Your File System

OpenAFS Linux Performance Improvements in 1.5/1.6

Simon Wilkinson
Your File System Inc.
20 April 2010

Memcache vs Diskcache

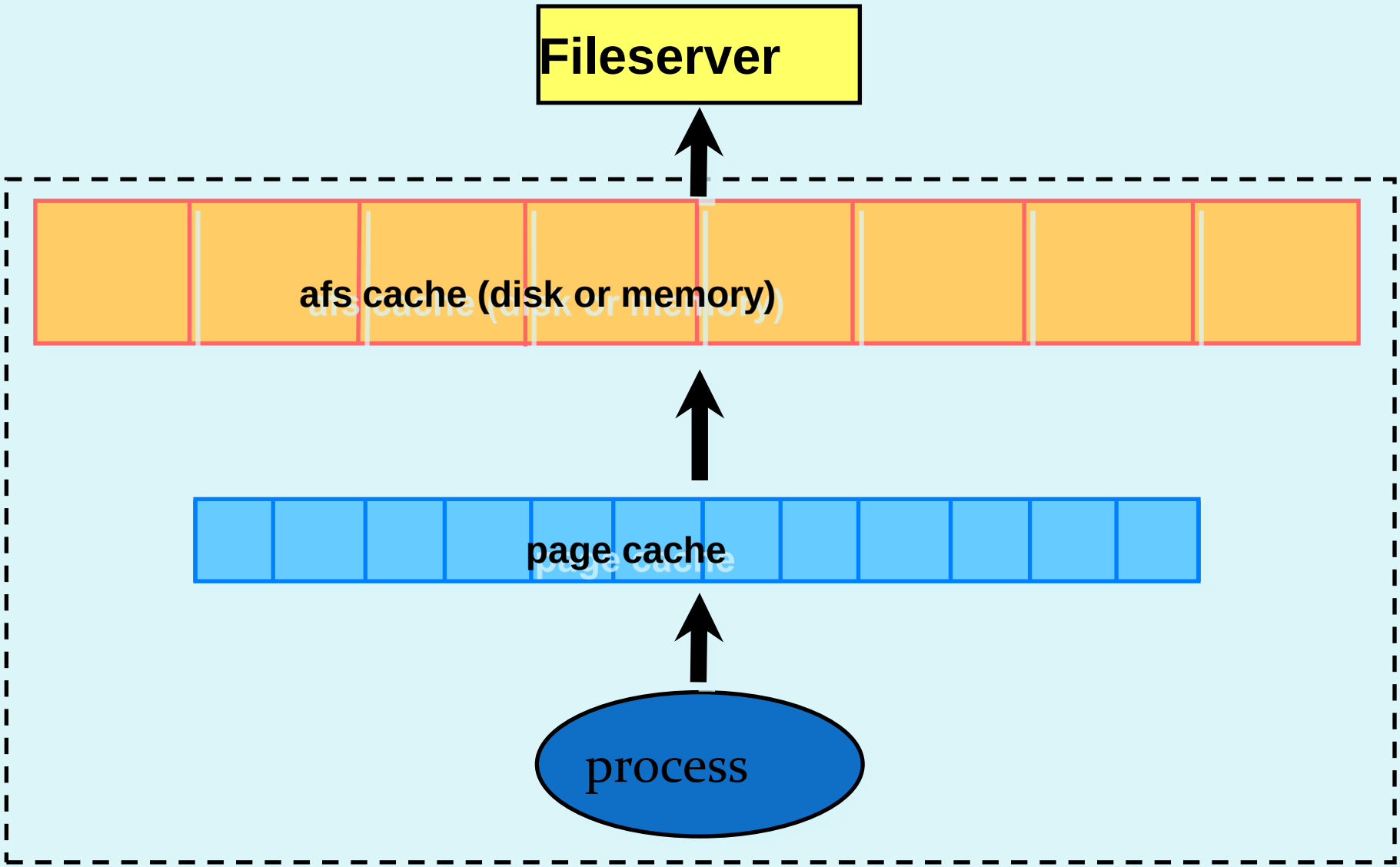
Assertion:

“On systems (such as Linux) with a file system page cache, it should be possible to improve AFS disk cache performance so it is similar to that of the AFS memory cache implementation”

Disk cache benefits over memory cache

- Can be larger
- Makes more efficient use of space
- Reduces potential for deadlocks
- Doesn't tie down kernel memory which hurts overall system performance

Architecture



Page Cache Improvements

- Reduce the number of redundant reads by correctly using the page dirty flag
- Enable readahead when filling the page cache from disk
- Remove duplicate writes of pages to disk by telling the kernel what we're doing
- Populate the page cache with a background thread, rather than doing it during requests

