Journaling vs. Soft Updates

Chris Stein

Harvard University

June 21, 2000

joint work with Margo Seltzer[†], Greg Ganger[*], Kirk McKusick [‡], Keith Smith[†], and Craig Soules[*]

[†] Harvard University, [*]Carnegie Mellon University,

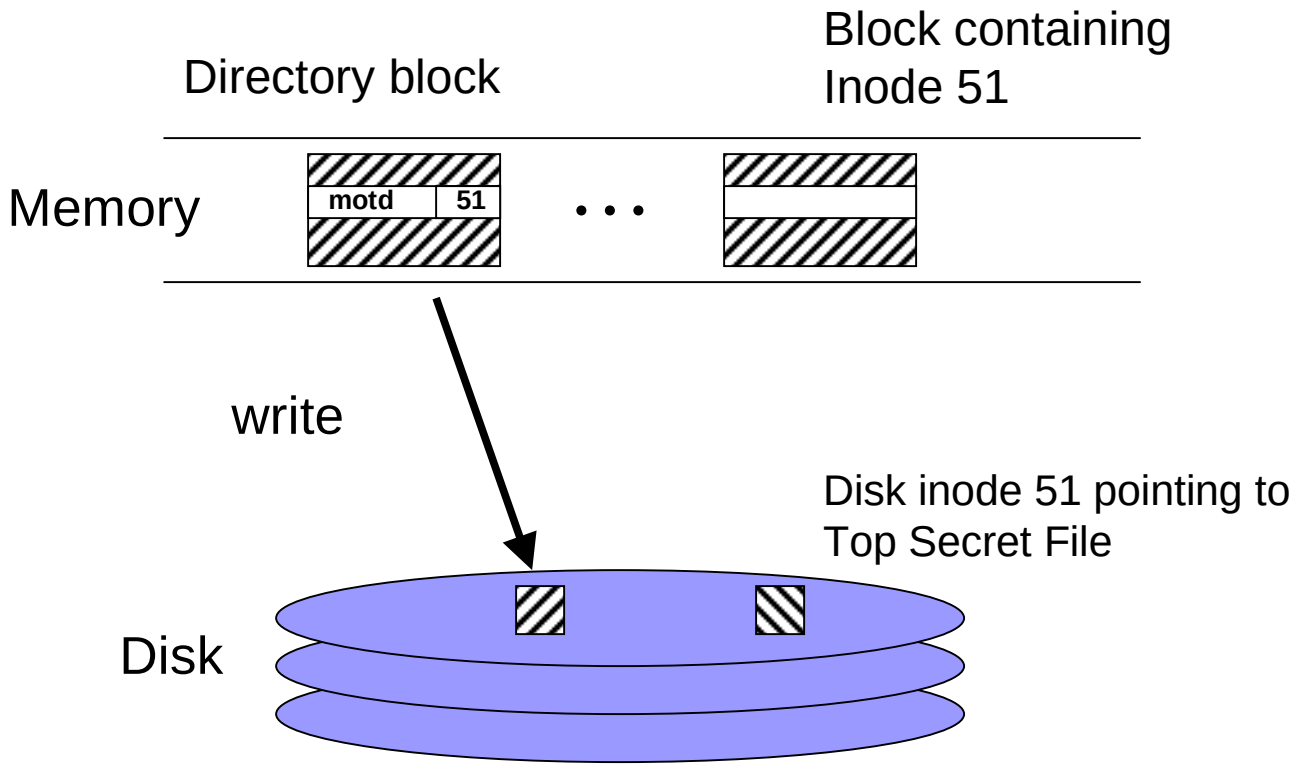
[‡]Author and Consultant

Talk Outline

- The Problem:
  - Meta-data consistency in file systems
- Two solutions:
  - Journaling and Soft Updates
- Evaluation
- Conclusions

Meta-Data Update Problem

- The file system meta-data contains inodes, directory blocks, and allocation bitmaps with interdependencies that must be cared for during disk updates.



Approaches to Meta-Data Management

- Synchronous Writes
  - FFS
- Ordered Writes
  - Soft Updates
- Logged Writes
  - Journaling

Properties of Meta-Data Ops

- Integrity:
  - The file system is always recoverable.
- Durability:
  - Updates are persistent once the call returns.
- Atomicity:
  - No partial meta-data operations are visible after recovery.

Soft Updates Overview

- Implementation:
  - Delayed meta-data writes.
  - Kernel maintains dependency information and uses it to order writes.
- Properties:
  - Meta-data operations are not durable or atomic.
  - Looser guarantees than FFS about when updates will reach disk.
  - No recovery necessary after a crash.

