

File System Challenges in Consumer Electronics Products

정찬균 (Chan Gyun Jeong)

SW Platform Lab., Corporate R&D
LG Electronics, Inc.

2014/10/31

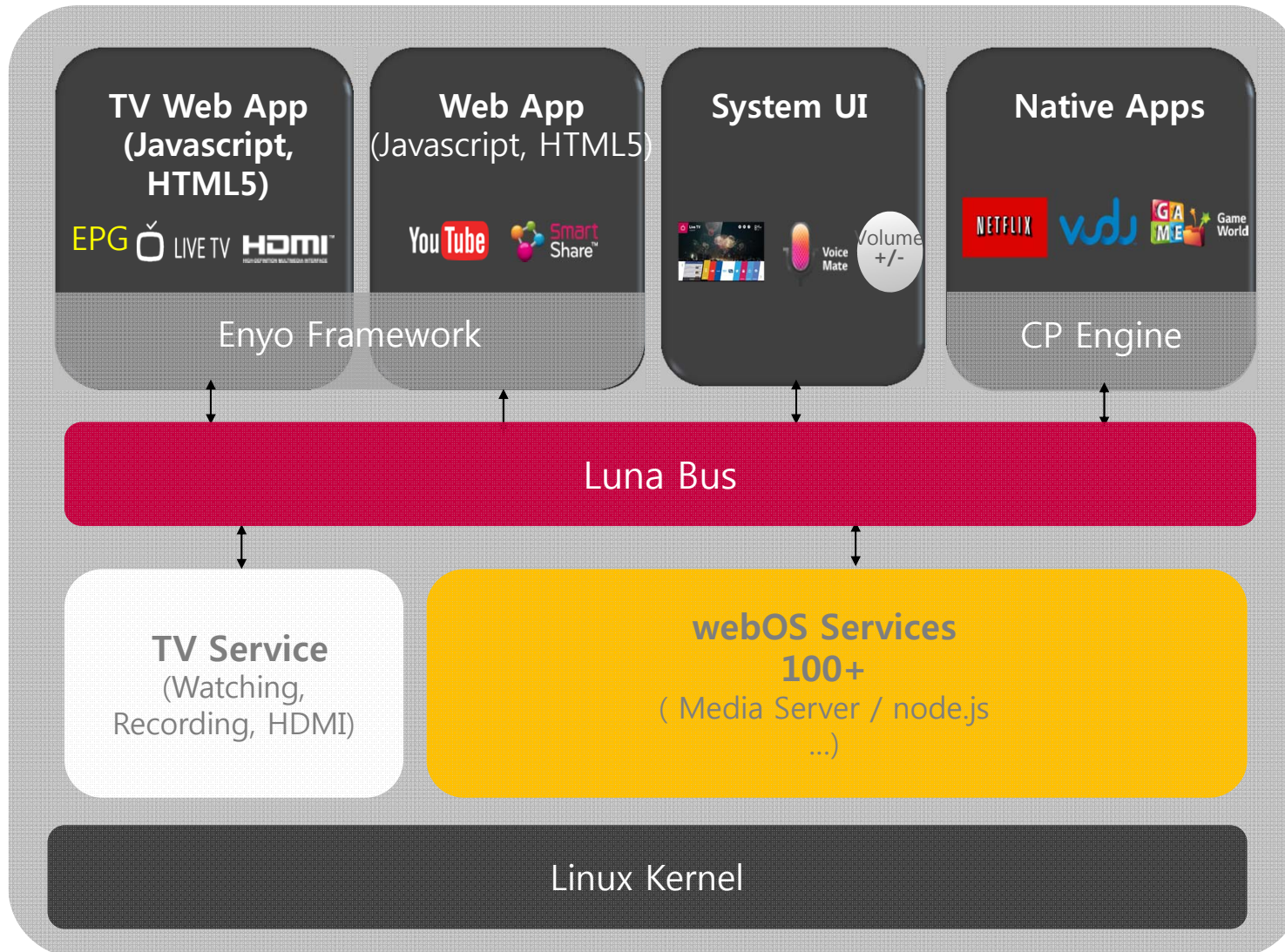


Contents

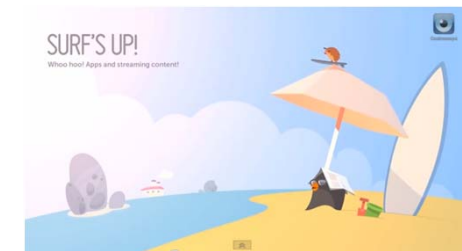
- LG webOS TV Overview
- File Systems in webOS TV
- Database in webOS TV
- Squashfs Challenges and Improvements
- File Truncation Performance
- FUSE Overhead in Android
- File System Patents Issue
- eMMC Lifetime Issue
- eMMC Performance Issues
- Expectations for NVRAM

