vPFS: Bandwidth Virtualization of Parallel Storage Systems

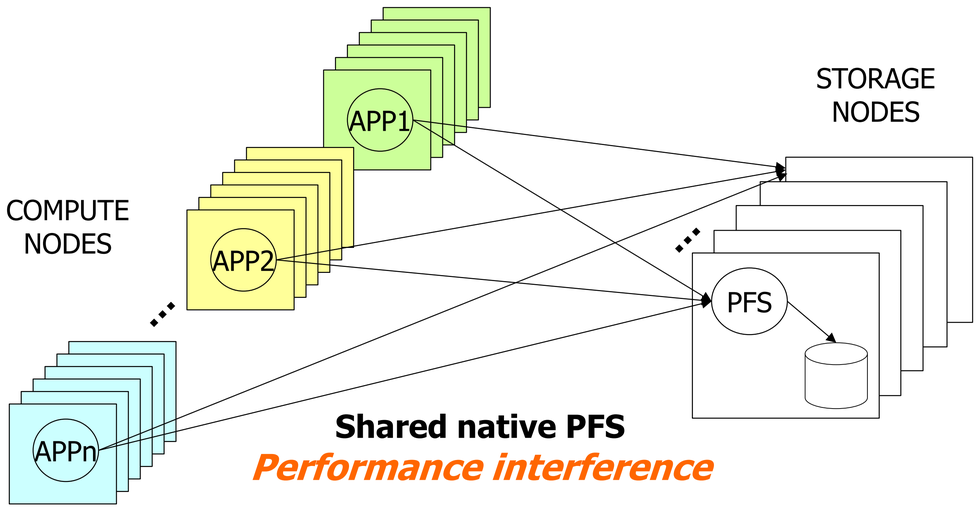
Yiqi Xu, Dulcardo Arteaga, Ming Zhao *Florida International University* Yonggang Liu, Renato Figueiredo *University of Florida*
Seetharami Seelam *IBM T.J. Watson Research Center*

Overview

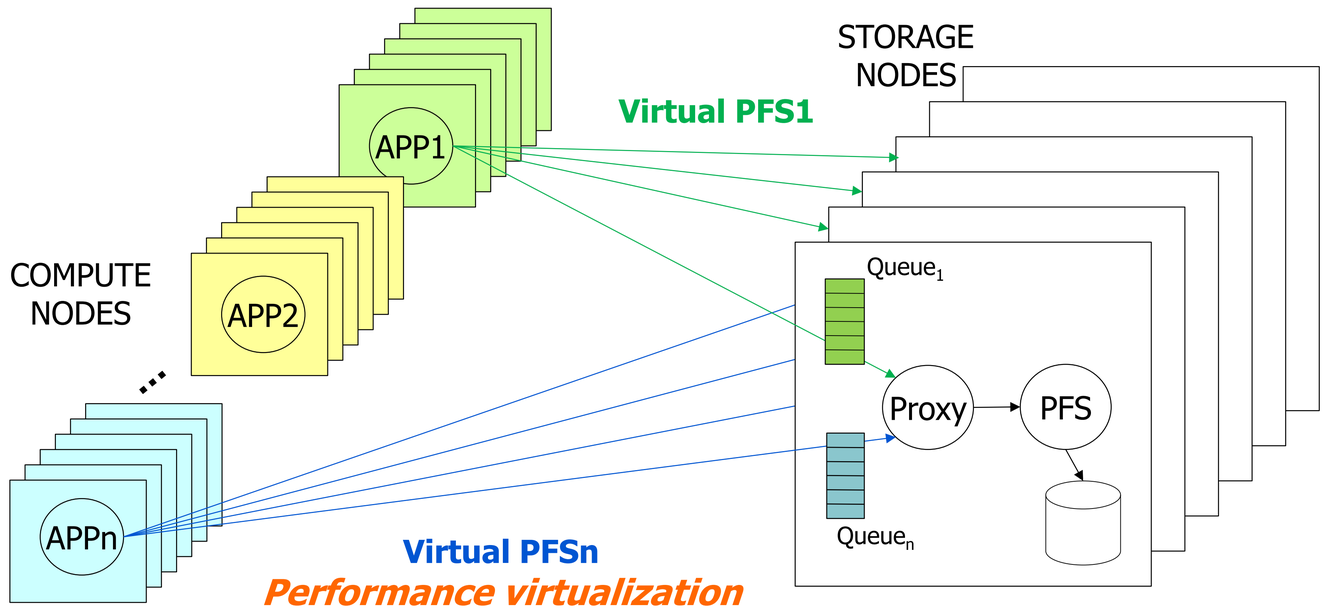
- **Goal:** Application Quality of Service (QoS) driven parallel storage bandwidth management
- **Challenges:**
  - The lack of QoS differentiation in typical high-performance computing (HPC) storage systems
  - The diversity in HPC applications' I/O access patterns and requirements
- **Solution:** vPFS based bandwidth management