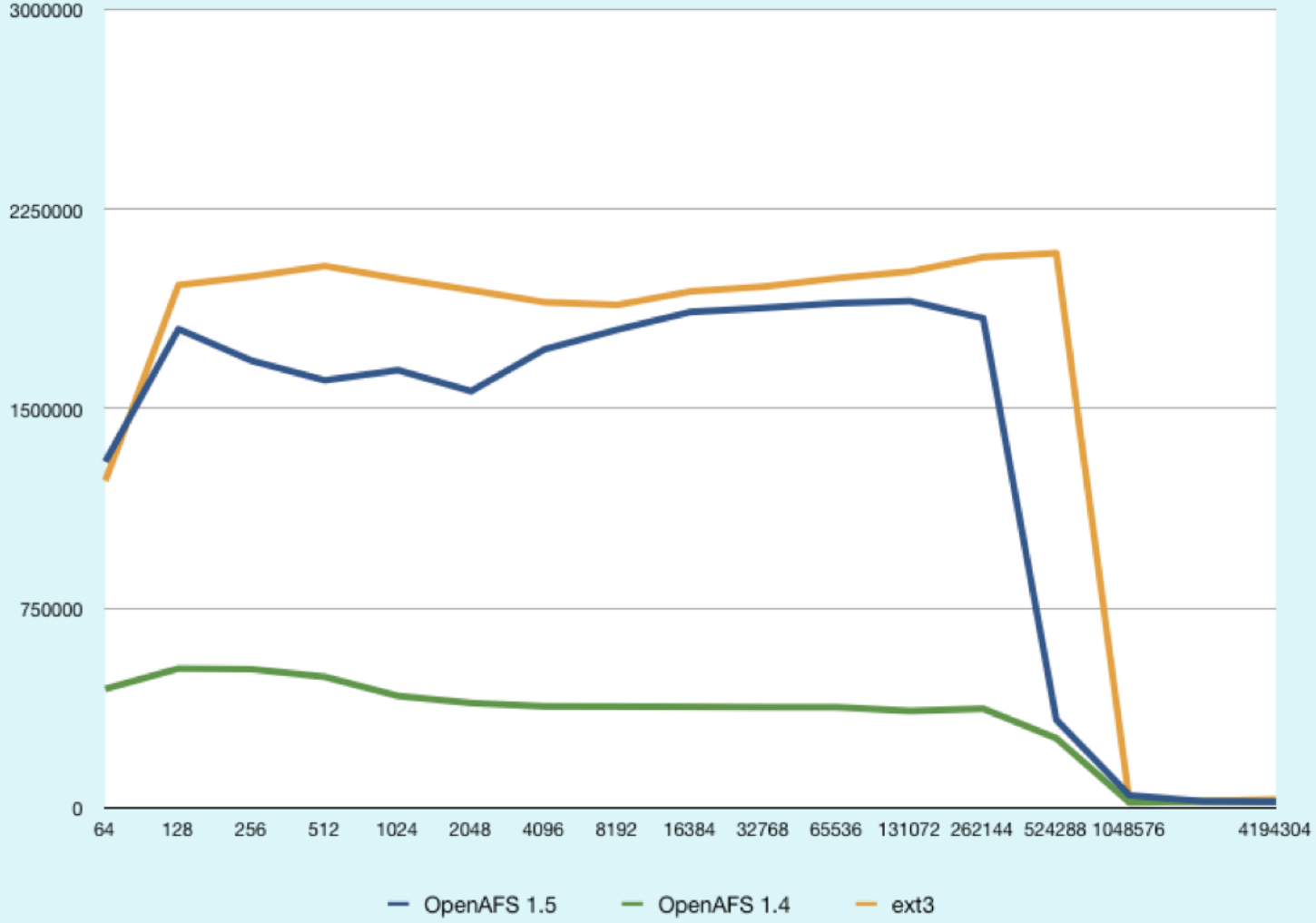
Minimize data copies

- Copying data is expensive
- Minimise the number of copies between the network, the various caches and user space
- Significant improvements made to write-on-close
- Other cases an ongoing project

Playing nicely with the other children

- Keyrings are now authoritative for PAG membership
- Many more memory management fixes
  - For the first time, it's possible to write to a mmap()'d file that's larger than the cache
  - Deadlocks are much harder to provoke in low memory situations

Cache read performance: AFS should match ext3 below 1GB



The long tail

- Significant real world performance improvements appear in the long tail
- Web server benchmarks suggest performance improvements of ~500% are obtainable for some readonly workloads.

Contact Info

- Simon Wilkinson
- Your File System Inc.
- sxw@your-file-system.com