LG webOS TV Overview

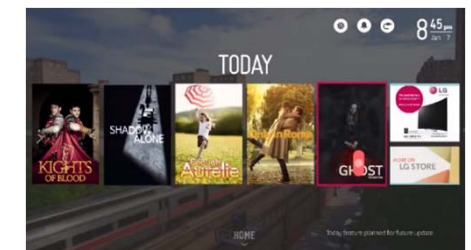
- Make TV **Simple** Again : 3S (Simple Switching/Connection/Discovery)



Simple Switching



Simple Connection



Simple Discovery

File Systems in webOS TV



- ext4 file system used for a partition requiring writable file feature in runtime
 - e.g. LG App Store partition

Root File System

TV Service

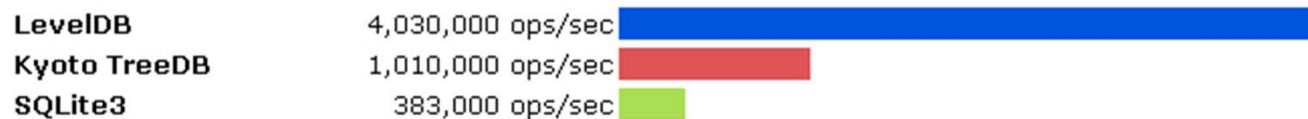
Fonts

Misc. Data

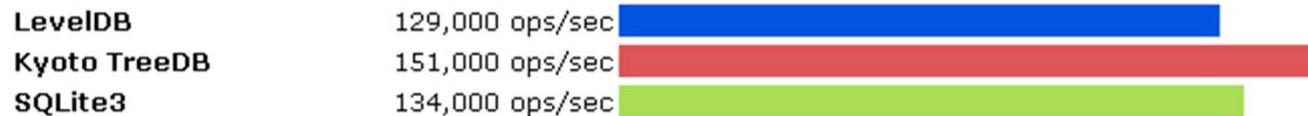
- Squashfs file system used for a partition not requiring writable file feature in runtime
 - e.g. rootfs, TV Service partitions
- Squashfs is a compressed read-only file system, and provides high performance with low overhead & size reduction

- DB8 Database Service
 - Fast and light Key-Value based DB service
 - Data stored as JSON(Java Script Object Notation) objects in collections
 - No SQL support
 - Using LevelDB as backend

