

# On the Performance Variation in Modern Storage Stacks

## 15<sup>th</sup> USENIX Conference on File and Storage Technologies 2017

Zhen Cao<sup>1</sup>, Vasily Tarasov<sup>2</sup>, Hari Prasath Raman<sup>1</sup>,  
Dean Hildebrand<sup>2</sup>, and Erez Zadok<sup>1</sup>

<sup>1</sup>Stony Brook University; <sup>2</sup>IBM Research – Almaden;



# 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**
- ◆ **(Coefficient of Variation)**

$$RSD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{\bar{x}}$$

- ◆ Understate impacts of outliers

- ◆ **Relative Range**

$$RelativeRange = \frac{max(X) - min(X)}{\bar{x}}$$

## Background (cont.)

### Measurements for variations

- ◆ **Relative Standard Deviation**
- ◆ **(Coefficient of Variation)**

$$RSD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{\bar{x}}$$

- ◆ Understate impacts of outliers

- ◆ **Relative Range**

$$RelativeRange = \frac{max(X) - min(X)}{\bar{x}}$$

## Background (cont.)

### Measurements for variations

- ◆ **Relative Standard Deviation**
- ◆ **(Coefficient of Variation)**

$$RSD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}}{\bar{x}}$$

- ◆ Understate impacts of outliers

- ◆ **Relative Range** ✓

$$RelativeRange = \frac{\max(X) - \min(X)}{\bar{x}}$$

## 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

1,700+ Configurations  
1.5 Years

## 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

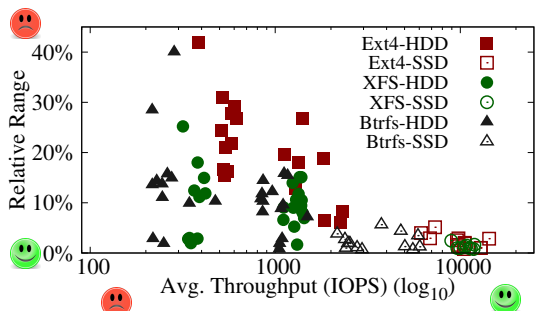
1,700+ Configurations  
1.5 Years

100+ Configurations  
1 Month

## Experiment Settings

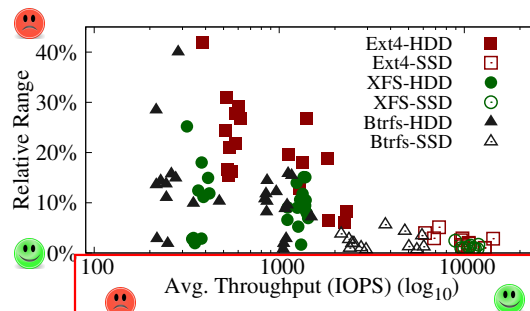
- Machines
  - ◆ 1 Intel Xeon quad-core 2.4GHz CPU, 4GB RAM
  - ◆ Devices: SAS, SATA, SSD
- Filebench
  - ◆ Workloads
    - Mailserver, Fileserver, Webserver
  - ◆ Fileset size: 10GB (2.5 × RAM size)
  - ◆ Running time
    - 2,000s for Mailserver; otherwise 800s
- 10+ runs for each experiment

