Yun TENG, Zhiyue LI, Jing HUANG, Guangyan ZHANG, 2022. ShortTail: taming tail latency for erasure-code-based in-memory systems. *Frontiers of Information Technology & Electronic Engineering*, 23(11):1646-1657.
https://doi.org/10.1631/FITEE.2100566

ShortTail: taming tail latency for erasure-code-based in-memory systems

Key words: Erasure code; In-memory system; Node fail-slow; Small write; Tail latency

Corresponding author: Guangyan ZHANG
E-mail: gyzh@tsinghua.edu.cn
ORCID: https://orcid.org/0000-0002-3480-5902

Motivation

1. In-memory systems with erasure coding (EC) enabled are widely used to achieve high performance and data availability. The fail-slow problem causes long tail latency due to the synchronous nature of multiple EC sub-operations.

2. It is diffcult to identify the fail-slow nodes which cause long tail latency.

3. It is a complex challenge to avoid accessing the fail-slow nodes.

Main idea

1. Identifying the fail-slow nodes by tracking their performance with a lightweight request monitor.

2. Avoiding reading the fail-slow nodes by performing degraded reads selectively.

3. Avoiding writing the fail-slow nodes by performing redirected writes.

4. Reducing the write amplification of small writes with adaptive write strategy.

Method

1. We propose ShortTail which tames tail latency for erasure-code-based in-memory systems.

2. ShortTail consists mainly of three techniques:

   Identifying fail-slow nodes;

   Sidestepping fail-slow nodes by avoiding reading or writing fail-slow nodes;

   Optimizing small writes using replication selectively.

Method (Cont’d)

Technique **1**: Identifying fail-slow nodes by tracking the node performance

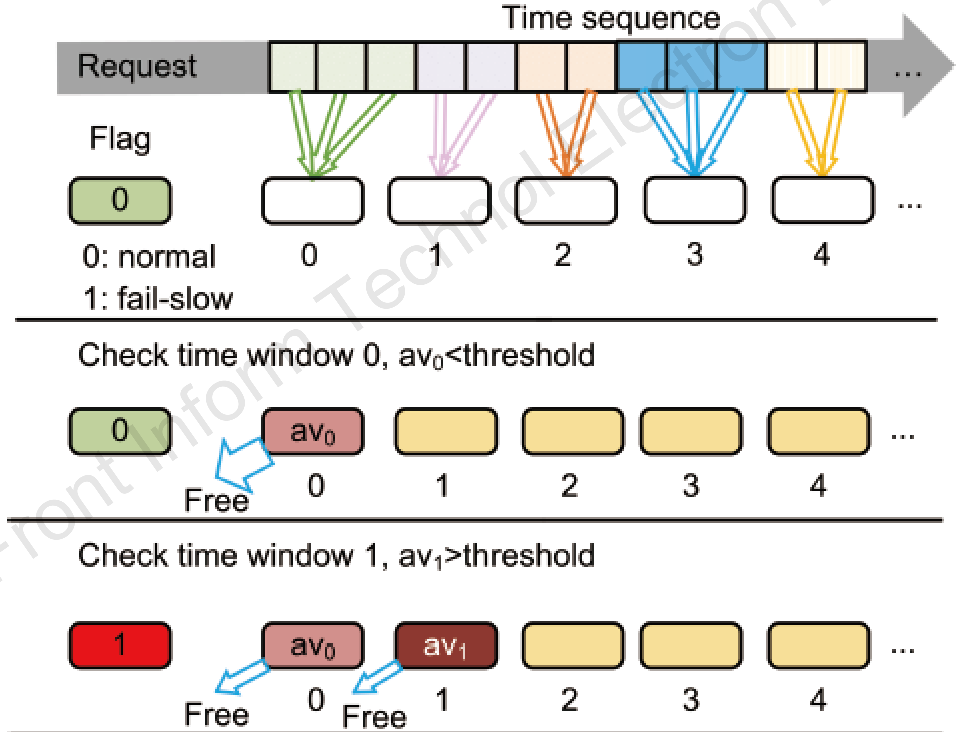

Fig. 2 Fail-slow node identification

Method (Cont’d)

Technique **2**: Sidestepping fail-slow nodes with degraded reads and redirected writes

