

---

## *A Highly Available Cluster Storage System Using Scavenging*

High Availability and Performance Computing Workshop (HAPCW'04), held with  
2004 Los Alamos Computer Science Institute Symposium (LACSI'2004)

**Xubin (Ben) He, Li Ou**  
{hexb, lou21}@tntech.edu

Storage Technology & Architecture  
Research(STAR) Lab  
Electrical and Computer Engineering  
Department



**Stephen L. Scott, Christian Engelmann**  
{scottsl, engelmann}@ornl.gov

Computer Science and Mathematics Division



## Outline

---

**Introduction**  
**Highly Available Metadata Management**  
**Metadata Management using Bloom Filters**  
**Conclusions and Future Work**

## Motivations

---

- Data intensive scientific applications → high performance computing.
- Distributed storage systems: high performance storage, wide area mass storage, cluster storage.
- Excellent performance, parallel support  
←→administration costs, central points of failure and control

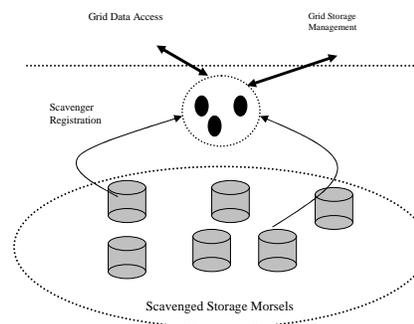
A Highly Available Cluster Storage System Using Scavenging, HAPCW'2004

3

## Existing work:

---

- HPSS
- Peer-to-Peer Storage
- Scavenging [Vazhkudai]



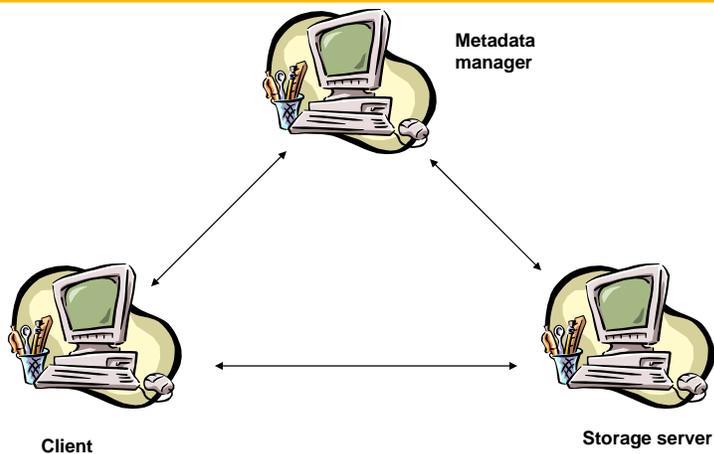
A Highly Available Cluster Storage System Using Scavenging, HAPCW'2004

4

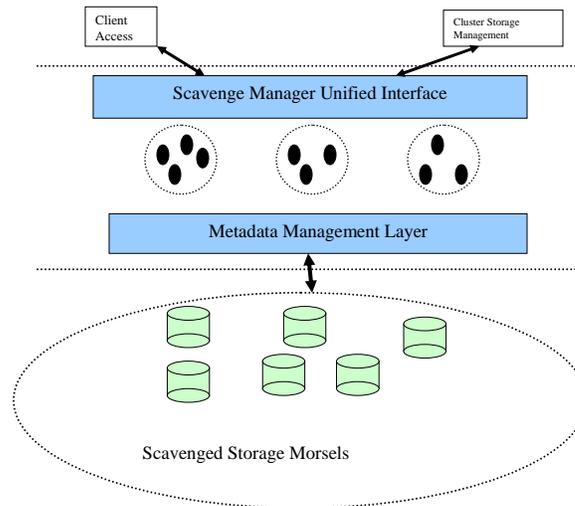
- 
- Introduction
- ✓ **Highly Available Metadata Management**
  - Metadata Management using Bloom Filters**
  - Conclusions and Future Work**

