

# ZoneTier: A Zone-based Storage Tiering and Caching Co-Design to Integrate SSDs with Host-Aware SMR Drives

Extended Abstract

Xuchao Xie <sup>† §</sup>, Liquan Xiao <sup>†</sup>, David H.C. Du <sup>§</sup>

<sup>†</sup> National University of Defense Technology, <sup>§</sup> University of Minnesota-Twin Cities  
{xiexuchao,xiaoliquan}@nudt.edu.cn, du@cs.umn.edu

## 1 INTRODUCTION

Deploying solid state drives (SSDs) as a cache layer (caching) or a fast storage tier (tiering) with hard disk drives (HDDs) has been prevalent in cost-effective high-performance storage systems [1]. In recent years, Shingled Magnetic Recording (SMR) is proposed to achieve higher areal density on disk platters to respond to the demand of exponentially increased storage capacity. Due to the higher storage capacity and lower cost, SMR drives have the potential to replace the legacy HDDs as a slower tier in hybrid storage systems.

Zone-based host-aware SMR drives are a promising type of SMR drive [2]. HA-SMR drives redirect non-sequential writes (NSWs) to internal media cache and trigger media cache cleaning to move NSWs to their targeted HA-SMR zones. During media cache cleaning, HA-SMR drives will endure precipitously degraded performance. However, existing SSD tiering and caching designs cannot well handle NSWs for HA-SMR drives due to HA-SMR zone information absence and extra generated NSWs during data migration.

We propose ZoneTier, a zone-based storage tiering and caching co-design, to alleviate the cleaning overhead of HA-SMR drives when integrating with SSDs. Our main contribution is leveraging the intrinsic host-aware property of HA-SMR drives to develop ways of controlling all the NSWs to HA-SMR drives in SSD tiering and caching policies.

## 2 ZONETIER OVERVIEW

In ZoneTier, SSD is partitioned into tiered storage and cache partitions. The boundaries of extents are aligned with HA-SMR zones for storage tiering. SSD zones and HA-SMR zones expose a linear array of logical blocks as a normal block device. The cache partition is used as cleaning-friendly cache (CF-Cache) to obtain cleaning-friendly NSW traffics to HA-SMR zones. There are four potential end points on the data path of an I/O request: SSD zone, CF-Cache, HA-SMR zone, and media cache. To be responsive to workload changes, ZoneTier dynamically moves data by four operations: *Promoting*, *Demoting*, *Evicting* and *Merging*.

Critical access frequency (CAF) is introduced in ZoneTier as the metric to evaluate zones for future placement. Zone placement decisions are regenerated periodically based on the combined read and write CAFs of each zone. To reduce zone promotion overhead and accommodate more valid data into SSD tier, ZoneTier only allocates an adequate number of fine-grained subzones to the valid blocks of a zone and filters out the intra-zone garbage. The intra-zone garbage is

the synthesized filling data used to represent the unwritten blocks in HA-SMR zones. As the boundaries of extents and HA-SMR zones are aligned, ZoneTier always resets the write pointers of the involved HA-SMR zones first to invalidate the stale data and initialize the write pointer positions. Thus, the consecutive writes from zone demotions are sequentially written to HA-SMR zones without polluting media cache.

In ZoneTier, we reshape the inevitable NSWs to HA-SMR zones to cleaning-friendly NSW traffics to alleviate the performance degradation of HA-SMR drives caused by cleaning. ZoneTier accurately identifies NSWs for caching and bypasses sequential writes by interacting with HA-SMR zones directly. In CF-Cache, cached blocks aiming to the same HA-SMR zone are always evicted together to reduce the number of non-sequential zones in HA-SMR drive and make each RMW operation reclaim more reusable media cache space.

## 3 PERFORMANCE EVALUATION

We implement a ZoneTier prototype that offers pseudo block device interface to accept I/O operations replayed from block-level trace files. Functions provided by open-source `libzbc` and `libaio` [3] are used to manipulate HA-SMR drives as zoned block device and normal block device respectively. Our experimental results show that the system performance of ZoneTier outperforms that of both the fine-grained tiering and the zone-based tiering that ignores the host-aware property. The average I/O response time reductions are up to 28.20% and 25.51% respectively. When demoting zones with consecutive 512KB writes, the response time of these internal writes is stable at about 13.65ms in ZoneTier while other designs appear having different levels of latency amplification. Besides, ZoneTier significantly reduces the 99th percentile latencies of the writes served by HA-SMR zones by up to 73.07% than that with LRU-based caching.

## REFERENCES

- [1] Weixia Xu, Yutong Lu, Qiong Li, Enqiang Zhou, Zhenlong Song, Yong Dong, Wei Zhang, Dengping Wei, Xiaoming Zhang, Haitao Chen, et al. Hybrid hierarchy storage system in milkyway-2 supercomputer. *Frontiers of Computer Science*, 8(3):367–377, 2014.
- [2] Fenggang Wu, Ming-Chang Yang, Ziqi Fan, Baoquan Zhang, Xiongzi Ge, and David H.C. Du. Evaluating host aware SMR drives. In *8th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 16)*, Denver, CO, 2016. USENIX Association.
- [3] Alireza Haghdoost, Weiping He, Jerry Fredin, and David H.C. Du. On the accuracy and scalability of intensive i/o workload replay. In *15th USENIX Conference on File and Storage Technologies (FAST 17)*, pages 315–328, Santa Clara, CA, 2017. USENIX Association.