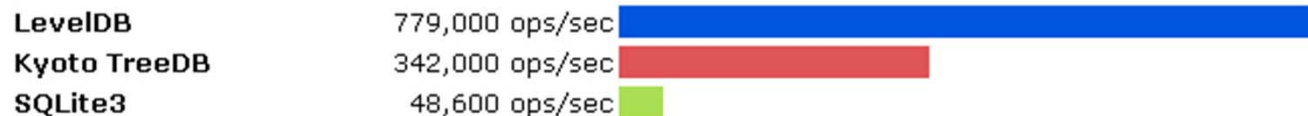
1) Sequential Reads



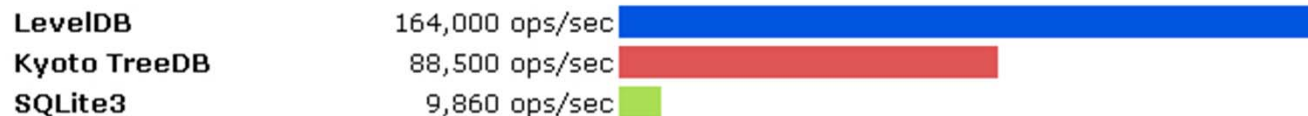
2) Random Reads



3) Sequential Writes

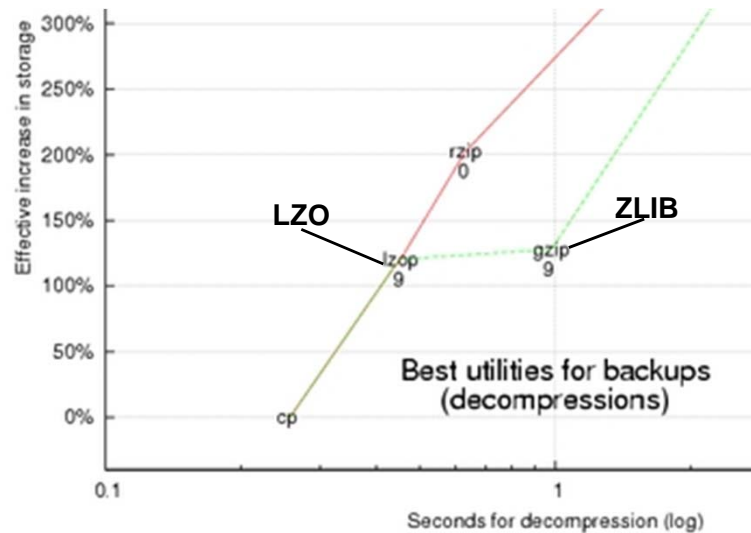


4) Random Writes



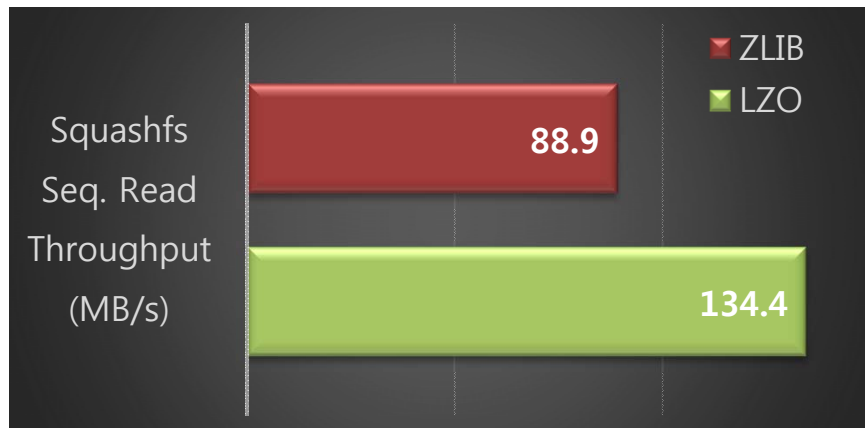
Squashfs Challenges and Improvements (1)

- Compression Algorithm Performance



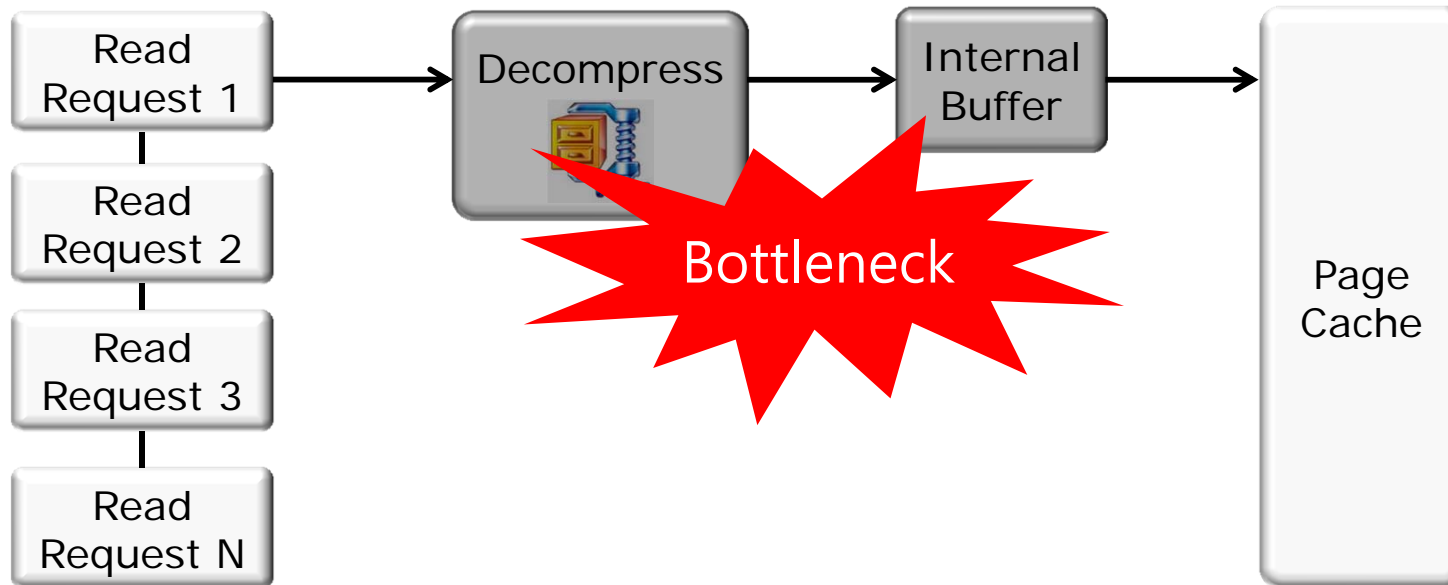
Source: <http://www.linuxjournal.com/node/8051/>

