

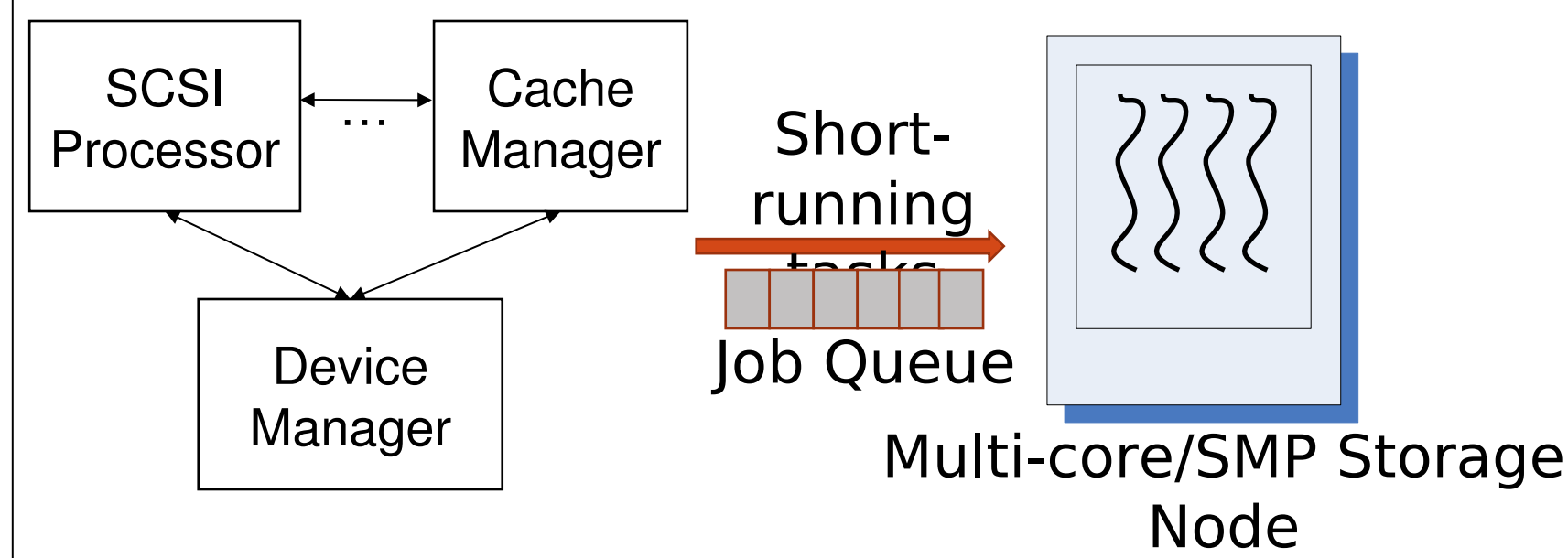
Enhancing Availability and Performance in Storage Systems and Services

Sangeetha Seshadri, Sankaran Sivathanu and Ling Liu
Distributed Data Intensive Systems Lab | Georgia Institute of Technology

