A Case for Network-Centric Buffer Cache Organization

Gang Peng, Srikant Sharma and Tzi-cker Chiueh

Computer Science Department
Stony Brook University

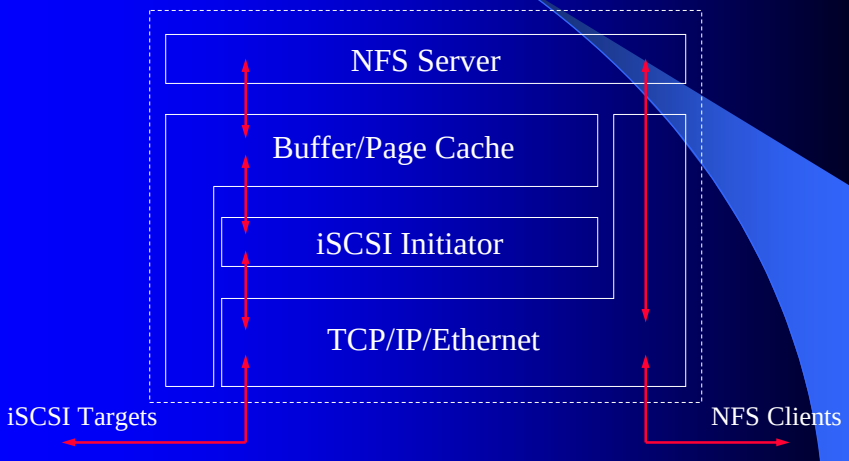
Driving Trends

- Storage Consolidation
  - Management, On-Demand computing
  - e.g., NAS, SAN, iSCSI
- Tiered Structure for Content Delivering
  - Storage servers, application servers and end users

Targeted Scenarios

- What is pass-through server
  - Channel data between external parties
  - Not interpret data most time
  - e.g., NFS server, VoD server with iSCSI storage
- To facilitate data transferring through pass-through servers

NFS Server with iSCSI storage



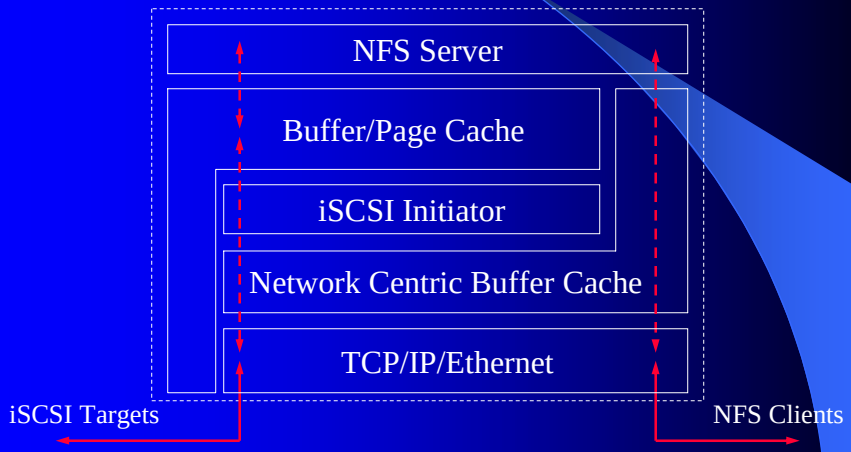
NFS Server with iSCSI storage

- Layered structure
- Data copying across different layers
- Two to three data copying each NFS read or write operation involves
- Overhead associated with network packet construction

Network-Centric Buffer Cache Organization (NCB)

- Data passing-through is stored in network-ready format
  - e.g., sk_buff list in Linux, mbuf list in FreeBSD
- Data is retrieved directly from NCB to deliver, instead of passing multiple layers

NFS Server with iSCSI Storage using NCB



Benefits of NCB

- Eliminate unnecessary data copying
- Remove overhead associated with constructing network packets
  - Network buffer allocation/de-allocation
  - Checksum calculation

Design Question to Answer

- Coupled with File Buffer
  - Data stored in one single place
  - Lack of portability
  - Intrusive to file buffer
- Independent Module
  - Easy to port
  - Intact to file buffer
  - May waste memory

Our Choice – NCB as Secondary Cache to File Buffer

- Independent of file buffer implementation
- Any cache miss in file buffer will be handled by NCB
- To alleviate memory waste
  - Limit file buffer space
  - Linux, FreeBSD

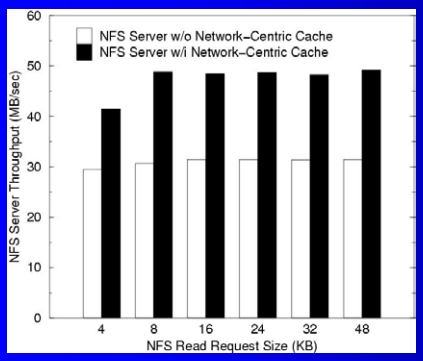
More Implementation Issues

- Logical Copy vs. Physical Copy
  - Logical Copy keeps the interface semantics between different layers unchanged
- Size mismatch between MTU and disk logical block
  - Software jumbo frame

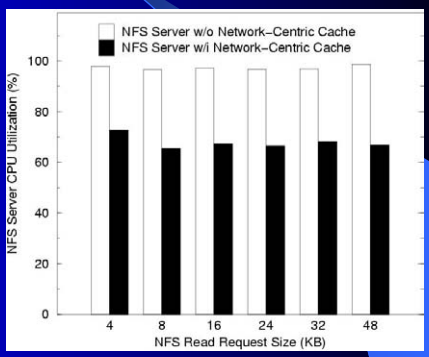
Prototype

- NFS server with Linux 2.4.19 kernel
- iSCSI reference implementation from InterOperability Lab
- Machines use Pentium-III 1 GHz with Intel PRO/1000 MT Gigabit NIC
- NetGear Gigabit switch connects all machines

Preliminary Results



Throughput



CPU Utilization

Related Work

- Zero copy schemes
  - Fbufs (Peter Druschel et al)
  - Zero-copy TCP in Solaris (H.K. Jerry Chu)
  - IO-Lite (Vivek S. Pai et al)
- High Speed Message Passing
  - Axon (James P.G. Sterbenz et al)
  - EMP (Piyush Shivam et al)

Summary

- Organize data passing through in network ready format
- Designed as independent module
- Preliminary result is promising
- Future work
  - Fully operational prototype development
  - Cache performance evaluation
  - Porting to other platforms, e.g., FreeBSD

Thank you for attendance

?