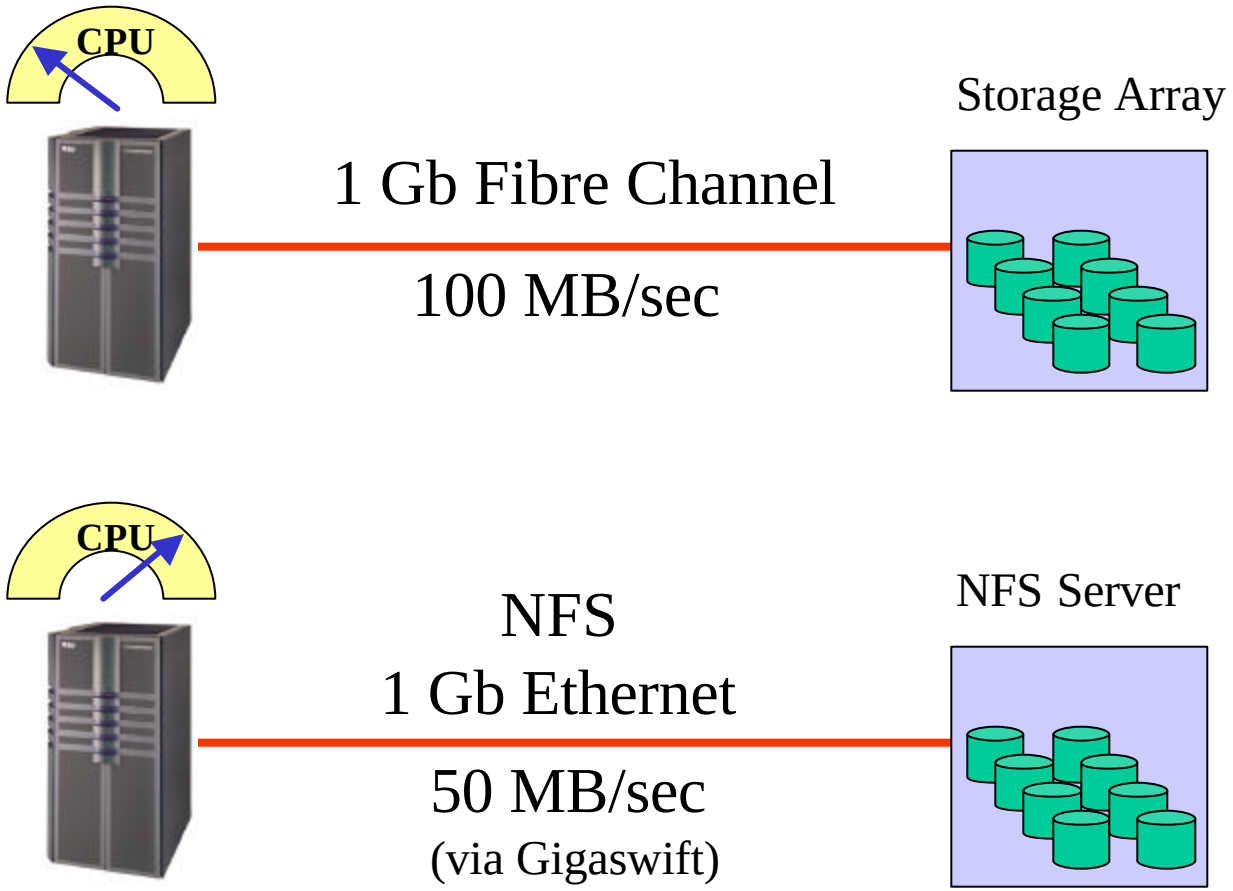
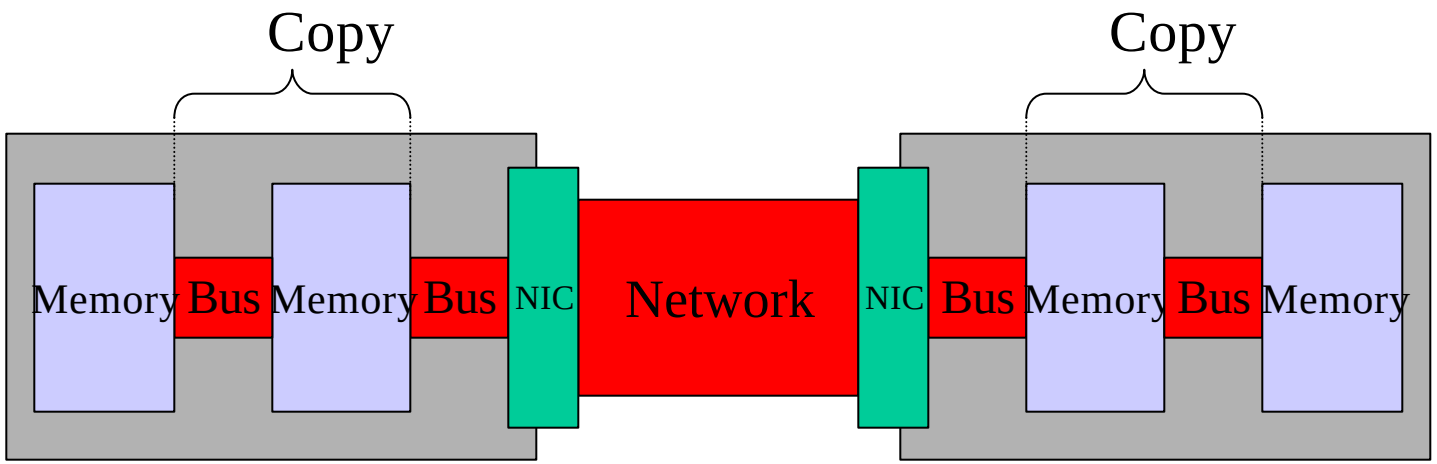
NFS over RDMA

