Analyzing Programming Languages’ Energy Consumption:
An Empirical Study

Stefanos Georgiou
Athens University of Economics and Business
Athens, Greece
sgeorgiou@aueb.gr

Maria Kechagia
Delft University of Technology
Delft, The Netherlands
m.kechagia@tudelft.nl

Diomidis Spinellis
Athens University of Economics and Business
Athens, Greece
dds@aueb.gr

ABSTRACT

Motivation: Shifting from traditional local servers towards cloud computing and data centers—where different applications are facilitated, implemented, and communicate in different programming languages—implies new challenges in terms of energy usage.

Goal: In this preliminary study, we aim to identify energy implications of small, independent tasks developed in different programming languages; compiled, semi-compiled, and interpreted ones.

Method: To achieve our purpose, we collected, refined, compared, and analyzed a number of implemented tasks from Rosetta Code, that is a publicly available Repository for programming chrestomathy.

Results: Our analysis shows that among compiled programming languages such as C, C++, Java, and Go offers the highest energy efficiency for all of our tested tasks compared to C#, VB.NET, and Rust. Regarding interpreted programming languages PHP, Ruby, and JavaScript exhibit the most energy savings compared to Swift, R, Perl, and Python.

CCS CONCEPTS
• Hardware → Power estimation and optimization; • Software and its engineering → Software libraries and repositories; Software design tradeoffs;

KEYWORDS
Energy Efficiency, Energy Optimization, Programming Languages

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

PCI ‘17, September 2017, Larissa, Greece
© 2017 Copyright held by the owner/author(s).
ACM ISBN 123-4567-24-567/08/06...
https://doi.org/10.475/123_4

1 INTRODUCTION

The increasing demands on services and computational applications from ICT-related products are major factors that contribute to the increase of energy consumption. Recent research conducted by Gelenbe and Caseau [7] and Van Heddeghem et al. [14] indicates a rising trend of the IT sector energy requirements. It is expected to reach 15% of the world’s total energy consumption by 2020.

Most of the studies, for energy efficiency, have considered energy consumption at hardware level. However, there is much of evidence that software can also alter energy dissipation significantly [2, 5, 6]. Therefore, many conference tracks (e.g. GREENS, eEnergy) have recognized the energy-efficiency at the software level as an emerging research challenge regarding the implementation of modern systems.

Nowadays, more companies are shifting from traditional local servers and mainframes towards the data centers. Some of the main characteristic of this approach is the pay-as-you-go feature, elasticity, scalability upon heavy workloads, and the hosting of any applications implemented in a variety of programming languages. However, there is a limited number of research works that examine the energy impact of programming tasks implemented in different programming languages.

To identify trends and possible gains regarding the reduction of energy consumption, during software development, we conducted an empirical study using data from the Rosetta Code Repository. Our goal is to elucidate energy usage from small tasks implemented in a variety of well-known and most used programming languages. To this end, our results show which of the interpreted and compiled programming languages offer more energy efficient implementations for specific tasks. Moreover, we show the negative impact if choosing an inefficient implementation.

The remainder of this paper is organized as follows. Section 2 describes: 1) our experimental setup, 2) our dataset’s cleaning method, 3) the software and hardware tools we used, and 4) our analysis methodology. In Section 3, we present our preliminary results and in Section 4 we discuss potential threats to validity. In Section 5, we list prior work done in the field and compare it with ours. Finally, we conclude in Section 6 and we present future research directions.

2 EXPERIMENTAL SETUP

In this Section, we describe our approach for conducting our experiment and for retrieving measurements. Initially, we provide information about the obtained dataset and the way we selected

1 Although in the physical sense energy cannot be consumed, we will use the terms energy “consumption”, “requirement”, and “usage” to refer to the conversion of electrical energy by ICT equipment into thermal energy dissipation to the environment. Correspondingly, we will use the term energy “savings”, “reduction”, “efficiency”, and “optimization” to refer to reduced consumption

2 http://greens.cs.vu.nl/
our tasks and refine it. Furthermore, we explain our experimental setup, the hardware and software tools we used.

### 2.1 Dataset

In the context of this study, we used the Rosetta Code,\(^4\) which is a publicly available programming chrestomathy site that offers 851 tasks, 230 draft tasks, and a collection of 658 different programming languages. In general, not all tasks are implemented, and not all tasks are possible to implement in all languages. For our study, we cloned a Github Repository\(^5\) that contains all the currently implemented tasks introduced in the Rosetta Code website.

