Name: Prashanth Mallyampatti
Unity ID: pmallya
Student ID: 200250501

CSC724- Advanced Distributed Systems
Paper Review

Insight: In-situ Online Service Failure Path Inference in Production Computing Infrastructures

Summary:

This paper mainly focuses on failure path inference inside the production environment. It particularly addresses non-crashing failures which tend to go unnoticed and the error might not always tell the root cause of the problem. Some tools have exceptional analysis approach, but overheads hinder them to be used. All these are considered in this paper at different levels of testing.

Insight infers the execution path of a failure, giving a small subset to look upon to solve the problem. Insight doesn't require any system instrumentation as such except for system call tracing. It considers both the environment data and the runtime outputs in different situations to get an optimal failure path. Insight system design contains Dynamic shadow server creation which clones the failure part to be analyzed and creates a shadow server to prepare for failure reproduction. Undesired interactions involving the shadow server is filtered by running a proxy outside the shadow server. False alarms are also handled and discussed in the paper. Next the shadow component is analyzed to find the failure path. The authors term this as Guided Binary Execution Exploration as they intercept the jump statements in the binary code and explore both true/false paths. Since each path is explored, the exact input is not required. The unmatched output is used a trigger to stop the current path traversal and rollback to previous branch point. Multi-path search is also proposed by the authors to speed up the failure reproduction and loops are handled. Lastly, the Runtime Output Matching considers both runtime outputs- if console log messages are too sparse and console log messages- which provide important information about the system states.

Insight has been tested using various cloud infrastructures and open source softwares. Evaluations on various environment data, and inputs has been done and it shows the high accuracies. As indicated in the paper, Insight can infer the root causes of the failure even if the branch points are in the middle of the execution path. Static analysis for failure reproduction, Insight has achieved 12 times faster than symbolic execution. System overheads are very low compared to the capacity of modern storage systems.

Strong Points:

1) Insight uses both console logs, and system calls to infer the failure path, when error messages are sparse- system calls are used. This increases the accuracy failure path inference.
2) Insight doesn't require the exact input which triggered the failure, as other factors are taken into consideration to find the failure path.
3) Since Insight considers the a partial set of data to infer paths, its overhead is greatly reduced, thus acting as a first step failure inference tool.
4) Case studies and real time examples presented shows clearly that Insight is capable of performing well.
5) Cases where Insight doesn't perform well has been shown, with a remedy solution to that.

Weak Points:

1) Firstly it depends on an external failure detector to detect a failure, which if delayed may change the state of the host after failure, thus making it harder for Insight to infer after cloning.
2) The way and type of outputs dropped from the shadow to the production server hasn't been discussed. Critical outputs must be taken care of.
3) Issues that could be raised, solutions to- if shadow system fails has not been discussed.