A Problem:
Data Center Performance



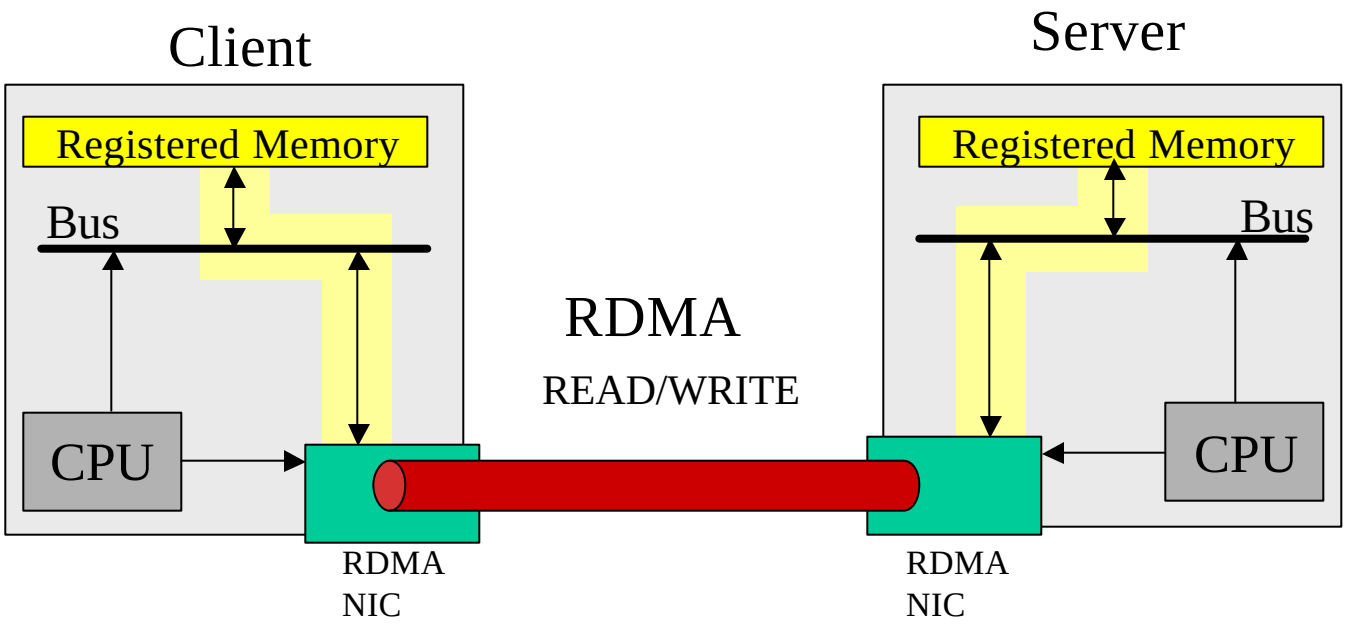
Network vs Bus Performance



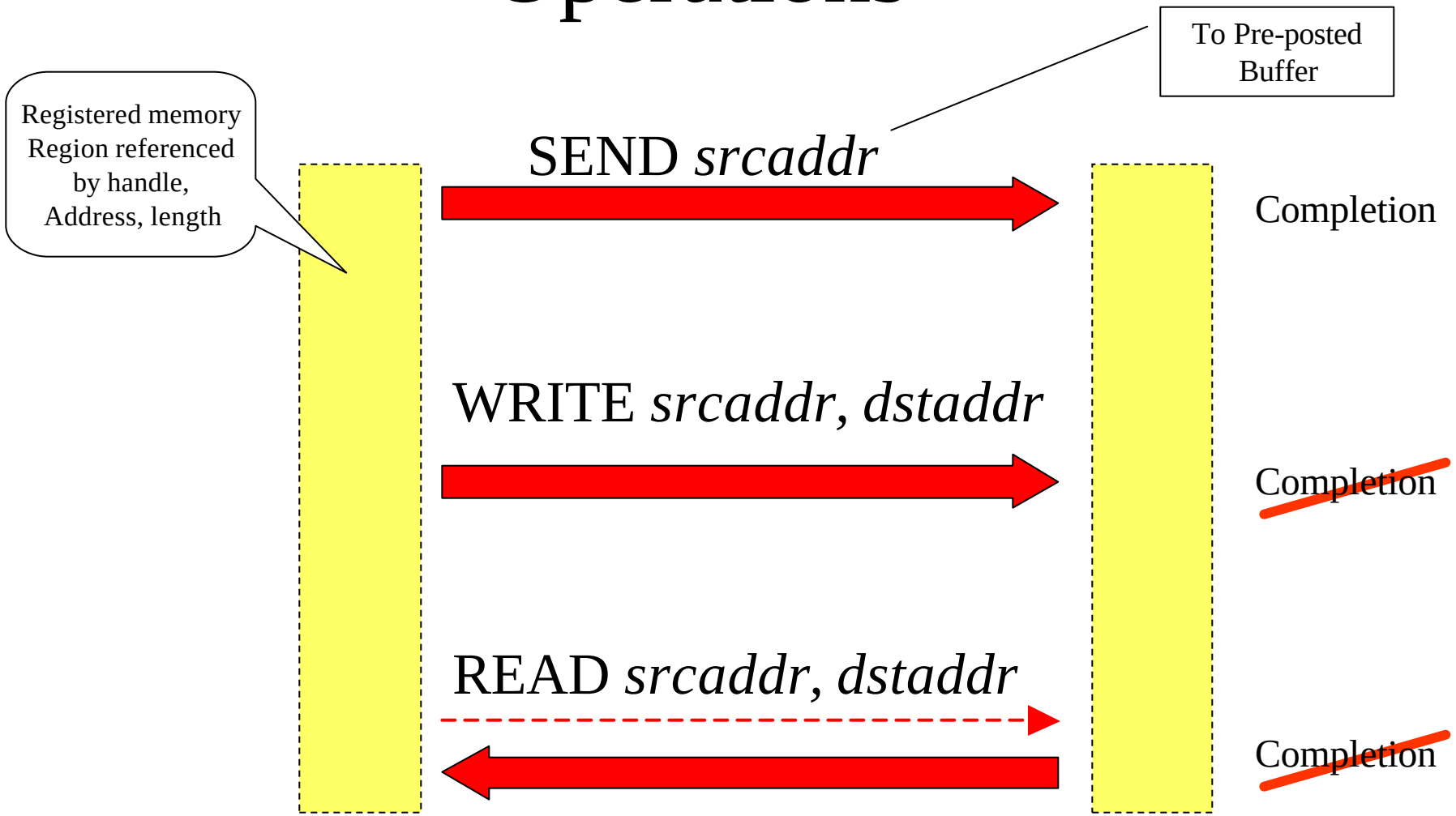
Latency gets worse with each memory copy which further loads the CPU.

What is RDMA ?

