Power-aware Storage-tiering Management for High-speed Tiered-storage Systems

Kazuhisa Fujimoto[†], Hirotoshi Akaike[††], Naoya Okada[†*], Kenji Miura[†], and Hiroaki Muraoka[†]
† Research Institute of Electrical Communication, Tohoku University. * Student
{fujimoto, kmiura, muraoka}@riec.tohoku.ac.jp, okada@kiroku.riec.tohoku.ac.jp
†† Systems Development Laboratory, Hitachi Ltd.
Hirotoshi.akaike.tv@hitachi.com

Large-scale high-speed mass-storage systems account for a large part of the energy consumed at data centers [1]. To save energy consumed by these storage systems, we propose a high-speed tiered-storage system with a power-aware method of storage-tiering management with minimum loss of performance, which we have called the energy-efficient High-speed Tiered-Storage system (eHiTS).

Our proposed method has two distinct features. The first is that eHiTS consists of a tiered-storage system with high-speed online storage as a first tier and low-power nearline storage with high capacity as the second tier. All files are always stored in nearline storage when it is created, in which the HDDs are usually left powered off. Only the volume that has stored the accessed files is copied from nearline to online storage before access. In our proposed method, data movement and its timing are managed inversely against ILM [2] where files are stored in online storage when it is created and relocated to the nearline storage when utilization becomes lower than a predetermined threshold based on user's policy. Even though the first-tier's online storage is used as a data cache in eHiTS, file requests in our proposed method are hit in online storage (cache storage) even during the first access, unlike in ordinary cache management. Moreover, the accessed files are copied back soon after the access period ends. This leads to the capacity of online storage being minimized with minimum loss of performance, resulting in energy savings.

The second feature is its ability to complete copying the volume that has stored the accessed files from nearline to online storage before access. The timing to start copying is predicted based on queueing theory with information about job-queue status in a batch-job scheduler of a batch-processing application.

We selected an HPC system as the first target among the batch-processing applications to evaluate eHiTS. Figure 1 outlines the architecture of eHiTS with an HPC system for scientific calculations. eHiTS is equipped with another feature where the HDD enclosure is powered off when there is no access to gain larger energy savings than in MAID systems [3].

By exploiting the features of job control in the HPC system, the accessed files and the timing to start copying them to online storage are predicted in eHiTS for each user-submitted job. A simple method of completing the copying of an accessed volume to online storage before access is to start the volume copy when the job is submitted into the job-queueing system. However, we found from the simulation results of the job-queueing system that there was not enough waiting time in the job queueing system to complete volume copy before access for lower utilization factor of the queueing system. To complete volume copy before access even in shorter waiting time cases, we propose a delayed method of job execution, where if the waiting time for the job to start to execute is shorter than a certain time, job execution is delayed.

Figure 2 compares an eHiTS's energy consumption with an ordinary tiered-storage system with ILM for system capacities of 256, 512, and 1024 TB. The preliminary results obtained from analytical studies with the measured parameters in our testbed revealed that the eHiTS's energy consumption decreased by as much as 46 % of that of the ordinary tiered-storage system for a system capacity of 1024 TB. We also found that the proposed method of delayed job execution could achieve a miss-prediction probability less than 10^{-3}.

We have been developing a testbed for eHiTS on which we intend to evaluate its energy conservation and predictable probability to demonstrate its effectiveness in conserving energy with minimum loss of performance in a real HPC environment.

Acknowledgements

This work was supported in part by Research and Development for the Next-Generation Information Technology of Ministry of Education, Culture, Sports, Science and Technology (MEXT). This work has been done in part as a joint research project with Hitachi Ltd.

References

[1] Q. Zhu, Z. Chen, L. Tan, Y. Zhou, K. Keeton, and J. Wikes. Hibernator: Helping Disk Arrays Sleep through the Winter. In Proc. ACM Symp. on Operating Syst. Principles, pp. 177–190, 2004.
[2] M. Peterson. ILM and Tiered Storage. Storage Networking Industry Association, 2006.
[3] F. Moore and A. Guha. Introducing COPAN Systems MAID Architecture (Massive Array of Idle Disks). White Paper, Copan Systems, 2004.

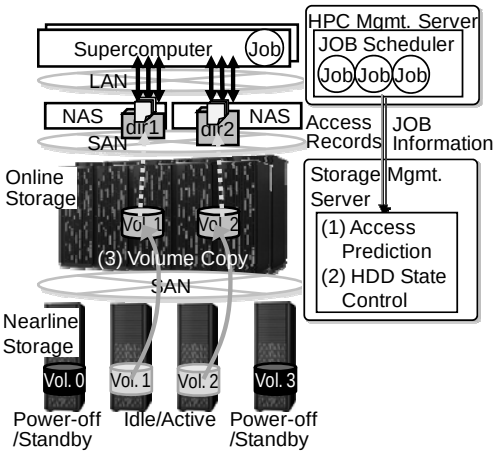

Fig. 1 System Architecture

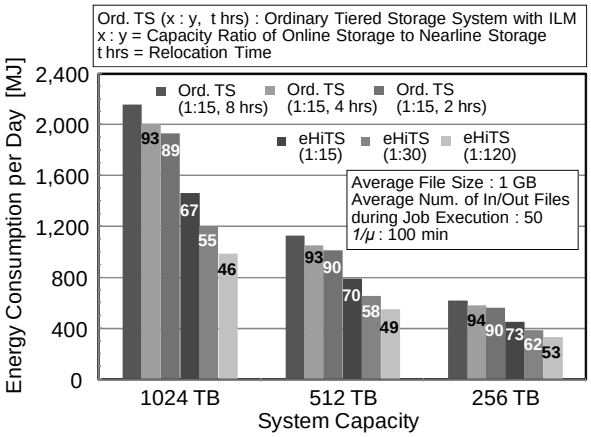

Fig. 2 Comparisons of eHiTS's Energy Consumption with Ordinary Tiered System