SFS: Random Write Considered Harmful in Solid State Drives

Changwoo Min[1,2], Kangnyeon Kim[1], Hyunjin Cho[2], Sang-Won Lee[1], Young Ik Eom[1]
[1]Sungkyunkwan University; [2]Samsung Electronics

Motivation and Goals

Improving random write performance of file systems on SSDs and extending life span of SSDs.

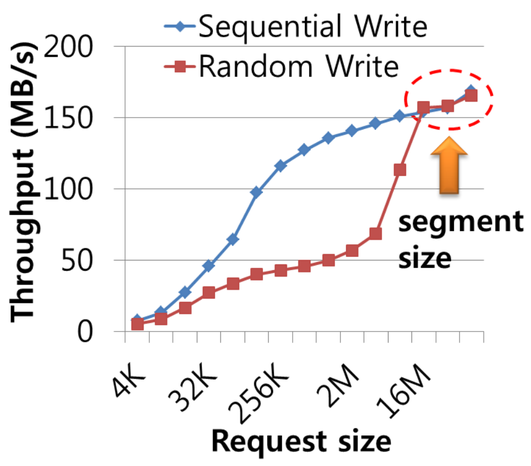
- Random writes cause internal fragmentation of SSDs and thus lead to performance degradation of more than 10x.
- In contrast to HDDs, the degradation caused by the fragmentation lasts for a while after write pattern is changed.
- Even worse, the fragmentation increases garbage collection cost in SSDs and thus reduces the life span of SSDs.

Our Approach: SFS

A log-structured file system with *on writing data grouping* and *cost-hotness segment cleaning*.

1. Why log-structured FS?

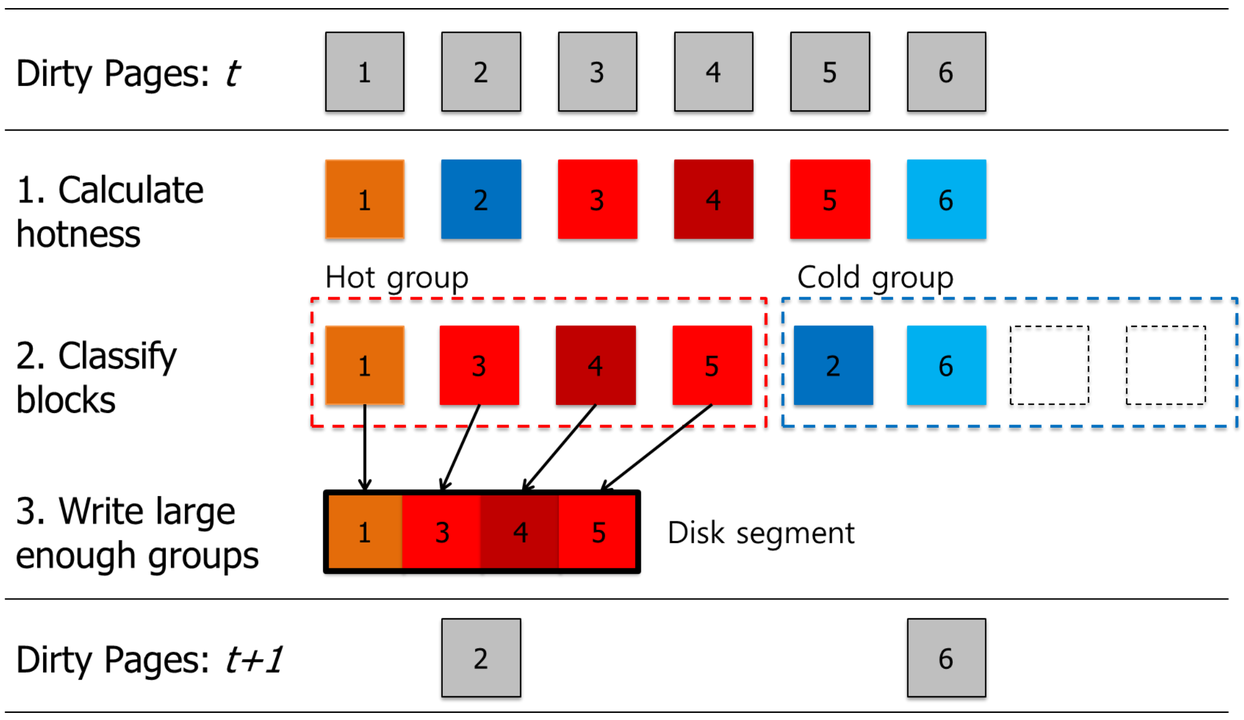
- Transforms the random writes at file system level into the sequential writes at the SSD.
- When the request size of random write is a multiple of clustered block size, random write bandwidth is converged to sequential write bandwidth.
- So, write performance can be mainly determined by sequential write performance of a SSD.



