Offline Deduplication with Lightweight Hash for Solid State Disk

Eunsoo Park, Dongkun Shin

College of Information & Communication Engineering, Sungkyunkwan University, Korea

E-mail : dongkun@skku.edu

Deduplication technique can expand the lifespan and capacity of flash memory-based storage device by eliminating duplicated write operations. The deduplication techniques can be classified into two approaches, i.e., online and offline approaches, based on the time of deduplication processing. During runtime, the online deduplication [1] checks each incoming page, and drops it if there is a same page in the storage. Instead, an address translation entry is registered in order to remap the read request on the deduplicated page. The online deduplication may increase the write latency since it should generate a complicated hash key for each page and should check whether there is a same page in the storage. Furthermore, at sudden power-off situation, the storage device cannot recover the deduplicated page if the corresponding address translation information is not flushed into non-volatile storage. The offline deduplication [2], on the other hand, finds the duplicated pages during idle time. It can hide the deduplication overhead by exploiting idle time, and it is not vulnerable to the sudden power-off. The offline dedplication can reduce the garbage collection overehad by eliminating copy operations on duplicated pags.

While the previous offline technique used a high-cost hash algorithm, our new approach uses only a lightweight hash algorithm such as CRC as shown in Fig1. Therefore, the memory space for caching hash keys can be removed, and more pages can be examined for deduplication during short idle intervals. As a result, it can reduce the write latency compared to online approach, and can reduce the garbage collection overhead compared to the previous offline deduplication technique.
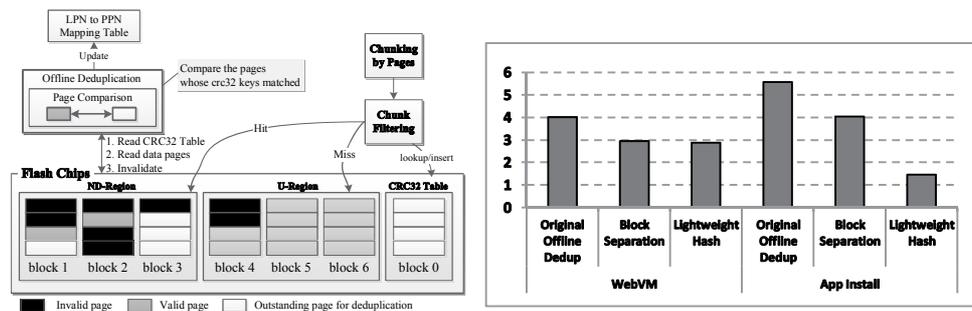

Fig 1 Architecture of suggested technique and normalized number of page copies in foreground GC

[1] F. Chen et al., "CAFTL: A Content-Aware Flash Translation Layer Enhancing the Lifespan of Flash Memory based Solid State Drives," FAST, 2011.

[2] A. Jeongcheol, S. Dongkun, "Offline Deduplication-Aware Block Separation for Solid State Disk," FAST, 2013.