BEPA: Bulk Erase based on Process Awareness for cloud SSD caches

Sang Jin Kim, Gi Lee, Byoung Jin Kim , and Eui-Young Chung
School of Electrical and Electronics Engineering, Yonsei University, Seoul 03722, Korea

Nowadays, as complex data processing based on big data is performed in a cloud environment, demands for high-speed input/output have been increased. As a result, the case of using a Solid State Drive (SSD) as a cache for a Hard Disk Drive (HDD) has been increased to improve storage performance rather than building a storage layer using only the existing HDD. However, current SSD cache performance reduces by keeping unused data for a long period of time and does not manage it efficiently, which reduces storage performance. In order to overcome these shortcomings, many techniques have been studied to process data by capturing the characteristics of the data and clustering it with hot/cold data[2].

This paper proposes Bulk Erase based on Process Awareness (BEPA) technique that manages storage by clustering in process units, further from clustering in data units. The data used during processes are clustered and collected, and the data clustered at the end of the process are bulk-erased at once (shown in Fig. 1). When one process ends, the data used in that process is hard to be used through cache, so the space can be bulk-erased to save enough free space for SSD. Through this technique, it is possible to improve performance of the SSD by reducing cache miss overhead. Compared to the conventional SSD management, BEPA technique reduced the read miss of SSD by about 40% for each process (shown in Fig. 2) and total operation time by an average of 11%. In the future, we intend to further develop a technique that enables clustering and management of unused data within the process.

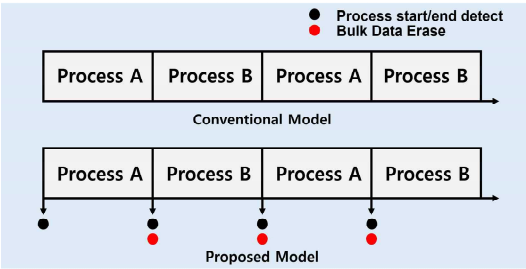

Fig 1. The concept of proposed scheme

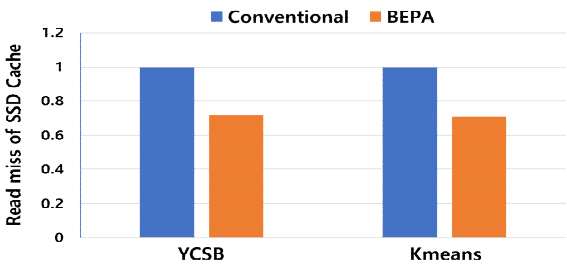

Fig 2. Comparison of read miss

Acknowledgments This work was supported by the Samsung Electronics Company, Ltd., Hwaseong, Korea.

References [1] Zhou, Ke, et al. "Efficient SSD cache for cloud block storage via leveraging block reuse distances." *IEEE Transactions on Parallel and Distributed Systems* 31.11 (2020): 2496-2509.

[2] Yang, Pan, et al. "Reducing garbage collection overhead in {SSD} based on workload prediction." *11th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 19)*. 2019.