Motivation

24,888 Configurations

23% unstable (> 5% std. dev.)

Motivation (cont.)

- Stable performance is critical
  - Benchmarking
  - Inaccurate conclusions
  - Production systems
    - Clients’ satisfaction
- Performance variation is complicated
  - Various sources
  - Time-consuming to measure
- Storage stacks are important contributors
- First systematic study

Background

- Measurements for performance
  - Throughput
    - Window size matters
  - Latency

Background (cont.)

Measurements for variations

- **Relative Standard Deviation**
  \[ RSD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}}{\bar{x}} \]
- Understate impacts of outliers

- **Relative Range**
  \[ \text{Relative Range} = \frac{\text{max}(X) - \text{min}(X)}{\bar{x}} \]
- Understate impacts of outliers

- **Relative Standard Deviation**
  \[ RSD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}}{\bar{x}} \]
- Understate impacts of outliers

- **Relative Range**
  \[ \text{Relative Range} = \frac{\text{max}(X) - \text{min}(X)}{\bar{x}} \]
Background (cont.)

Measurements for variations

- Relative Standard Deviation
- (Coefficient of Variation)

\[ RSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2} \]

- Understate impacts of outliers

\[ \text{Relative Range} = \frac{\text{max}(X) - \text{min}(X)}{\overline{x}} \]

Methodology

- Important parameters
  - File Systems
    - Ext4, XFS, Btrfs
  - F/S parameters
    - Block Size, Inode Size, etc.
  - I/O Schedulers
  - Devices: HDDs, SSDs
  - Latin Hypercube Sampling
  - Explore many-factor systems

Experiment Settings

- Machines
  - 1 Intel Xeon quad-core 2.4GHz CPU, 4GB RAM
  - Devices: SAS, SATA, SSD
- Filebench
  - Workloads
    - Mailserver, Fileserver, Webserver
  - Fileset size: 10GB (2.5 × RAM size)
  - Running time
    - 2,000s for Mailserver; otherwise 800s
- 10+ runs for each experiment

Performance Variation: Overview

- Workload: Mailserver
On the Performance Variation in Modern Storage Stacks (FAST'17)

3/2/2017

Workload: Mailserver

Ext4
- Up to 42%
- Most above 20%

XFS
- Most under 20%

Btrfs
- Most under 15%

Sampled Configurations
- Mailserver
- Fileserver
- Webserver
Performance Variation: Overview (cont.)

Mailserver  Fileserver  Webserver

Sampled Configurations

Relative Range

0% 10% 20% 30% 40%

1 2 3 4 5 6 7 8 9 10

3/2/2017  On the Performance Variation in Modern Storage Stacks (FAST'17)
Ext4 Case Study

**Throughput (IOPS)**

<table>
<thead>
<tr>
<th>baseline</th>
<th>+no_lazy</th>
<th>mount</th>
</tr>
</thead>
<tbody>
<tr>
<td>2048-2048-writeback-noatime-noop-SATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22%</td>
<td>47%</td>
<td></td>
</tr>
</tbody>
</table>

Workload: Fileserver

3/2/2017 On the Performance Variation in Modern Storage Stacks (FAST'17)

---

Ext4 Case Study

**Throughput (IOPS)**

<table>
<thead>
<tr>
<th>baseline</th>
<th>+no_lazy</th>
<th>mount</th>
</tr>
</thead>
<tbody>
<tr>
<td>2048-2048-writeback-noatime-noop-SATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22%</td>
<td>47%</td>
<td></td>
</tr>
</tbody>
</table>

Workload: Fileserver

3/2/2017 On the Performance Variation in Modern Storage Stacks (FAST'17)

---

Ext4 Case Study

**Throughput (IOPS)**

<table>
<thead>
<tr>
<th>baseline</th>
<th>+no_lazy</th>
<th>mount</th>
</tr>
</thead>
<tbody>
<tr>
<td>2048-2048-writeback-noatime-noop-SATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22%</td>
<td>47%</td>
<td></td>
</tr>
</tbody>
</table>

Workload: Fileserver

3/2/2017 On the Performance Variation in Modern Storage Stacks (FAST'17)

---
Ext4 Case Study (cont.)

Workload: Fileserver

- baseline
- +no_lazy
  - mkfs.ext4 -E lazy_itable_init=0
- +umount
  - sync → umount + mount
- +alloc
  - Preserve layout
  - e.g., dd if=/dev/sde1

Throughput (IOPS)

Block Number

default alloc. policy

3/2/2017 On the Performance Variation in Modern Storage Stacks (FAST'17)
Ext4 Case Study (cont.)

Workload: Fileserver

3/2/2017  On the Performance Variation in Modern Storage Stacks (FAST'17)

Throughput vs. Time

Workload: Fileserver
Window Size: 120s

3/2/2017  On the Performance Variation in Modern Storage Stacks (FAST'17)

Throughput vs. Time

Workload: Fileserver
Window Size: 120s

3/2/2017  On the Performance Variation in Modern Storage Stacks (FAST'17)
Throughput vs. Time

Workload: Fileserver
Window Size: 120s

Throughput vs. Time (cont.)

Cum. Pct. of Time Windows (%)
Relative Range (%)

60s 120s 180s 240s

Ext4
Workload: Fileserver
Throughput vs. Time (cont.)

Cum. Pct. of Time Windows (%)

Relative Range (%)

60s
120s
180s
240s
400s

Ext4 Workload: Fileserver

3/2/2017 On the Performance Variation in Modern Storage Stacks (FAST'17)

Throughput vs. Time (cont.)

Cum. Pct. of Time Windows (%)

Relative Range (%)

60s
120s
180s
240s
400s

Ext4 Workload: Fileserver XFS

3/2/2017 On the Performance Variation in Modern Storage Stacks (FAST'17)
Conclusions

- First systematic study of storage performance variations
- Characterization from various perspectives
- Randomization in file systems
- Lazy initialization and background activities contribute to instability
- We will make datasets public

Future Work

- Aged file systems
- More workloads/benchmark tools
- More Devices
  - Near full-capacity SSDs
  - PCM, SMR, etc.
- More storage layers
  - RAID, LVM, etc.
On the Performance Variation in Modern Storage Stacks

Zhen Cao¹, Vasily Tarasov², Hari Prasath Raman¹, Dean Hildebrand², and Erez Zadok¹

Thank You

Q&A