## Performance Variation: Overview



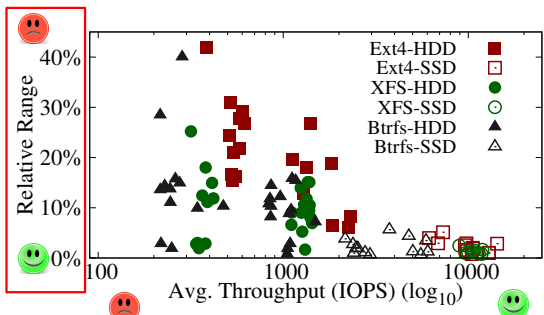
Workload: Mailserver

## Performance Variation: Overview



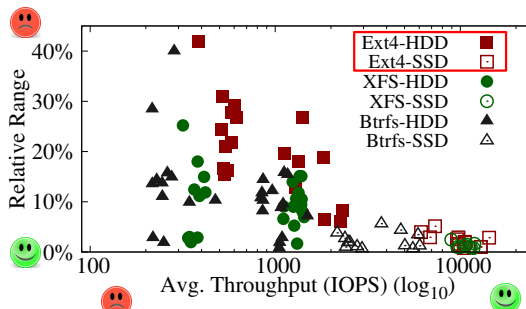
Workload: Mailserver

# Performance Variation: Overview



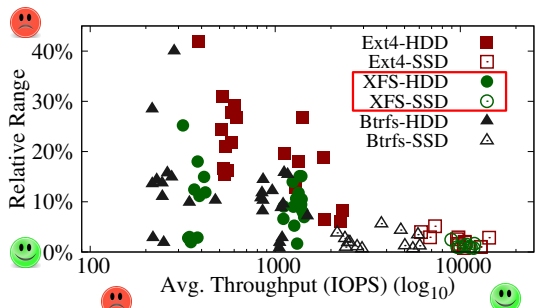
Workload: Mailserver

# Performance Variation: Overview



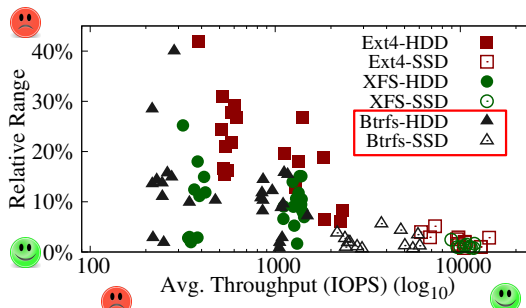
Workload: Mailserver

# Performance Variation: Overview



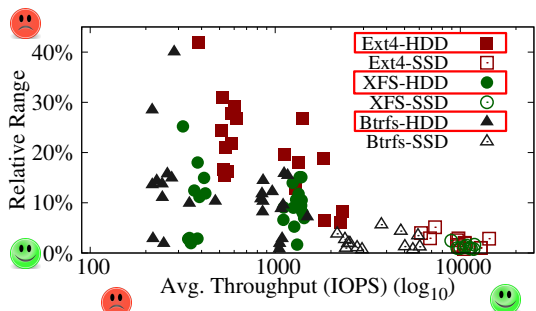
Workload: Mailserver

# Performance Variation: Overview



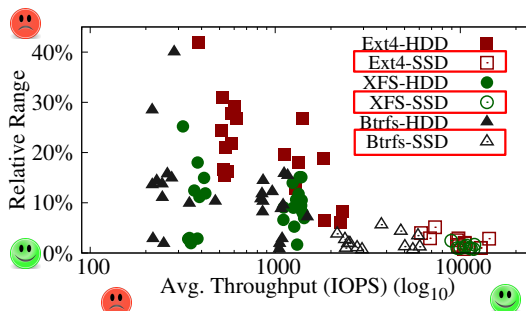
Workload: Mailserver

# Performance Variation: Overview



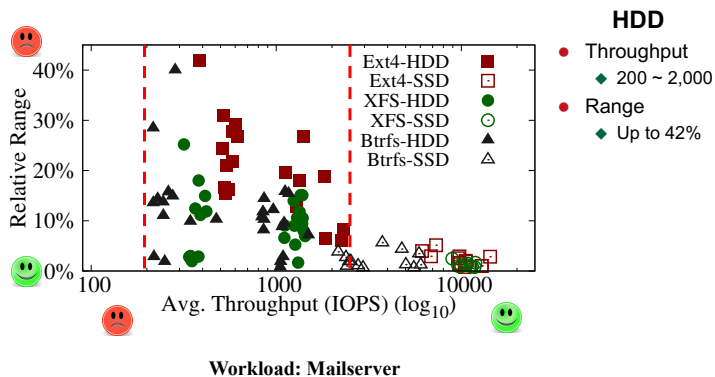
Workload: Mailserver

# Performance Variation: Overview

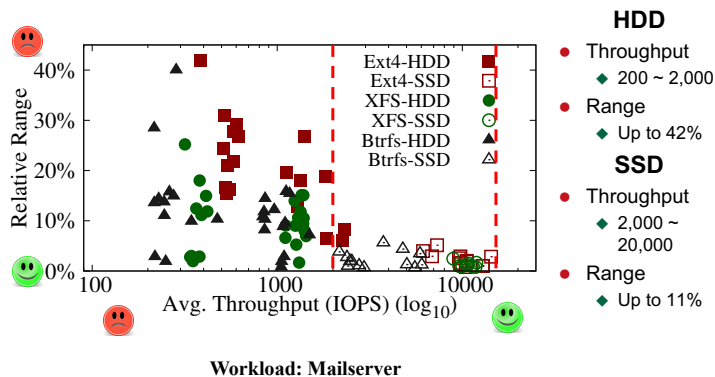


Workload: Mailserver

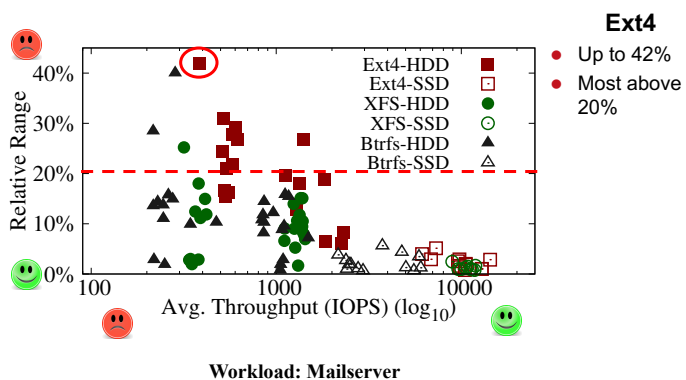
