Automated Generation of Tailored Load Tests for Continuous Software Engineering

Henning Schulz | Nov. 12, 2020
Load tests detect load-related problems

(Jiang and Hassan, 2015)
Load tests detect load-related problems
(Jiang and Hassan, 2015)

Amazon's website crashed as soon as Prime Day began
By Nick Stell | @brickettall | Jul 16, 2018, 3:16pm EDT

https://www.theverge.com/2018/7/16/17577654/amazon-prime-day-website-down-deals-service-disruption
“The main objective in selecting a workload should always be representativeness” (Ferrari, 1972)
Modern software engineering is continuous

Bezemer et al., 2019
Eismann et al., 2020

DevOps team

Menascé et al., 1999
Krishnamurthy et al., 2006
Vögele et al., 2018
Modern software engineering is continuous

Bezemer et al., 2019
Eismann et al., 2020

DevOps team

build ---|--- test

plan ---|--- monitor

release

session logs

Menascé et al., 1999
Krishnamurthy et al., 2006
Vögele et al., 2018

frontend service
course schedule service
exams service
Existing load test generation process
Our contribution (I): automation

session logs ➔ clustering ➔ parameterization ➔ workload model ➔ load test

parameterization model
Our contribution (2): service-tailoring

- Trace logs
- Log-based service-tailoring
- Clustering
- Model-based service-tailoring
- Workload model
- Parameterization
- Load test

Service-tailoring description
Our contribution (III): context-tailoring

trace logs → log-based service-tailoring → clustering → Workload Model Repository → context-tailoring → model-based service-tailoring → parameterization → workload model → load test

context information → context-tailoring description

Repository context-tailoring model-based service-tailoring parameterization

Telescope or PROPHET
We generated and executed 44 load tests
We generated and executed 44 load tests
Clustering and context improve accuracy

- "distance to reference workload"

- "baseline"

- "perfect forecasting"  "Telescope"  "Prophet"

- "total/no context"  "indiv/no context"  "indiv/context"
Issue #1: the user behavior changes
Issue #2: forecasting has sharp jumps
Issue #2: forecasting has sharp jumps
We can generate tailored load tests automatically...

... but the existing workload characterization & forecasting methodologies are not (fully) ready for this

1. think times not integrated into clustering
2. missing prediction of curve shape
References (I)

A Survey on Load Testing of Large-scale Software Systems
Z. M. Jiang and A. E. Hassan
IEEE Transactions on Software Engineering, vol. 41, no. 11, 2015

Workload characterization and selection in computer performance measurement
D. Ferrari
IEEE Computer, vol. 5, no. 4, pp. 18–24, 1972

How is performance addressed in DevOps?
C.-P. Bezemer et al.
Proceedings of the 10th ACM/SPEC International Conference on Performance Engineering (ICPE 2019)

Microservices: a performance tester’s dream or nightmare?
S. Eismann, C.-P. Bezemer, W. Shang, D. Okanović, and A. van Hoorn
Proceedings of the 11th ACM/SPEC International Conference on Performance Engineering (ICPE 2020)
References (II)

A methodology for workload characterization of E-Commerce sites
D. A. Menascé, V. A. F. Almeida, R. Fonseca, and M. A. Mendes

A synthetic workload generation technique for stress testing session-based systems
D. Krishnamurthy, J. A. Rolia, and S. Majumdar

WESSBAS: extraction of probabilistic workload specifications for load testing and performance prediction - a model-driven approach for session-based application systems
C. Vögele, A. van Hoorn, E. Schulz, W. Hasselbring, and H. Krcmar
Software and System Modeling, vol. 17, no. 2, 2018
Publications

Towards Automating Representative Load Testing in Continuous Software Engineering
Henning Schulz, Tobias Angerstein, and André van Hoorn
Companion of the 9th ACM/SPEC International Conference on Performance Engineering (LTB@ICPE 2018)

Reducing the Maintenance Effort for Parameterization of Representative Load Tests Using Annotations
Henning Schulz, André van Hoorn, and Alexander Wert
Journal of Software Testing, Verification and Reliability, Special Issue on Testing Extra-functional Properties, 2019

Microservice-tailored Generation of Session-based Workload Models for Representative Load Testing
Henning Schulz, Tobias Angerstein, Dušan Okanović, André van Hoorn

Behavior-driven Load Testing Using Contextual Knowledge – Approach and Experiences
Henning Schulz, Dušan Okanović, André van Hoorn, Vincenzo Ferme, Cesare Pautasso
Proceedings of the 10th ACM/SPEC International Conference on Performance Engineering (ICPE 2019)