SmartDedup: Optimizing Deduplication for Resource-constrained Devices

Qirui Yang, Runyu Jin, Ming Zhao

Arizona State University

http://visa.lab.asu.edu
Resource Management on Edge and IoT

- Limited on-device storage
  - Limited I/O performance
  - Limited capacity
  - Limited endurance

- Deduplication can
  - Eliminate redundant I/Os → improve performance
  - Reduce flash writes → improve endurance
  - Remove redundant data → improve utilization

- But
  - Is there enough data duplication in device workloads?
  - How to exploit it using limited resources on the device?
SmartDedup—A Smart Deduplication Solution for Smart Devices

- Cohesively designed in-memory and on-disk fingerprint stores
  - On-disk store complements the small in-memory store

- Synergistically integrated in-line and out-of-line deduplication
  - Out-of-line deduplicates data skipped by in-line

- Adaptive deduplication according to resource and battery availability
  - Dynamically enabled/disabled, dynamic processing rate
Trace Replay (on Nexus)

- Real-world device workloads have a good level of duplicates

- SmartDedup achieves
  - Up to 51.1% write speedup
  - Up to 70.9% write reduction

<table>
<thead>
<tr>
<th>Segment</th>
<th>Write (GB)</th>
<th>Duplication ratio (%)</th>
<th>Read/write ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.2</td>
<td>75.8</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>12.4</td>
<td>47.9</td>
<td>2.2</td>
</tr>
<tr>
<td>3</td>
<td>9.1</td>
<td>26.4</td>
<td>6.8</td>
</tr>
</tbody>
</table>

![Graph showing deduplication ratio, write speedup, and write reduction]
Welcome to Our Talk

SmartDedup

Thursday 3:50 pm

Track II: Deduplication Session