## Performance Variation: Overview



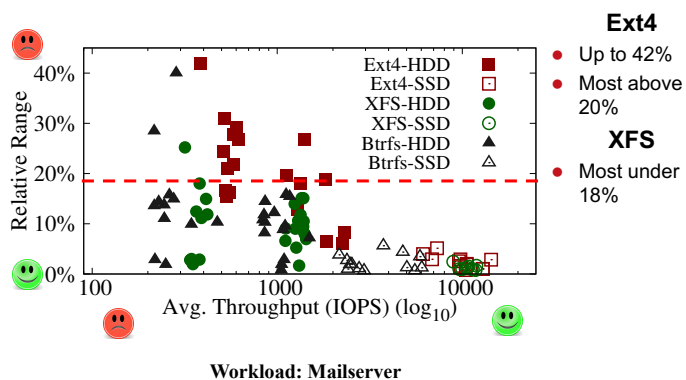
## Performance Variation: Overview



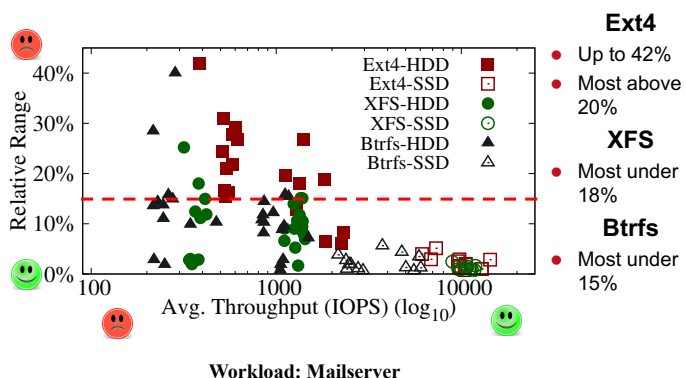
## Performance Variation: Overview



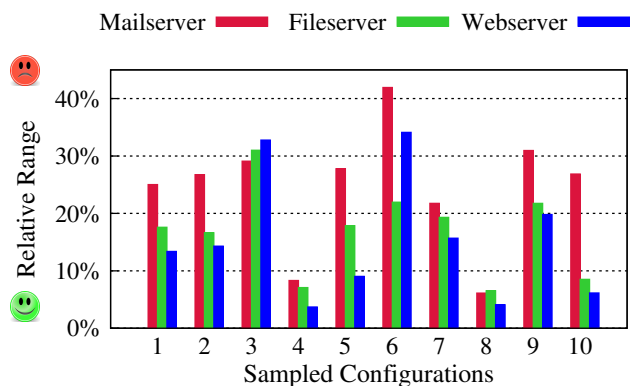
## Performance Variation: Overview



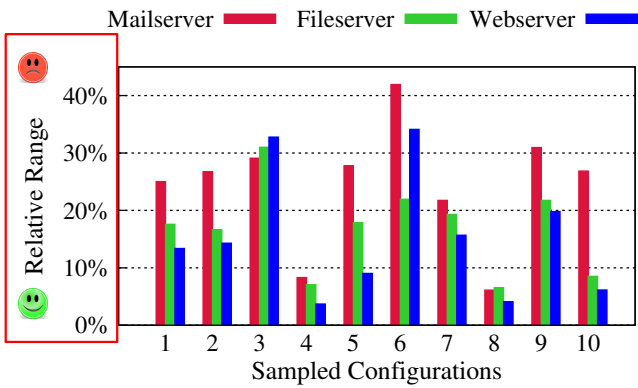
## Performance Variation: Overview



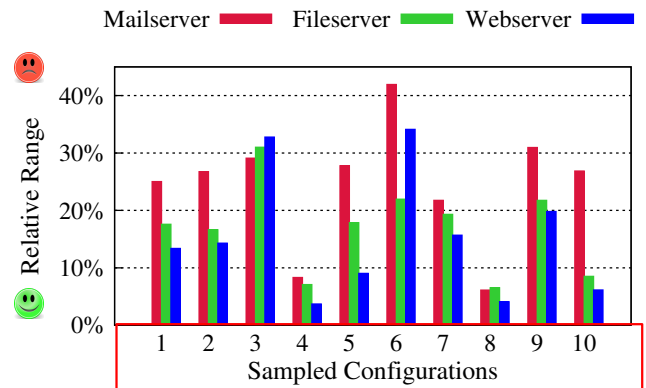
## Performance Variation: Overview (cont.)



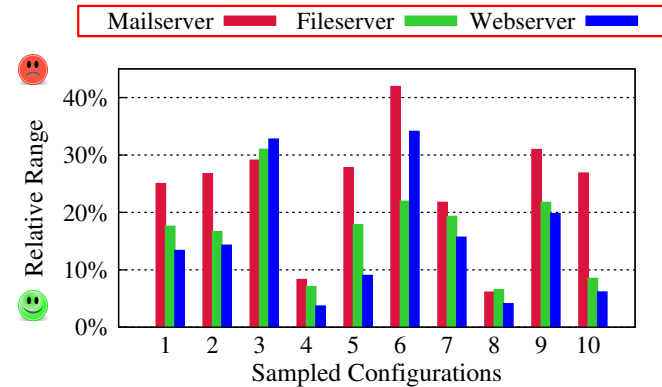
## Performance Variation: Overview (cont.)



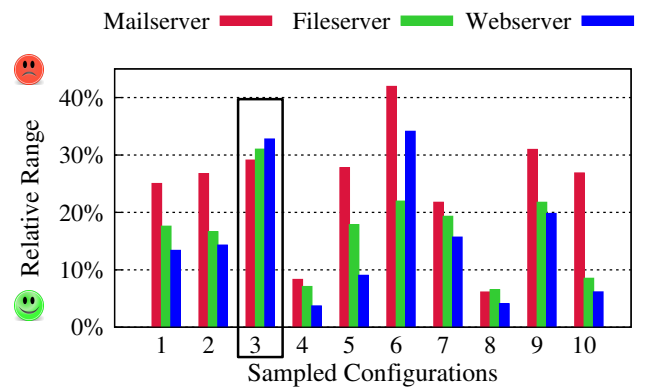
## Performance Variation: Overview (cont.)



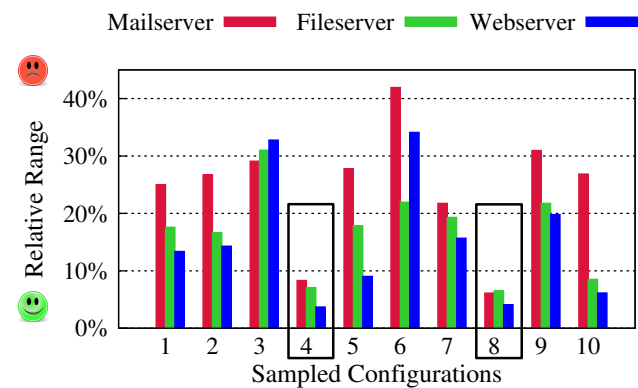
## Performance Variation: Overview (cont.)



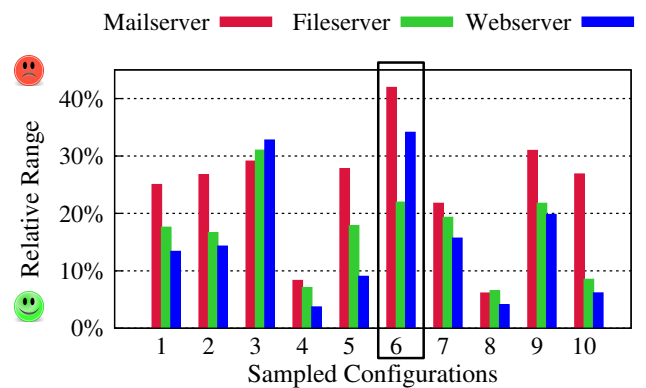
## Performance Variation: Overview (cont.)



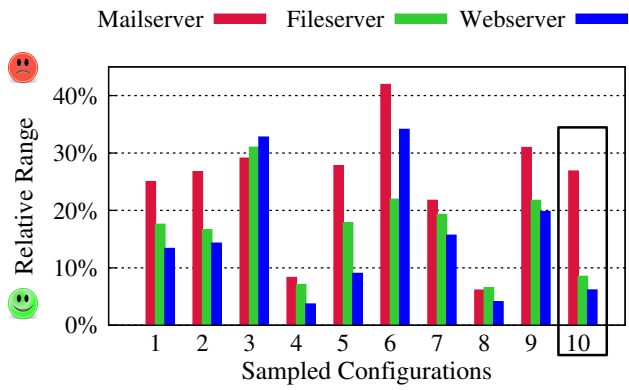
## Performance Variation: Overview (cont.)



