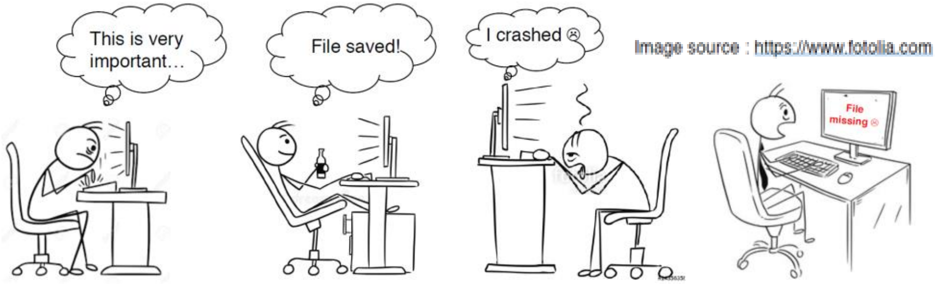
Iowa State University
Data Storage Lab

HotStorage '20
usenix ASSOCIATION
JULY 13–14, 2020

On Failure Diagnosis of the Storage Stack

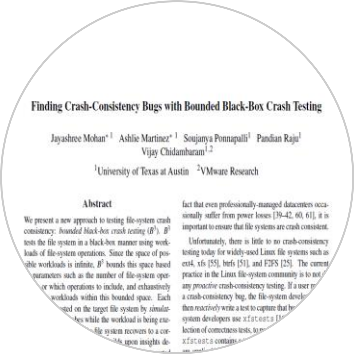
Duo Zhang, Om Rameshwar Gatla, Runzhou Han, Mai Zheng
Iowa State University

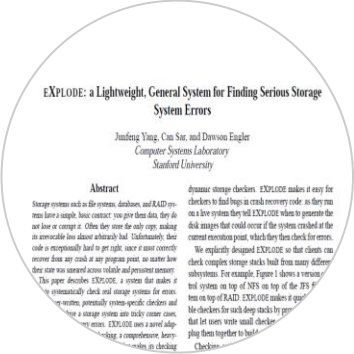
Storage System Failures Are Troublesome

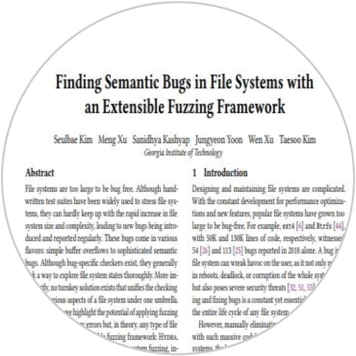


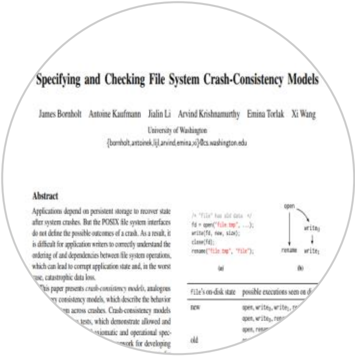
Existing Efforts Are Not Enough

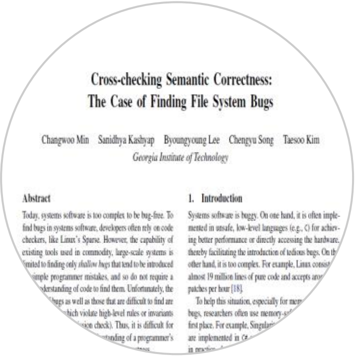
- Mostly focus on *testing*
  - Require a special *testing environment*
    - e.g., a customized kernel
  - Still cannot prevent all failures in *production environment*