3. Cost-hotness cleaning

- Natural extension of cost-benefit policy.
- Since hotness directly represents the update likelihood of segment, we use segment hotness instead of age.
- More proper victim segment selection.
- $cost - hotness = \frac{free\ space\ generated}{cost * segment\ hotness}$

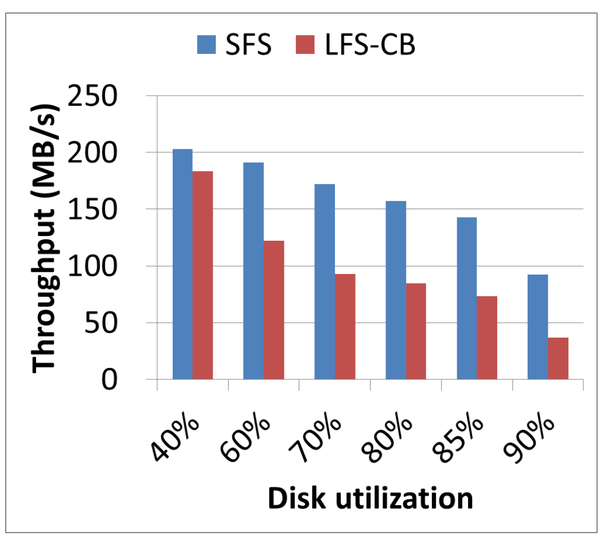
2. On writing data grouping



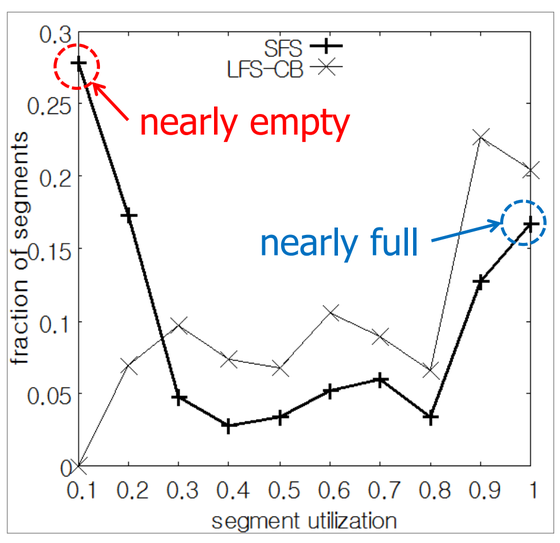
- Eager block grouping on writing to reduce segment cleaning overhead.
- Colocate blocks with similar update likelihood (*hotness*) into the same segment to form a sharp bimodal distribution of segment utilization.
- *Hotness* is determined by write frequency and temporal locality: `write count / age`
- Criteria for block grouping is determined by a proposed *iterative segment quantization algorithm*, which finds natural hotness groups across segments in disk.

Evaluation

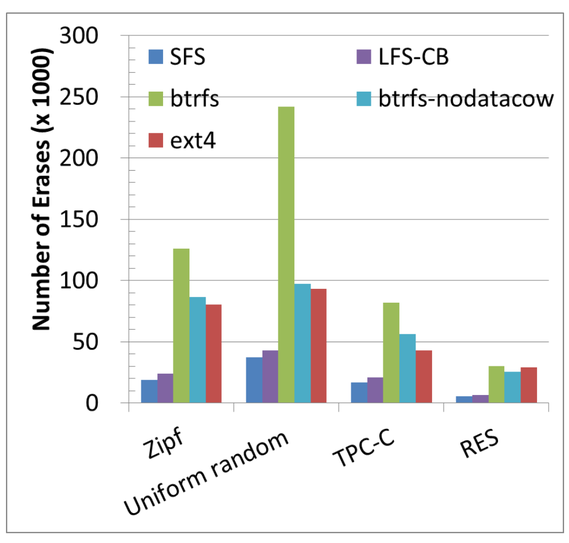
SFS outperforms the traditional LFS by 2.5x in terms of throughput.
Comparing to ext4 and btrfs, block erase count inside the SSD is reduced by 7.5x.
Even in HDDs, SFS outperforms state-of-the-art file systems by 39x in terms of throughput.



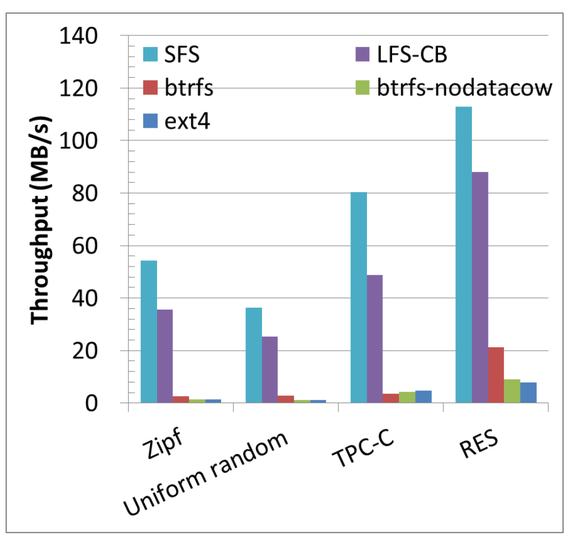
SSD-M, TPC-C



SSD-M, TPC-C



FAST FTL



HDD-M, 85% disk utilization