

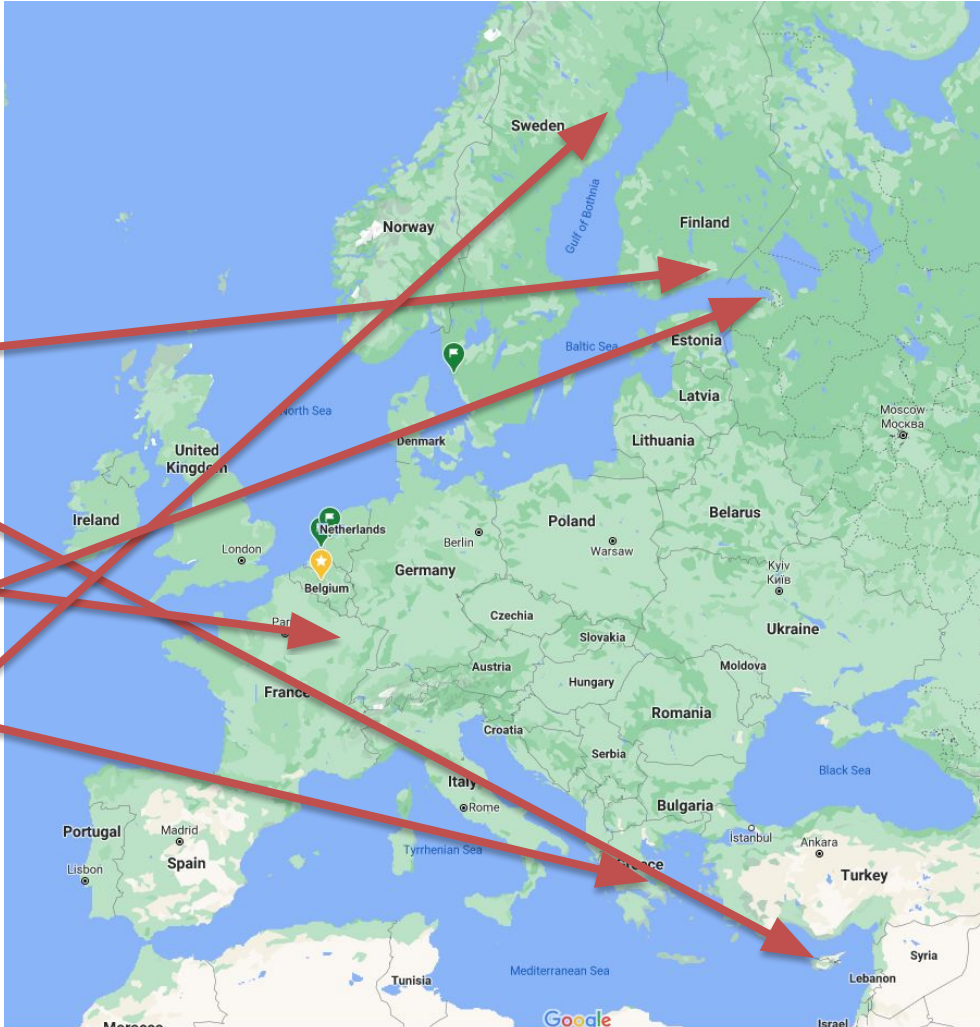


Saving Energy in Software Development by Making the Right Choices

By **Stefanos Georgiou**

4th of March 2022
Sustainable Software Engineering

Before PhD



Next to my PhD studies

Singular Logic

Innovation at your

<> Allcancode

We specialize in super-fast
web & mobile app development
for Advertising & Marketing.

📱 🖥️ 🌐 📺 📺 📺 iOS



Post Doctoral Fellow at Queen's University



Spare time

33 followers · 14 following · 60

Athens University of Economics and Bu...
Budapest, Hungary
sgeorgiou@aueb.gr
https://stefanos1316.github.io/my_port...
@StefanosGeorgi1

Highlights

Arctic Code Vault Contributor

PRO

Organizations



8,389 contributions in the last year

Contribution settings



Learn how we count c...

@AUEB-BALab

Activity overview

Contributed to
stefanos1316/my_c
stefanos1316/prog
stefanos1316/PhD
and 5 other repos





Agenda

- Introduction
- Key terms
- Existing Challenges
- Programming Languages
- Inter-Process Communication Technologies
- Security Mechanisms
- Deep Learning
- What are we missing?

