Ji-guang Wan, Da-ping Li, Xiao-yang Qu, Chao Yin, Jun Wang, Chang-sheng Xie. A reliable and energy-efficient storage system with erasure coding cache. *Frontiers of Information Technology & Electronic Engineering*, 2017, **18**(9):1370-1384. http://dx.doi.org/10.1631/FITEE.1600972

A reliable and energy-efficient storage system with erasure coding cache

Key words: Reliability; Energy-efficient; Storage system; Erasure coding; Cache management

Corresponding author: Ji-guang Wan
E-mail: jgwan@mail.hust.edu.cn

ORCID: http:// orcid.org/0000-0003-0489-6783

Motivation

- Due to the continuously increasing scale of parallel I/O systems, the proportion of the energy consumption of I/O systems to the total cost of ownership (TCO) becomes larger and larger.
- Existing methods only store the up-to-date writes in the primary disk group, the reliability of the new data will be degraded when the non-primary disk groups are spun down, until the new data is written back to all the non-primary replicas.
- To save energy without degrading reliability and performance, we propose to use erasure coding to encode the up-to-date data.

Main idea

- Provide ideal reliability of the new data and old data, improve the performance, decrease the energy consumption.
- To save energy, we can power off as many disks as possible, but should take the reliability into consideration.
- To improve the performance, we can design some methods to accelerate the speed of small writes and the speed of destage.

Main method

- RERAID employs part of the free space in the primary disk group and uses erasure coding to construct a code cache at the front end to absorb new writes.
- Use a buffer to cluster many small random writes into a few large writes, and then write them to the code cache in a parallel fashion sequentially
- When the workload is light, all the non-primary replicas will be turned down to save energy.

