CS 739 – WiscAFS

Deep Jiten Machchhar
Varun Kaundinya
Vishnu Ramadas

Introduction

- Build AFS-like filesystem using FUSE – used unreliableFS
- Communicate through gRPC with server
- Maintain a client object linked with FUSEInit
  - Client makes calls to server using gRPC stub
- Client and server implemented in C++
- Close to open consistency

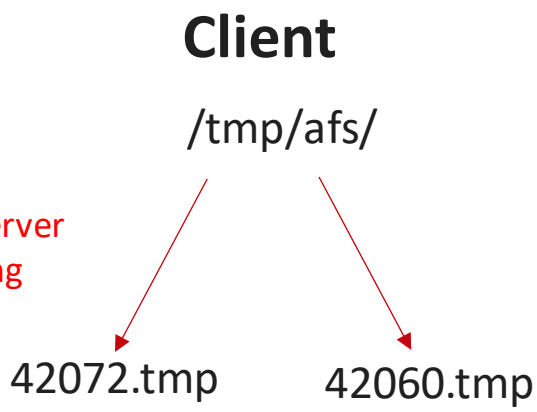
Design

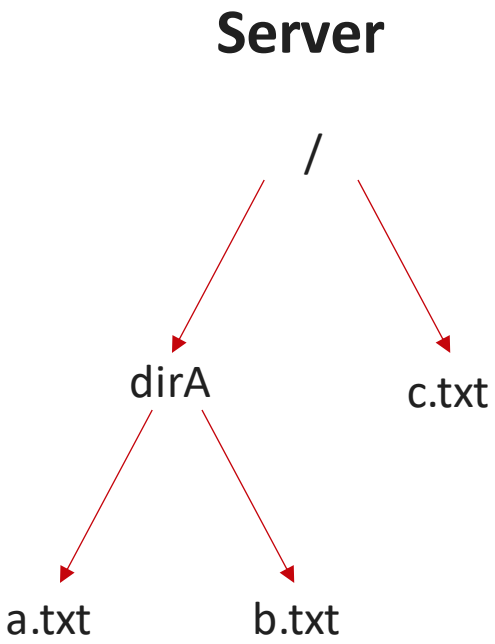
- Read whole file during open
- Maintain mapping of paths and cache files locally
- Server reflects changes only after a close – last writer wins
  - Every open sends a request to the server for a fresh copy of the file
  - Every write to the file sets a dirty bit in the metadata
  - Flush writes file to server only if it is dirty

Design

Path+Name	Dirty	Inode	Attributes+Permissions
/c.txt	0	42060	RDONLY, 0644
/dirA/a.txt	1	42072	RDWR, 0777

Open("/c.txt", O_RDONLY, 0644)
Read("/c.txt")
Open("/dirA/a.txt", O_RDWR, 0777)
Write("/dirA/a.txt") <- Set dirty
Close("/dirA/a.txt") <- Flush file to server
Close("/c.txt") <- Not dirty, do nothing





Design

- Ops implemented

File basic	Directory related	Others
Open/Create	Mkdir	Lstat (getattr)
Close	rmdir	statfs
Read	Readdir	Chmod/chown
Write	Opendir	access
Unlink		Truncate
Rename		

Design - Crashes

- Client crash
  - Cache meta-data updated only after file is written
  - New re-open brings fresh data from server
  - On restart, reset cache meta-data
- Server crash
- During close, atomic rename after writing to a temp file
  - Client time-out if no response within set seconds
    - Application responsibility to re-try

Results - Filebench

- file create.mkv

Results - Filebench

Benchmark	Operations	Operations/sec	Rd/Wr	Bandwidth (mbs/sec)	Time per Operation (ms/op)
create	269	26.897	0/27	26.8	37.093
createrand	635	63.493	0/58	28.8	15.719
rread	0	0	0	0	0
rwritedsync	0	0	0	0	0
seqread	0	0	0	0	0
seqwrite	325	32.497	0/32	32.4	29.608
writefsync	23575	2357.301	0/2355	18.4	0.422
mongo	248	24.797	3/4	0.1	40.042

Results - Filebench

Benchmark	Number of files	Operations	Operations/sec	Rd/Wr	Bandwidth (mbs/sec)	Time per Operation (ms/op)
createfiles	4000	859	85.892	0/29	0	11.624
delete	500	66	65.991	0/0	0	124.941
statfile	2000	150440	150432.988	0/0	0	0.099
fileserver	2000	381	38.096	3/9	1.3	1095.754
varmail	200	12883	214.697	33/33	0.8	74.288
webserver	2000	1754	175.384	57/4	1.2	240.7

Results

- make xv6-2.mkv
- make leveldb.mkv

Consistency

Scenario	Description
Default	Last writer wins as per expectation
Delete	Last writer wins despite intermediate deletes
Rename	Renamed file and the original file both exists if last writer was not the rename
Chmod	Permission on server should match the permission locally cached files
Server crash	Server crash, doesn't affect the data

Consistency

- Consistency.mp4
- serverCrashWrite.mp4

Durability

- Tried adding a queue to delay write operatios till next fsync
- Reading before fsync returns stale data to the user

THANK YOU

Questions?