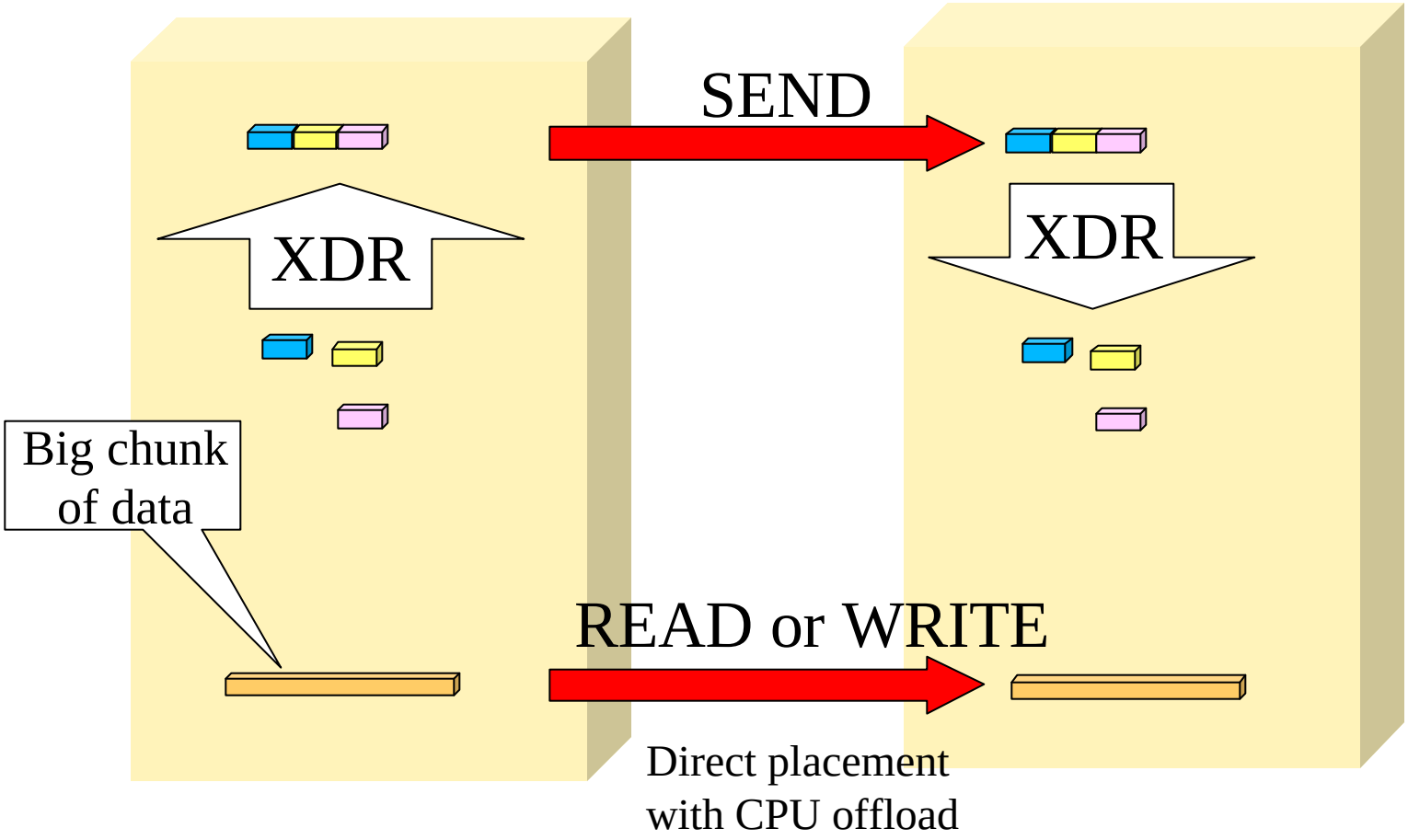
- DMA: Direct Memory Access
- RDMA: *Remote* Direct Memory Access
- Supports Direct Placement
- Networking offload for CPU



