Empirical Analysis of Flaky Tests in Practice - An Automotive Case Study

Bachelor's Thesis, Master's Thesis

Supervisor: Prof. Dr. Alexander Pretschner
Advisor: Claudius Jordan
Email: claudius.jordan@tum.de
Phone: +49 (89) 289 - 17314
Starting date: Any time

Informatik 4 – Lehrstuhl für Software und Systems Engineering
Prof. Dr. Alexander Pretschner
Fakultät für Informatik
Technische Universität München

Boltzmannstraße 3
85748 Garching bei München

https://www.in.tum.de/i04/

sponsored by tracetronic GUIDING YOUR VISION

Context

In the context of vehicle function development, Hardware-in-the-Loop test benches are used for functional validation on integration and system level. As the complexity of such modern cyber-physical systems (CPS) like vehicles increases, the necessary test infrastructures (TI) equally grow in complexity. Such TIs are engineered in parallel to the system under test (SUT). As a consequence, they too need to be considered error-prone, which only recently gained attraction in the literature [1, 2, 3, 4].

One problem that arises from error-prone TIs is that the produced test results cannot be regarded *reliable* and the actual goal of testing - identifying faults in the SUT - is impeded. In the software engineering (SE) domain, such unreliable tests are called *flaky* ([5, 6, 7, 8, 9]). We know that flakiness *is* a problem. What we are missing though, is a thorough analysis of the problem's extent across different projects at our industry partner.

Goal

In this thesis an empirical analysis is to be performed. We want to quantify impact of test flakiness in the context of system-level testing in the automotive domain by virtue of selected, real-world validation & verification (V&V) projects at our industry partner.

Working Plan

1. Familiarize with the literature on flaky tests
2. Familiarize with different project settings at our industry partner
3. Plan the case study[1]
4. Write the exposé
5. Perform the case study
6. Write the thesis report

Deliverables

- Exposé (about 6 weeks after kick-off)
- Final thesis report written in English and in conformance with TUM guidelines
- Presentation of the work at the chair (2-3 weeks after submission)

Application:
Please apply via email to claudius.jordan@tum.de. Your email should explain your interest in the topic and contain your current transcript of records. The most promising candidates will be invited for an informal interview. Upon mutual agreement, the thesis will be performed in cooperation with TraceTronic GmbH.

References

[1] Kristian Wiklund et al. "Impediments for software test automation: A systematic literature review". In: *Softw. Test. Verif. Reliab.* 27.8 (2017), pp. 1–20.

[2] Isabel Evans et al. "Stuck in Limbo with Magical Solutions: The Testers' Lived Experiences of Tools and Automation". In: *Proc. 15th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl.* SCITEPRESS - Science and Technology Publications, 2020, pp. 195–202.

[3] Kim Herzig and Nachiappan Nagappan. "Empirically Detecting False Test Alarms Using Association Rules". In: *2015 IEEE/ACM 37th IEEE Int. Conf. Softw. Eng.* Vol. 2. IEEE, May 2015, pp. 39–48.

[4] Claudius V. Jordan et al. "Framework for Flexible, Adaptive Support of Test Management by Means of Software Agents". In: *IEEE Robot. Autom. Lett.* 4.3 (2019), pp. 2754–2761.

[5] Qingzhou Luo et al. "An empirical analysis of flaky tests". In: *Proc. ACM SIGSOFT Symp. Found. Softw. Eng.* 16-21-November-2014 (2014), pp. 643–653.

[1] https://github.com/acmsigsoft/EmpiricalStandards/blob/master/docs/CaseStudy.md

[6] Celal Ziftci and Diego Cavalcanti. “De-Flake Your Tests : Automatically Locating Root Causes of Flaky Tests in Code At Google”. In: *2020 IEEE Int. Conf. Softw. Maint. Evol.* IEEE, Sept. 2020, pp. 736–745.

[7] Wing Lam et al. “Root causing flaky tests in a large-scale industrial setting”. In: *ISSTA 2019 - Proc. 28th ACM SIGSOFT Int. Symp. Softw. Test. Anal.* 2019, pp. 204–215.

[8] Jesús Morán et al. “FlakyLoc: Flakiness Localization for Reliable Test Suites in Web Applications”. In: *J. Web Eng.* (June 2020).

[9] Per Erik Strandberg et al. “Intermittently Failing Tests in the Embedded Systems Domain”. In: *ArXiv e-prints* (2020). arXiv: 2005.06826.

Informatik 4 – Lehrstuhl für Software und Systems Engineering
Prof. Dr. Alexander Pretschner
Fakultät für Informatik
Technische Universität München

Boltzmannstraße 3
85748 Garching bei München

https://www.in.tum.de/i04/

sponsored by