



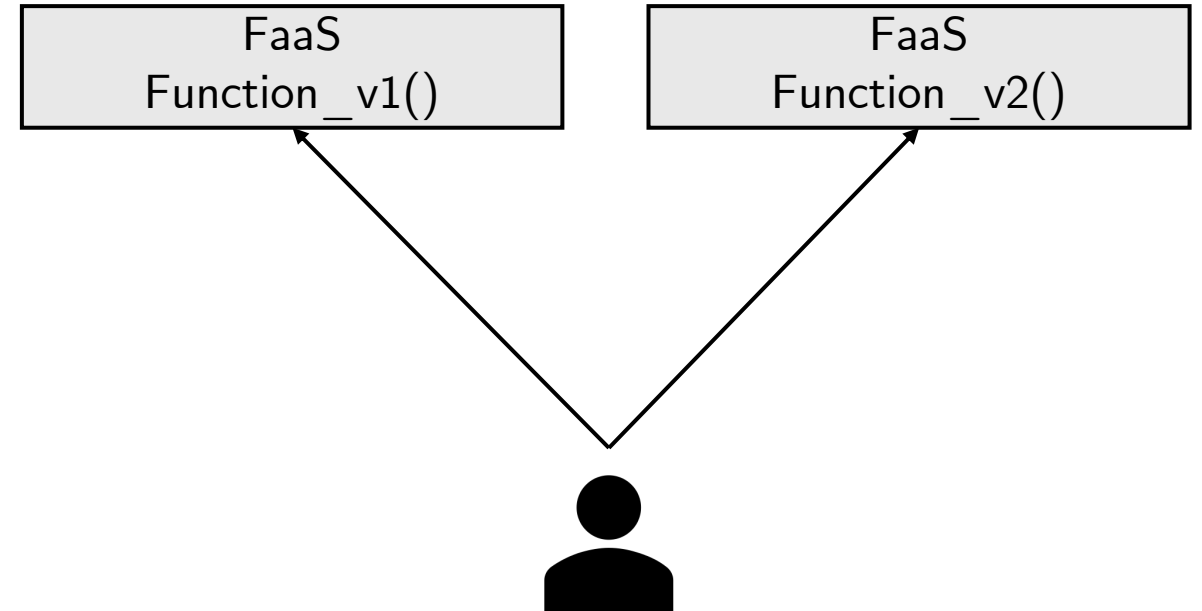
Increasing Efficiency and Result Reliability of Continuous Benchmarking for FaaS Applications

Tim C. Rese, Nils Japke, Sebastian Koch, Tobias Pfandzelter, David Bermbach
| Scalable Software Systems