- High decompression performance needed for CE products rather than compression speed
- LZO (Lempel-Ziv-Oberhumer) outperforms ZLIB(aka. GZIP) in decompression performance
- Squashfs LZO support contributed to mainline Linux kernel by LG
- LG contributed support for new LZ4 compression to mainline as well
- LZ4 outperforms LZO when unaligned memory access is enabled in ARM
- Squashfs LZ4 ready to upstream into mainline



Squashfs Challenges and Improvements (2)

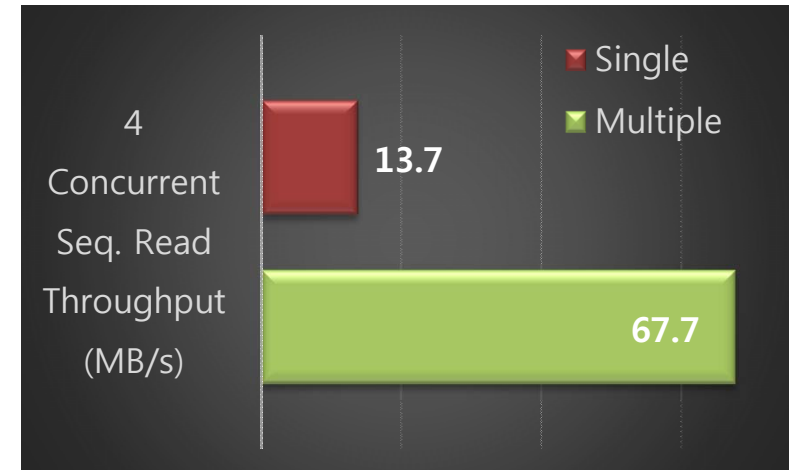
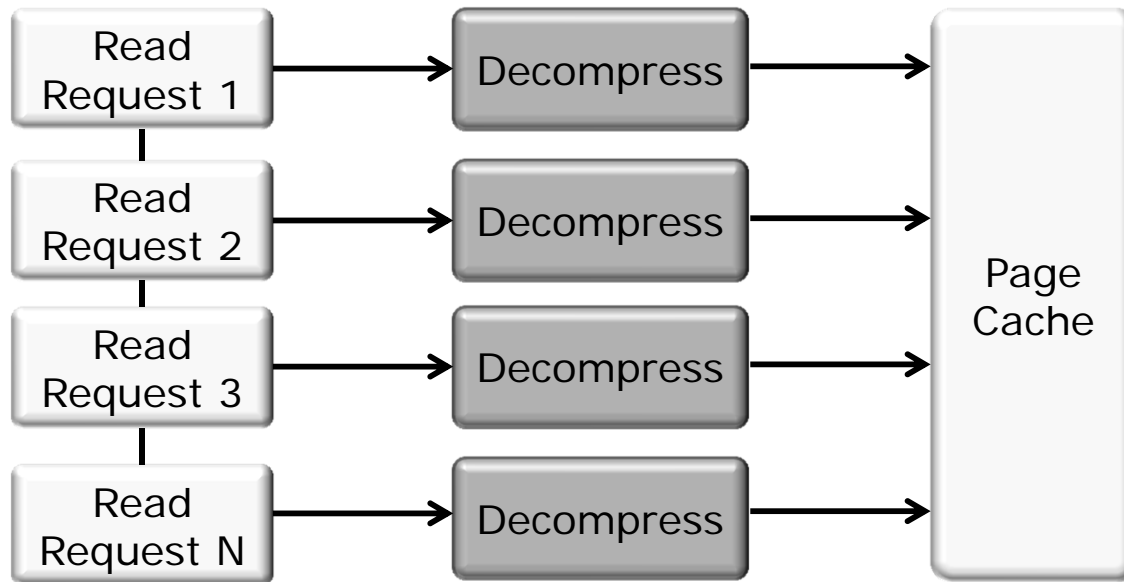
- Single Decompressor Problem