Journaling Overview

- Implementation:
  - Log logical meta-data operations.
  - Write meta-data in-place asynchronously.
  - Write-ahead logging (WAL) protocol guarantees recoverability.
- Properties:
  - Log is scanned for recovery.
  - Meta-data operations are atomic.
  - Durability can be toggled on/off.

Feature Comparison

	Integrity	Durability	Atomicity	Recovery
FFS	X	X		fsck disk scan (minutes)
Sync Journaling	X	X	X	log-based (seconds)
Async Journaling	X		X	log-based (seconds)
Soft Updates	X			immediate
FFS-async				may be impossible

Experimental Setup

- Software:
  - Modified FreeBSD kernel. Taken from the current tree on Jan. 26th, 2000.
  - 2 journaling file system implementations (LFFS-WAFS, LFFS-file).
- Hardware:
  - 500 MHz Xeon Pentium III
  - 512 MB RAM
  - 3 x 9GB 10,000 RPM Seagate Cheetahs

Microbenchmarks

- Create, Write, Read, Delete.
- Results
  - Read/write performance identical for all systems.
  - All async systems have similar create throughput.
  - Soft Updates has great delete performance due to its ability to background work.

Macrobenchmarks

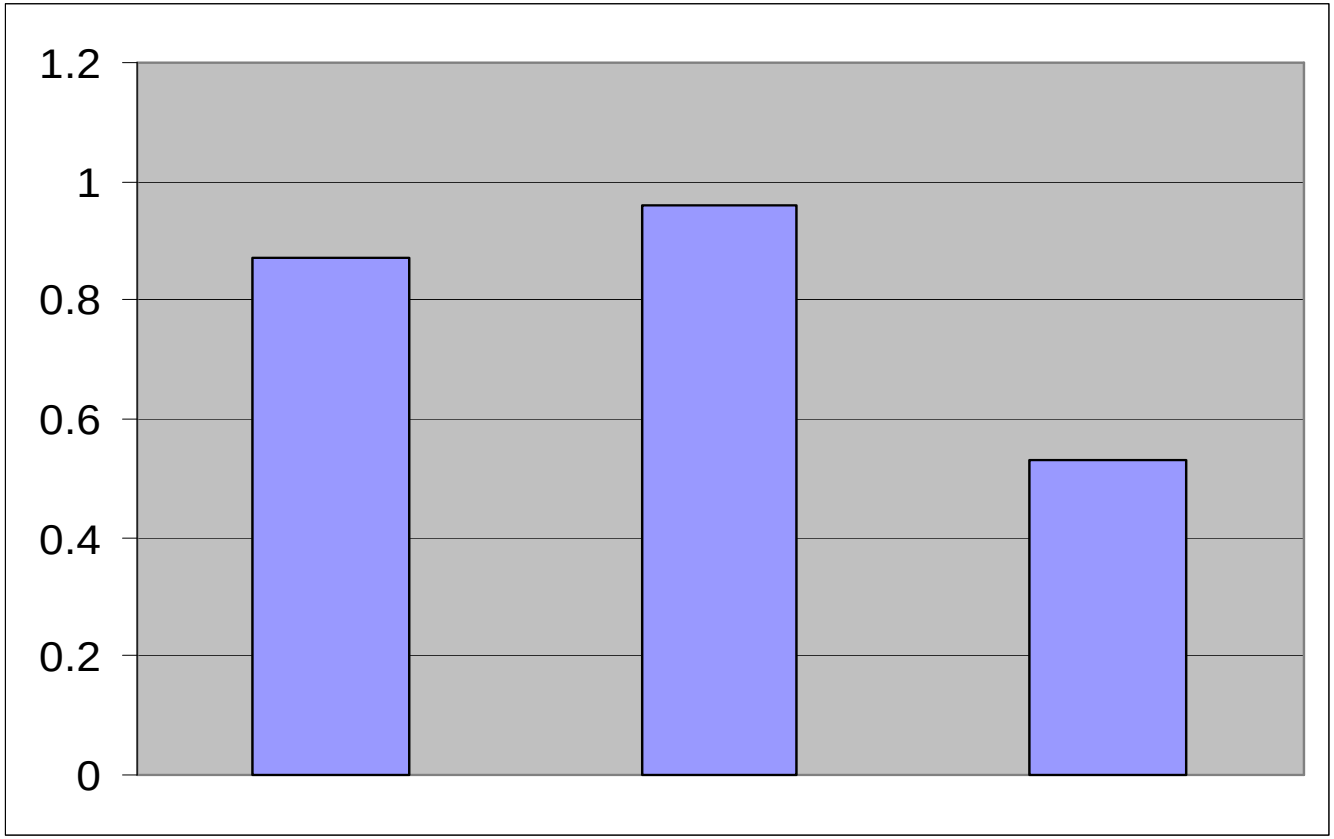
- SSH-build
  - Unpacks, configures, and builds ssh.
- NetNews
  - Simulates the work of a news server.
- SDET
  - Emulates user interactive software development workload.
- PostMark
  - Designed to model the workload seen by ISPs under heavy load. Combination of e-mail, news, and e-commerce transactions.

