Improving Flash Resource Utilization at Minimal Management Cost in Virtualized Flash-based Storage Systems
Jianzhe Tai¹ Deng Liu² Zhengyu Yang¹ Xiaoyun Zhu³ Jack Lo³ Ningfang Mi¹
Northeastern University¹ Twitter Inc² VMware Inc³

Abstract
Effectively leveraging Flash resources has emerged as a highly important problem in enterprise storage systems. Our new Flash resource managers adopt the ideas of thermodynamic heating and cooling to identify data blocks that can benefit the most from being put on Flash and migrate data blocks between Flash and magnetic disks in a lazy and asynchronous way. Experimental evaluation of the prototype shows that both VFRM and GLB-VFRM achieve better cost-effectiveness than traditional caching solutions, i.e., obtaining IO hit ratios even slightly better than some of the conventional algorithms as Flash size increases yet costing orders of magnitude less IO bandwidth.

Background
- **Flash Resources Benefits:**
  SSD (Solid State Disk) has low latency, high IOPS and I/O concurrency, low energy consumption.

- **How to Utilize SSD in Virtualization Environment?**
  (1) **Storage Caching Approach:** Design a cost-effective use of Flash resources in the virtualized environment for both enhancing performance and minimizing Flash operational cost.
  (2) **Tiering Storage Approach:** Design a automated tiering method across different storage tiers, aiming to guarantee multiple SLAs for applications with dynamic workloads. Need data redundancy, like RAID or mirroring.

- **Limitation of Existing Approaches:**
  (1) Very fine-grained temporal and special granularities, i.e., updating the SSD and MD per IO request (4KB) per IO epoch, which triggers high memory overhead, high computing cost, and high IO cost for updating contents.
  (2) Difficult to resize the cache on time according the incoming traces and low utilization efficiency of SSD.

Motivation

![Fig 1. Cache Approach](image)

![Fig 2. Tiering Approach](image)

Fig 3. IO Popularity Analysis of Microsoft Cambridge traces

- **Obv. 1** For some workload, the block access frequency shows a bimodal distribution pattern: A small amount of SSD can absorb most of the repeated IO requests.
- **Obv. 2** The distribution of I/O popularity varies slowly over time.
- **Obv. 3** Block access patterns vary across different workload.

Goal & Method

- **Goals:**
  (1) Automatic & Dynamic Flash resource provisioning based on actual needs of the workload.
  (2) Better Flash resource utilization and performance enhancement.
  (3) Minimal cost for managing/operating Flash (e.g., CPU, Memory, I/O bandwidth).

- **Solution:**
  (1) Enlarge spatial granularity (1MB) and temporal granularity (5min).
  (2) Divide SSD into two zones. Private Zone has isolated and fixed size space for each VM. Public Zone is free to compete among VMs.
  (3) Use temperature-based model for predicting the IO popularity of bins:

\[
T = \sum_{i=1}^{N} \left( H(M_{i}) \cdot C(i) \right) \cdot \sum_{j=1}^{M_{i}} \frac{S_{i,j}}{N+1}
\]

Conclusion

For both single and multiple VMs cases, VFRM saves at least 44.38% of IO costs (up to 89.90%) while achieves similar or even higher hit ratios compared with traditional algorithms.

Reference