- Squashfs provides only one decompression stream buffer which incurs single-threaded decompression for concurrent requests
- Gives poor performance on parallel I/O workload of multicore systems
- Additional memory copy needed due to internal buffer

Squashfs Challenges and Improvements (3)

- Multiple Decompressor Solution



- Gives great performance for parallel I/O workloads on Multi-core systems
- Eliminates a memory copy by directly decompressing into page cache
- But requires more CPU and memory usage than single thread
- LG submitted a patch set and made a contribution to mainline kernel

File Truncation Performance

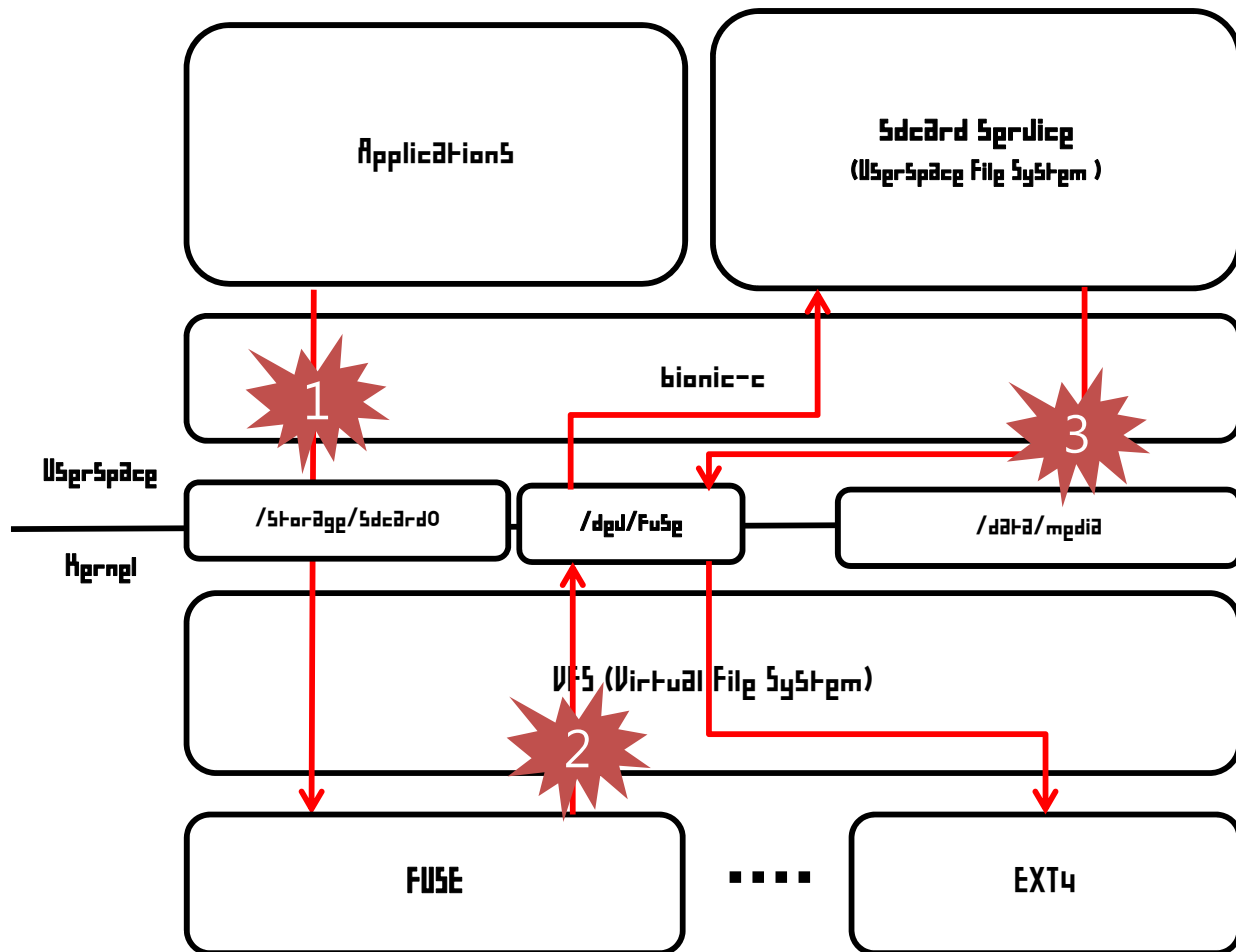
- How about performance if we need to cut data in the middle of a video file ?