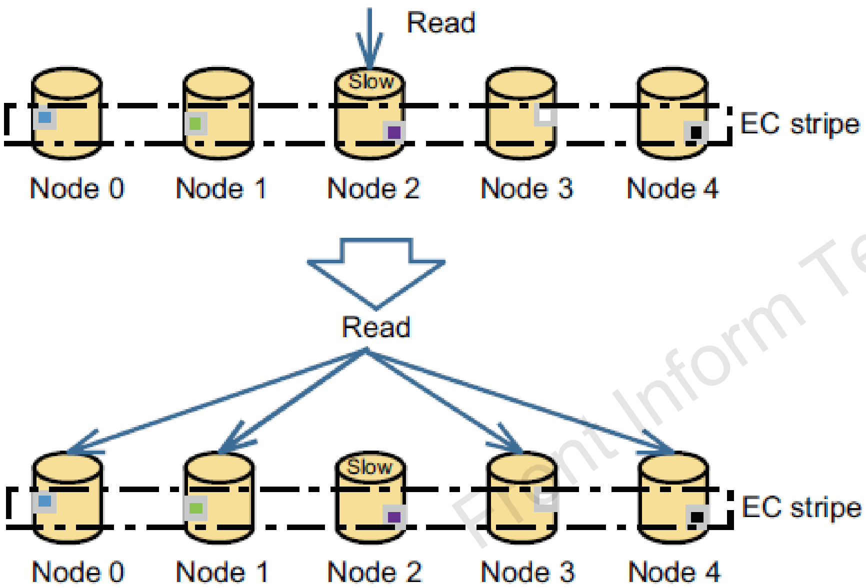

Fig. 3 Fail-slow node sidestepping for read requests

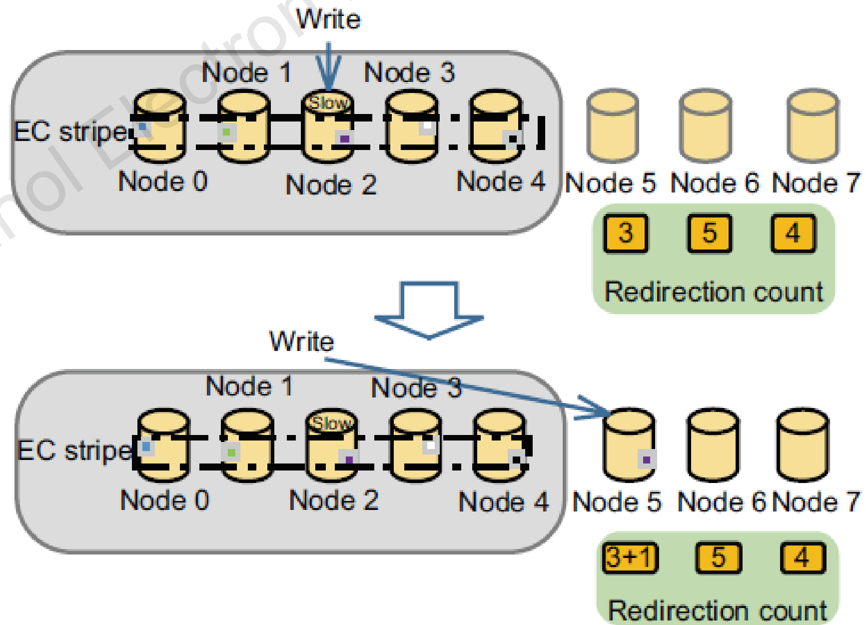

Fig. 4 Fail-slow node sidestepping for write requests

Method (Cont'd)

Technique **3**: Optimizing small writes with adaptive replication-and-striped strategies