To select popular programming languages, we consulted the website of tiobe,\(^6\) a software quality company. Tiobe uses a search query for index rating of the most popular programming languages around the web on a monthly basis. This query is based on a formula\(^7\) that uses the highest ranked search engines (according to Alexa)\(^8\) and a number or requirements enlisted for the programming languages. We decided to choose the top 15 programming languages enlisted for June 2017. From the current list, we excluded programming languages such as Delphi (not available for Linux OS we are using) and Assembly (different implementations between processor architectures). In contrast, we included Rust in our dataset that is a memory safe programming language and is gaining vast popularity. In the web. Therefore, we ended up with 14 programming languages as it is illustrated in Table 1.

<table>
<thead>
<tr>
<th>Programming Languages</th>
<th>Compilers and Interpreters version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compiled</td>
<td>gcc version 6.3.1 20161221-</td>
</tr>
<tr>
<td></td>
<td>go version go1.7.5</td>
</tr>
<tr>
<td></td>
<td>rustc version 1.18.0</td>
</tr>
<tr>
<td>Semi-Compiled</td>
<td>vb.net mono version 4.4.2.0 (vbnc)</td>
</tr>
<tr>
<td></td>
<td>C# mono version 4.4.2.0 (mics)</td>
</tr>
<tr>
<td></td>
<td>Java javac version 1.8.0_131</td>
</tr>
<tr>
<td>Interpreted</td>
<td>JavaScript node version 6.10.3</td>
</tr>
<tr>
<td></td>
<td>Perl perl version 5.24.1</td>
</tr>
<tr>
<td></td>
<td>PHP php version 7.0.19</td>
</tr>
<tr>
<td></td>
<td>Python python version 2.7.13</td>
</tr>
<tr>
<td></td>
<td>R Rscript version 3.3.3</td>
</tr>
<tr>
<td></td>
<td>Ruby ruby version 2.3.3p222</td>
</tr>
<tr>
<td></td>
<td>Swift swift version 3.0.2</td>
</tr>
</tbody>
</table>

The selected tasks were: `array-concatenation`, `classes` (creating an object and calling a method to print a variable’s value), `url-encoding` and `decoding`, `bubble-`, `quick-`, `insertion-`, `merge-`, and `selection-` sorting algorithms. Moreover, to further refine our dataset we used the following steps:

- Some of the tasks offered more than one implementation for the same programming language. Thus, we had to browse manually through each directory and remove them until we had only one that is consistent with the other implementation. For instance, when most of the implemented tasks used iterative implementation, we removed the ones using recursion.
- The Java file names and their public names where different which resulted in compilation error. Thus, we had to manually change them.
- Some of the implementations did not have main classes, or the same data with other tasks. Therefore, we changed the source code to offer consistency.
- For some programming languages that do not offer the class option such as C and Go, we used `structs`.
- Some of the tasks were relatively small and finished faster than a second which makes it impossible for our power analyzer to capture those results. Therefore, we added all the selected tasks in an iteration loop of a million times.

After applying the above modifications on our dataset, we categorized our programming languages in three main categories, namely, compiled, semi-compiled, and interpreted (see Table 1). For the programming languages which offer a semi-compiled approach such as Java, vb.net, and C#, we added them under the category of compiled languages for our experiments. In addition, we compared the compiled and semi-compiled implementations while having scenarios with and without compiler optimizations.

### 2.2 Hardware and Software components

#### 2.2.1 Hardware Components

The physical tools we used comprise: 1) portable personal computer (HP EliteBook 840 G3),\(^9\) 2) real-time electricity usage monitoring tool (Watts Up Pro (wup)),\(^10\) and 3) embedded device (Raspberry Pi 3b).\(^11\)

In general, there are two venues for retrieving energy consumption from a computer-based system. On the one hand, this is achievable by indirect energy measurements through estimation models or performance counters, core component of software monitoring tools. On the other hand, it is done via direct measurement, hardware power analyzers and sensors. However, each of these approaches has its own pitfalls. The direct approach, i.e., hardware components, offers coarse-grained measurements for the whole systems’ energy consumption and low sampling rate. The indirect approach, i.e., software components, suffers from inaccuracy, lack of interoperability, and additional system overhead. In our research, we decided to use a direct approach such as `wup` since it does not have software constrains and is relatively cheap to buy.

