

A Reliability Analysis Framework for Cloud Storage Systems

Mai Zheng[†], Joseph Tucek[‡], Feng Qin[†], Mark Lillibridge[‡]

[†] *The Ohio State University*



[‡] *HP Labs*



Reliability of Storage Systems is Important



Storage Devices are Imperfect

- Both traditional hard drives and modern solid-state drives (SSDs) may fail in unexpected ways
 - e.g., latent sector errors, torn writes, checksum mismatch, unserializable writes, metadata corruption, etc.

[Bairavasundaram et al. SIGMETRICS'07, FAST'08],

[Jiang et al. FAST'08],

[Zheng et al. FAST'13], etc.



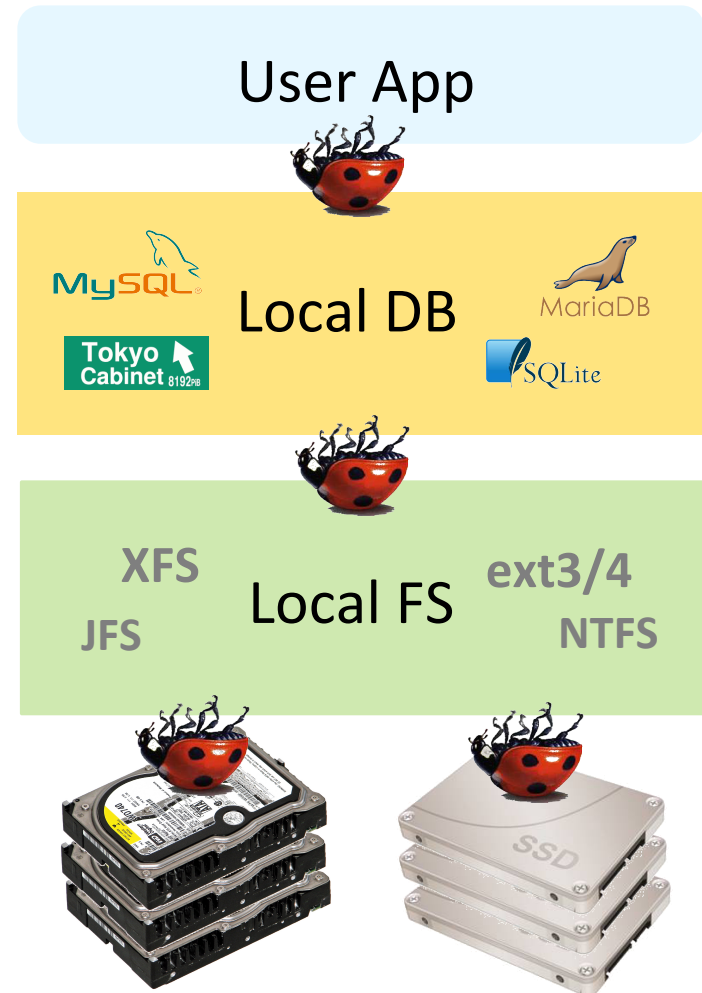
Local Storage Software is Imperfect, too

- Big gaps of understanding/assumptions between different layers
 - e.g., behavior of different configurations/journaling modes
 - e.g., behavior of *mmap* and *fdatasync*

[Pillai *et al.* OSDI'14],

[Zheng *et al.* OSDI'14],

etc.



We have found a way to thoroughly analyze local storage systems under failures

Carefully designed workloads to stress certain functionality (e.g., concurrency control)