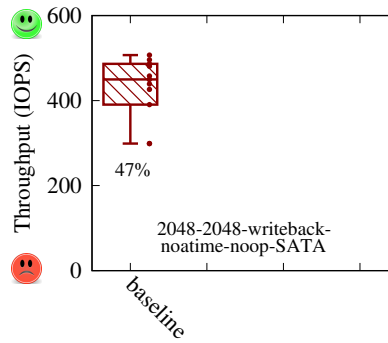
## Performance Variation: Overview (cont.)



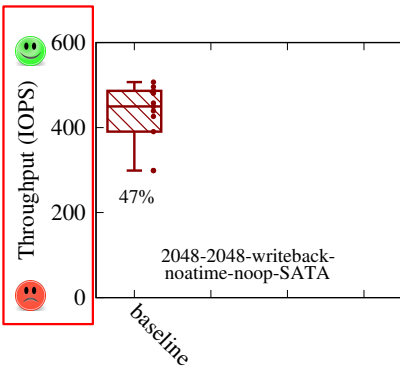
# Performance Variation: Overview (cont.)



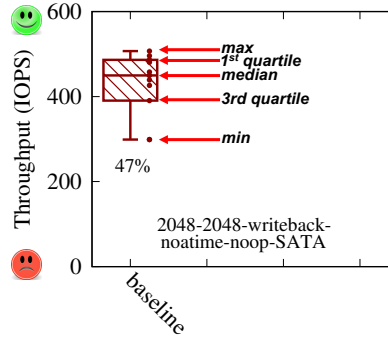
# Ext4 Case Study



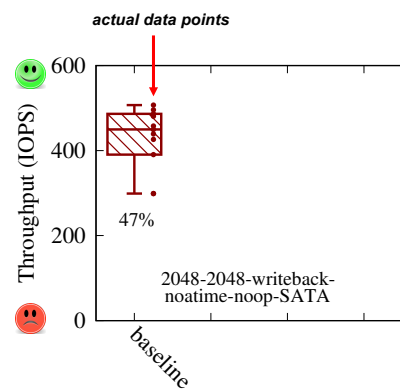
# Ext4 Case Study



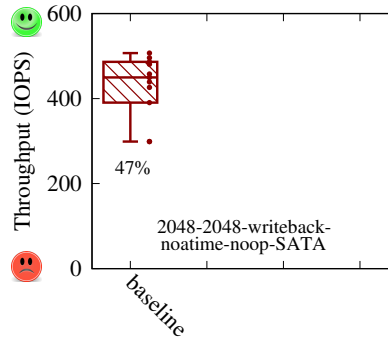
# Ext4 Case Study



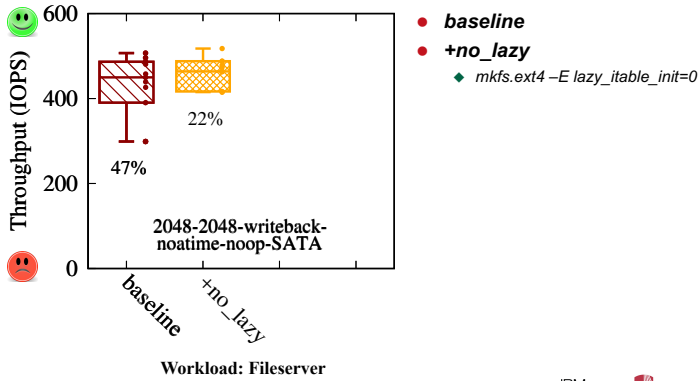
# Ext4 Case Study



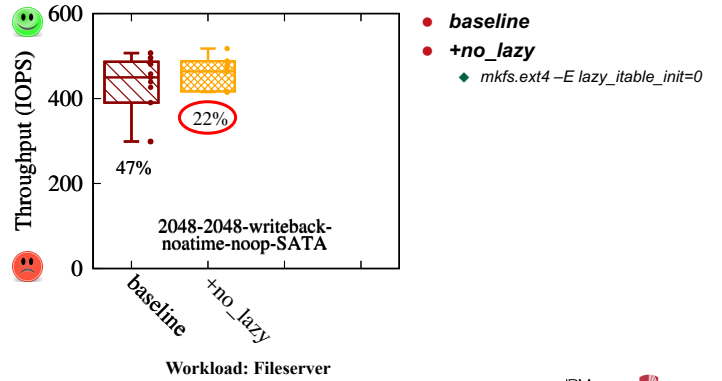
# Ext4 Case Study



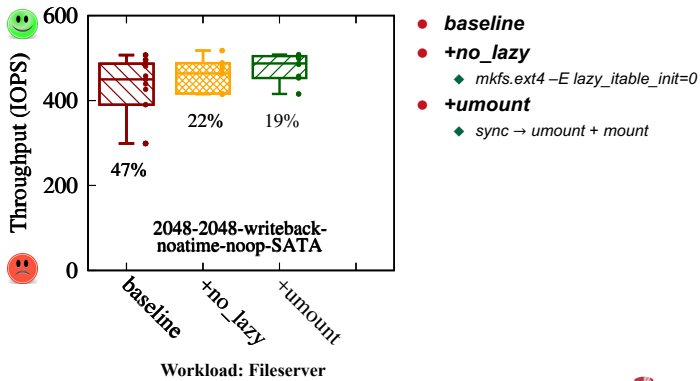