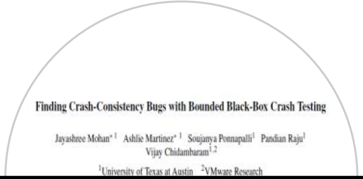






Existing Efforts Are Not Enough

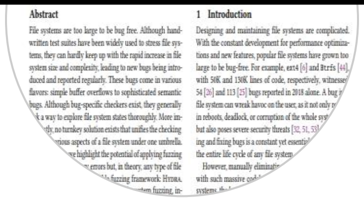
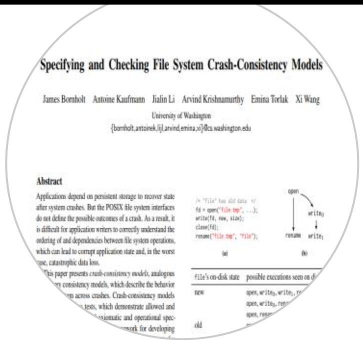
- Mostly focus on *testing*
  - Require a special *testing environment*
    - e.g., a customized kernel
  - Still cannot prevent all failures in *production environment*


Finding Crash-Consistency Bugs with Bounded Black-Box Crash Testing

Jayashree Mohan[*1] Ashlie Martinez[*1] Soujanya Ponnapalli[1] Pandian Raju[1] Vijay Chidambaram[1,2]

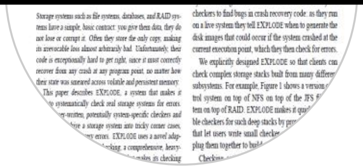
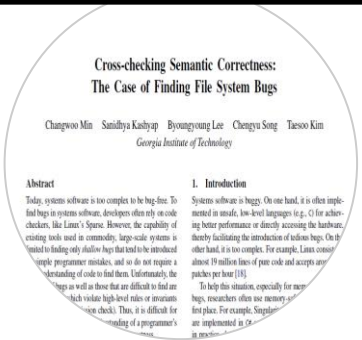
[1]University of Texas at Austin [2]VMware Research

What to do if failures happen ?



Specifying and Checking File System Crash-Consistency Models

James Bornholt Antoine Kaufmann Jialin Li Arvind Krishnamurthy Emina Torlak Xi Wang

University of Washington

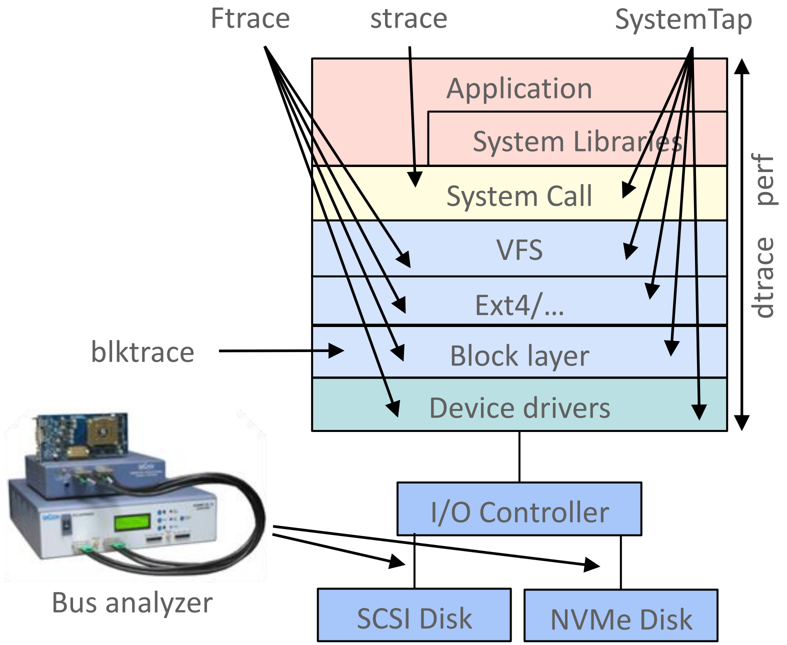


Cross-checking Semantic Correctness: The Case of Finding File System Bugs

Changwoo Min Sanidhya Kashyap Byoungyoung Lee Chengyu Song Taesoo Kim

Georgia Institute of Technology

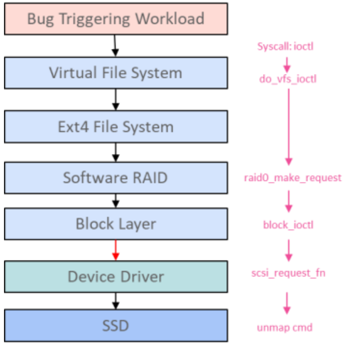
Practical Diagnosis Tools & Limitations

- Practical diagnosis tools
  - Software-based
    - e.g., GDB, SystemTap, Ftrace
  - Hardware-based
    - e.g., Bus analyzer
- Limitations
  - Require substantial manual efforts
    - e.g., GDB single-stepping
  - Require special hardware
  - Only cover partial storage stack