---

\(^4\)<http://rosettacode.org/wiki/Rosetta_Code>
\(^5\)<https://github.com/numeism/RosettaCodeData>
\(^6\)<https://www.tiobe.com/tiobe-index/>
\(^7\)<https://www.tiobe.com/tiobe-index/programming-languages-definition/>
\(^8\)<http://www.alexa.com/>
\(^10\)<https://www.wattsupmeters.com/secure/products.php?pn=0>
\(^11\)<https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>
Analyzing Programming Languages’ Energy Consumption: An Empirical Study

2.2.2 Software Components. To extract data, manage, and use our Rosetta Code Repository, we developed a number of shell scripts as enlisted below, which are publicly available on our Git Repository.13

- **script.cleanAll**: removes the current instance of tasks in the current working directory and copies the new one found from the parent directory.
- **script.findCommonTasksInLanguages**: provides a list of tasks with the number of existing implementations in different languages.
- **script.createNewDataSet**: filters the Rosetta Code current dataset and removes programming languages and tasks not added as command line arguments.
- **script.fromUpperToLower**: changes the current instance of task directories and files from upper to lower case.
- **script.compileTasks**: compiles all tasks found under the tasks’ directory and produces error reports if a task fails to compile.
- **script.executeTasksRemotely**: executes all the tasks’ implementations found under tasks’ directory. Moreover, it sends command to wup (retrieves measurements and stores them on remote host, through ssh) in order to start retrieving measurements for each test case.
- **script.createPlottableData**: creates a single file that enlists all the executed tasks with the energy consumption for each implementation. In addition, we used ntp14 to synchronize both system clocks which helped us to map our results of run-time performance and energy consumption.
- **script.plotGraphs**: after retrieving our data we use this script to plot our graphs. For plotting our graphs we used Gnuplot,15 an open-source general purpose pipe-oriented plotting tool.

Note that most of the scripts offer the –help option that shows a list of available command line arguments and options. In addition, we provide a `readme.md` file, available in our Repository, as a guideline for using our scripts and reproducing the obtained results.

2.3 Retrieving Energy Measurements

After rebooting a computer system, it is necessary to wait for a short period of time to reach a stable condition (where the energy usage is stable) in order to prevent additional noise in our results.

As an secondary step for our experiment, we shut down background processes, as suggested by Hindle [8], found in modern OS (Operating System) such as disk defragmentation, virus scanning software, cron jobs, automatic updates, disk indexing, document indexing, rss feed updates, and so on to minimize possible noise interferences in our measurements. Making the following steps, we reduced our platform’s idle power consumption from 8.6 to 5.8 watts on average.

After reaching the stable condition, we launched our main script i.e., **script.executeTasksRemotely**, that executes all the tasks implemented in different programming languages. Before executing a task, the execution script sends a command to the remote host, through ssh to retrieve power consumption from the remote host.

---

13https://github.com/stefanos1316/Rosetta-Code-Research
14http://www.ntp.org/
15http://www.gnuplot.info/
i.e., Raspberry Pi, through a password-less ssh connection to start collecting power consumption measurements from wup for the currently executing task. In addition, the local host retrieves run-time performance measurements through the `time`\(^{16}\) command and stores them in timestamped directories which we analyze later. Between each execution of a task, we added a `sleep`\(^{17}\) period of three minutes. The time gap exists to ensure that our experimental platform reached a stable condition and to avoid unnecessary noise in our measurements. For example, to ensure the platform’s CPU is cooled down and the fan is no longer consuming more power.