NetNews

- Simulates the work of a news server.
- Tremendous load, both in terms of the amount of data and the number of meta-data operations.

NetNews: Results

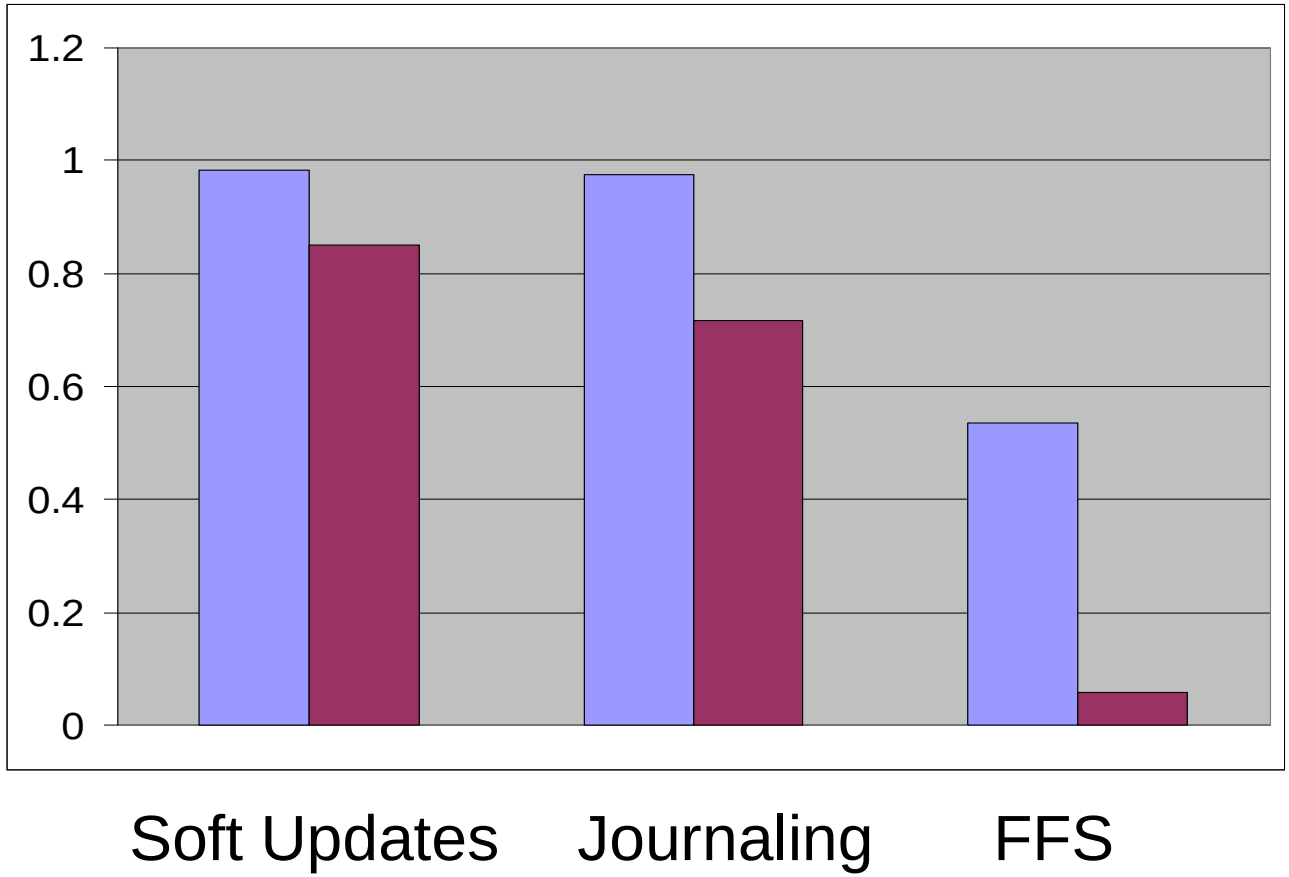
Throughput Relative to FFS-async



PostMark

- Designed to model the workload seen by ISPs under heavy load.
- Simulates a combination of e-mail, news, and e-commerce transactions.
- Different results for small and large file sets.

PostMark: Results



Conclusions

- Durability is expensive, integrity need not be.
- Configuration changes can have a significant impact on performance, with no change in features.