The architecture of RERAID

- Primary replica is divided into data area and code cache area

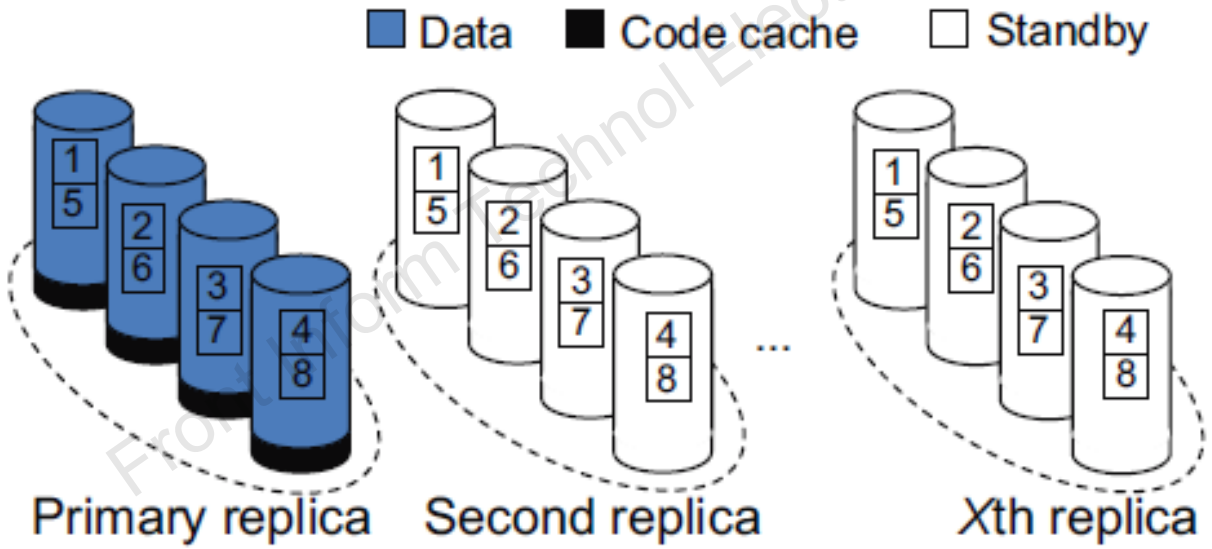

Fig. 1 The architecture of RERAID

ECW algorithm

- Cluster many small writes into large writes and write to the code cache

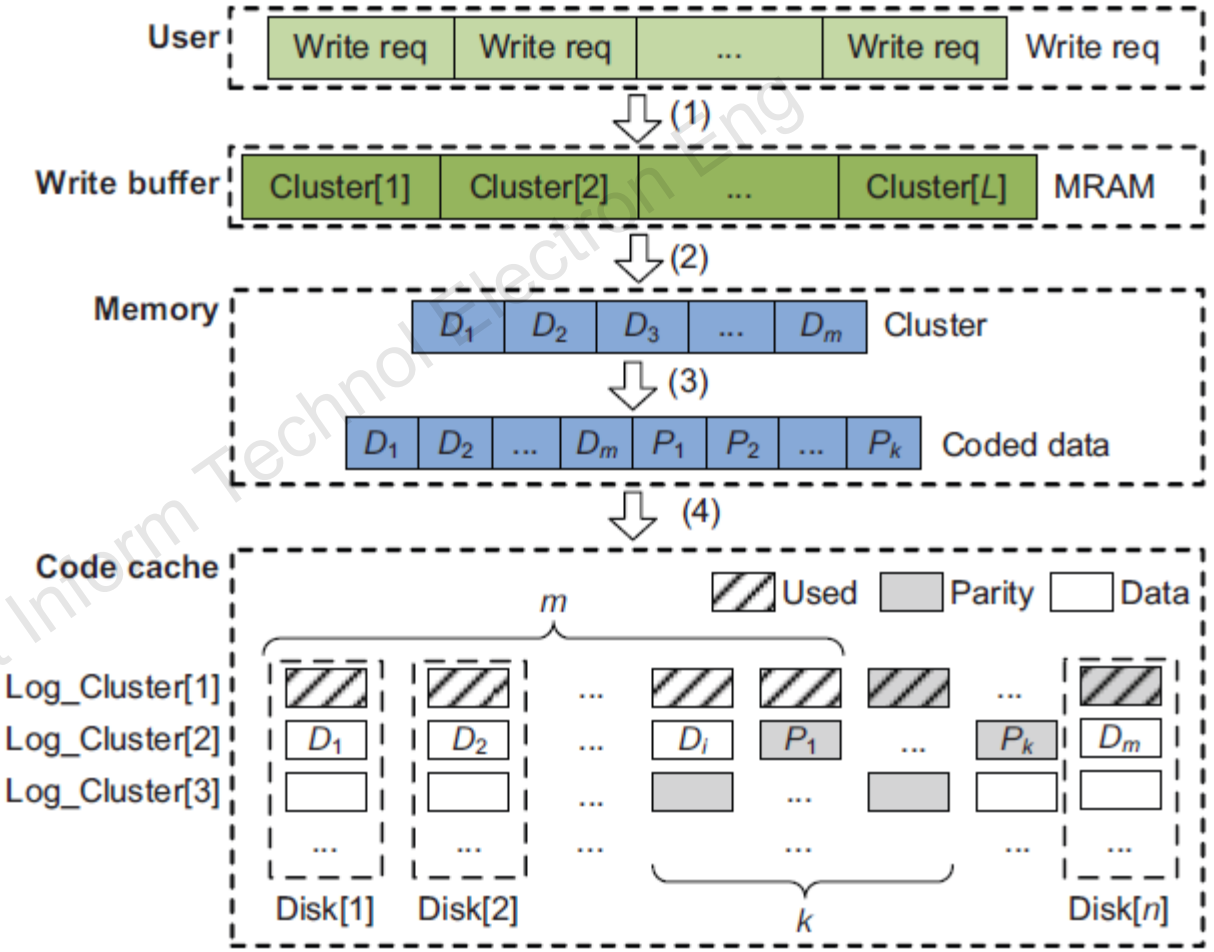

Fig. 2 The erasure coding writing (ECW) algorithm of RERAID

Flush to the data area

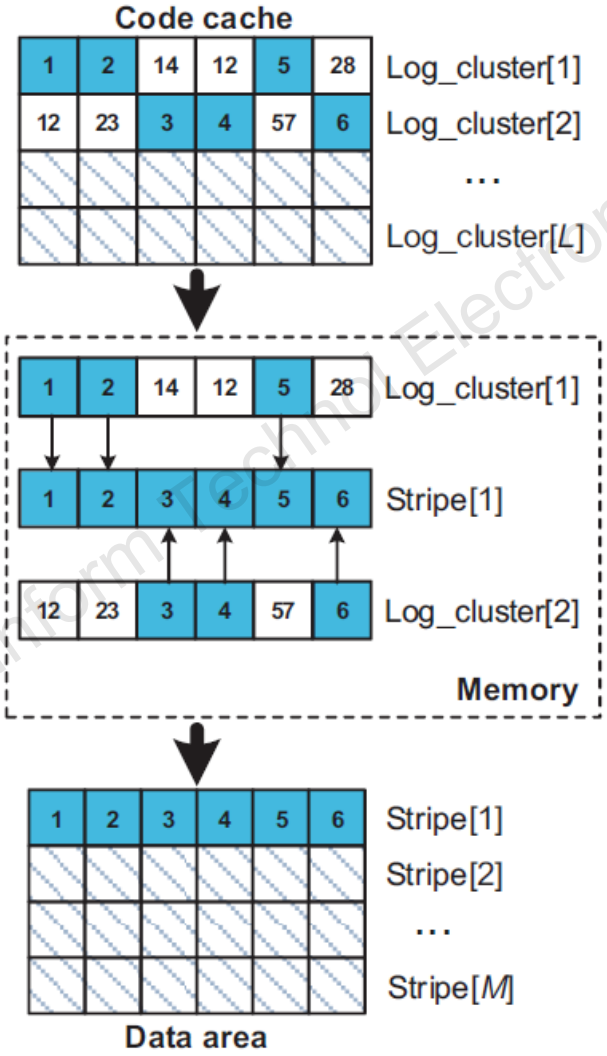

Fig. 4 The destaging process of RERAID

Major results

- RERAID has the best energy conservation.