## Ext4 Case Study



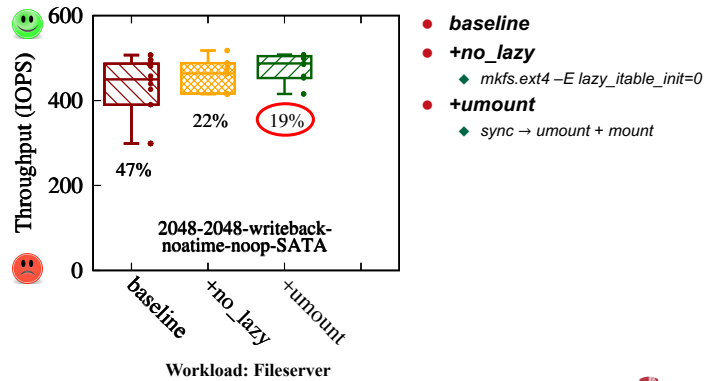
## Ext4 Case Study



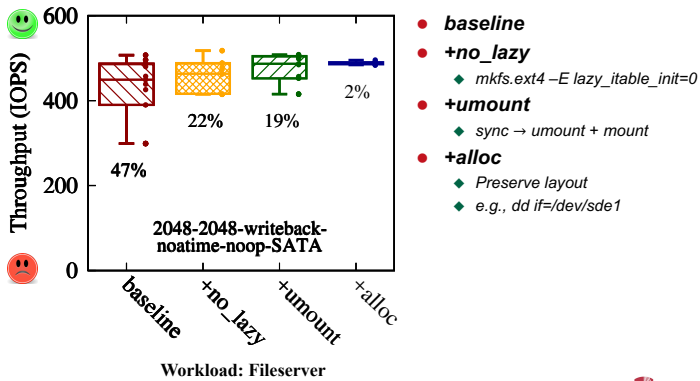
## Ext4 Case Study



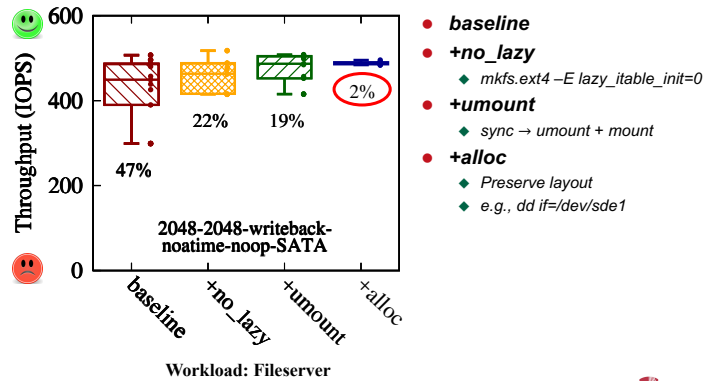
## Ext4 Case Study



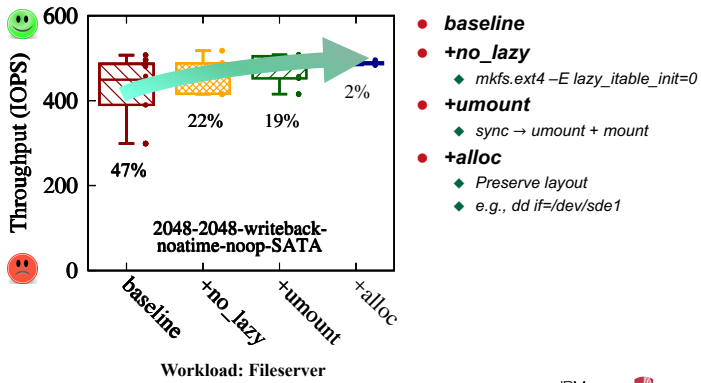
## Ext4 Case Study



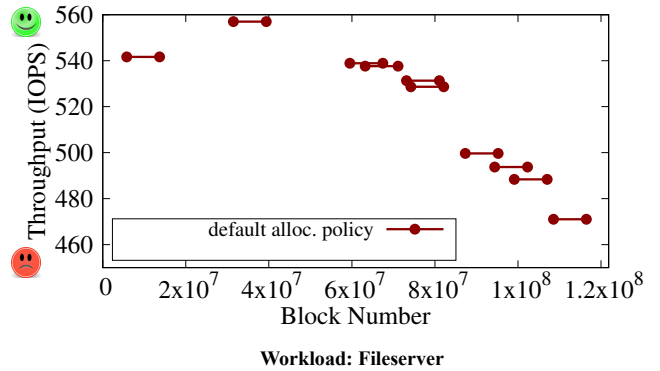
## Ext4 Case Study



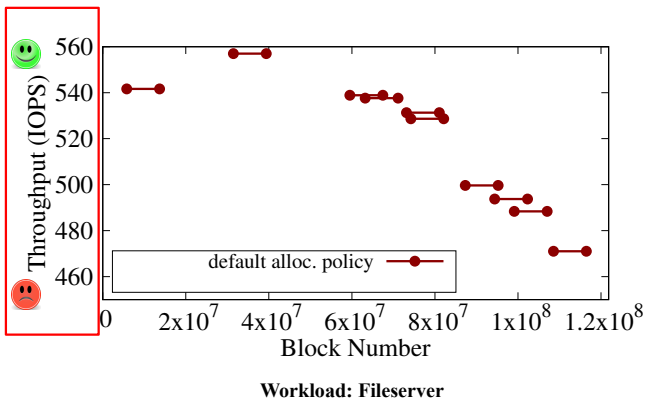
## Ext4 Case Study



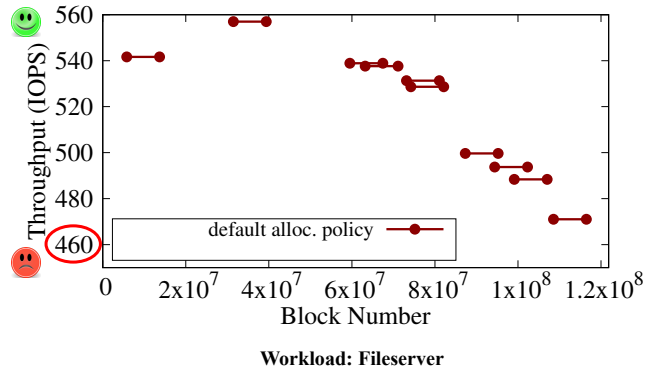
## Ext4 Case Study (cont.)