## User Data vs. Metadata

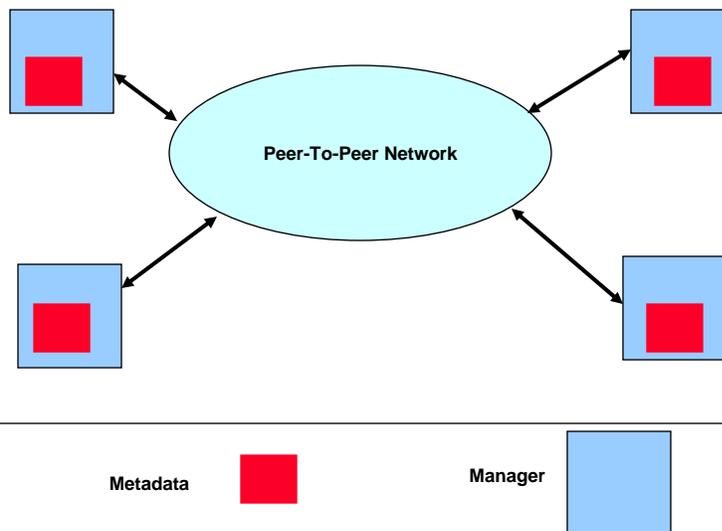
---



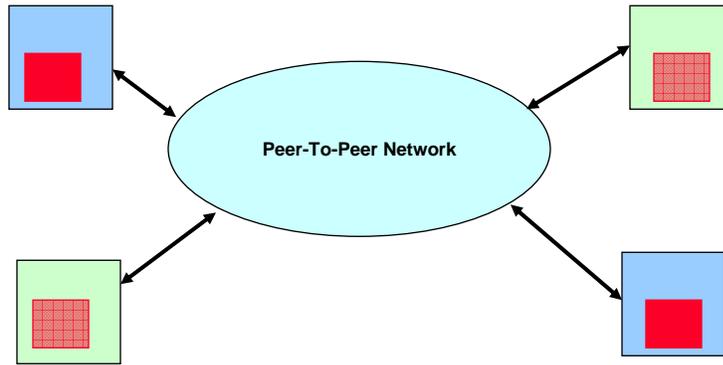
## Distributed Metadata Management



## High Availability: Scheme #1 (Pure P2P)



### High Availability: Scheme #2 (active/hot-standby)



Metadata



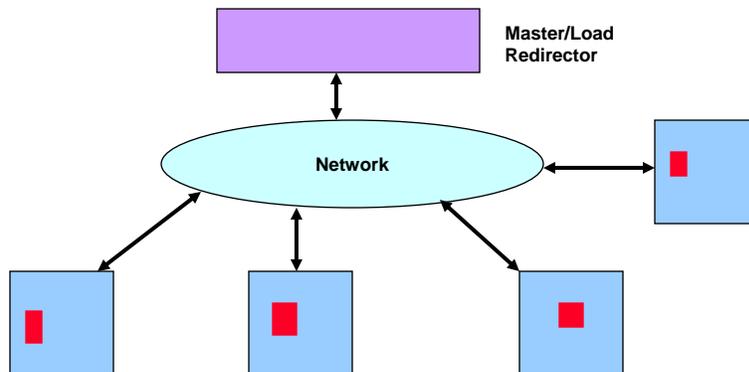
Active Manager



Hot-Standby  
Manager



### High Availability: Scheme #3 (Partitioning)



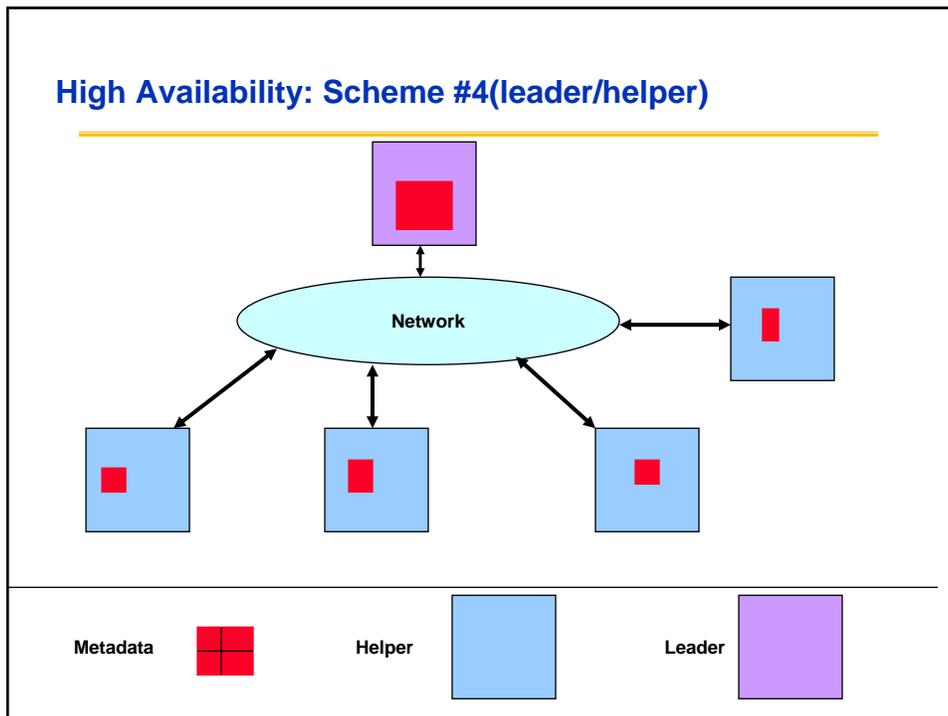
Metadata



Manager



## High Availability: Scheme #4(leader/helper)



## Group Communications

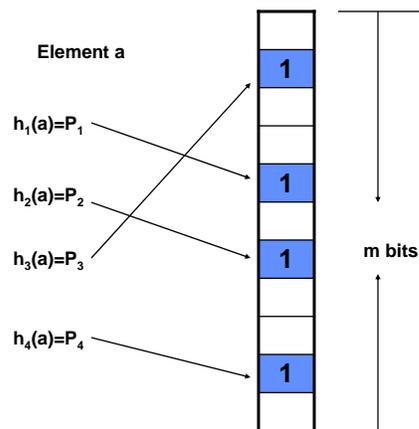
- Peer-to-Peer distributed control
- Reliable broadcast, Atomic Broadcast
- Atomic transactions guarantee metadata integrity

- 
- Introduction
  - Highly Available Metadata Management
  - ✓ **Metadata Management using Bloom Filters**
  - Conclusions and Future Work

## Bloom Filter

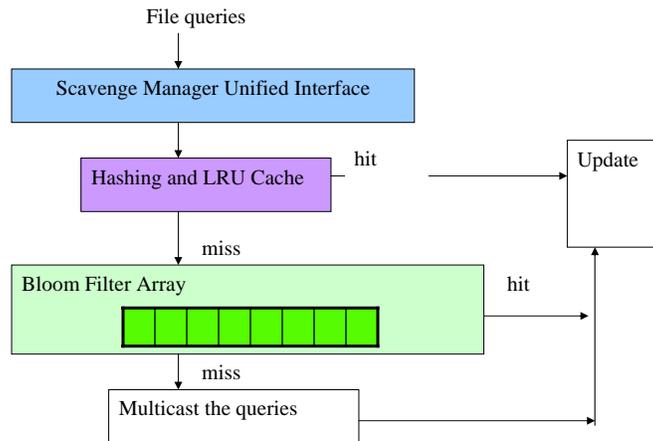
---

- A Bloom filter is a fast and efficient method for representing a set  $A = \{a_1, a_2, \dots, a_n\}$  of  $n$  elements to support membership queries.



## Metadata Management using Bloom Filters

---



15

- 
- Introduction
  - Highly Available Metadata Management
  - Metadata Management using Bloom Filters
  - ✓ **Conclusions and Future Work**

## Conclusions

---