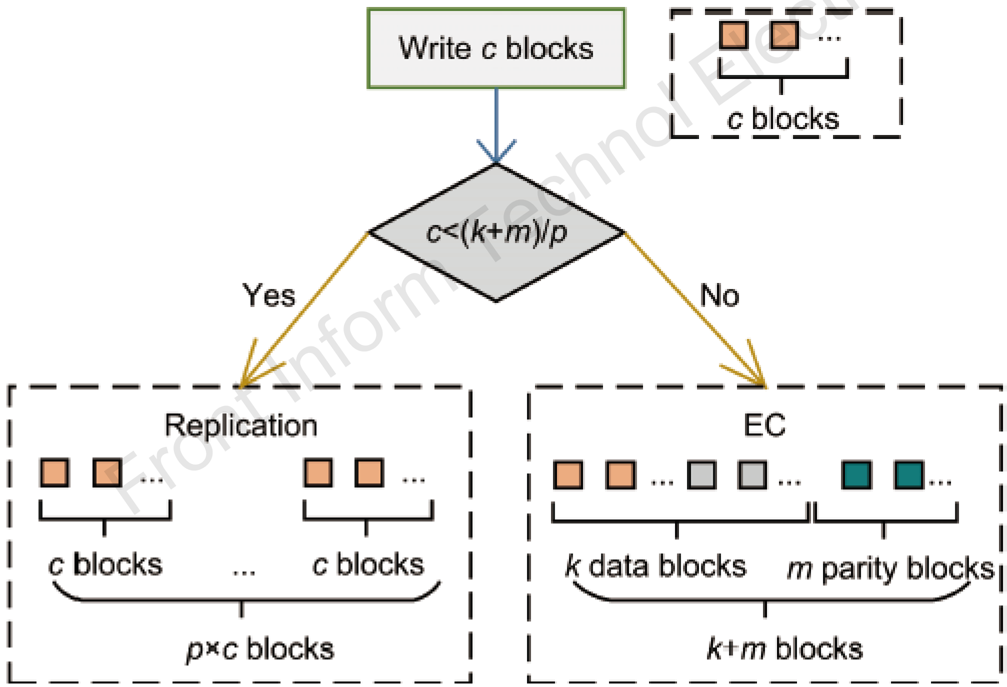

Fig. 5 Optimization for small writes

Major results

Overall performance under different workloads: ShortTail can reduce the P99 tail latency by up to 63.77%.

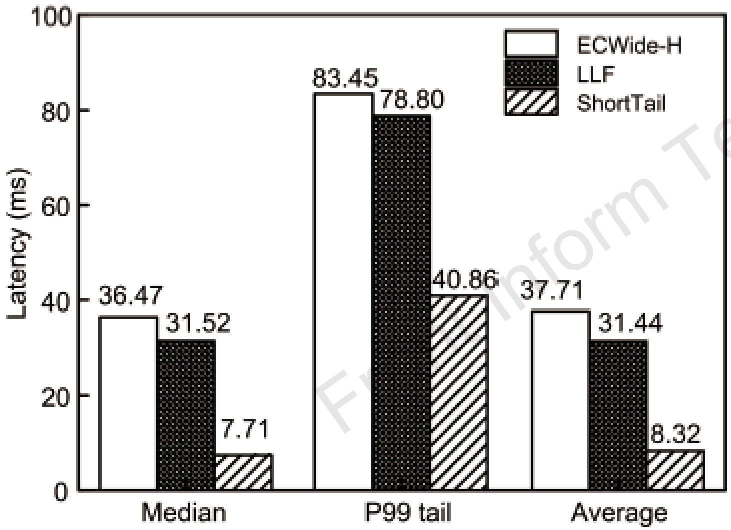

Fig. 6 Performance comparison under the YCSB workload

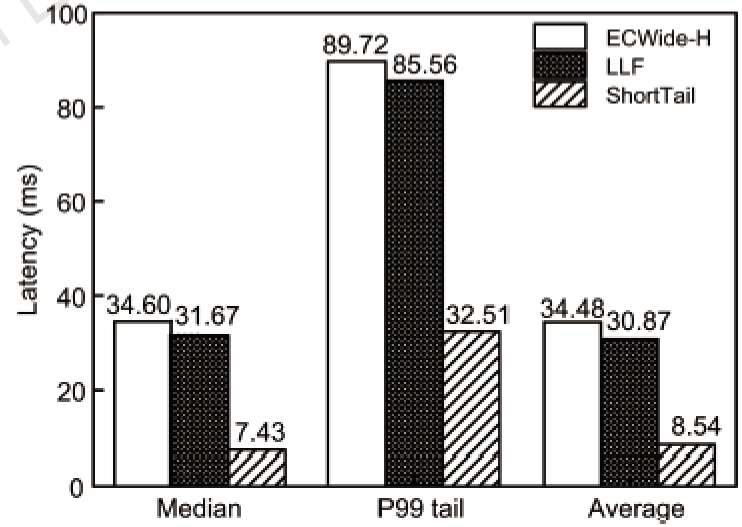

Fig. 7 Performance comparison under the Usr_1 workload

Major results (Cont'd)

Effects of individual techniques: each technique has obvious effect on tail latency reduction.