Introduction



23%



14%

Annualized Total Bitcoin Footprints

Carbon Footprint

114.06 Mt CO₂



Comparable to the carbon footprint of
Czech Republic.

Electrical Energy

204.50 TWh



Comparable to the power
consumption of Thailand.

Electronic Waste

32.14 kt



Comparable to the small IT equipment
waste of the Netherlands.

Single Bitcoin Transaction Footprints

Carbon Footprint

1278.36 kgCO₂



Equivalent to the carbon footprint of
2,833,277 VISA transactions or 213,059
hours of watching Youtube.

Electrical Energy

2291.94 kWh



Equivalent to the power consumption
of an average U.S. household over
78.56 days.

Electronic Waste

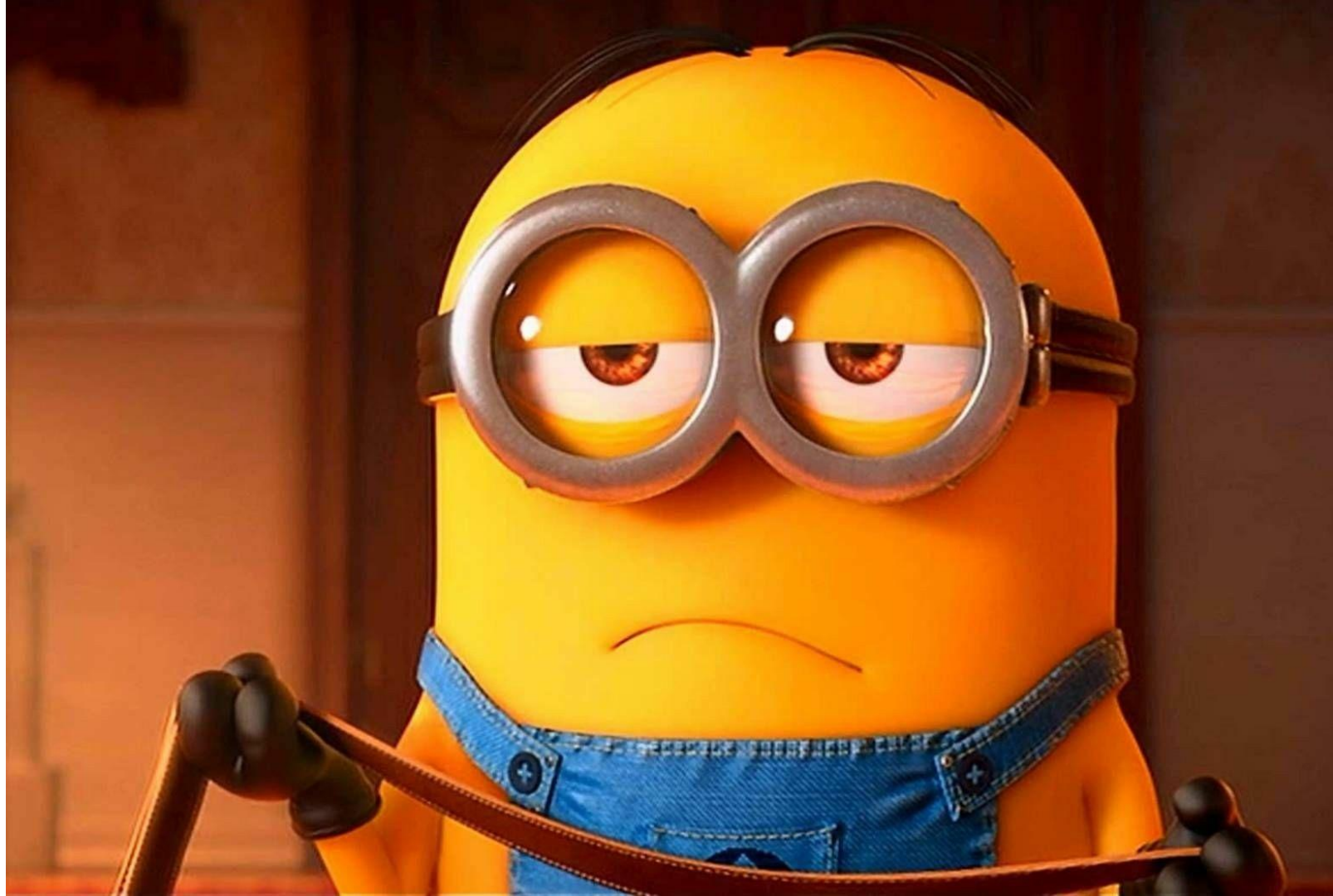
360.20 grams



Equivalent to the weight of 2.20
iPhones 12 or 0.74 iPads. (Find more
info on e-waste [here](#).)

