Twins or False Friends? A Study on Energy Consumption and Performance of Configurable Software (Replication Package)

Max Weber*, Christian Kaltenecker†, Florian Sattler†, Sven Apel†, Norbert Siegmund*
* Leipzig University, Germany
† Saarland University, Saarland Informatics Campus, Germany

***Abstract*—The artifacts described in this document refer to the study, data, and implementation of the paper ´Twins or False Friends? A Study on Energy Consumption and Performance of Configurable Software". In this document, we describes how the energy and performance measurement data set was collected and how the analysis of the correlation between energy consumption and performance can be repeated. Specifically, we describe the steps and tools that we have used to implement our approach, the data we have obtained, and the evaluation setup. We further list the available artifacts, such as raw measurements, configurations, and scripts which are available in a long-term data archive.**

I. Overview

This document describes the artifacts of our paper "Twins or False Friends? A Study on Energy Consumption and Performance of Configurable" [1], in which we answer the question on, whether runtime performance is a viable proxy for energy consumption. We found that although performance generally correlates with energy consumption, this correlation can break down for individual configurations, depending on distinct configuration options.

Figure 1 presents our empirical study, including all data sets (i.e., feature models, configurations, and measurements), tools, and scripts (e.g., measurement scripts) that have been produced in this process. All artifacts are provided as a Zotero archive [1] and in a GitHub repository [2].

II. Data Acquisition

Our study contained two data collection phases, one for the literature study and one for measuring performance and energy consumption. Next, we describe how we collected and classified the literature and explain the full measurement setup.

A. Literature Study Data

Our study data refers to the process described in Section 3.A in the correspoding paper. We provide all literature study data in a table format on our supplementary web page [3], providing a csv file and a Google docs table. The data is structured as follows: we state the title of the analyzed paper, the classification of the correlation mode (i.e., positive correlation, negative correlation, no correlation), and the causes of correlation if provided.

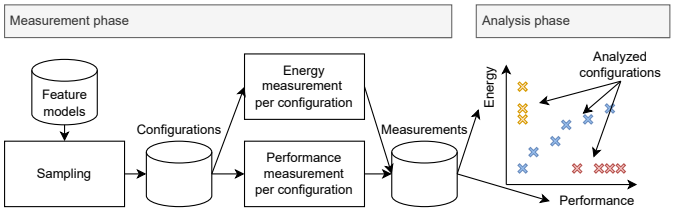

Fig. 1: Overview of the artifacts used in the main paper.

B. Measurement Infrastructure

To minimize measurement bias for our performance and energy measurements, we carefully prepared the measurement setup in the following way: (1) Measurements for a single software system are carried out on devices with similar hardware, having a minimal installation of Ubuntu 18.04. (2) We systematically defined the experiments for scalability and reproducibility with the low-overhead workload manager Slurm [4], making sure that no concurrent tasks are executed on the same machine.

All system-level measurements are carried out on two clusters, each equipped with a dedicated energy measurement device. We measured the runtime performance directly on each machine. Every machine is either plugged into a GUDE 8045- 1 or IPT iPower P1 PDU (power distribution unit), both measuring the power draw with an interval of one second per outlet. The machines of the cluster are composed of Intel NUCs each equipped with i7-8559U CPU, 32 GB DDR4-2666, and 500 GB NVME SSD.

All function-level measurements are conducted on a dedicated single machine running an Intel Core i5-7500 with 4 GB RAM and 120 GB SSD. This machine contains a 2.5 kHz energy measurement system [2], which is based on the INA226 power measurement chip. We provide detailed instructions on how to set up this energy measurement system [5], including a set-up overview, a detailed accuracy estimation, measurement software, and a billing list for the devices. We used the PERF profiling tool[6] for measuring the runtime within a function.

[1]Zotero: https://zenodo.org/record/7647208
[2]GitHub: https://github.com/AI-4-SE/TwinsOrFalseFriends
[3]Supplementary Web page literature: https://github.com/AI-4-SE/TwinsOrFalseFriends#literature-study
[4]Slurm: https://slurm.schedmd.com/documentation.html
[5]Function-level energy measurement system: https://github.com/AI-4-SE/TwinsOrFalseFriends/tree/main/fast-energy-measurements
[6]perf: https://perf.wiki.kernel.org/index.php/Main_Page

III. Artifacts

We produced two types of artifacts: measurement data and analyses scripts, which we explain next.

A. Measurement Data

In total, we invested 2.5 years of CPU time obtaining the largest performance–energy data set on configurable software systems we are aware of. To reproduce our results and facilitate future research, we provide all data on the supplementary Web page and in a Zenodo archive. This data includes all system-level and function-level energy and performance measurements [7].

For each subject system we provide the following files:

- *FeatureModel.xml*: A feature model containing a tree-based representation of the configuration space of the subject system. We used the feature model to sample configurations that we include in our analysis, in the SPL Conqueror format [3].
- *measurements.csv*: A csv file containing all configuration option values and the corresponding measured performance and energy consumption values. A row represents a single measurement, columns represent options and the last columns the measured values (i.e. energy consumption and runtime performance).
- *measurements_ml.csv*: A csv file containing all configuration option values and the corresponding measured performance and energy consumption values for all functions of the software systems BROTLI, LRZIP, and X264.
- *[system]_prof_configurations.cfg*: A text file containing all profiled configurations for the systems BROTLI, LRZIP, and X264.

B. Analyses Scripts

We provide all scripts that we used to conduct our analysis [8]. We have automated the installation of all third-party software and the setup of all scripts in a Dockerfile. We further give instructions for a manual setup. Once the system is set up, the whole analysis pipeline is executed with a single python command [9].

Acknowledgments

Apel's and Siegmund's work has been funded by the German Research Foundation (SI 2171/3-1, SI 2171/2-2, AP 206/11-1, AP 206/11-2, and Grant 389792660 as part of TRR 248 – CPEC). Siegmund's work has been supported by the Federal Ministry of Education and Research of Germany for the "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig" (ScaDS.AI).

[7] Energy and performance measurement data: https://github.com/AI-4-SE/TwinsOrFalseFriends/tree/main/data

[8] Analysis scripts: https://github.com/AI-4-SE/TwinsOrFalseFriends/tree/main/scripts

[9] Script execution: https://github.com/AI-4-SE/TwinsOrFalseFriends#usage

References

[1] M. Weber, C. Kaltenecker, F. Sattler, S. Apel, and S. Norbert, "Twins or False Friends? A Study on Energy Consumption and Performance of Configurable Software," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023.

[2] D. Bedard, R. Fowler, M. Lim, and A. Porterfield, "PowerMon 2: Fine-grained, Integrated Power Measurement," *RENCI Technical Report Series*, pp. 1–15, 2009.

[3] N. Siegmund, A. Grebhahn, S. Apel, and C. Kästner, "Performance-Influence Models for Highly Configurable Systems," in *Proceedings of the Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2015, pp. 284–294.