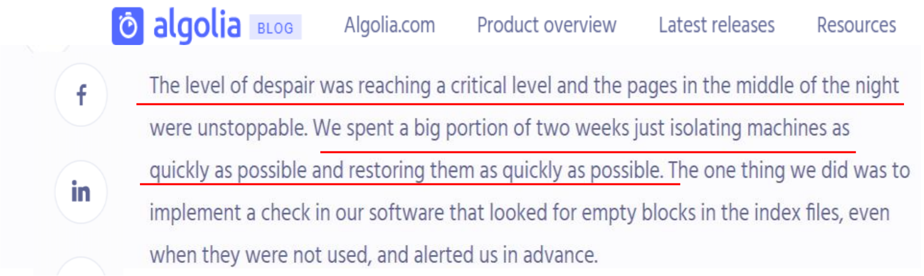
A Real-World Case: Diagnosis Is Challenging

- Algolia data center incident:
  - Servers crashed and files corrupted for unknown reason
  - After weeks of diagnosis, Samsung SSDs were mistakenly blamed
  - After one month, a Linux kernel bug was identified as root cause



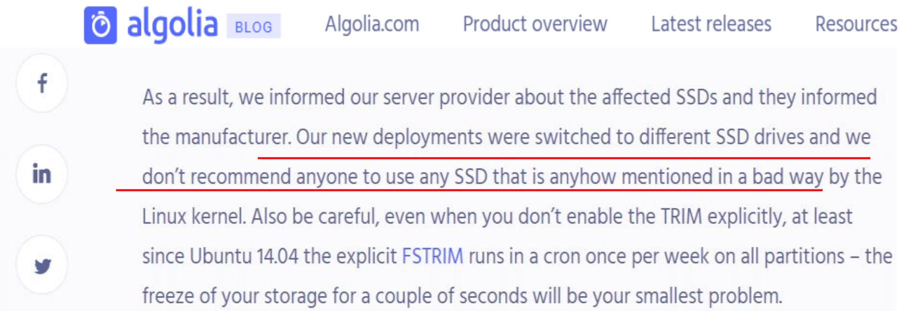
When Solid State Drives are not that solid

Adam Surak | Jun 15th 2015 | 12 min read | Engineering



algolia BLOG Algolia.com Product overview Latest releases Resources

The level of despair was reaching a critical level and the pages in the middle of the night were unstoppable. We spent a big portion of two weeks just isolating machines as quickly as possible and restoring them as quickly as possible. The one thing we did was to implement a check in our software that looked for empty blocks in the index files, even when they were not used, and alerted us in advance.