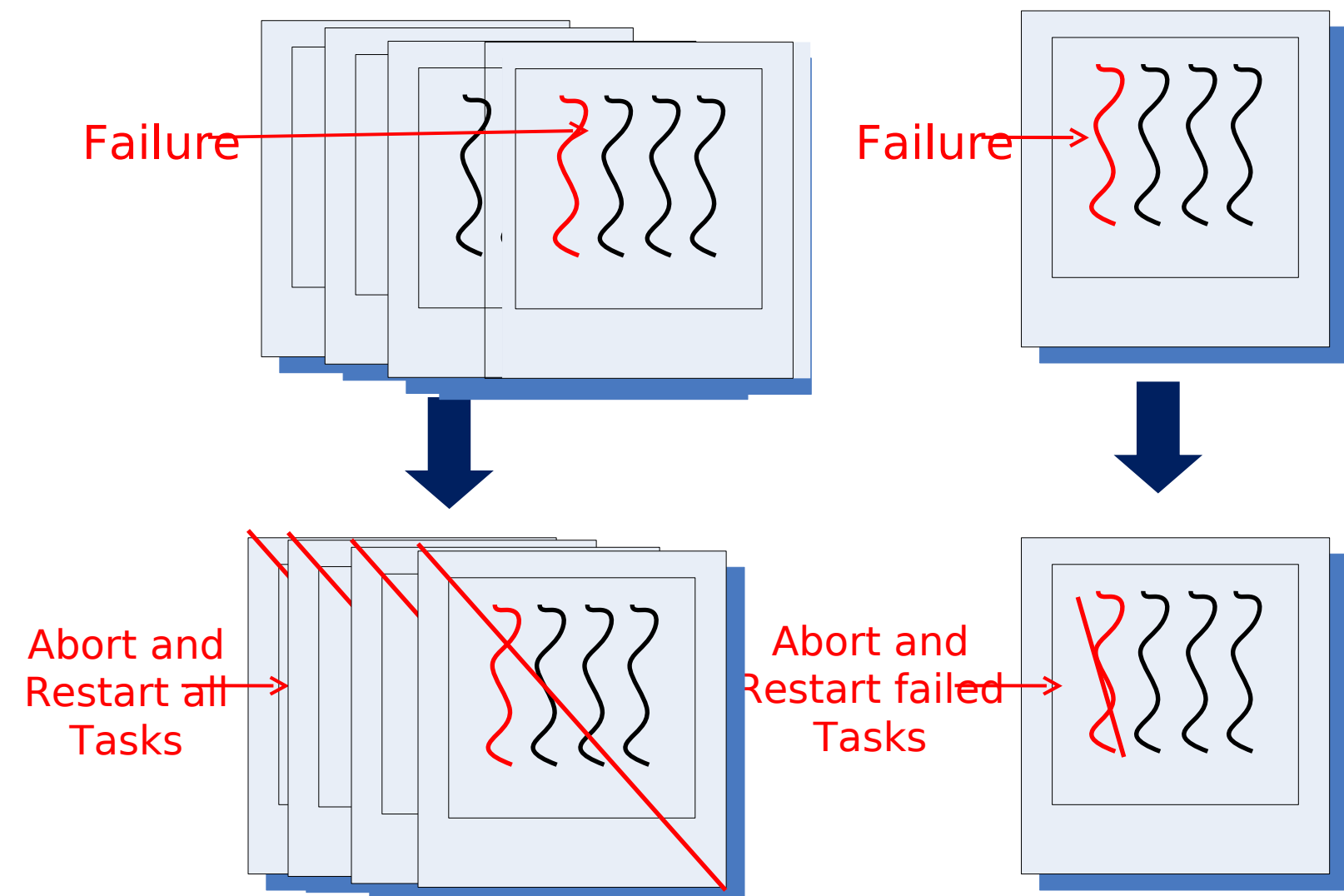
Recovery-Conscious Scheduling

Overview:

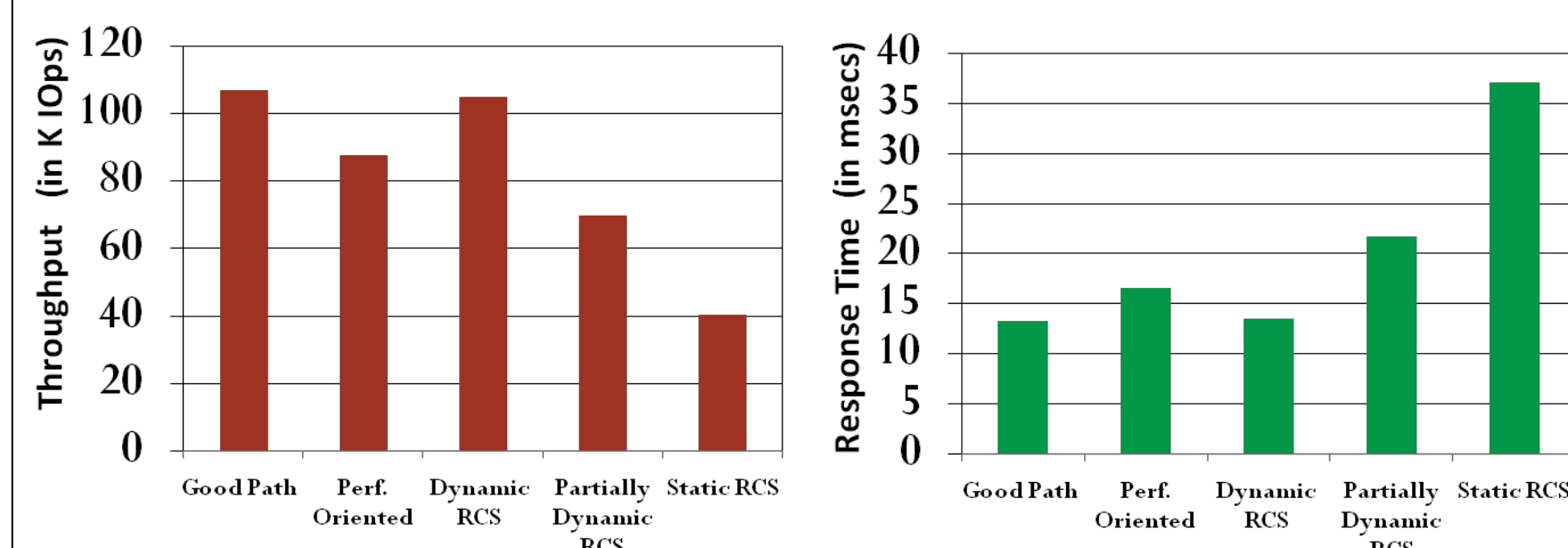
- Storage systems - foundations of modern data centers.
 - Extremely high availability expectation.
 - Issues:
 - Complex, legacy architectures.
 - Concurrent development, quality assurance processes.
 - Large scale installations – 1000s of components.
 - Failures are the norm, not exception.
- Goal: Enhancing availability in large scale storage systems and services.