Function-as-a-Service (FaaS) & Continuous Benchmarking

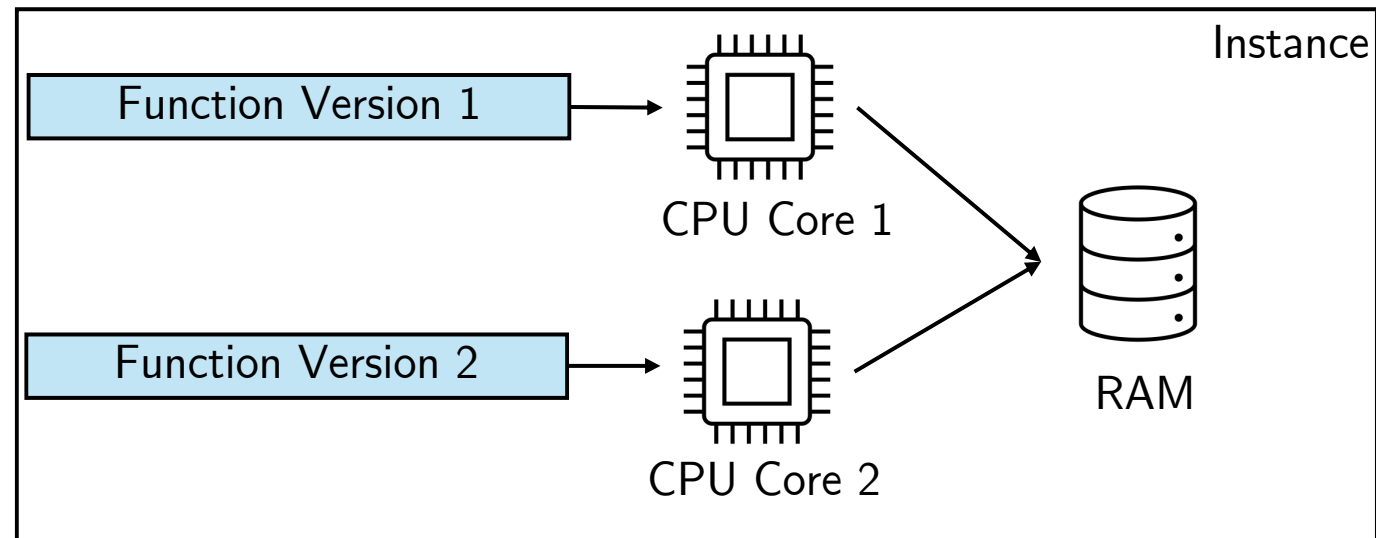
- FaaS has various benefits
- FaaS applications receive frequent updates in modern development cycles
- Continuous benchmarking is essential to track performance
- “Traditional” approach has flaws[1-2]



“Traditional” Benchmarking Approach

Duet Benchmarking

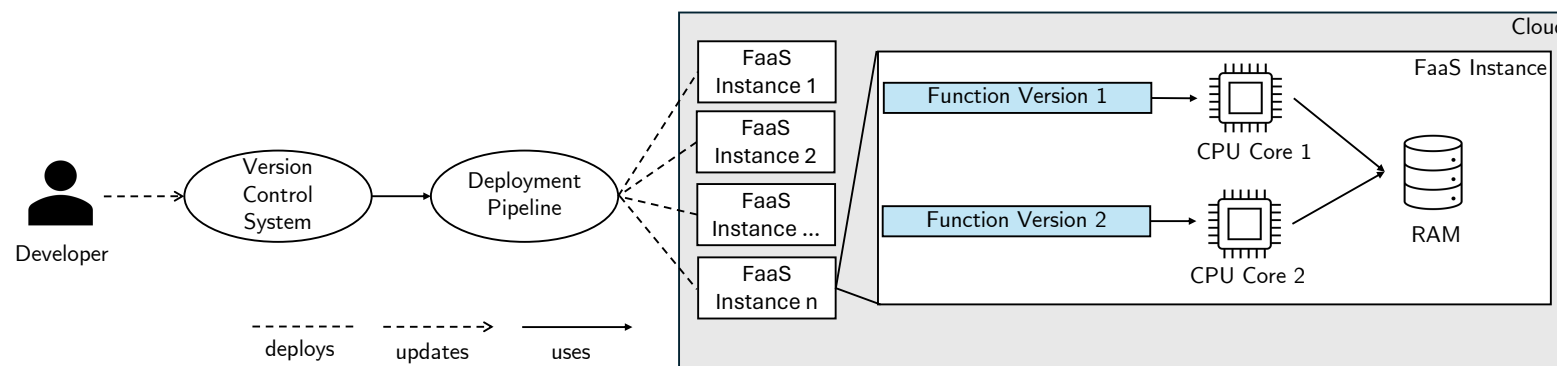
- Previous research has shown the effectiveness of the duet benchmarking for microbenchmarks and benchmark suites[3-4]
- Run artifacts on the same instance
- Isolate and provide the same amount of resources to each version
- Run in them in parallel



