

Threats to validity

This document tackles the threats to the validity of the study by following the recommendations in [1] and how we have tried to minimize their effects:

Construct validity. The first point concerns the reliability of the measurements. We have used two different hardware measurement devices (for Tablet and PC, called EET) to measure the consumption, which allows us to obtain accurate measurements of the energy consumed by the tablet and the PC in a small-time interval (approximately one hundred samples per second).

Obviously, the measurements obtained are specific to the PC and Tablet where the studies are run and may differ if we use others. Also, the results would be different if other measurement mechanisms as estimation would be used. However, the hardware measurement devices used have already been validated and shown to be reliable [2] and have been used in the past in other measurement studies of this type.

Internal validity. Among the uncontrolled factors that could affect the results of the experiment, the most notable are those related to the conditions under which the measurements were made. We carried out all the measurements under the same temperature conditions and at the same time of day to avoid any influence on the results due to the Internet charge or the temperature in the laboratory.

In addition, several runs were performed to mitigate possible consumption-related outliers. The same DUT (Tablet or PC) has been used to perform the runs and capture the energy consumption and measures have been taken to ensure that the DUTs were always in the same conditions for the execution of each of the different functionalities. To avoid possible background processes running, before starting each measurement, all programs that could cause interference were closed and the base consumption corresponding to each DUT was subtracted.

Regarding the number of measurements performed, 10 measurements were carried out. Although some authors such as [3] recommend performing at least 30 repetitions to evaluate software power consumption in a controlled environment, the reliability of our measurement devices has shown a low variability among the results of each run (unless when some kind of error occurs), so we think the number of repetitions to be sufficient to consider the results valid.

External validity. Finally, about the generalisability of the results obtained in this experiment, the results are based on the analysis of the functionalities available in the Instagram application at the time of the study. Currently, Instagram has added new functionalities and modified others, such as the reel templates, which could alter its energy consumption. It could be the case that the last included functionalities could affect to the analysed ones, but we think it probably will change the consumption but not the classifications obtained. However, the aim of this study is to make users aware that not all features consume the same amount of energy, and to offer some tips for more responsible use, so the changes made by Instagram do not affect the objective of our work.

[1] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, y A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[2] J. Mancebo, H. O. Arriaga, F. García, M. Á. Moraga, I. García-Rodríguez de Guzmán, y C. Calero, «EET: A Device to Support the Measurement of Software Consumption», en 2018

IEEE/ACM 6th International Workshop on Green And Sustainable Software (GREENS), 2018, pp. 16-22.

- [3] E. Kern *et al.*, «Sustainable software products—Towards assessment criteria for resource and energy efficiency», *Future Gener. Comput. Syst.*, vol. 86, pp. 199-210, 2018.