System-Level VS. Task-Level Recovery



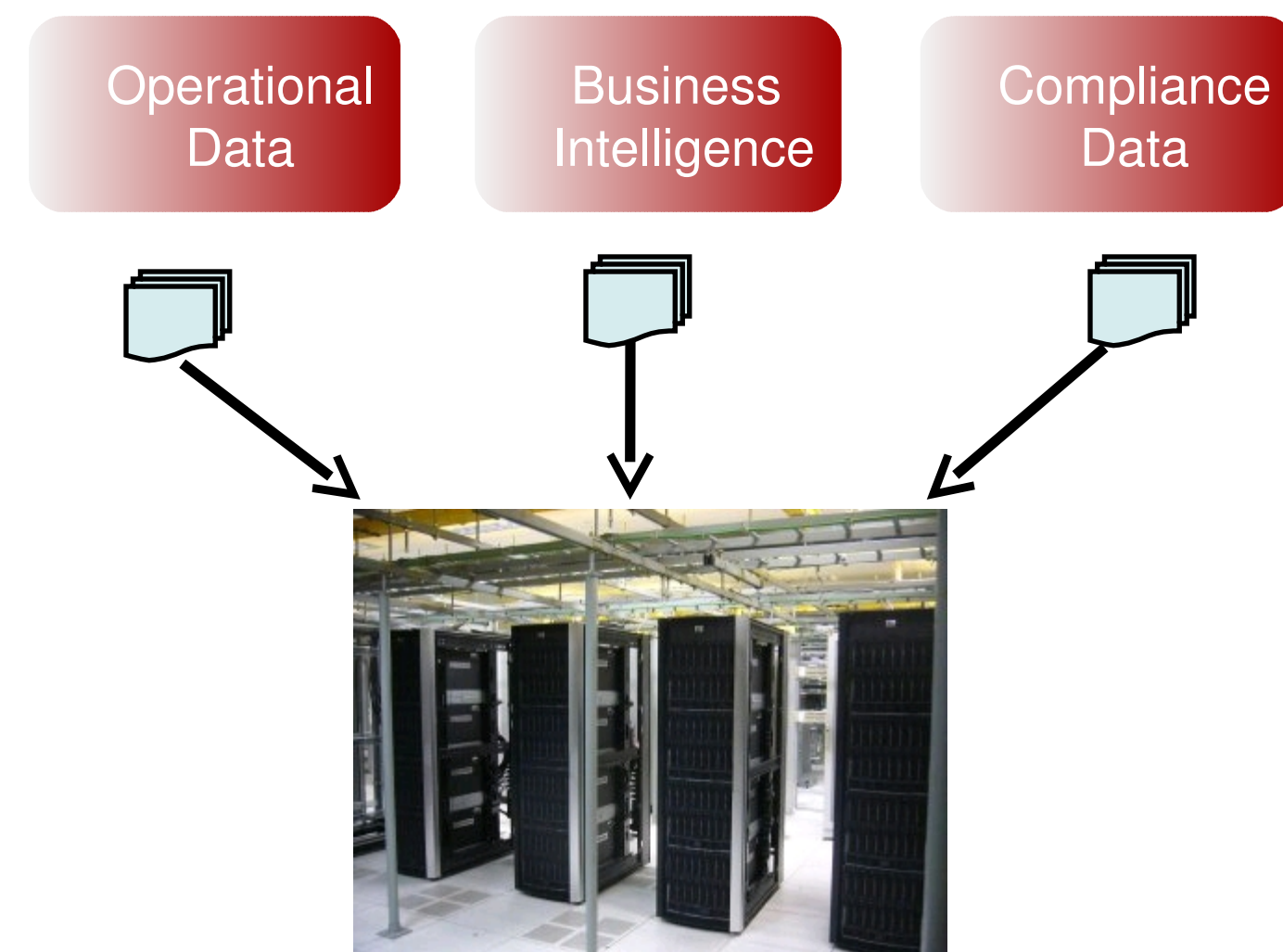
Experimental Results



Publications

[1] S. Seshadri, L. Chiu, C. Constantinescu, S. Balachandran, C. Dickey, L. Liu, and P. Muench. Enhancing storage system availability on multi-core architectures using recovery conscious scheduling. In USENIX FAST, 2008.