Parallel File System Virtualization based Bandwidth Management

- **vPFS: Parallel File System Virtualization**
  - Enable per-application virtual PFSs upon shared physical PFS deployment (e.g., PVFS2, Lustre, GPFS, PanFS, etc.)
  - Allow virtual PFSs to be dynamically created and destroyed based on application lifecycles
  - Allocate parallel storage bandwidth across virtual PFSs per application demand
- **Proportional Sharing of Parallel Storage**
  - Distributed parallel I/O scheduling upon vPFS using enhanced DSFQ scheduler
  - Low cost total-service proportional sharing
    - **Threshold-driven**: broadcast only when local service exceeds a threshold
    - **Layout-driven**: use layout to approximate total service without synchronization



Traditional HPC storage with shared parallel file system

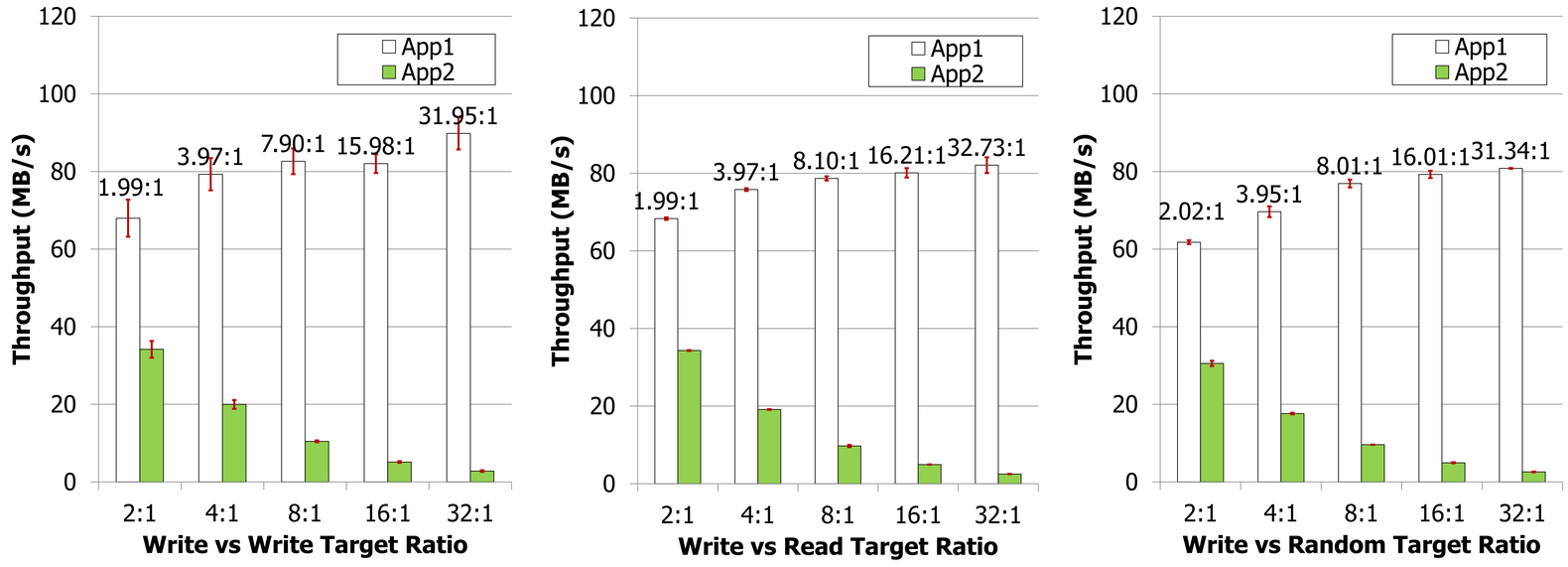


Virtualized parallel storage with per-application virtual PFSs

Prototype and Evaluation

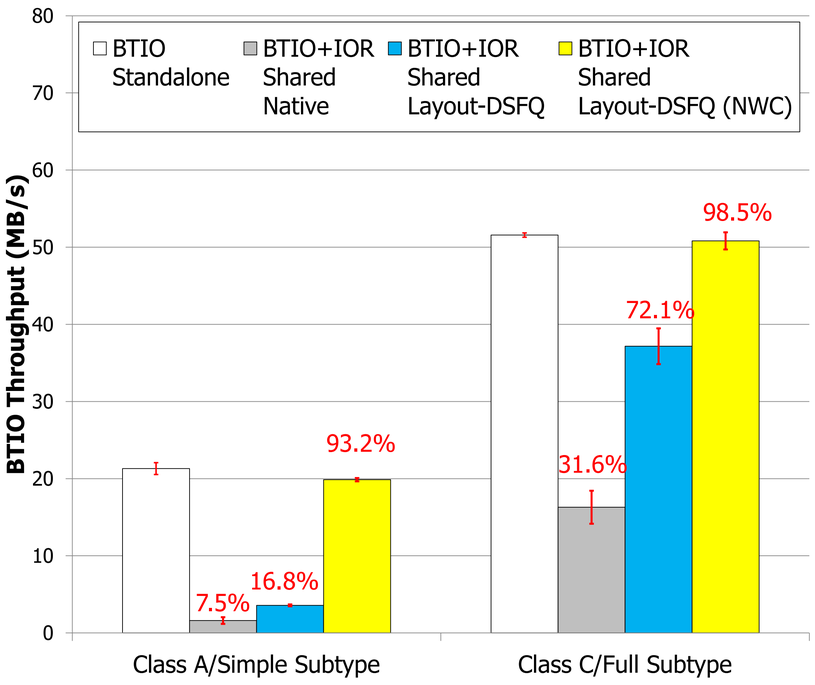
- **PVFS2-based proxy prototype**
  - Interpose and virtualize a deployed physical PVFS2 system
  - Schedule parallel I/Os using enhanced DSFQ
- **Experiment setup**
  - Up to 256 parallel processes on 8 physical nodes