- Using normal file truncation works, but it takes very long time depending on the video file size
- We modified some kernel file systems to help file truncation performance and split a large video into smaller files in DVR-enabled products
- Recently, fallocate(FALLOC_FL_COLLAPSE_RANGE) feature merged in mainline kernel

FUSE Overhead in Android (1)

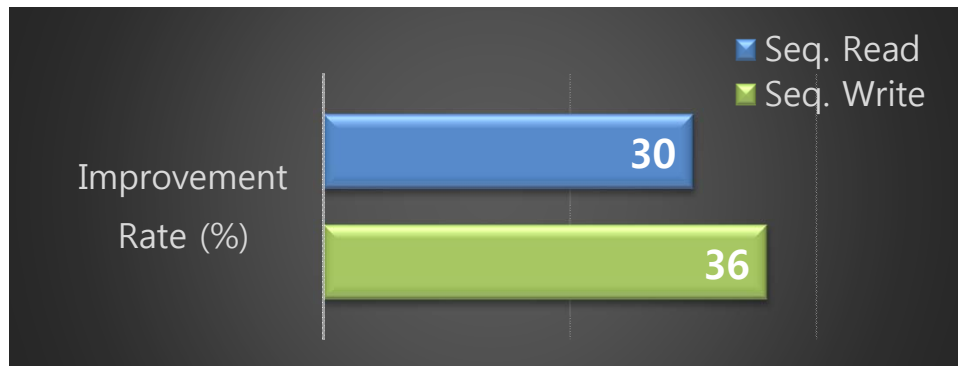
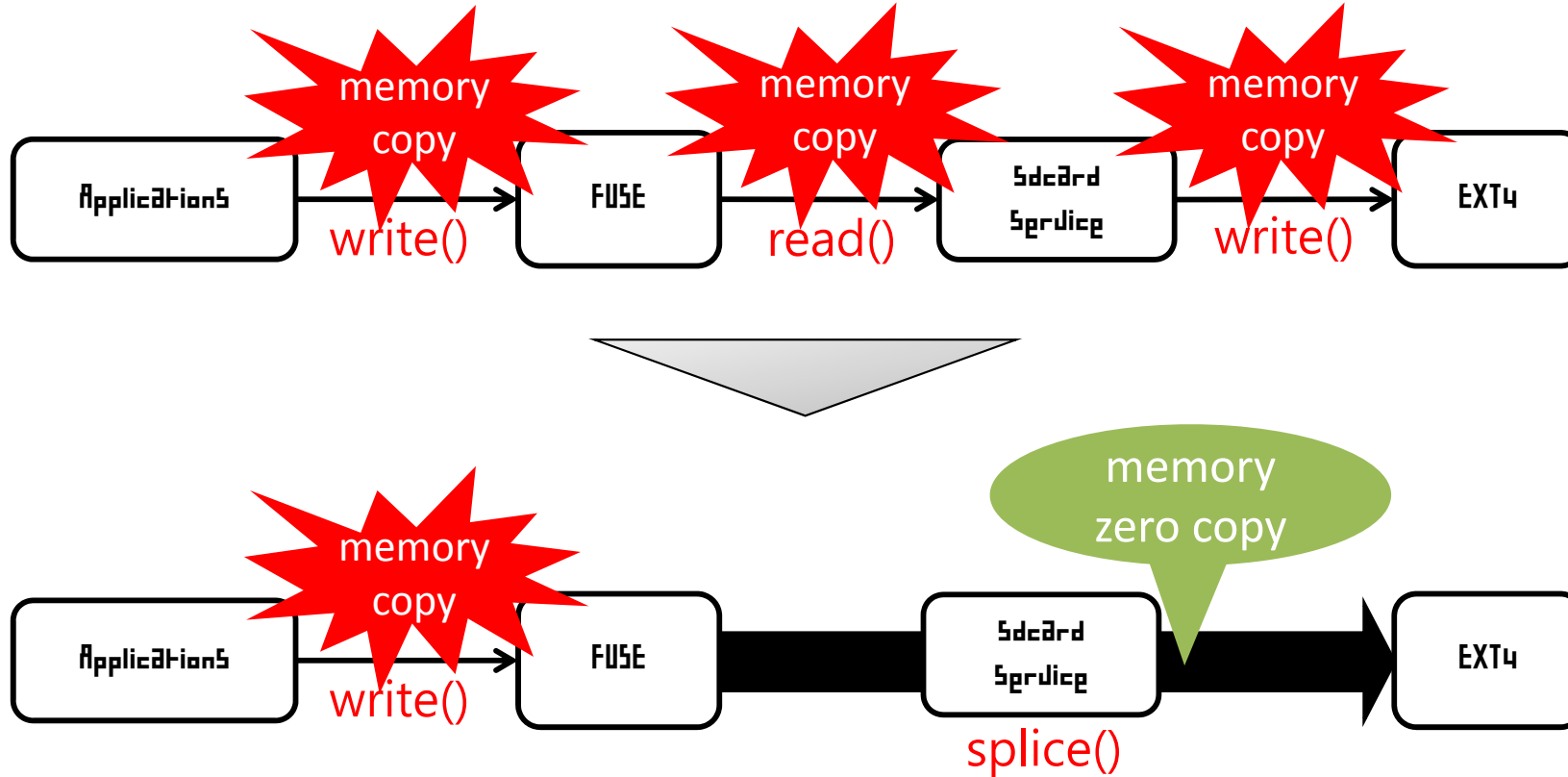
- FUSE (File System In Userspace)
 - Let applications create their own file systems without modifying kernel code
 - Used to emulate external storage and ensure security in Android