RDMA Operations



RDMA RPC Data Movement



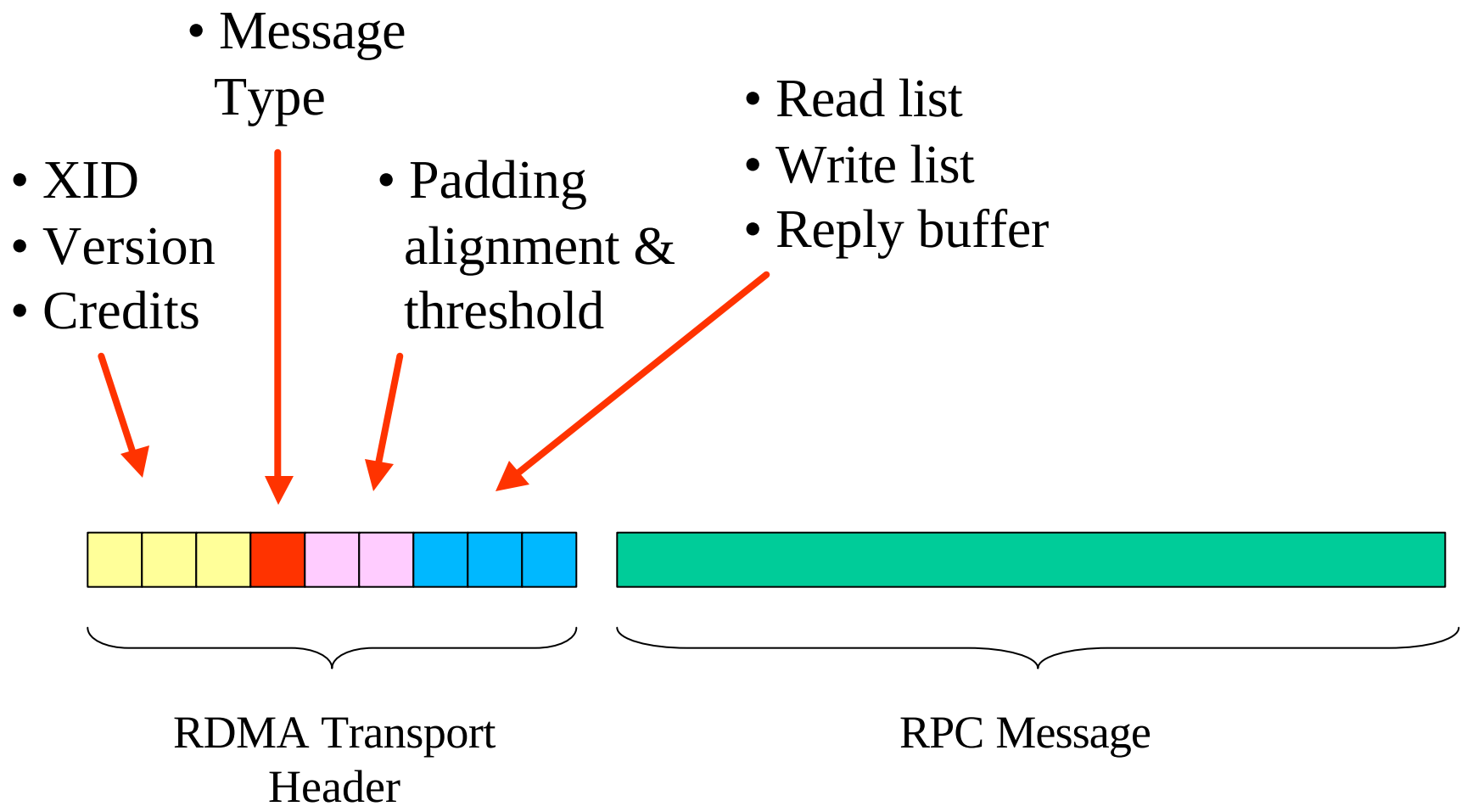
RDMA Technology

- RDMA is native to Infiniband HBAs
- Coming soon to Ethernet via RDDP WG
  - Remote Direct Data Placement
- Will be implemented in NICs along with iSCSI and Ipsec
- APIs: VIA, DAPL, ITAPI