  - Up to 8 PVFS2 servers
  - Typical parallel I/O benchmarks
    - IOR (seq. read, seq. write, rand. read/write)
    - NPB BTIO (class A and class C, collective and non-collective)

Proportional Sharing between two IOR Instances (each w/ 128 processes)



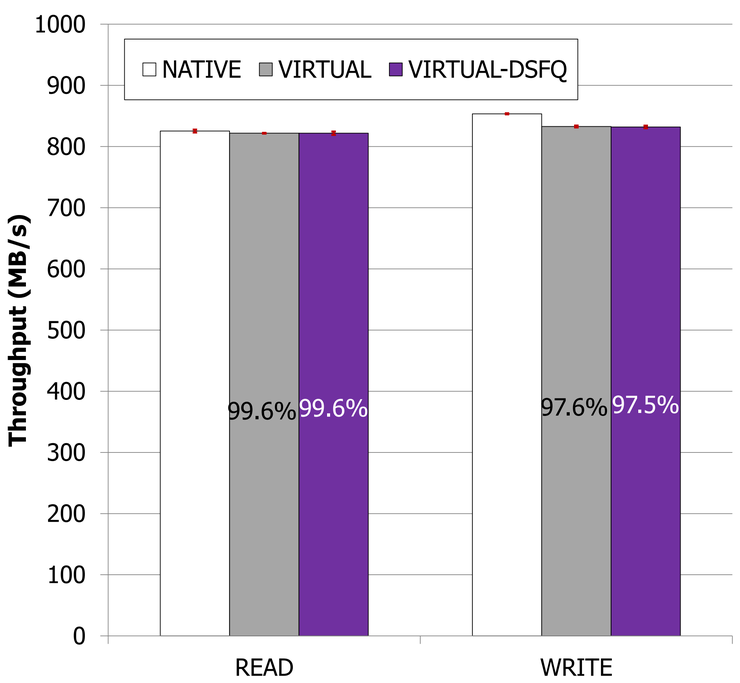
Good proportional sharing achieved for different intensive parallel I/O patterns

Proportional Sharing between BTIO & IOR



BTIO well isolated from IOR

Virtualization Overhead



Less than 3% throughput overhead

Conclusion and Future Work

- **Conclusions**
  - vPFS enables flexible bandwidth allocation and strong performance isolation
  - Enhanced DSFQ supports good total-service fairness on parallel storage systems
  - Performance and resource overhead of proxy-based vPFS implementation is small
- **Future Work**
  - Study latency-driven scheduling for applications sensitive to response times
  - Consider autonomic I/O optimization upon the vPFS framework

FIU VISA

Virtualized Infrastructure, Systems & Applications (VISA) Research Laboratory (http://visa.cis.fiu.edu)

Effort sponsored by National Science Foundation under grants CCF-0937973 and CCF-0938045. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.