[Extended Abstract] Tailored Load Testing for Continuous Software Engineering: Approaches, Experiences, and Outlook

Henning Schulz
henning.schulz@novatec-gmbh.de
Novatec Consulting GmbH, Germany

Context

Continuous software engineering (CSE) aims to produce high-quality software through frequent and automated releases of concurrently developed services. By replaying workloads representing the behavior of the production environment's users, load testing can identify quality degradation, including performance regressions, under realistic conditions [1]. The literature proposes several approaches that extract representative workload models from recorded data, e.g., based on Markov chains [2, 3]. However, these approaches contradict CSE's high pace and automation in three aspects. First, they require manual parameterization, such as the specification of input data. Their focus on the system level furthermore results in resource-intensive load tests, especially for distributed microservice applications. Finally, due to variations in the production workload, there is a lack of means to select appropriate workload periods that justify being replayed in a time-consuming test.

Objective

In this talk, we present the results of a dissertation addressing the automated generation of tailored load tests. This research aimed to reduce the time and resources required for CSE-integrated testing by tailoring load tests to the services of interest and the workload-influencing context. An exemplary context is a special sale a webshop offers to increase the traffic. We also targeted a high degree of automation to fit CSE and a suitable level of abstraction to support users of different expertise.

Method

We developed and evaluated description languages, algorithms, and an automated load test generation approach. The approach integrates existing workload model extraction [3], modified clustering based on an established methodology [2], and time-series forecasting approaches [4, 5]. The evaluation comprises laboratory experiments, industrial case studies, an expert survey, and formal proofs.

Results

In this talk, we summarize the results of multiple previous publications. We showed that representative context-tailored load tests can be generated by learning a workload model incrementally, enriching it with contextual information, and predicting the expected workload using time series forecasting [6]. For further tailoring the load tests to services, we proposed extracting call hierarchies from recorded invocation traces [7]. Dedicated models of evolving manual parameterizations automate the generation process and restore the representativeness of the load tests [8]. Furthermore, the integration of our approach with the BenchFlow framework for automated test execution enables load testing for non-experts [9].

Conclusion

While we assessed the proposed approach as a suitable solution for the described problem, we identified limitations of its building blocks relevant to the community's future work. These blocks are the clustering [2, 3] and forecasting [4, 5] techniques from existing work, which we have assessed to be limited for predicting sharply fluctuating workloads, such as load spikes.

References

[1] Z. M. Jiang and A. E. Hassan, "A survey on load testing of large-scale software systems," IEEE Trans. Software Eng., vol. 41, no. 11, pp. 1091–1118, 2015.
[2] D. A. Menascé, V. A. F. Almeida, R. Fonseca, and M. A. Mendes, "A methodology for workload characterization of E-Commerce sites," in Proceedings of EC 1999, pp. 119–128.
[3] C. Vögele, A. van Hoorn, E. Schulz, W. Hasselbring, and H. Krcmar, "WESSBAS: extraction of probabilistic workload specifications for load testing and performance prediction - a model-driven approach for session-based application systems," Software and System Modeling, vol. 17, no. 2, pp. 443–477, 2018.
[4] A. Bauer, M. Züfle, N. Herbst, S. Kounev, and V. Curtef, "Telescope: An automatic feature extraction and transformation approach for time series forecasting on a level-playing field," in Proceedings of ICDE 2020, pp. 1902–0905.

[5] S. J. Taylor and B. Letham, "Forecasting at scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.
[6] H. Schulz, T. Angerstein, and A. van Hoorn, "Towards automating representative load testing in continuous software engineering," in Companion of ICPE 2018, pp. 123–126.
[7] H. Schulz, T. Angerstein, D. Okanović, and A. van Hoorn, "Microservice-tailored generation of session-based workload models for representative load testing," in Proceedings of MASCOTS 2019, pp. 323–335.
[8] H. Schulz, A. van Hoorn, and A. Wert, "Reducing the maintenance effort for parameterization of representative load tests using annotations," Softw. Test. Verification Reliab., vol. 30, no. 1, 2020.
[9] H. Schulz, D. Okanović, A. van Hoorn, V. Ferme, and C. Pautasso, "Behavior-driven load testing using contextual knowledge - approach and experiences," in Proceedings of ICPE 2019, pp. 265–272.