- ✓ Needs 3 memory copies for read() or write()
- ✓ Could be overhead in some systems

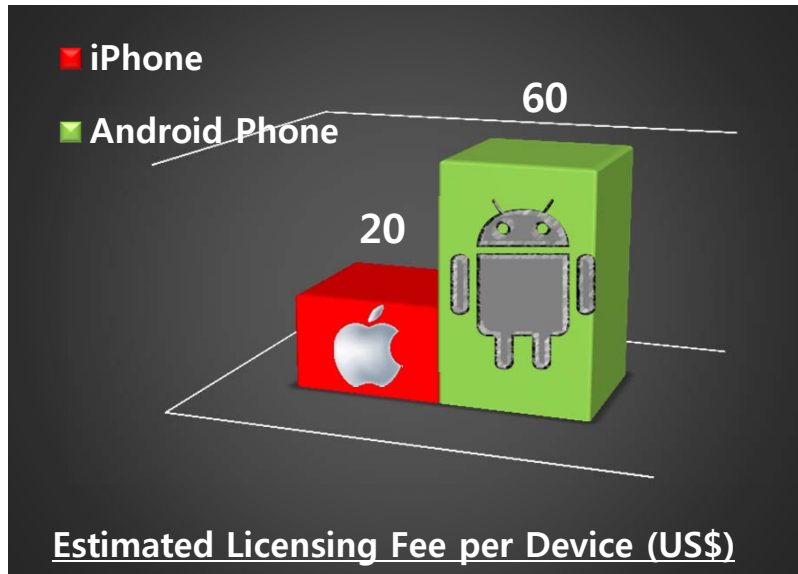
FUSE Overhead in Android (2)

- Removes unnecessary memory copies by splice in Linux kernel



File System Patents Issue

- BOM(Bill of Materials) cost is very important in CE products, BUT:



Source: digitaltrends.com, unwiredview.com

BUSINESS INSIDER

Microsoft Is Making An Astonishing \$2 Billion Per Year From Android Patent Royalties

 JAY YAROW
NOV. 6, 2013, 12:36 PM

Microsoft is generating \$2 billion per year in revenue from Android patent royalties, says Nomura analyst Rick Sherlund in a new note on the company.

He estimates that the Android revenue has a 95% margin, so it's pretty much all profit.