Reducing energy consumption to combat Climate change





Sustainable software development?



Energy-efficient vs Run-time-efficient

Computer	A	B
Energy (in Joules)	30	20
Time (in seconds)	10	20

Software Development Life Cycle for Energy-Efficiency: Techniques and Tools

1

Software Development Life Cycle for Energy-Efficiency: Techniques and Tools

STEFANOS GEORGIU, Athens University of Economics and Business, Singular Logic S.A.
STAMATIA RIZOU, Singular Logic S.A.
DIOMIDIS SPINELLIS, Athens University of Economics and Business

Motivation: In modern IT systems the increasing demand for computational power is tightly coupled with ever higher energy consumption. Traditionally, energy efficiency research has focused on reducing energy consumption at the hardware level. Nevertheless, the software itself provides numerous opportunities for improving energy efficiency.

Goal: Given that energy efficiency for IT systems is a rising concern, we investigate existing work in the area of energy-aware software development and identify open research challenges. Our goal is to reveal limitations, features, and trade-offs regarding energy-performance for software development and provide insights on existing approaches, tools, and techniques for energy-efficient programming.

Method: We analyze and categorize research work mostly extracted from top-tier conferences and journals concerning energy efficiency across the software development life cycle phases.

Results: Our analysis shows that related work in this area has focused mainly on the implementation and verification phases of the software development life cycle. Existing work shows that the use of parallel and approximate programming, source code analyzers, efficient data structures, coding practices, and specific programming languages can significantly increase energy efficiency. Moreover, the utilization of energy monitoring tools and benchmarks can provide insights for the software practitioners and raise energy-awareness during the development phase.

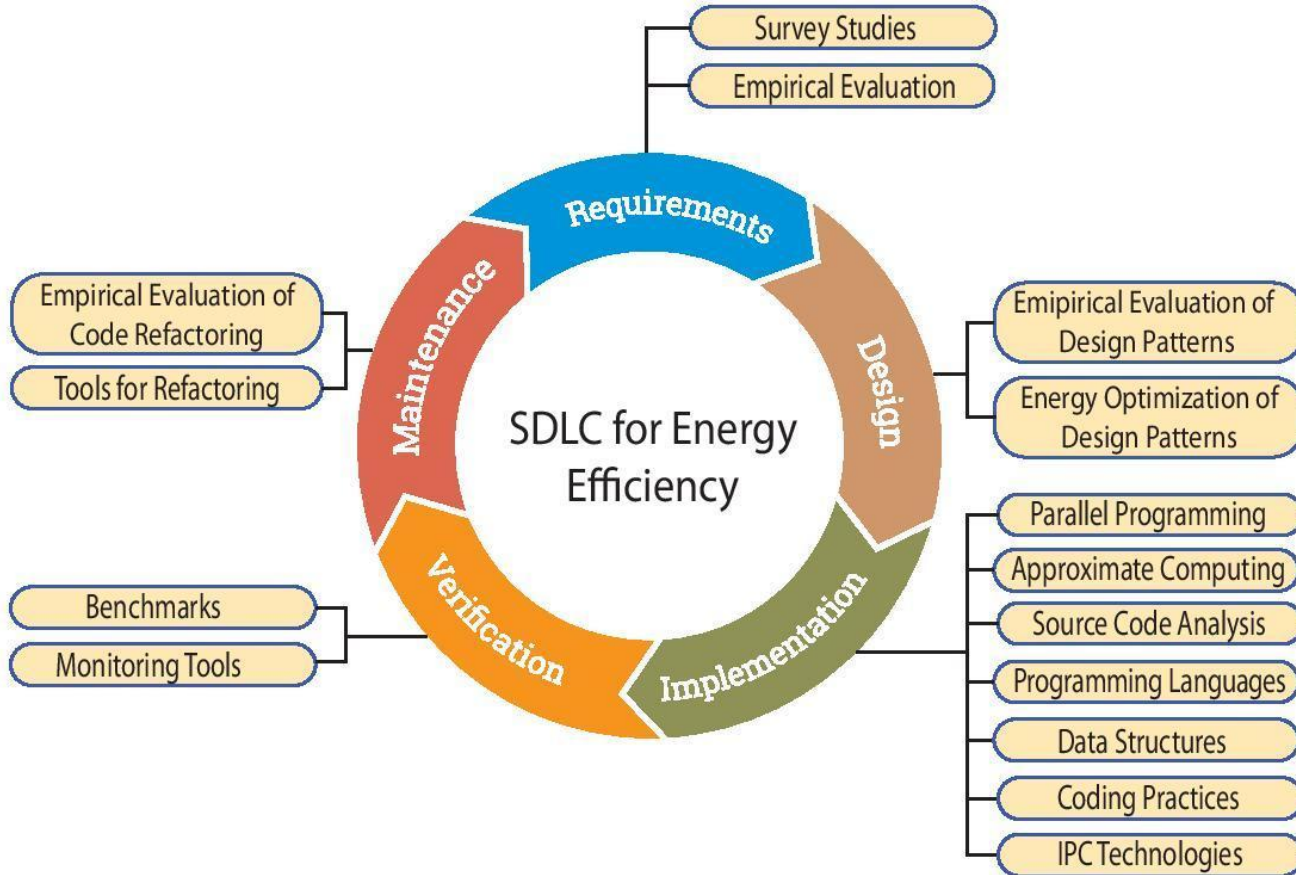
CCS Concepts: • **Software and its engineering** → **Software maintenance tools**; **Requirements analysis**; **Software design engineering**; **Software verification and validation**; *Extra-functional properties*; *Data types and structures*; *Classes and objects*; *Software design tradeoffs*; *Concurrent programming structures*; *Patterns*; *Software implementation planning*; • **Computing methodologies** → *Parallel computing methodologies*;