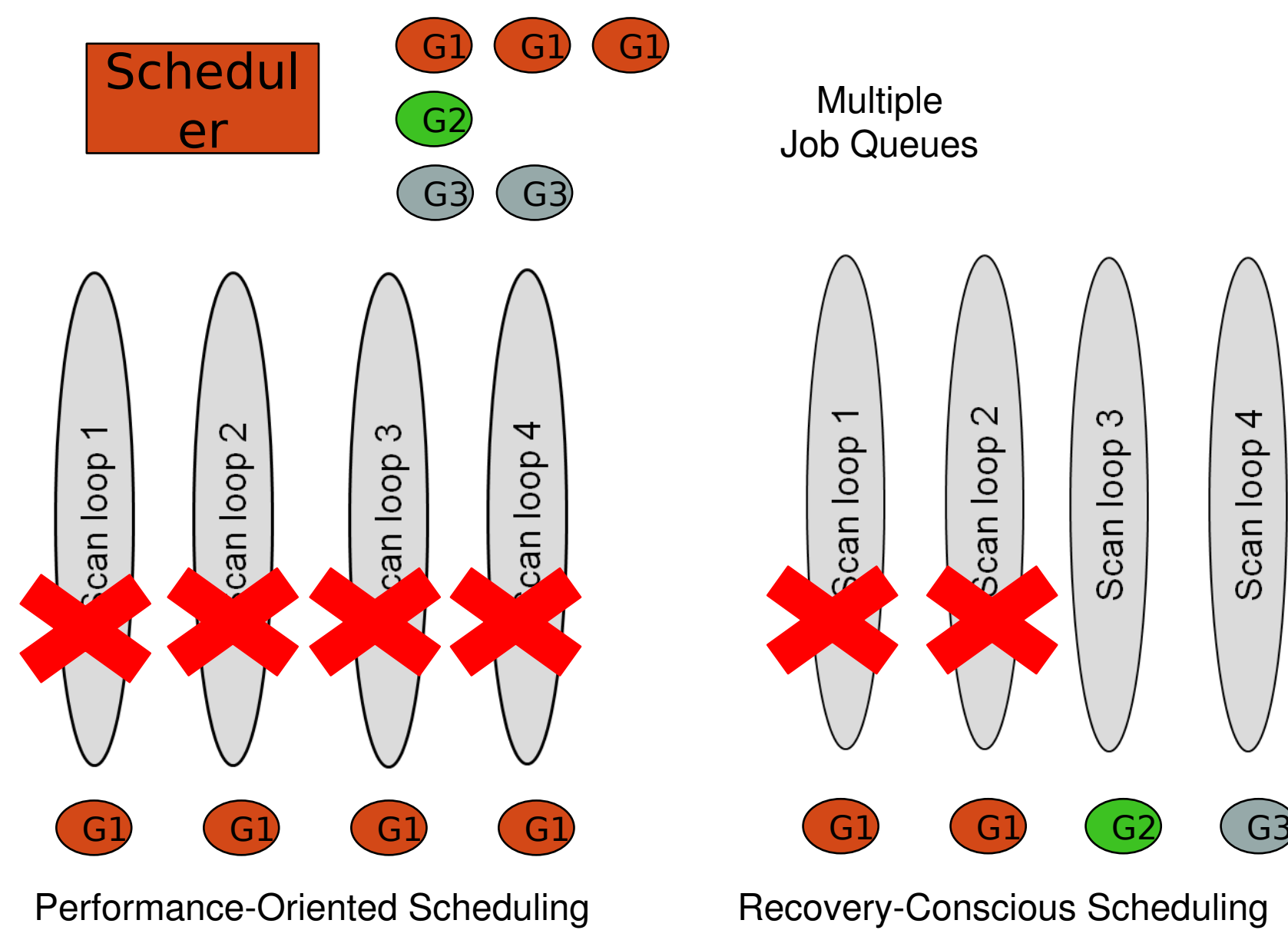
Contact: sangeeta@cc.gatech.edu, lingliu@cc.gatech.edu