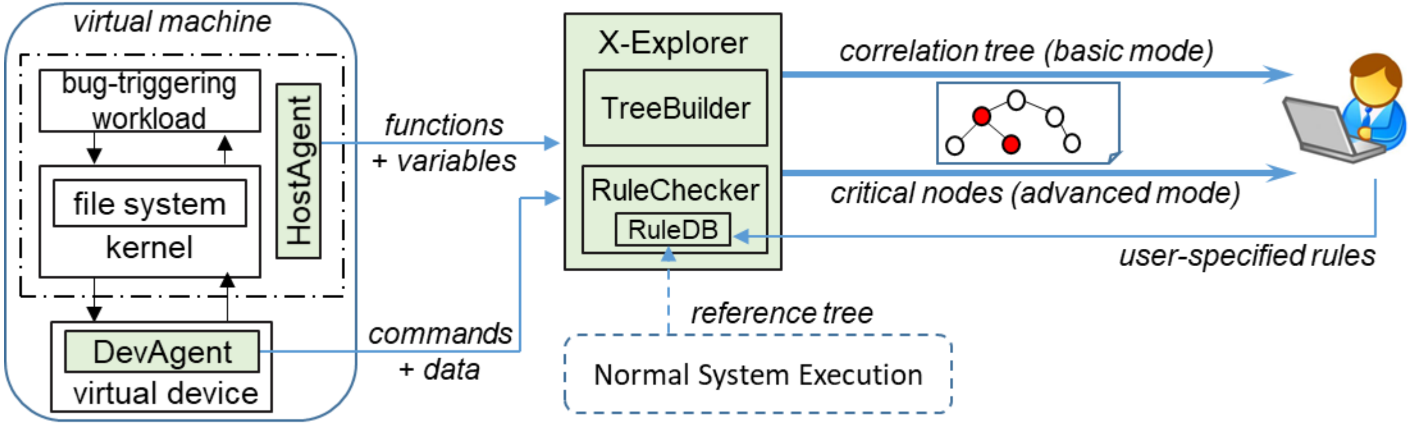
algolia BLOG Algolia.com Product overview Latest releases Resources

As a result, we informed our server provider about the affected SSDs and they informed the manufacturer. Our new deployments were switched to different SSD drives and we don't recommend anyone to use any SSD that is anyhow mentioned in a bad way by the Linux kernel. Also be careful, even when you don't enable the TRIM explicitly, at least since Ubuntu 14.04 the explicit FSTRIM runs in a cron once per week on all partitions – the freeze of your storage for a couple of seconds will be your smallest problem.

Our Approach

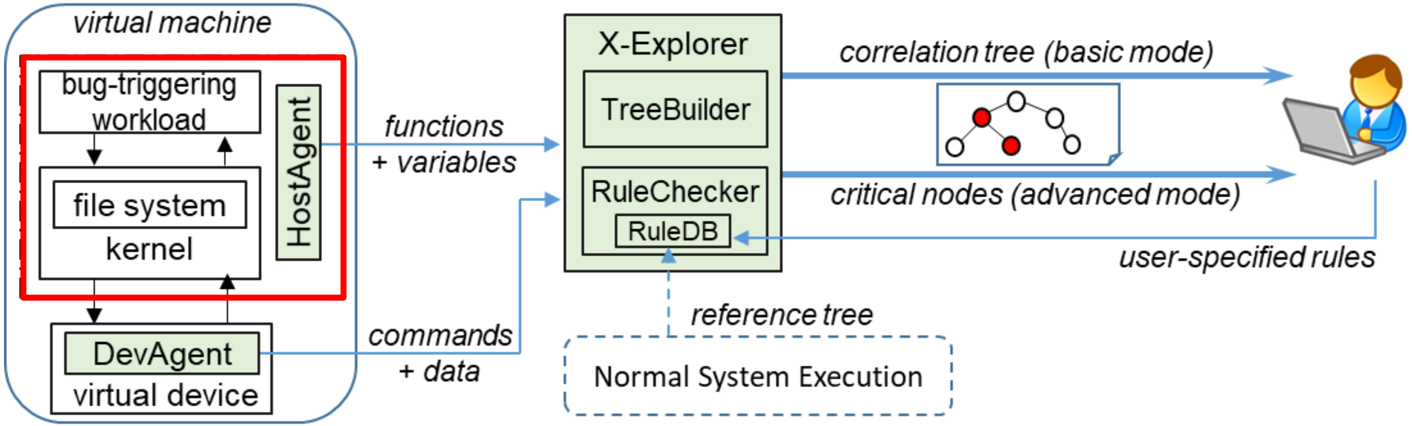
X-Ray: A Cross-Layer Approach

- Support unmodified software stack
- Intercept device activity without relying on kernel or special hardware
- Visualize multi-layer correlation
- Narrow down root cause (semi)automatically