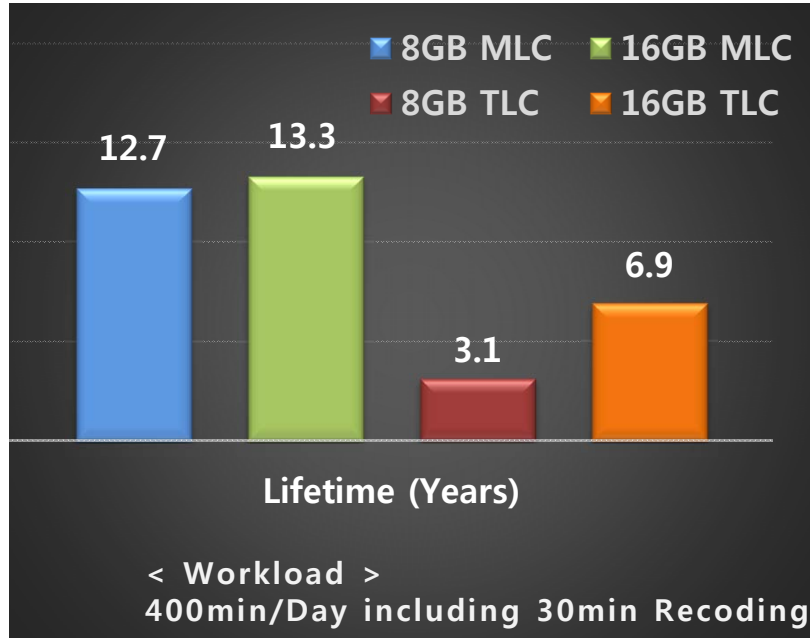
This money, says Sherlund, helps Microsoft hide the fact that its mobile and Xbox groups are burning serious cash.



- Even patents cost for file systems matters a lot !
- SDXC includes M\$'s exFAT file system as a mandatory feature
- Eventually adopting open standard file systems will benefits manufacturers and end users

eMMC Lifetime Issue

- TV lifetime issue when using DVR(Digital Video Recording) function

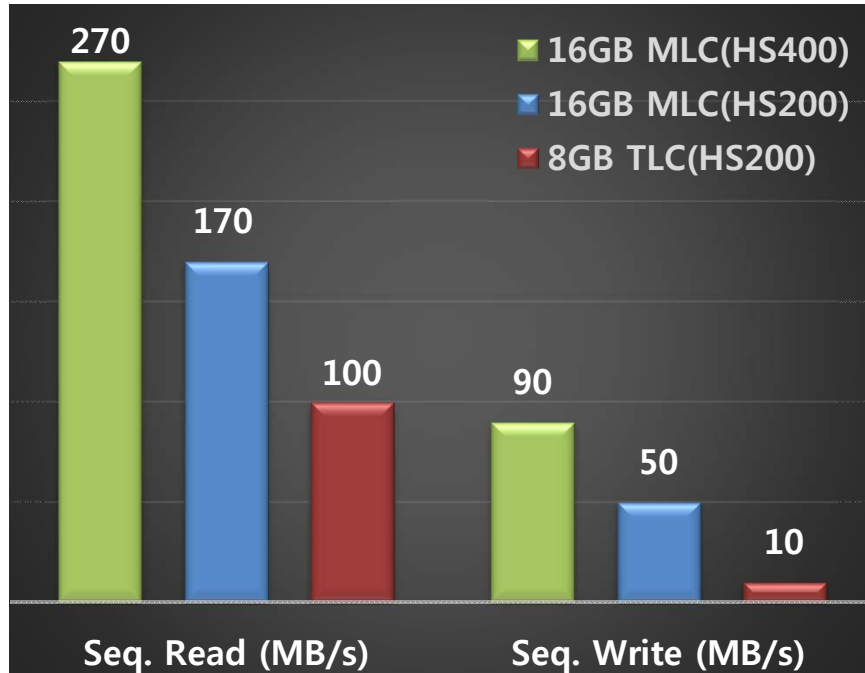


- Huge amount of writing data incurred by DVR enabled workload
- Even Timeshift (aka. Live Playback) feature spec out for eMMC
 - ✓ e.g. 5.7 GB for daily workload with Full HD video (19.39 Mbps in Korea)
- How to optimize WAF(Write Amplification Factor) in file system and block device driver layer ?
- How to improve eMMC lifetime in FTL ?



eMMC Performance Issue

- MLC vs TLC on Performance



- Not much performance issues for MLC-type eMMC flash memory
- But performance lacks for TLC-type eMMC in some workloads
- Any chances to improve performance of TLC-type eMMC ?

Expectations for NVRAM

- Workload aware or File System aware FTL
 - There are various workloads in smart phone and smart TV
 - How about FTL customization for per-partition workload ?
 - If FTL could handle I/O based on workload characteristic per-partition ?
 - ✓ DVR partition : 4MB large sequential I/O and write priority
 - ✓ Database partition : 1 ~ 4KB small random I/O
- Byte-addressable Persistent Memory
 - Cost Innovation
 - Mass Producibility
 - Performance



Q&A