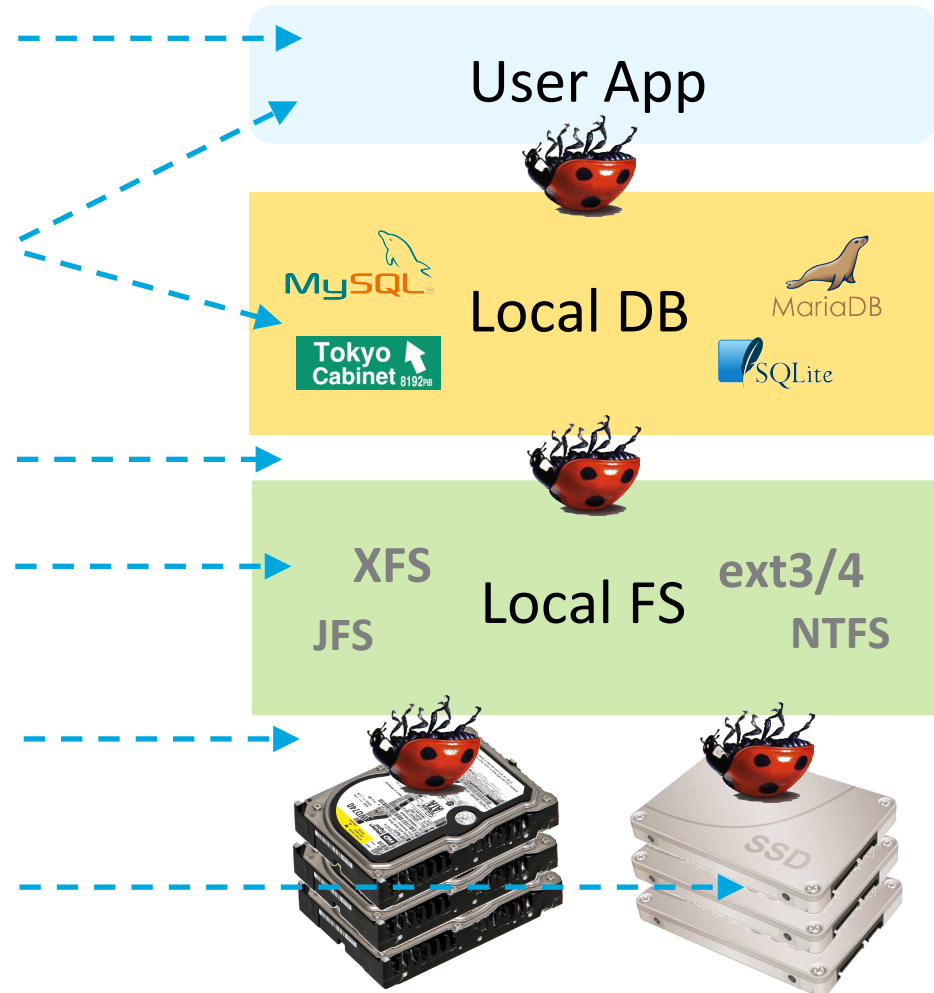
Instrumentation to gather **function call** trace

Instrumentation to gather **system call** trace

probe FS to gather **file** info.

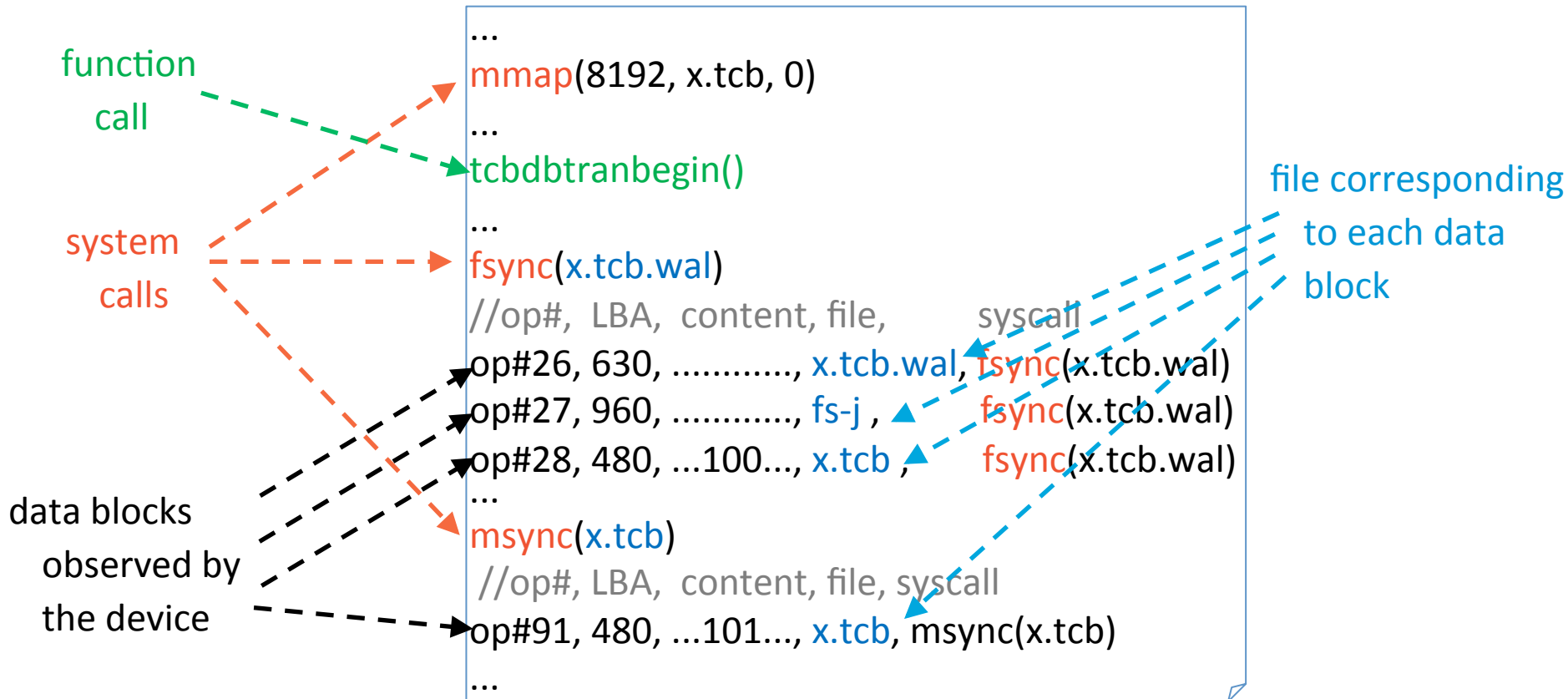
Customized driver to gather **SCSI commands**; replay to simulate failure patterns

