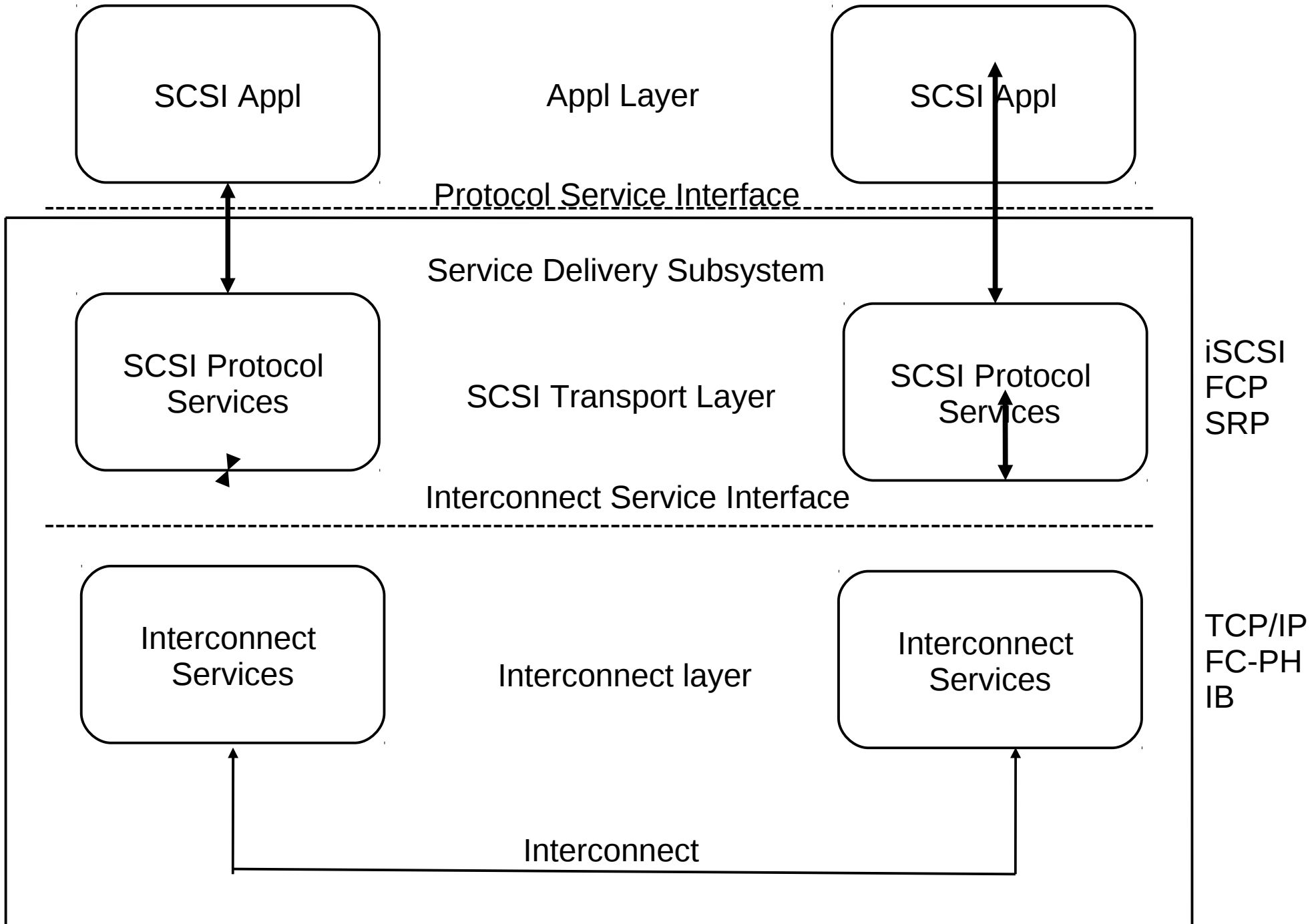
Write with 3-way transfers + SAR



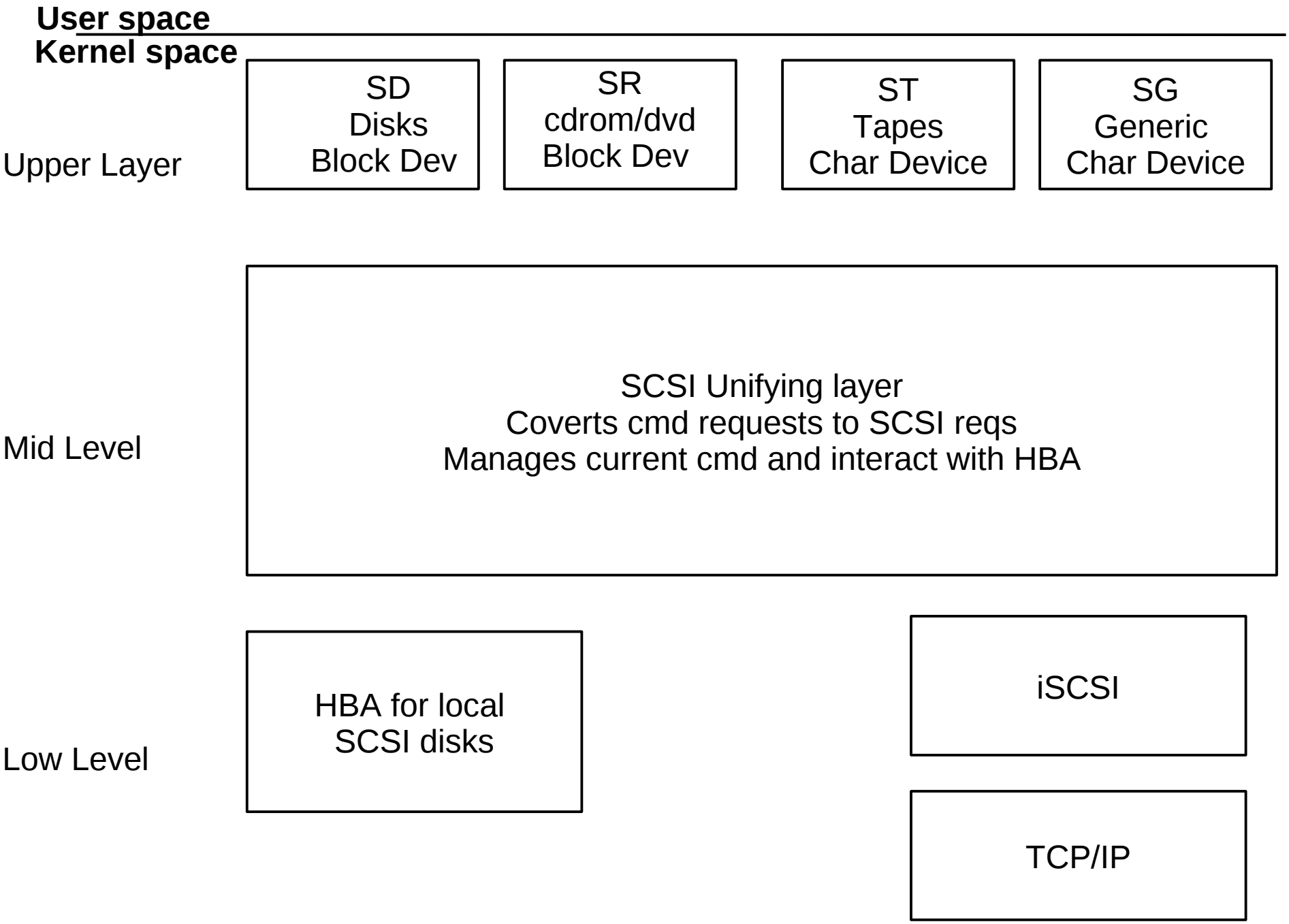
Read with 3-way transfers + SAR



SCSI-3 Arch


SCSI Appl
Appl Layer
SCSI Appl
Protocol Service Interface
Service Delivery Subsystem
SCSI Protocol Services
SCSI Transport Layer
SCSI Protocol Services
iSCSI
FCP
SRP
Interconnect Service Interface
Interconnect Services
Interconnect layer
Interconnect Services
TCP/IP
FC-PH
IB
Interconnect

SCSI layers in Linux



More SCSI Terms

- SCSI-1: Linked commands: Eg: seek+read
  - Many cmds executed in seq in 1 I/O process (IOP)
  - One entry in IOP Q
- SCSI-2: IOP Qing:
  - Untagged: allows a target to accept one IOP from each initiator for each logical unit or target routine
  - target can accept cmds from an initiator while executing IOPs from another
  - Tagged: target can accept a series of IOPs from the same or diff initiators. Executed according to a Q mgmt alg or in specified order
  - Tagged queuing allows drive to accept multiple commands from each initiator

SCSI Command Processing/Concurrency Models

- Tagging
- Ordering wrt diff tags
- Mixing tagged and untagged
- Error Models

SCSI-2 Models

- Simple Q tag
  - IOP added to cmd Q; target decides when to execute (disks can use elevator alg)
  - Commands from other initiators also executed in an order selected in the same manner
- Head of Q tag
  - Add to beginning of Q
  - Multiple Head of Q tags in LIFO order
- Ordered Q tag
  - In order of entry except Head of Q tag processes

Ordering wrt diff tags

- All cmds received with a SIMPLE QUEUE TAG msg prior to a cmd received with an ORDERED QUEUE TAG msg, regardless of initiator, executed before that command with the ORDERED QUEUE TAG msg