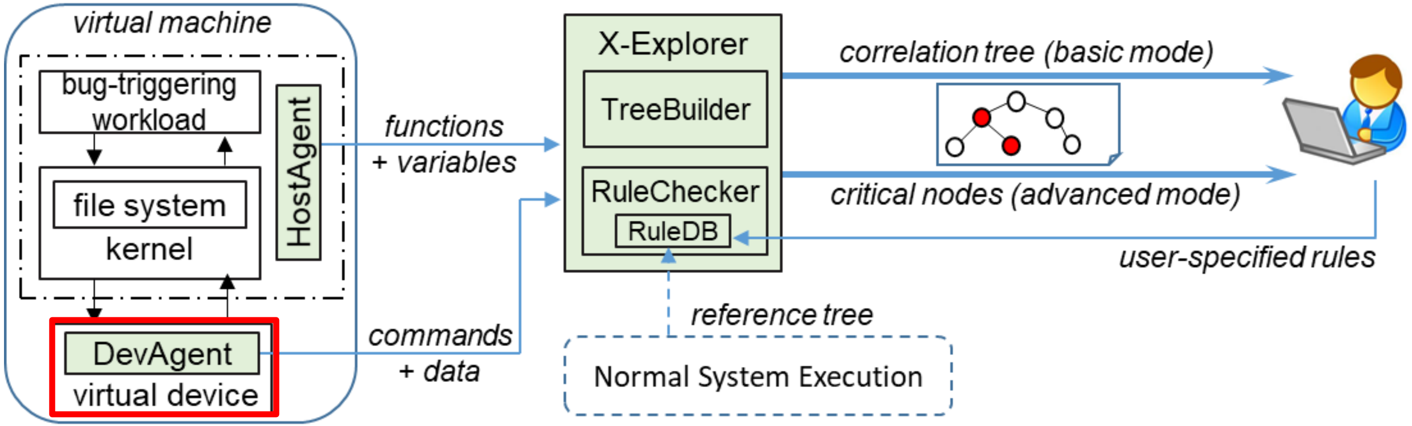
X-Ray: A Cross-Layer Approach

- HostAgent: help understand host-side system activities
  - Trace host-side events
    - e.g., syscalls, kernel functions



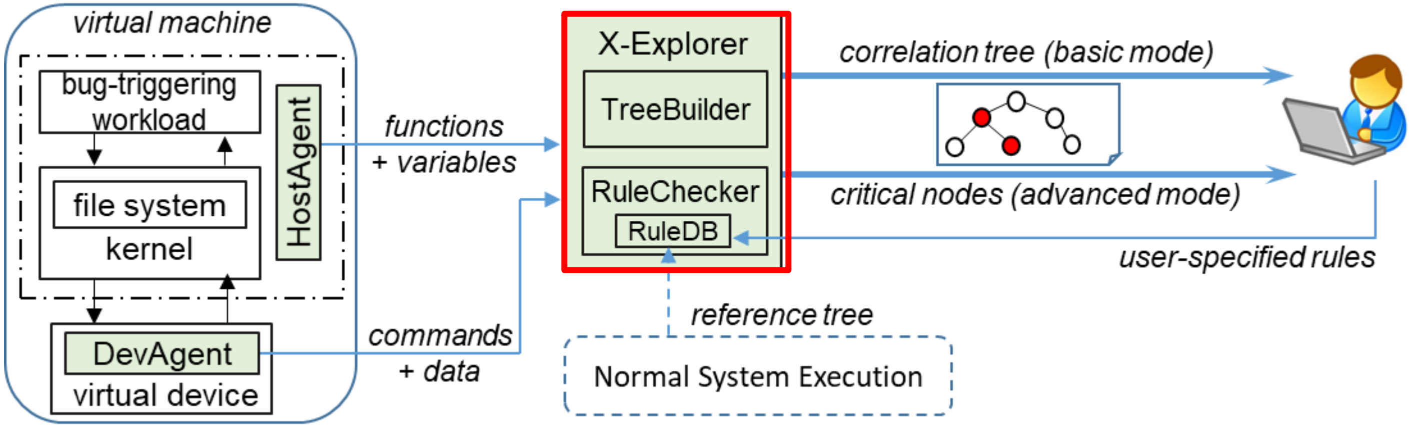
X-Ray: A Cross-Layer Approach

- DevAgent: help understand changes of persistent states
  - Trace device commands
    - e.g., SCSI, NVMe