\(^{16}\)https://linux.die.net/man/1/time
\(^{17}\)http://man7.org/linux/man-pages/man3/sleep.3.html
Table 2: Comparative Results show the percentage of increased energy usage while using the inefficient implementation in comparison to the efficient

<table>
<thead>
<tr>
<th>Tasks Name</th>
<th>Implementations Efficient</th>
<th>Implementations Inefficient</th>
<th>Comparative Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array-Concat</td>
<td>PHP, Ruby</td>
<td>Swift</td>
<td>888%</td>
</tr>
<tr>
<td>Classes</td>
<td>PHP, Python</td>
<td></td>
<td>1616%</td>
</tr>
<tr>
<td>Bubble</td>
<td>JavaScript, Swift</td>
<td></td>
<td>12694%</td>
</tr>
<tr>
<td>Insertion</td>
<td>JavaScript, Perl</td>
<td>R</td>
<td>9430%</td>
</tr>
<tr>
<td>Merge</td>
<td>JavaScript, R</td>
<td></td>
<td>2894%</td>
</tr>
<tr>
<td>Quick</td>
<td>PHP, Swift</td>
<td></td>
<td>1212%</td>
</tr>
<tr>
<td>Selection</td>
<td>JavaScript, R</td>
<td></td>
<td>6657%</td>
</tr>
<tr>
<td>Url-Decode</td>
<td>PHP, Python</td>
<td>R</td>
<td>2963%</td>
</tr>
<tr>
<td>Url-Encode</td>
<td>PHP, R</td>
<td></td>
<td>3239%</td>
</tr>
</tbody>
</table>

is kept relatively high the run-time performance is not affected negatively like in Java.

Figure 3 shows the total energy dissipation from the beginning until the end of a task (while making use of the compiler optimizations). The obtained results show the use of compiler optimizations reduces energy usage in most of the cases while it increases it in others. For C#, the energy usage of quick-, insertion-, bubble-sort and url-decode increased in range of 1% to 10% whereas for merge-, selection-sort, classes and url-encode reductions were between 3.8% and 36.6%. C’s gcc -O3 achieved energy reductions ranging from 43.36% to 99.6% for the majority of tasks apart from url-decoding where the energy increased to 14.7%. In the case of C++, the g++ -O3 resulted in energy savings between 0.9% to 84.71% for all tasks except for the url-decoding that introduced increased energy usage of 33.33%. For Go most of sorting algorithms and url-decoding energy usage reductions were ranging from 14.39% to 31.35% while for insertion-sort and url-encoding increased between 15.38% to 62%.

Java was the only programming language with energy reduction for all tasks, in range of 6% to 98.4%. In the case of Rust, energy requirements for all tasks, reduced in range of 15.8% to 97.7% apart from the classes task where the energy usage increased to 15.4%. Regarding vb.net, the compiler optimization had very small impact on energy consumption; less than 10%.

For the interpreted programming languages, we can see that the energy consumption among them is significantly different as depicted in Table 2. The most inefficient case is Swift that consumes 12694% more energy compares to JavaScript. In general, PHP, Ruby, and JavaScript achieved the most energy efficiency compared to Swift, R, Perl, and Python where their implementations contributed to the highest energy consumption. In terms of run-time performance, Figure 6 results show the energy consumed by the interpreted tasks (see Figure 5) have a relationship with the execution time.

4 THREATS TO VALIDITY
Internal: In order to avoid additional overhead in our experimental setup, we used a remote host to collect our results. Therefore, the need of a wireless connection was necessary, which might result on additional energy requirements (by making use of the ssh to start and stop the wup). Moreover, we cannot have full control of our os workloads and background operations. Therefore, it is possible that some daemons might start running while testing our experiment.

External: Our real-time power analyzer offers minimum sampling interval of a second. In Figures 1, 3, and 5 the energy dissipation results to zero can be interpreted as follows. When the tasks execution is less than a seconds, this makes it impossible for wup to capture such measurements.

5 RELATED WORK
Most empirical studies evaluate software projects from particular programming language families. Here, we count the energy consumption of programming tasks across 14 programming languages. To the best of our knowledge, this is the first study that assesses the energy consumption in different programming languages using the Rosetta Code Repository. In the following, we present related work to our topic and compare our results with the results from previous studies.

5.1 Programming Languages
Studies regarding the strengths and weaknesses of different programming languages can help developers to decide which programming language they will use to perform specific programming tasks. For instance, if programmers aim at the scalability and performance of their systems, they use functional programming most of the times. On the other hand, when they want to develop programs with high modularity, they use object-oriented programming languages.

Closest to our paper is the empirical study that Nanz and Furia conducted on the Rosetta Code Repository to compare the efficiency of eight popular programming languages, including C, Go, C#, Java, F#, Haskell, Python, and Ruby [11]. Contrary to this work, we used a power analyzer to run programming tasks on 14 different programming languages in order to compare the energy consumption at runtime.

