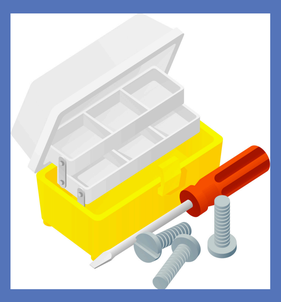


CASE: Exploiting Content Redundancy for Improving Space Efficiency and Benchmarking Accuracy in Storage Emulation

Lei Tian, Hong Jiang

University of Nebraska-Lincoln

INTRODUCTION

Timing-Accurate Storage Emulation

Memulator: Pioneer work

Space-Efficient Storage Emulation

David: saves space by omitting file data but storing file metadata only

Content-Retained and Space-Efficient Storage Emulation

CASE: incorporates data deduplication to eliminate redundancy

MOTIVATION STUDY

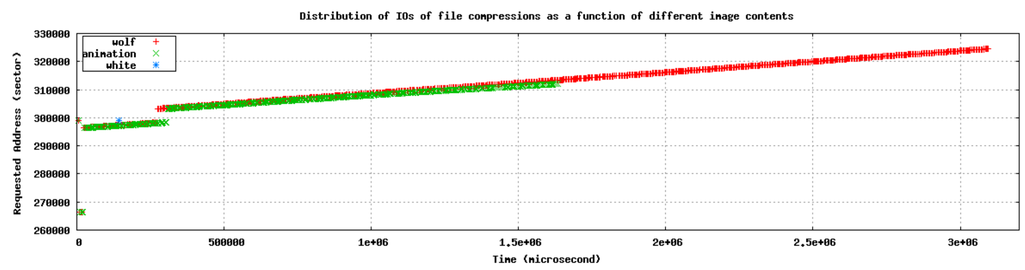
Study the impact of data content on benchmarking accuracy

1. Compress three bitmap images

Resolution	2400 × 1920
File Size	13,824,054 bytes

wolf animation white

2. Use *blktrace* to trace IOs during saving compressed files



3. Observations and Implications

- I/O patterns significantly differentiate from each other
- Prone to either overestimating or underestimating the real performance without storing the exact data contents

APPROACH & ARCHITECTURE

CASE: a flexible content-aware and space-efficient storage emulator for benchmarking

Idea: deploying data deduplication in storage emulation

Goals

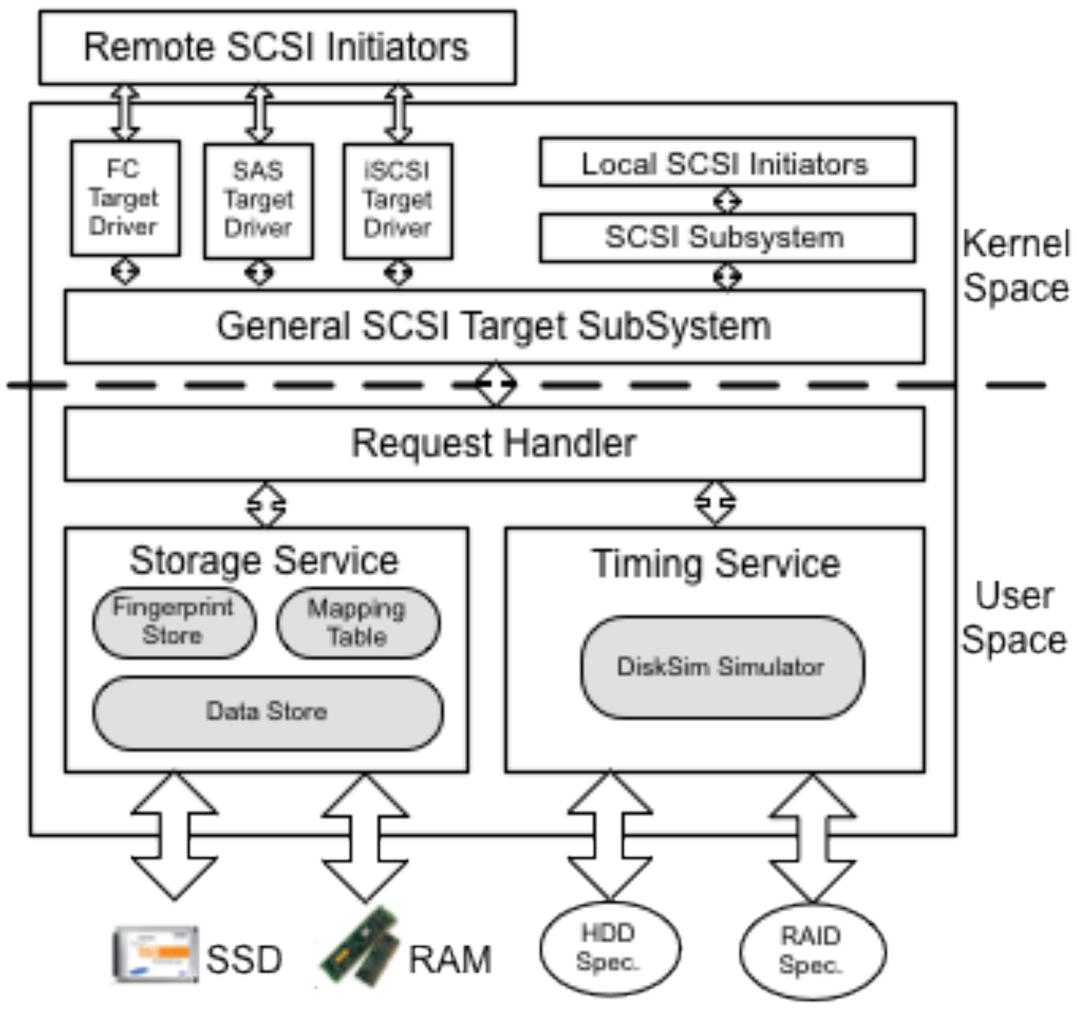
- Timing accuracy
- Space efficiency through redundancy elimination
- No modification to FS, DB and interfaces

Design

- Request Handler: receives and forwards IOS
- Timing Service: Computes the response time for every IO
- Storage Service
  - Fingerprint Store: a RAM-resident index facility
  - Mapping Table: LBA <==> PBA
  - Data Store: stores and retrieves data chunks

Implementation

Fixed-size chunking, a user-space and pure block-level implementation



PRELIMINARY RESULTS

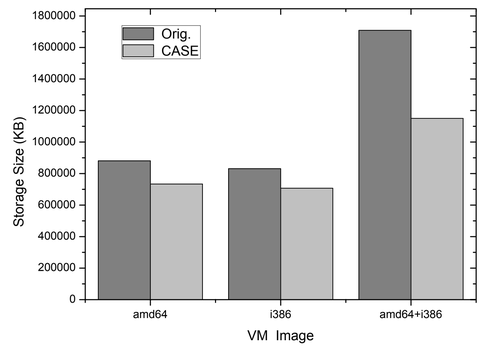
Trace-driven Evaluation

Saves space by 2 orders of magnitude

Trace	FS Size(MB)	CASE Storage(MB)	Storage Savings
web-vm	71,680	720.4	99.0%
mail	512,000	1148.1	99.8%
homes	481,280	1451.0	99.7%

Real-world workload-based Evaluation

Saves space by up to 33% if we copy both VM images



Acknowledgements: This work was supported by the US NSF under Grants IIS-0916859, CCF-0937993, CNS-1016609, and CNS-1116606.

Contacts: Lei Tian: tian@cse.unl.edu Hong Jiang: jiang@cse.unl.edu