Duet Benchmarking Methodology

Evaluation

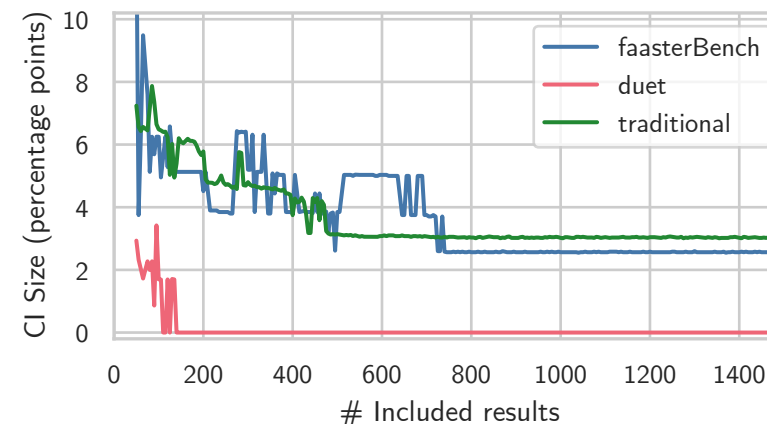
- Compare Duet Benchmarking FaaS approach to traditional and randomized sequential (faasterBench) benchmarking methodologies
- Deploy two functions (CPU- and MEM-intensive), inject artificial performance change, and run 1500 experiments (A/A and A/B Configurations).
- Regard interval at max. sample size and interval size development for all 3 approaches
- Source: <https://github.com/timchristianrese/DuetFaaS-code>



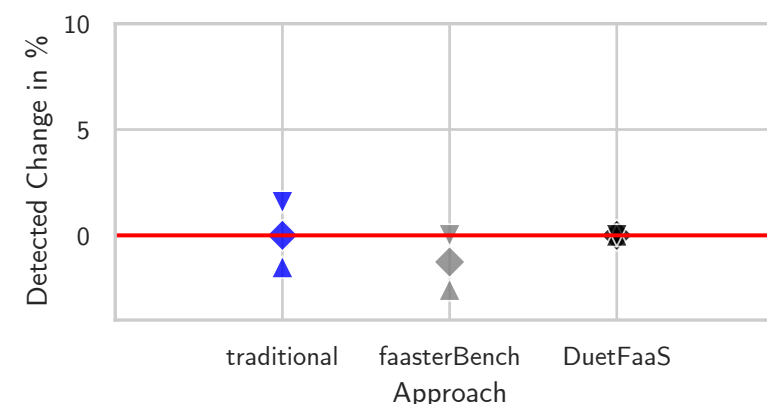
DuetFaaS Architecture

Results

- Duet Benchmarking provides more accurate results with fewer calls, making it ideal for deployment pipelines
- Function type had little impact
- A/A and A/B results were similar
- Full results found in our paper (under review) :
 - <http://arxiv.org/abs/2405.15610>



MEM A/A CI Size Development



MEM A/A Percentile Interval (1500 Results)

References

1. Ali Abedi and Tim Brecht. 2017. Conducting repeatable experiments in highly variable cloud computing environments. *In Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering (ICPE '17)*. 287–292.
2. Martin Grambow, Tim Dockenfuß, Trever Schirmer, Nils Japke, and David Bermbach. 2023. Efficiently Detecting Performance Changes in FaaS Application Releases. *In Proceedings of the 9th International Workshop on Serverless Computing (WoSC '23)*. 13–17.
3. Lubomír Bulej, Vojtěch Horky, Petr Tuma, François Farquet, and Aleksandar Prokopec. 2020. Duet benchmarking: Improving measurement accuracy in the cloud. *In Proceedings of the ACM/SPEC International Conference on Performance Engineering (ICPE '20)*. 100–107.
4. Nils Japke, Christoph Witzko, Martin Grambow, and David Bermbach. 2023. The Early Microbenchmark Catches the Bug – Studying Performance Issues Using Micro- and Application Benchmarks. *In Proceedings of the 16th IEEE/ACM International Conference on Utility and Cloud Computing*. 1–10