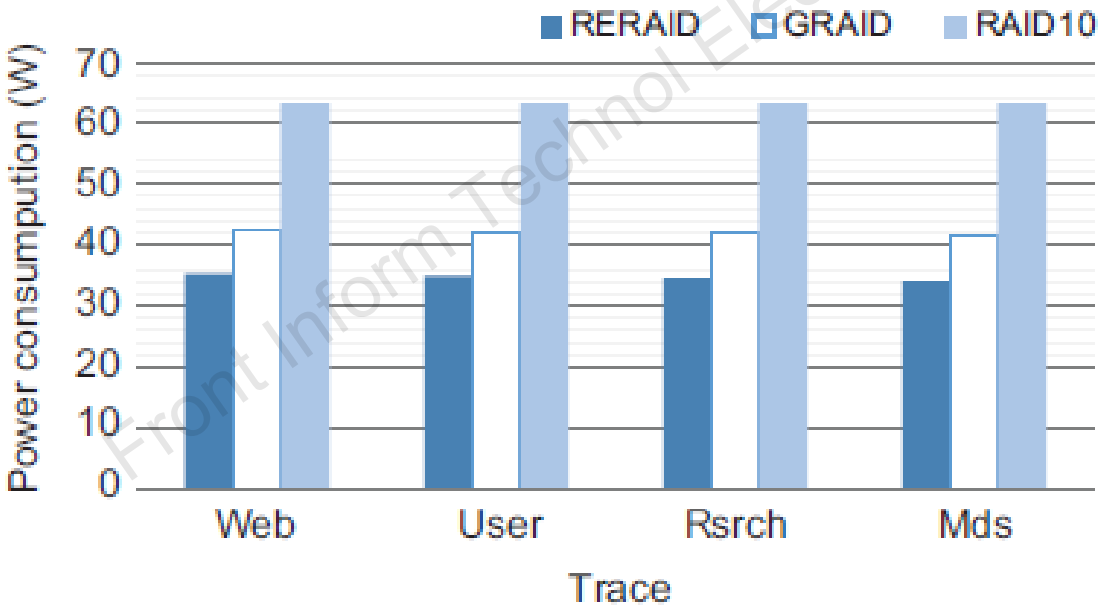

Fig. 11 Total energy consumption under different traces

Major results (Cont'd)

- RERAID performs better than the others under Web, Mds, and Rsrch using the ECW algorithm

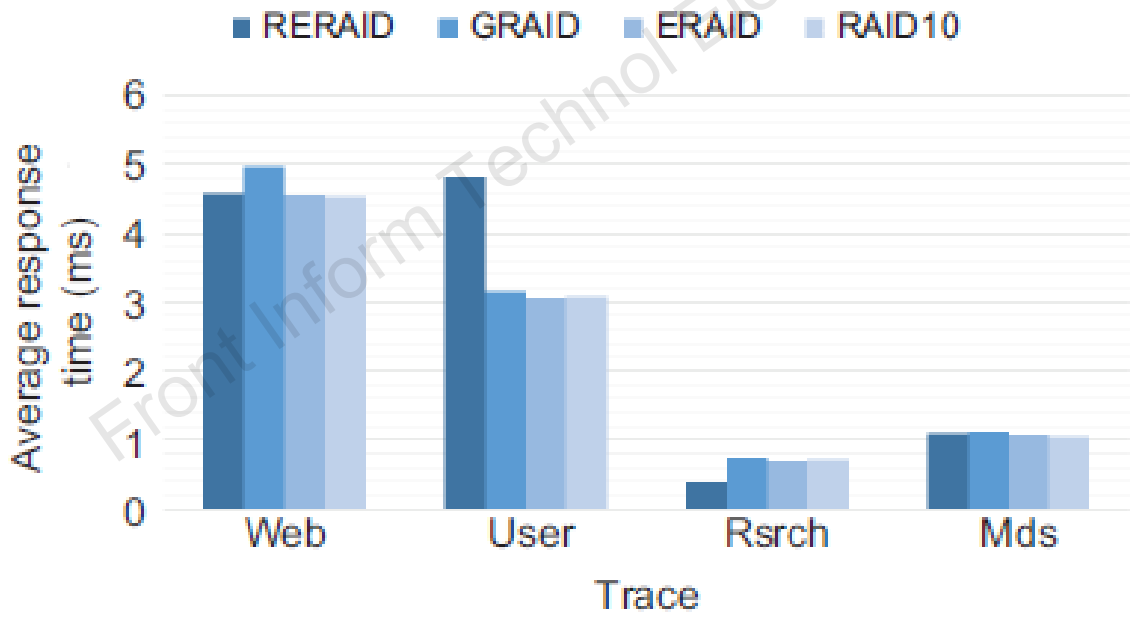

Fig. 12 Total average response time of different traces

Conclusions

- RERAID saves energy by spinning down the replica groups without migrating data or imposing extra requirements.
- Its code cache comprises part of the free space in the primary disk group and uses erasure coding to ensure the reliability of the data .
- It provides a ECW algorithm to accelerate the write speed and a flush disk mechanism to accelerate the flush
- Experiments demonstrated RERAID is effective in improving performance and saving energy without reducing the reliability