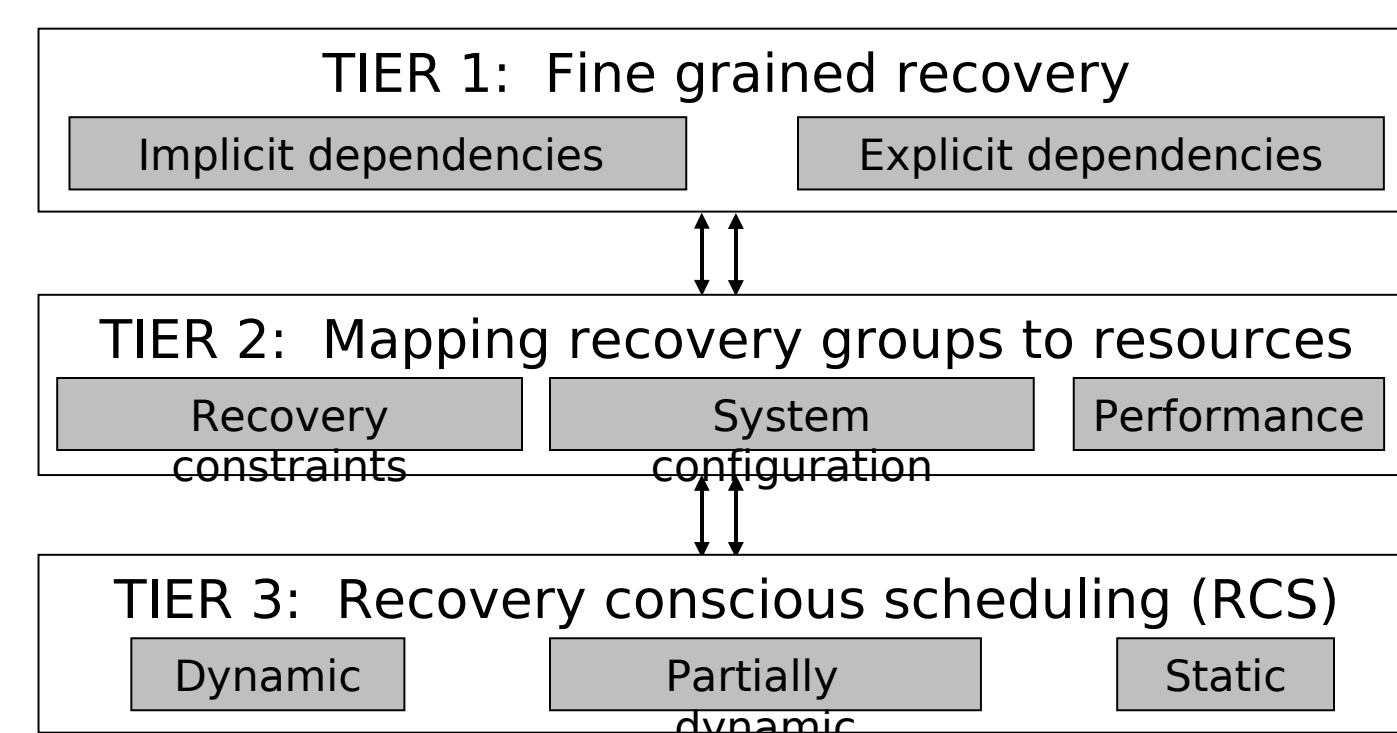
Recovery-conscious scheduling (RCS)

- Goal:
- Reduce the ripple effect of software failures.
 - Improve the availability of the system.
- Approach:
- Exploring trade-offs between recovery time and system performance.
 - Enforcing some serializability of recovery-dependent tasks.
 - Bounding resource consumption of the recovery process.