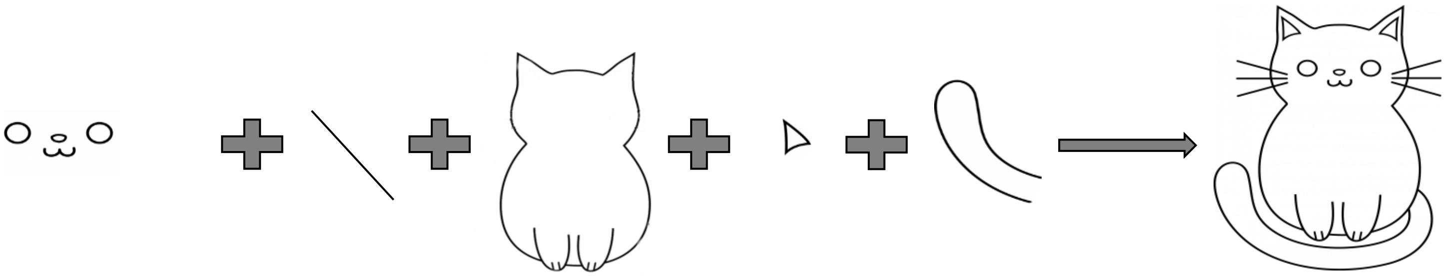
X-Ray: A Cross-Layer Approach

- X-Explorer: facilitate diagnosis in two ways
  - Build and visualize multi-layer correlation (i.e., correlation tree)
  - Highlight critical nodes/paths based on rules



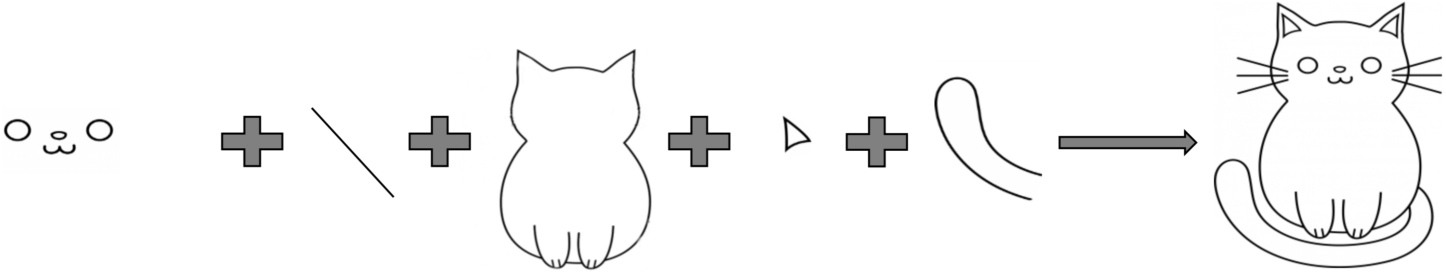
Key Challenge #1

- How to correlate information across layers ?



Key Challenge #1

- How to correlate information across layers ?
  - Cannot use SCSI/NVMe hints ☹
    - Require modification to workload/OS
  - Use timestamp ☺
    - Customized Ftrace frontend
      - Convert execution time to epoch time
    - NTP(Network Time Protocol) based synchronization
      - Solve accuracy problem caused by virtualization



Key Challenge #2

- How to reduce manual efforts ?