real failure patterns collected from devices; allow reproducing real failures



We have found a way to thoroughly analyze local storage systems under failures

- Combined info. provides a clear whole picture of complicated scenarios



We are extending to cloud

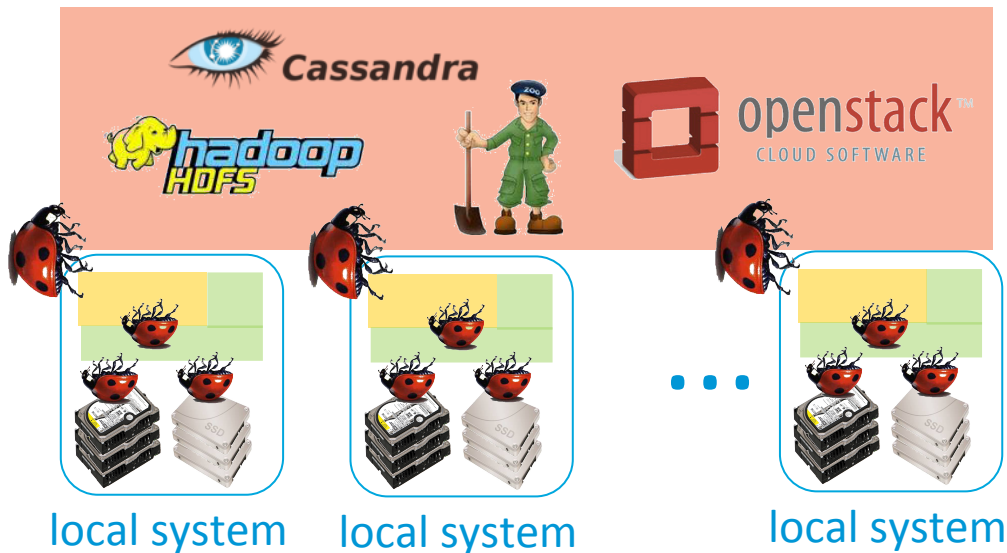
- Existing work has found many protocol-level bugs in cloud systems
 - mainly under three simulated failure modes: message reordering, machine crashes/reboot, network partitioning
 - focus on testing



[Yang *et al.* NSDI'09],
[Ju *et al.* SOCC'13],
[Leesatapornwongsa *et al.*
OSDI'14], etc.

We are extending to cloud

- We are different
 - more fine-grained local failure modes based on previous studies
 - failure propagation chain along data path from local to cloud
 - whole picture to help diagnosis/tuning/etc. (e.g., complete tracing down to local SCSI commands)



What we need form a testbed

- Hardware
 - different types of local storage devices: hard disks, SSDs, hybrid
 - hardware RAID (and the privilege to change configuration)
- Software
 - capability to install/config/instrument software along the data path
(E.g., iSCSI driver, software RAID, local file system, local backend databases, cloud management systems)
 - synchronization service

What we need form a testbed

- Hardware
 - different types of local storage devices: hard disks, SSDs, hybrid
 - hardware RAID (and the privilege to change configuration)
- Software
 - capability to install/config/instrument software along the data path (E.g., iSCSI driver, software RAID, local file system, local backend databases, cloud management systems)
 - synchronization service



Thank you!