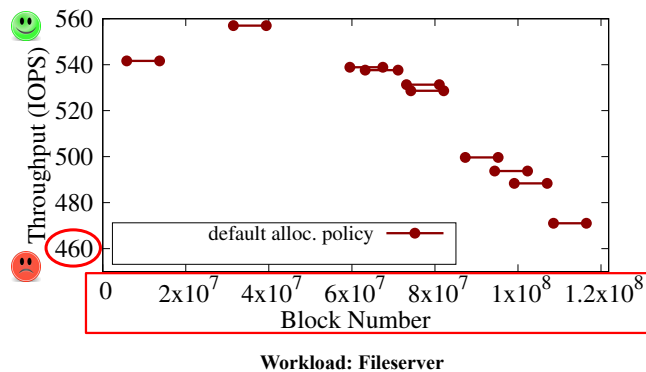
## Ext4 Case Study (cont.)



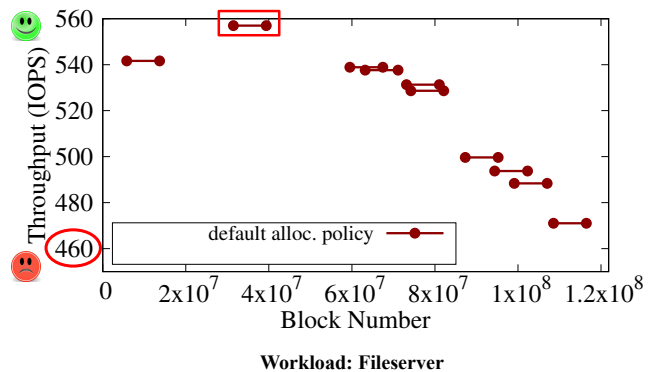
## Ext4 Case Study (cont.)



## Ext4 Case Study (cont.)

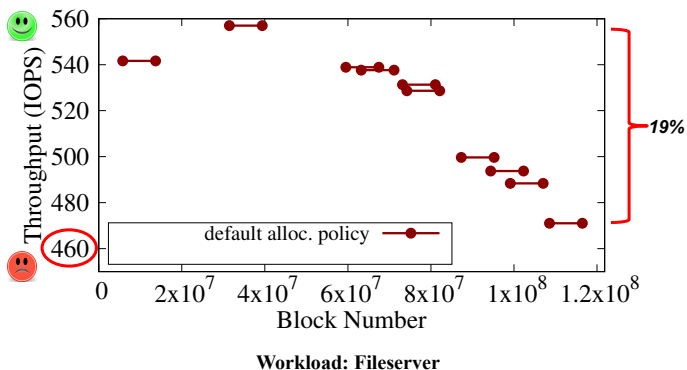


## Ext4 Case Study (cont.)





## Ext4 Case Study (cont.)



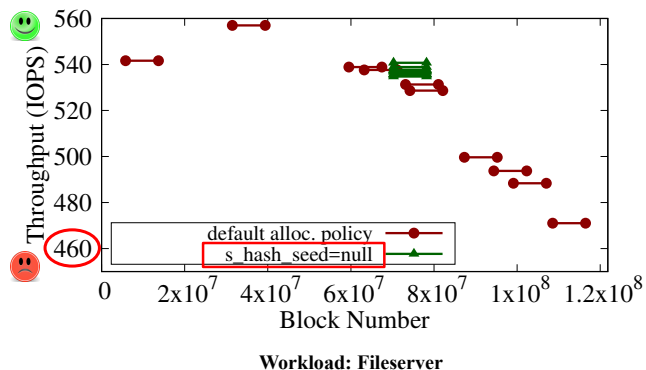
3/2/2017

On the Performance Variation in Modern Storage Stacks (FAST'17)

49



## Ext4 Case Study (cont.)



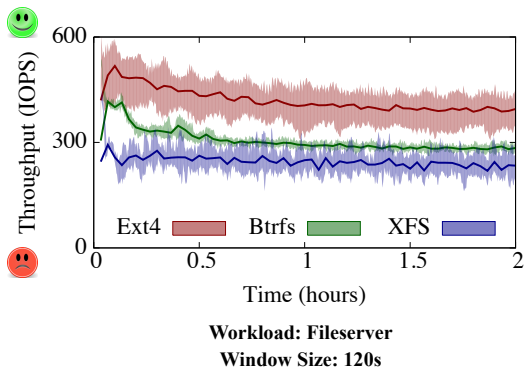
3/2/2017

On the Performance Variation in Modern Storage Stacks (FAST'17)

50



## Throughput vs. Time



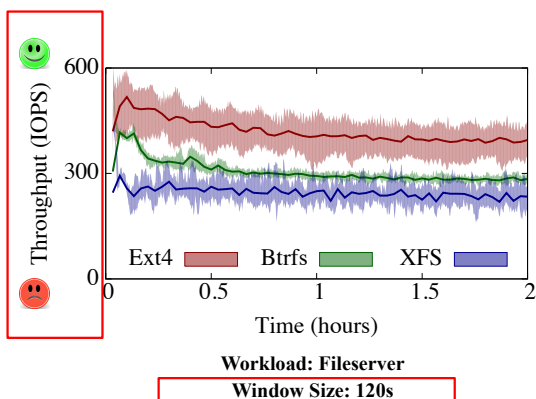
3/2/2017

On the Performance Variation in Modern Storage Stacks (FAST'17)

51



## Throughput vs. Time



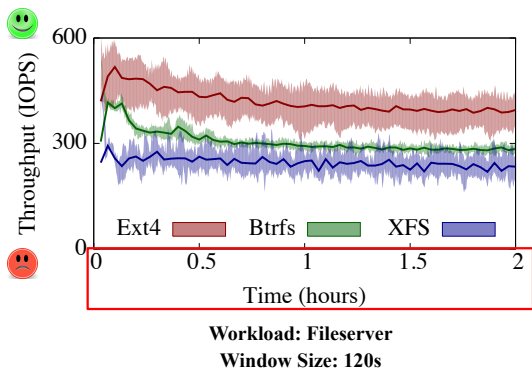
3/2/2017

On the Performance Variation in Modern Storage Stacks (FAST'17)