Performance-Oriented VS. Recovery-Conscious



Three-tier Recovery Conscious Framework



Pro-Active Disks

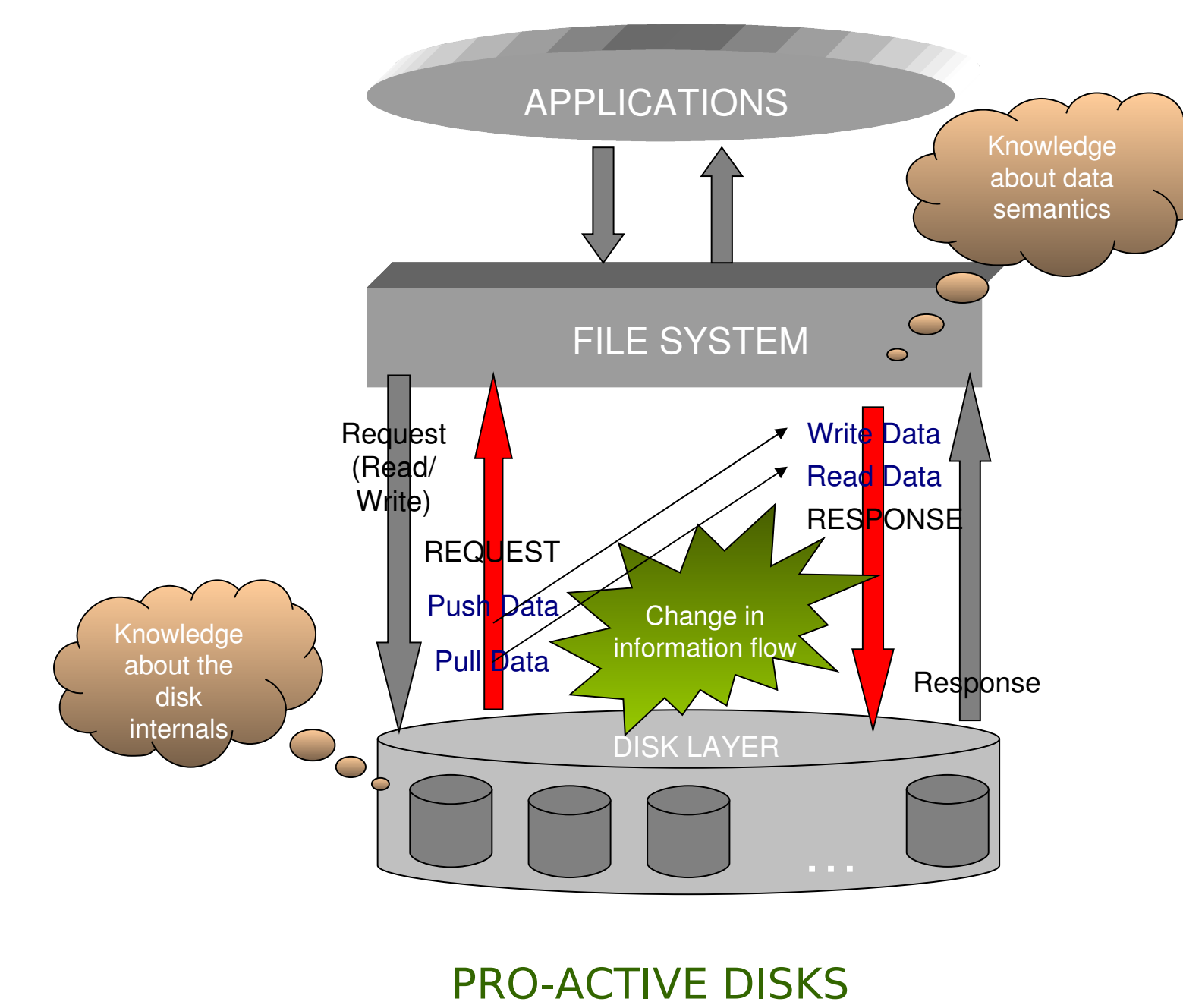
Problem:

- File system can distinguish data blocks, inode blocks and blocks corresponding to a file.
- Disk is aware of head position, track boundaries, overall disk load, etc.
- Interface is less expressive: file system issues read_block() / write_block() requests and disk services it.
- No knowledge of data placement, access method at file system level.
- No knowledge of data semantics, block correlation, etc. at disk level.

Result : Information gap and limited functionality!!