  - User-level and kernel

Applying RDMA to NFS

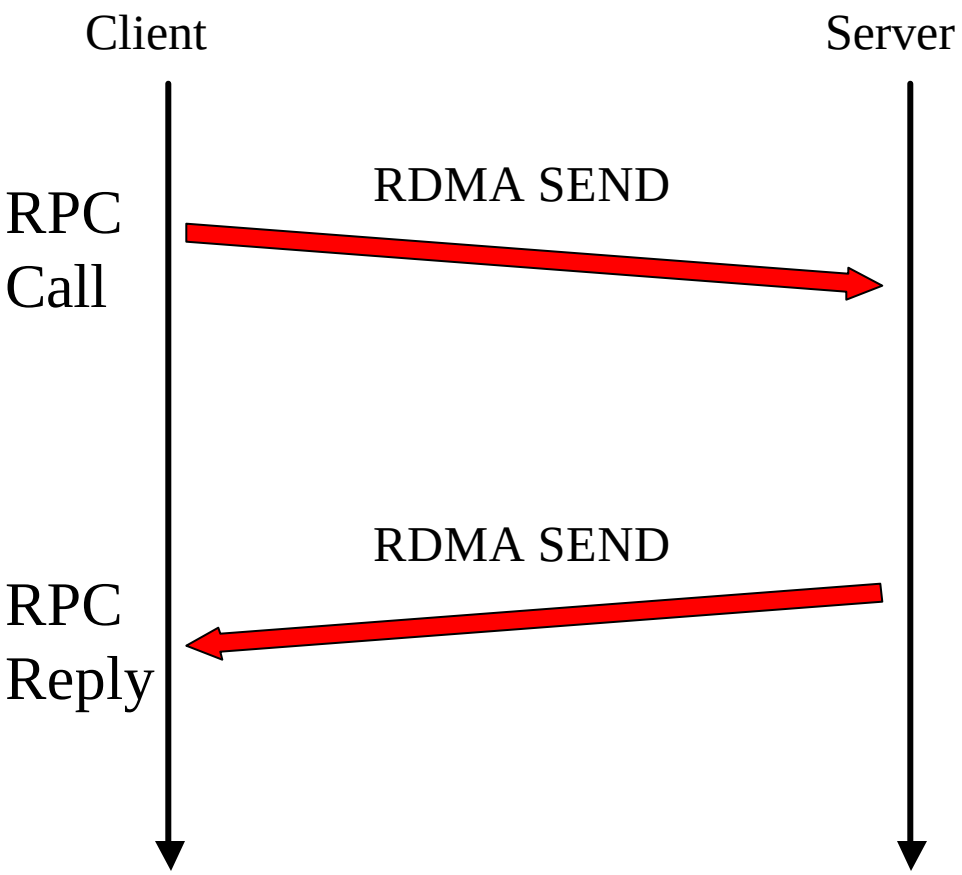
- A new ONC RPC transport
  - RDMA Transport for ONC RPC
  - NFS Direct Data Placement
  - Connection Configuration Protocol
- NFSv4 RDMA and Session Extensions
  - For optimal use of the new RDMA transport