52



## Throughput vs. Time



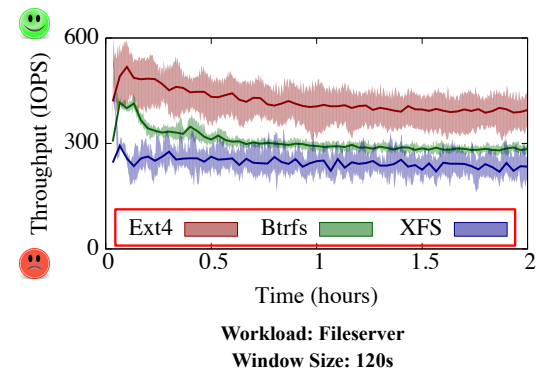
3/2/2017

On the Performance Variation in Modern Storage Stacks (FAST'17)

53



## Throughput vs. Time



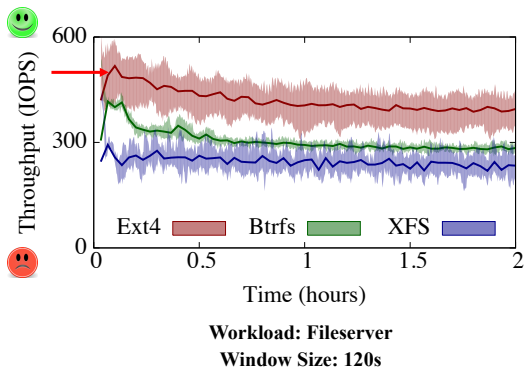
3/2/2017

On the Performance Variation in Modern Storage Stacks (FAST'17)

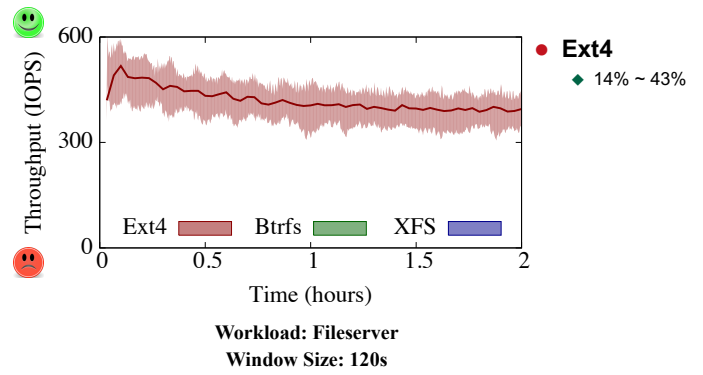
54



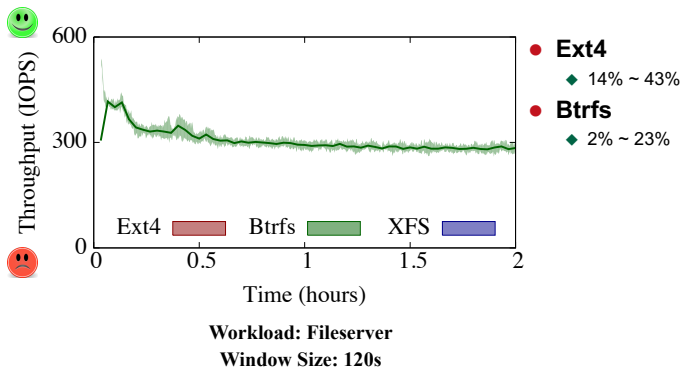
## Throughput vs. Time



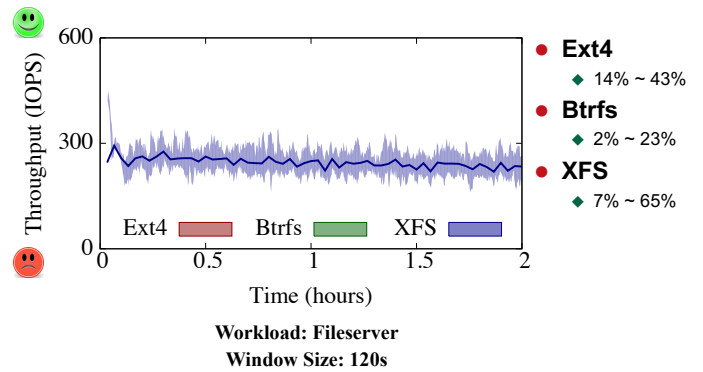
## Throughput vs. Time



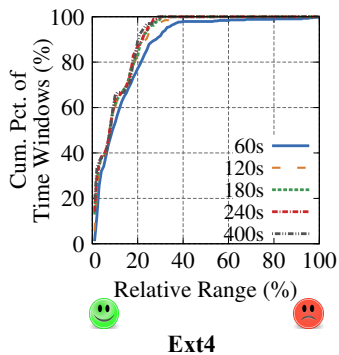
## Throughput vs. Time



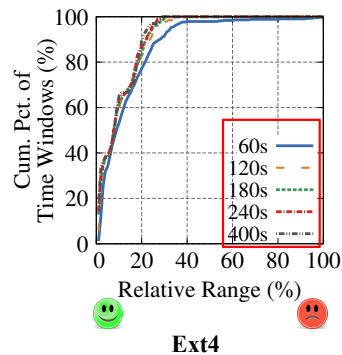
## Throughput vs. Time



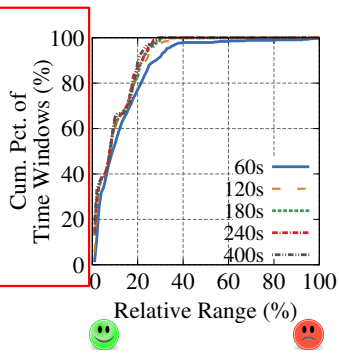
## Throughput vs. Time (cont.)



## Throughput vs. Time (cont.)



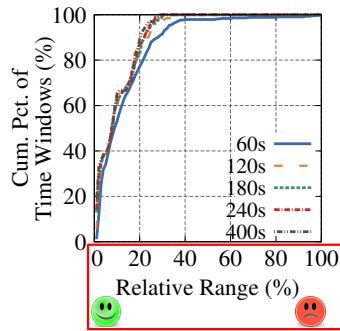
## Throughput vs. Time (cont.)