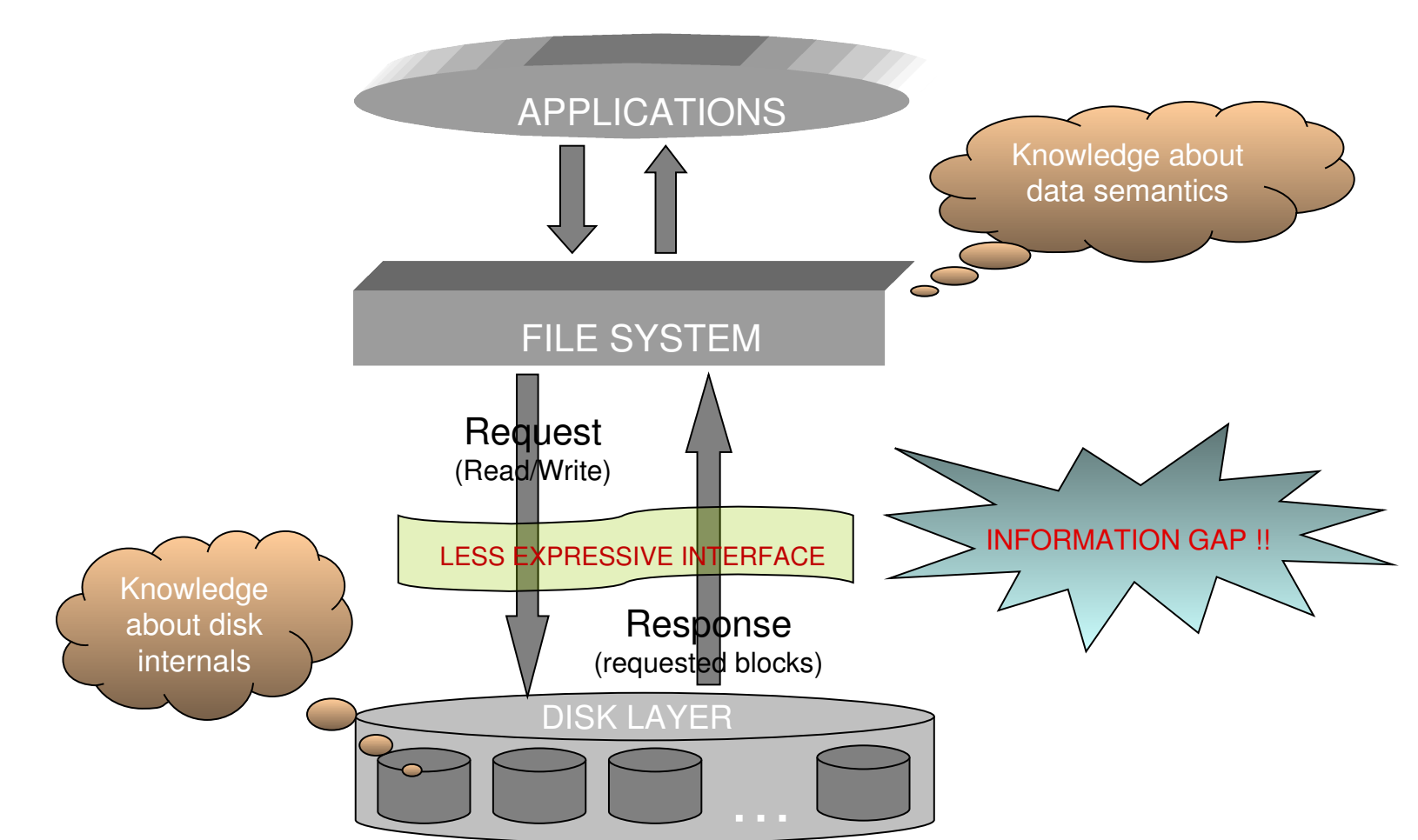
Benefits of Reduced Information Gap

- Increased read/write performance, if file system knows track boundaries and issues track-aligned reads.
- Decreased disk latency if file system schedules requests with knowledge of the current head position.
- Disk idle time utilized by file system to flush blocks.
- Ability to offer different reliability levels by controlling the degree of replication.



Pro-Active Disks

- Pro-Active Disk decides on when the file system should read/write data
- Uses the knowledge of disk internals to
 - Capture the opportunity, the file system can utilize it to read/write data.
 - Make decision if captured opportunity is beneficial.
 - Signals file system to read/write data if it decides the opportunity is beneficial.
- Therefore file system utilizes the benefits of knowing the disk internals – reduced information gap !



TRADITIONAL DISKS

Possible Solutions:

- More expressive interface:
- File system can have separate interface to get the information about disk and also to tell the disk the semantic information of data.
 - Why not ? File system and the disk becomes more inter-dependent. One has to evolve to changes in the other.
- Reduce the disk-file system boundary:
- Make file system understand the disk characteristics through inference from the behavior of probing workloads.
 - Why not ? Complex code needs to be implemented in the file system. Need separate code for each type of disk.
- Our solution: Make disks do it !

Examples

- A Pro-Active Disk can capture the idle time in the disk and ask the file system write some of its dirty buffers in this idle bandwidth.
- Disk-initiated pre-fetching.
 - Track buffers hold a set of recently accessed tracks.
 - Pro-Active Disk can measure the significance of the track buffer for the current workload.
 - If found to be significant, the to-be-reclaimed track can be sent to the file system cache instead of just discarding.
- Co-operative Caching can be implemented completely at the disk level without the client knowing it. No complex book-keeping needed in client.

Further Information

Contact: sankaran@cc.gatech.edu, lingliu@cc.gatech.edu