Additional Key Words and Phrases: GreenIT, Energy Efficiency, Energy Optimization, Energy Profiling, Design Patterns, Parallel Programming, Code Refactoring, Source Code Analysis, Coding Practices, Approximate Programming, Software Development Life Cycle

ACM Reference Format:

Stefanos Georgiou, Stamatia Rizou, Diomidis Spinellis, 2019. Software Development Life Cycle for Energy Efficiency: Techniques and Tools. *ACM Comput. Surv.* 1, 1, Article 1 (January 2019), 35 pages.
DOI: 10.1145/3337773

Survey Study: Context



Survey Study: Research Challenges

RC1. Limited investigation on diverse programming languages.

RC2. Limited investigation on diverse remote Inter-Process Communication technologies.

RC3. Limited tooling support for finding which data structures are the most energy-efficient for specific case.

RC4. Selection of configurations and parameters (parallel and approximate programming).

What are your Programming Language Energy-Delay Product Implications?

What Are Your Programming Language's Energy-Delay Implications?

Stefanos Georgiou

Athens University of Economics and Business
sgeorgiou@aub.gr

Panos Louridas

Athens University of Economics and Business
louridas@aub.gr

Maria Kechagia

Delft University of Technology
m.kechagia@tudelft.nl

Diomidis Spinellis

Athens University of Economics and Business
dds@aub.gr

ABSTRACT

Motivation: Even though many studies examine the energy efficiency of hardware and embedded systems, those that investigate the energy consumption of software applications are still limited, and mostly focused on mobile applications. As modern applications become even more complex and heterogeneous a need arises for methods that can accurately assess their energy consumption.

Goal: Measure the energy consumption and run-time performance of commonly used programming tasks implemented in different programming languages and executed on a variety of platforms to help developers to choose appropriate implementation platforms.

Method: Obtain measurements to calculate the Energy Delay Product, a weighted function that takes into account a task's energy consumption and run-time performance. We perform our tests by calculating the Energy Delay Product of 25 programming tasks, found in the Rosetta Code Repository, which are implemented in 14 programming languages and run on three different computer platforms, a server, a laptop, and an embedded system.

Results: Compiled programming languages are outperforming the interpreted ones for most, but not for all tasks. C, C#, and JavaScript are on average the best performing compiled, semi-compiled, and interpreted programming languages for the Energy Delay Product, and Rust appears to be well-placed for I/O-intensive operations, such as file handling. We also find that a good behaviour, energy-wise, can be the result of clever optimizations and design choices in seemingly unexpected programming languages.