- All cmds received with a SIMPLE QUEUE TAG msg after a cmd received with an ORDERED QUEUE TAG msg, regardless of initiator, executed after that cmd with ORDERED QUEUE TAG msg

Mixing tagged and untagged

- An I/O process received from an initiator without a queue tag message while there are any tagged I/O commands in the command queue from that initiator is an incorrect initiator connection, unless there is a contingent allegiance condition.
- An I/O process received from an initiator with a queue tag message while there is an untagged command in the command queue from that initiator is also an incorrect initiator connection.
- In either of these cases the drive removes all commands in the queue from that initiator, aborts the command in process if it is from that initiator, and sets the Sense Key to Aborted Command and the Sense Code to Overlapped Commands Attempted.

SCSI Error Models

- Bus reset:
  - Hard: kill all active and outstanding IOP and reset state of target
  - Soft: try to complete outstanding IOP and preserve current state of target
- Clear Q: (needs tagging) kills only active IOP and all those waiting for this target routine
- Abort Tag: kills only active IOP within ordered tagged Qs
- Abort: Kills all running IOP processes and those in Q
- Asynch Event Notification: from target to initiator
- ACA: sense info that has to be read before next request

Summary

- SCSI Client Server Arch
  - Longevity of protocol
- SCSI protocol takes into acct slowness of devices
  - Also has to consider parallelism when possible
  - Also model errors
  - No security model
- Just as the SCSI layer handles the above issues, every other layer may have to handle the same issues