Key Challenge #2

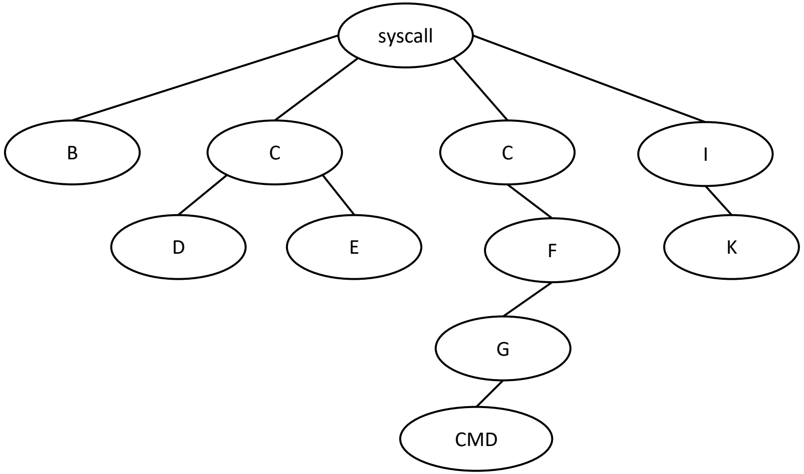
- How to reduce manual efforts ?
  - Visualize cross-layer events & dependencies in a correlation tree

Dependency

Syscall	→	B
Syscall	→	C
C	→	D
C	→	E
Syscall	→	C
C	→	F
F	→	G
G	→	CMD
Syscall	→	I
I	→	K

Tracing log

☹



Cross-layer tree

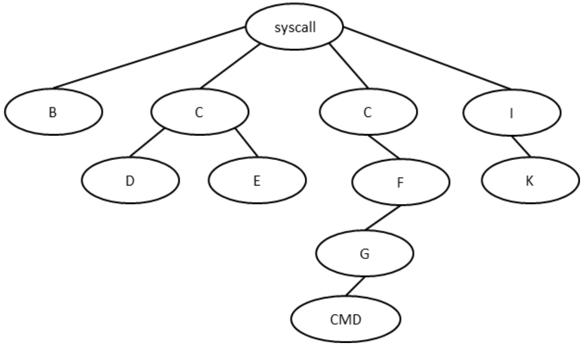
☺

Key Challenge #2

- How to reduce manual efforts ?
  - Visualize cross-layer events & dependencies in a correlation tree
  - Automatically narrow down the root cause via rules

Key Challenge #2

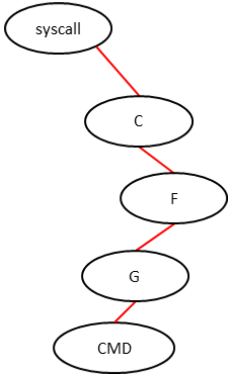
- How to reduce manual efforts ?
  - Visualize cross-layer events & dependencies in a correlation tree
  - Automatically narrow down the root cause via rules
    - Rules specified by users (e.g., “ancestors of device commands”)



Correlation tree

☹

Rule specified by users →

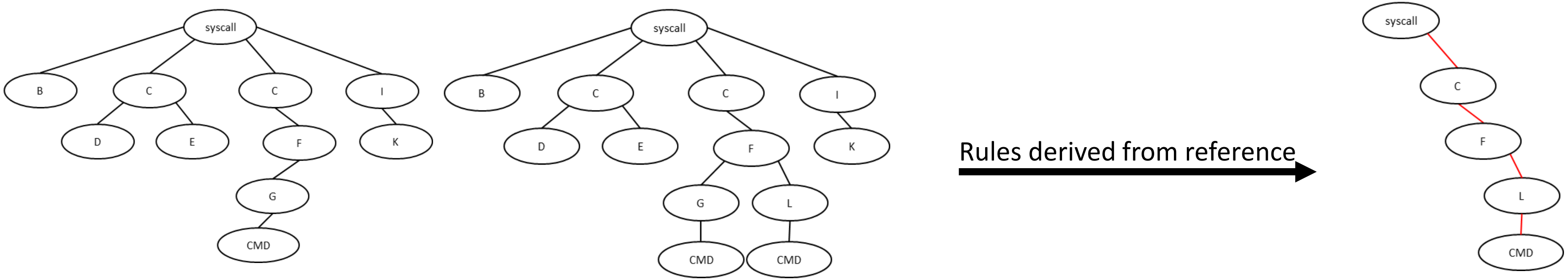


Critical part

☺

Key Challenge #2

- How to reduce manual efforts ?
  - Visualize cross-layer events & dependencies in a correlation tree
  - Automatically narrow down the root cause via rules
    - Rules specified by users (e.g., “ancestors of device commands”)
    - Rules derived from reference execution (e.g., non-failure run due to different kernel version)