- **Investigate Availability issues Scavenged Storage Systems, propose 4 solutions for maintaining multiple metadata managers:**
  - P2P
  - Active/hot-standby
  - Partitioning
  - Leader/Helper
- **Speed up the metadata searching:**
  - Bloom Filters

A Highly Available Cluster Storage System Using Scavenging, HAPCW'2004 17

## Future Work

---

- **Comparing the proposed 4 schemes**
- **Scalability**
- **Metadata Cache**

A Highly Available Cluster Storage System Using Scavenging, HAPCW'2004 18

## Acknowledgements

---

- Research Office and Center for Manufacturing Research, Tennessee Technological University
- Ralph E. Powe Junior Faculty Enhancement Award by Oak Ridge Associated Universities (ORAU).
- Mathematics, Information and Computational Sciences Office, Office of Advanced Scientific Computing Research, Office of Science, U. S. Department of Energy.

## Questions and Comments?

---



---

## *A Highly Available Cluster Storage System Using Scavenging*

High Availability and Performance Computing Workshop (HAPCW'04), held with  
2004 Los Alamos Computer Science Institute Symposium (LACSI'2004)

**Xubin (Ben) He, Li Ou**

{hexb, lou21}@tntech.edu

**Storage Technology & Architecture  
Research(STAR) Lab**  
Electrical and Computer Engineering  
Department



**Stephen Scott, Christian Engelmann**

{scottsl, engelmann}@ornl.gov

Computer Science and Mathematics Division