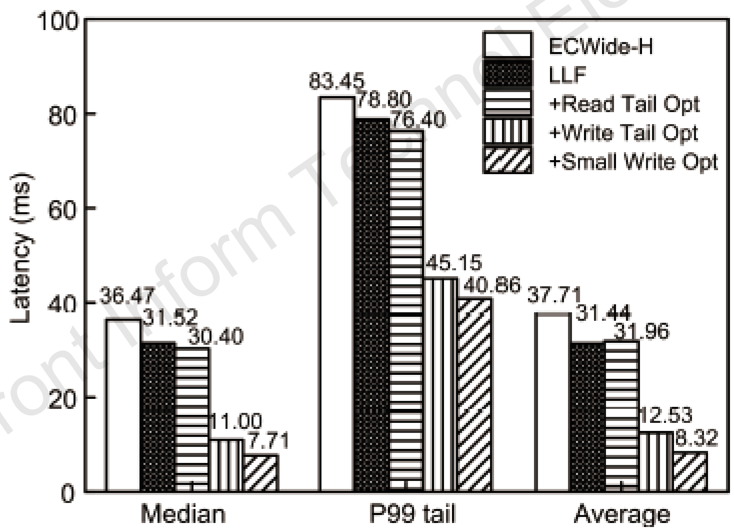

Fig. 8 Performance contributions of individual techniques

Major results (Cont’d)

Sensitivity to internal parameters: the parameters selected by ShortTail are effective.

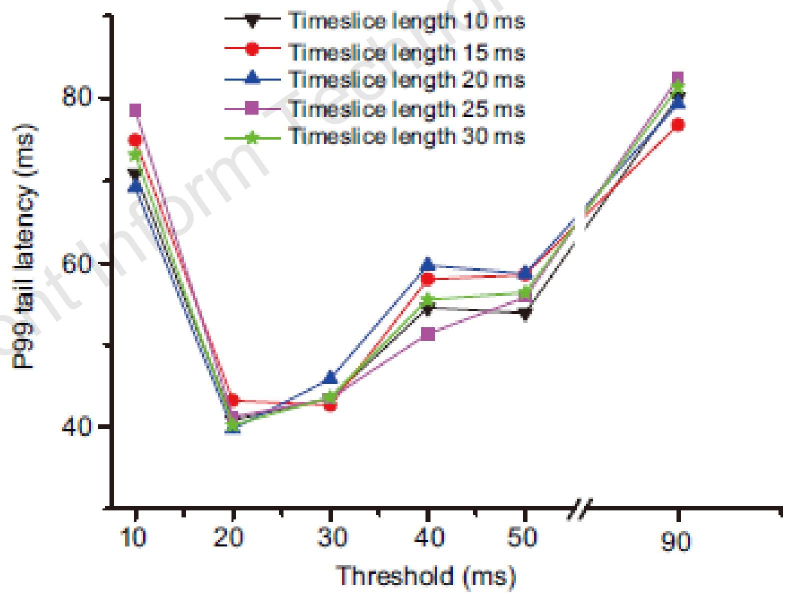

Fig. 9 Influence of changing parameters

Conclusions

1. We propose ShortTail to tame tail latency for erasure-code-based in-memory systems.
2. It consists of three techniques: **identifying fail-slow nodes**, **sidestepping accessing fail-slow nodes,** and **optimizing small writes.**
3. ShortTail reduces the tail latency in EC-based in-memory systems significantly, and also brings significant improvements with respect to the median latency and average latency.

Jing HUANG received her BS degree from the Information College, North East University, China, in 1998, and her MS and PhD degrees from the Computer College, Jilin University, China, in 2003 and 2008, respectively. She is now a professor with the College of Computer Science and Technology, Jilin University. Her research interests include social network analysis, data mining, and multi-agent systems.

Guangyan ZHANG received his PhD degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2008. He is now an associate professor with the Department of Computer Science and Technology, Tsinghua University. His current research interests include big data computing, network storage, and distributed systems.