RDMA Transport Header



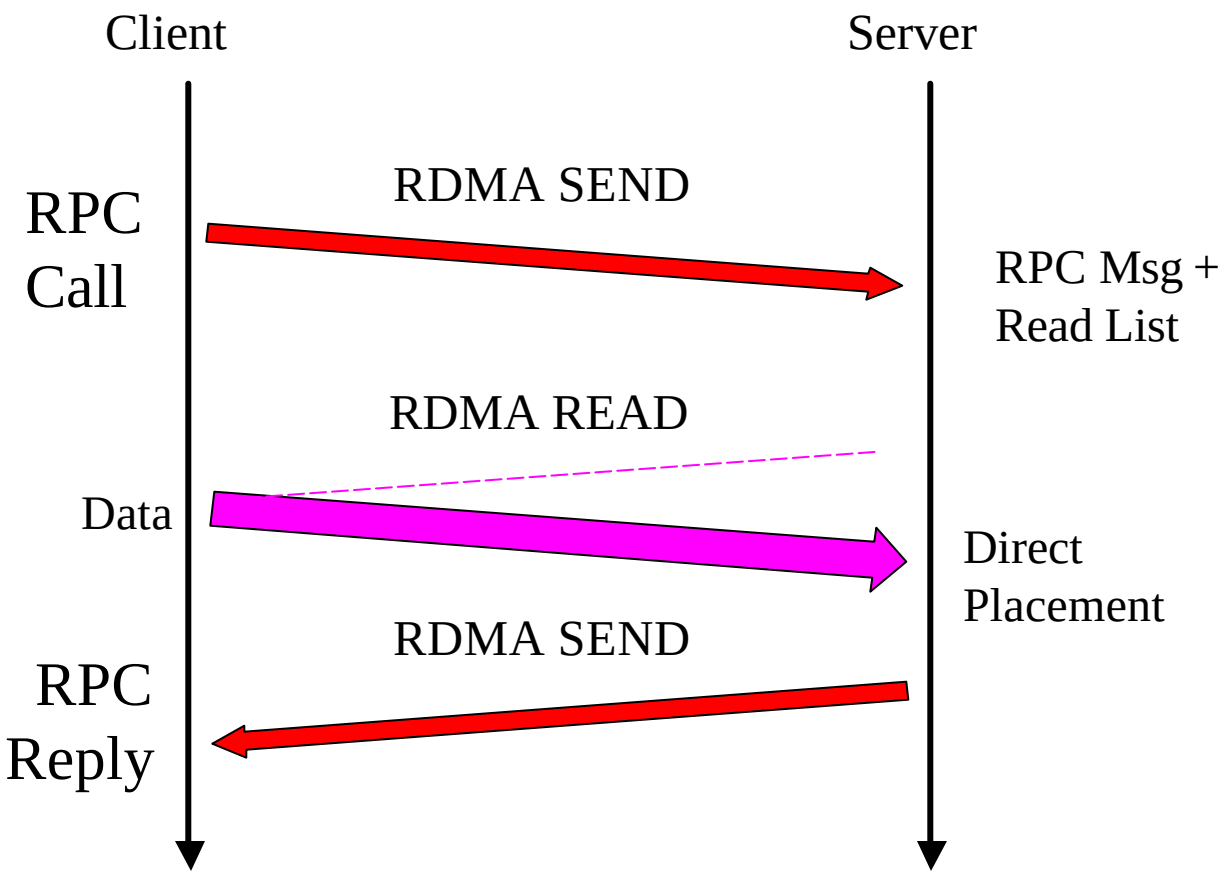
Small RPC Messages

Most RPC Messages are small.
Examples:
- LOOKUP
- GETATTR



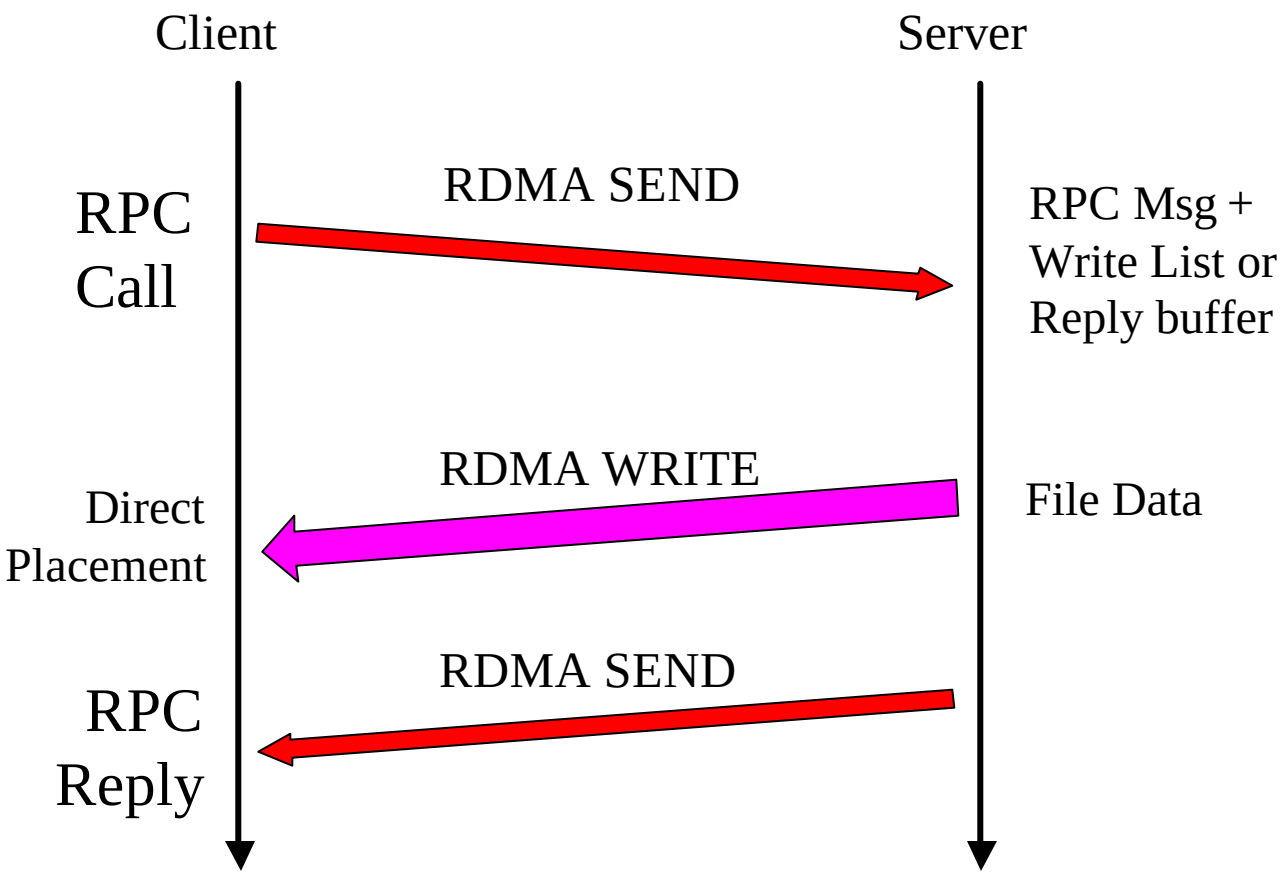
Big RPC Call

Example:
NFS WRITE

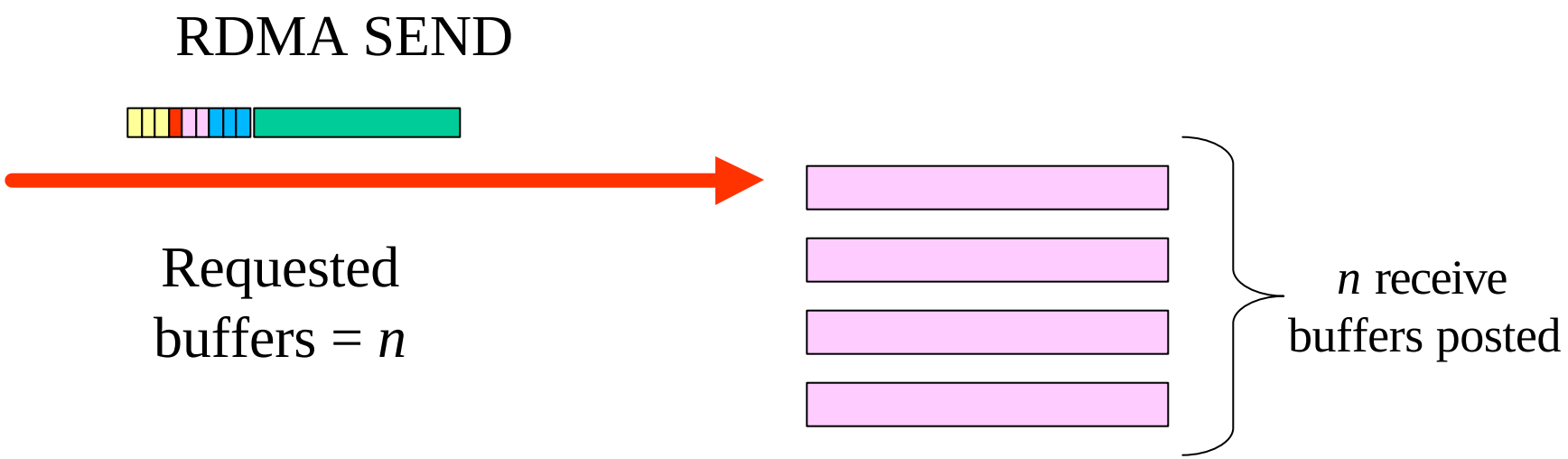