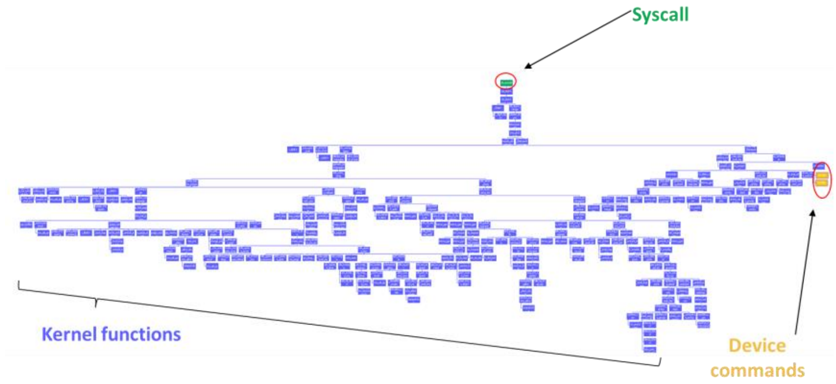
Tree from Abnormal execution ☹ Tree from reference execution

Difference part ☺

Preliminary Results

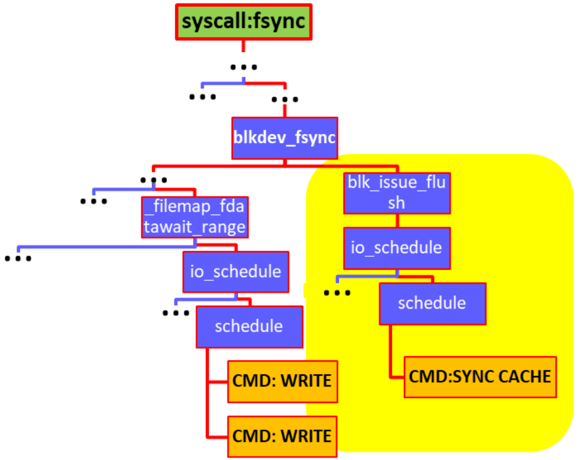
Preliminary Results

- Case Study
  - A kernel bug manifested as serialization errors on SSDs [Zheng *et. al*.@TOCS'16, FAST'13]
  - The problem can be observed in the correlation tree clearly
  - Rules can help narrow down the root cause quickly



Tree from abnormal execution

Rules



Pinpointed root cause

Preliminary Results

- Result summary
  - 5 failure cases reported in the literature
  - 3 simple rules to define critical parts of the correlation trees
  - Reduce the search space for root causes effectively
    - 0.06% - 4.97% nodes of the original trees

Case ID	node count in original tree	node count by Rule#1	node count by Rule#2	node count by Rule#3
1	11,353 (100%)	704 (6.20%)	571 (5.03%)	30 (0.26%)
2	34,083 (100%)	697 (2.05%)	328 (0.96%)	22 (0.06%)
3	24,355 (100%)	1254 (5.15%)	1210 (4.97%)	/
4	273,653 (100%)	10230 (3.74%)	/	/
5	284,618 (100%)	5621 (1.97%)	5549 (1.95%)	/

Conclusion and Ongoing Work

- X-Ray: A cross-layer approach for failure diagnosis
  - Support unmodified software stack
  - Intercept device activity without relying on kernel or special hardware
  - Visualize multi-layer correlation
  - Narrow down root cause (semi)automatically
- Explore more real-world failure cases
- Derive more diagnosis rules
- Automate the comparison based on reference tree

Thanks !

Duo Zhang
duozhang@iastate.edu
https://www.ece.iastate.edu/~mai/lab/dsl.html