In addition, Meyerovich and Rabkin conducted an empirical study by analyzing 200,000 SourceForge projects and asking almost 13,000 programmers to identify characteristics that lead the latter to select appropriate programming languages in business level [10]. However, this study is a survey on the adoption of programming languages in the industry. Our goal here is different. We compare the energy consumption of programming tasks performed in several programming languages.

5.2 Energy Consumption and Performance
Several researchers have investigated the energy efficiency and run-time performance impact over different programming languages. Also, a significant amount of works have taken into account the execution environment where the programs can run efficiently.

In particular, Abdulsalam et al. conducted experiments on workstations [1], whereas Rashid et al. on an embedded system [12] and Chen and Zong on smart-phones [3]. Abdulsalam et al. evaluated the energy effect of four memory allocation choices (malloc, new, array, and vector) and they showed that malloc is the most efficient in terms of energy and performance [1]. Chen and Zong showed by using the Android Run Time environment (instead of Dalvik), that the energy and performance implications of Java are...
similar to C and C++ [3]. Finally, Rashid et al. compared the energy and performance impact of four sorting algorithms written in three different programming languages (ARM assembly, C/C++, and Java). They found that Java consumes the most energy [12]. From all these studies it seems that Java and Python consume a lot of energy and perform slowly in comparison with C/C++ and Assembly.

Additionally, many empirical studies have assessed the impact of coding practices (e.g. the use of for loops, getters and setters, static method invocation, views and widgets, and so on) regarding energy consumption. Characteristically, Tonini et al. conducted a study on Android applications and found that the use of for loops with specified length and the access of class variables without the use of getters and setters can reduce the amount of the energy that the applications consume [13]. Furthermore, in their study, Linares-Vásquez et al. performed analysis over 55 Android applications from various domains and they reported the most energy consuming API methods [9]. For instance, they found that the 60% of the energy-greedy APIs, 37% were related to the graphical user interface and image manipulation, while the remaining 23% were associated with the database.

Contrary to previous works, here we compare energy consuming programming tasks in more than 14 programming languages. Our results show that significant diverge with respect to energy consumption exist for interpreted programming languages. Moreover, we provide comparison in programming languages such as Go, Rust, VB.NET, and C# which is not available in prior works.

6 CONCLUSIONS AND FUTURE WORK

Nowadays, applications have been consolidated from local servers and mainframes towards data centers. This imposes the energy efficient development of independent and reusable applications in a variety of programming languages. Goal of this paper is to compare the energy consumption of specific programming tasks in different programming languages and identify which languages are appropriate to be used in modern services. In brief, we conducted an empirical study using a power analyzer to measure the energy consumption of several programming tasks found in the Rosetta Code Repository, for 14 popular programming languages. According to our findings, the total energy consumption—for the tested tasks—of compiled programming languages seems to be much lower compared to this of the interpreted ones. Overall, our experiments revealed that VB.NET and Swift are the most inefficient programming languages among the compiled and interpreted, respectively. Prior works [1, 4] also evaluated the energy efficiency of compiled programming languages and showed that C and C++ were the most energy efficient compared to Java. In contrast, the tasks we compared show that Java’s energy consumption does not diverge significantly from C and C++. Moreover, in tasks such as url -encode and decode, Java achieves more energy efficiency than C and C++. In addition, we also compared Go which results to even higher energy savings compared to C and C++ for all tasks apart from insertion-sort. Regarding interpreted programming languages, JavaScript indicates the highest level of energy-efficiency. On the other hand, Swift seems to consume the most energy in total.

As far as future work is concerned, we would like to test all the 29 collected tasks and, furthermore, to implement and evaluate additional ones, including exception handling, and particular tasks for functional programming. Moreover, we will test the collected tasks in different CPU architectures, such as AMD and ARM. We, also, plan to collect resource usages to identify possible relationships between programming languages and resources. To this end, we expect that the obtained results can shed light on how to efficiently develop larger and more complex applications.

ACKNOWLEDGMENTS

This work is partially funded by the seneca project, which is part of the Marie Skłodowska-Curie Innovative Training Networks (Itn-eid). Grant agreement number 642954.

The authors would also like to thank Dr. Vasiliki Efstathiou for her valuable help to complete this paper.

REFERENCES