Big RPC Reply

Example:
NFS READ



Credits for Flow Control



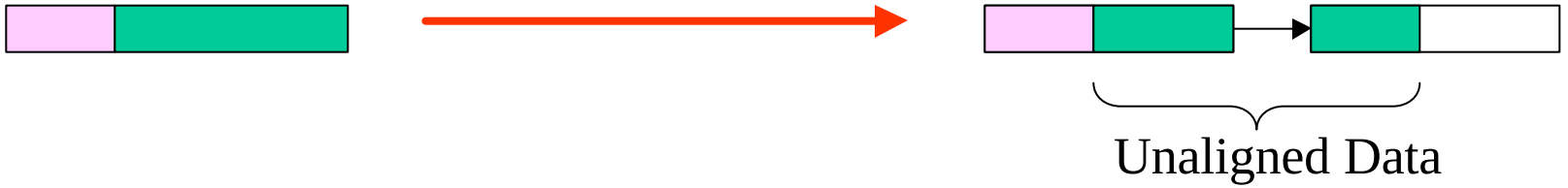
Credits determine concurrency on a single connection. Client requests server maintain *n* pre-posted buffers. Client must not exceed number of buffers confirmed by server.

Send Alignment Padding

Without Padding

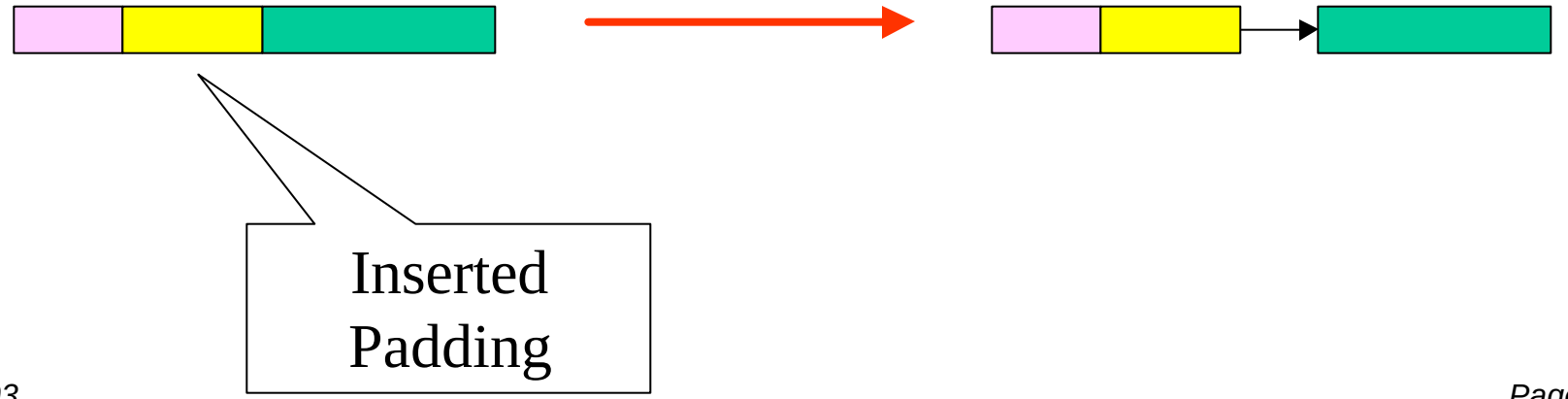
SEND

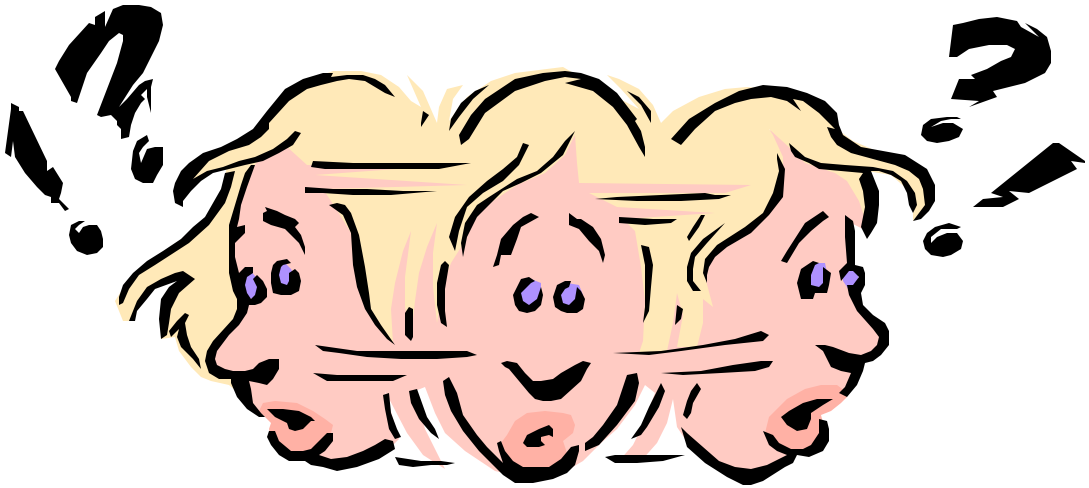
Receive Buffer Chain



With Padding

SEND





Questions & Answers