Ext4

Workload: Fileserver

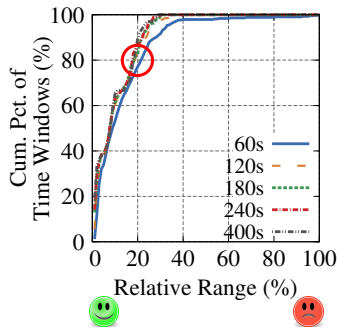
## Throughput vs. Time (cont.)



Ext4

Workload: Fileserver

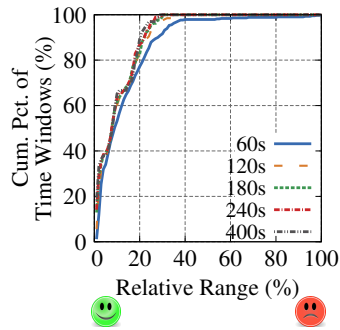
## Throughput vs. Time (cont.)



Ext4

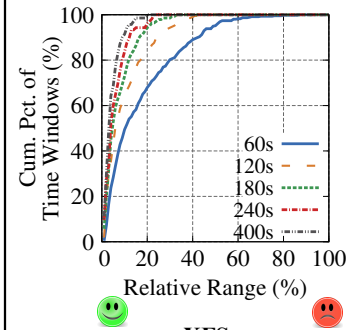
Workload: Fileserver

## Throughput vs. Time (cont.)



Ext4

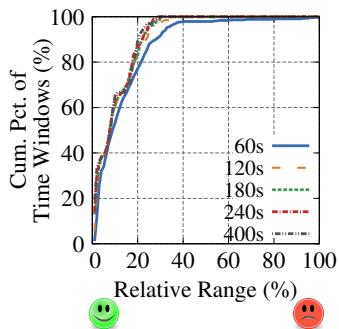
Workload: Fileserver



XFS

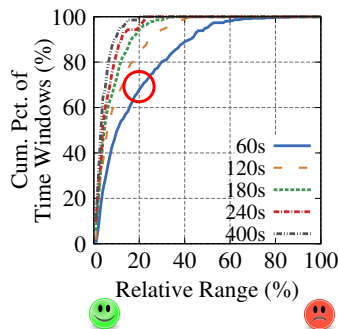
Workload: Fileserver

## Throughput vs. Time (cont.)



Ext4

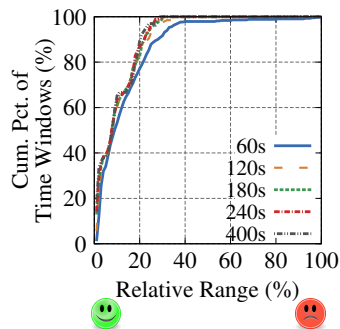
Workload: Fileserver



XFS

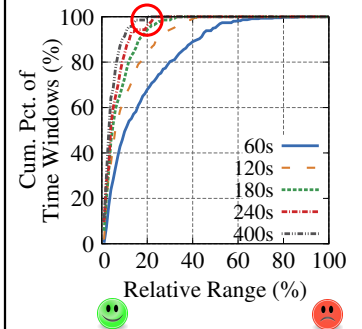
Workload: Fileserver

## Throughput vs. Time (cont.)



Ext4

Workload: Fileserver

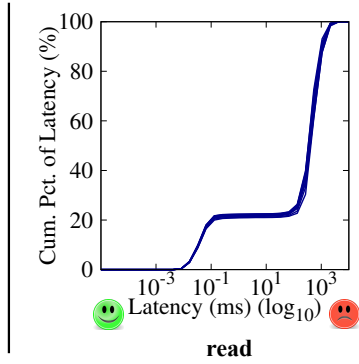


XFS

Workload: Fileserver

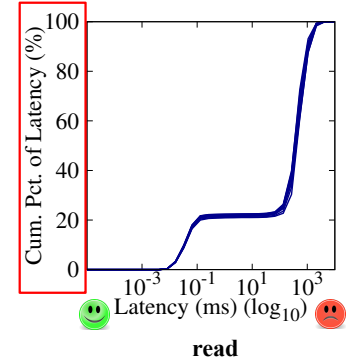
## Latency Variation

- Ext4 - HDD
  - ◆ 32%
- Fileserver
  - ◆ create
  - ◆ delete
  - ◆ open
  - ◆ read
  - ◆ stat
  - ◆ write



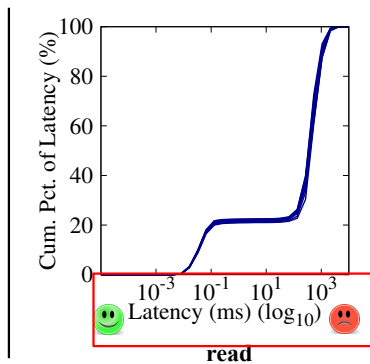
## Latency Variation

- Ext4 - HDD
  - ◆ 32%
- Fileserver
  - ◆ create
  - ◆ delete
  - ◆ open
  - ◆ read
  - ◆ stat
  - ◆ write

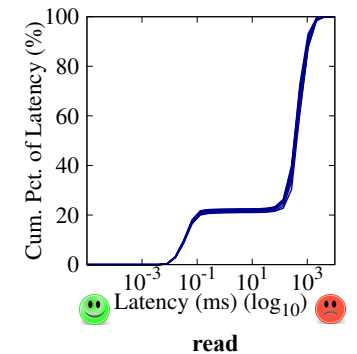
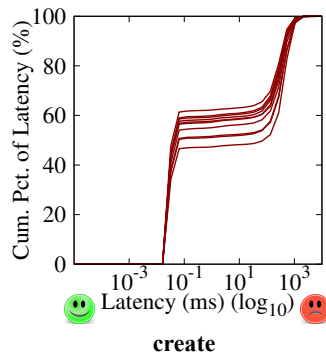


## Latency Variation

- Ext4 - HDD
  - ◆ 32%
- Fileserver
  - ◆ create
  - ◆ delete
  - ◆ open
  - ◆ read
  - ◆ stat
  - ◆ write



## Latency Variation



## 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 Cao<sup>1</sup>, Vasily Tarasov<sup>2</sup>, Hari Prasath Raman<sup>1</sup>, Dean Hildebrand<sup>2</sup>, and  
Erez Zadok<sup>1</sup>

# Thank You Q&A