CCS CONCEPTS

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

KEYWORDS

Programming Languages; Energy-Delay-Product; Energy-Efficiency;

ACM Reference Format:

Stefanos Georgiou, Maria Kechagia, Panos Louridas, and Diomidis Spinellis. 2018. What Are Your Programming Language's Energy-Delay Implications?. In *MSR '18: MSR '18: 15th International Conference on Mining Software Repositories*, May 28–29, 2018, Gothenburg, Sweden. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3196398.3196414>

1 INTRODUCTION

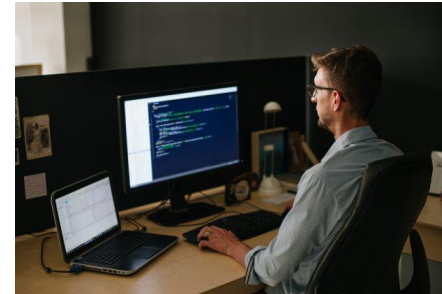
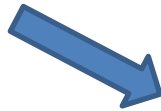
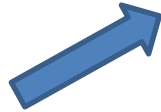
Nowadays, energy consumption¹ matters more than ever before—given that modern software applications should be able to run on devices with particular characteristics (e.g., regarding their main memory and processor). Although hardware design and utilization is undoubtedly a key factor affecting energy consumption, there is much evidence that software can also significantly influence the energy usage of computer platforms [7, 16, 18].

Today, software practitioners can select from a large pool of programming languages to develop software applications and systems. Each of these programming languages comes with a number of features and characteristics that can affect the energy consumption and run-time performance of programming tasks implemented in such languages. With the advent of cloud computing, data centers, and mobile platforms, the same programming tasks can run on distinct platforms consuming energy in different ways. In this context, there is limited work available that examines the energy and performance implications of particular programming tasks that are written in different languages and run on different platforms.

In this paper, we measure the Energy Delay Product (EDP), a weighted function of the energy consumption and run-time performance product, for a sample of commonly used programming tasks. We do this to identify which *programming language implementations* (i.e. programming tasks developed in particular programming

Motivation

Ruby
C# Rust Go
Python Java
Visual Basic .NET
R JavaScript
Perl Swift
C++ C PHP



Energy Delay Product

$$\text{EDP} = E \times T^w$$

$$w = 1$$

$$w = 2$$

$$w = 3$$



Research Questions



RQ1: Which programming languages are the most EDP efficient and inefficient for particular tasks?

RQ2: Which type of programming languages are, on average, more EDP efficient and inefficient for each of our selected platforms?

RQ3: How much does the EDP of each programming language differ among the selected platforms?

Programming Languages



- Monthly index rating based on languages popularity
- Data retrieved from 25 search engines using search query
- Programming Languages criteria:
 1. At least, 5000 hits on Google
 2. Turing complete
 3. Wikipedia page

Selected Programming Languages

Categories	Programming Languages	Compilers & Interpreters		
		Embedded	Laptop	Server
Compiled	C	6.3.0	6.4.1	6.4.1
	C++	6.3.0	6.4.1	6.4.1
	Go	1.4.3	1.7.6	1.7.6
	Rust	1.20.0	1.18.0	1.21.0
	Swift	3.1.1	3.0.2	3.0.2
Semi-Compiled	C#	4.6.2	4.6.2	4.6.2
	VB.NET	4.6.2	4.6.2	4.6.2
	Java	1.8.0	1.8.0	1.8.0
Interpreted	JavaScript	9.0.4	8.9.3	8.9.3
	Perl	5.24.1	5.24.1	5.24.1
	PHP	5.6.30	7.0.25	7.0.25
	Python	2.7.23	2.7.13	2.7.13
	R	3.3.3	3.4.2	3.4.2
	Ruby	2.4.2	2.4.1	2.4.1

Rosetta Code Repository



ROSETTACODE.ORG

[Create account](#) [Log in](#)



Community ▾

Explore ▾

[Main page](#) [Discussion](#) [View source](#) [History](#)

I'm working on modernizing Rosetta Code's infrastructure. Starting with communications. Please accept [this time-limited open invite to RC's Slack..](#) --Michael Mol (talk) 20:59, 30 May 2020 (UTC)

Rosetta Code

Rosetta Code is a [programming chrestomathy](#) site. The idea is to present solutions to the same task in as many different languages as possible, to demonstrate how languages are similar and different, and to aid a person with a grounding in one approach to a problem in learning another. Rosetta Code currently has 1,083 [tasks](#), 226 [draft tasks](#), and is aware of 813 [languages](#), though we do not (and cannot) have solutions to every task in every language.

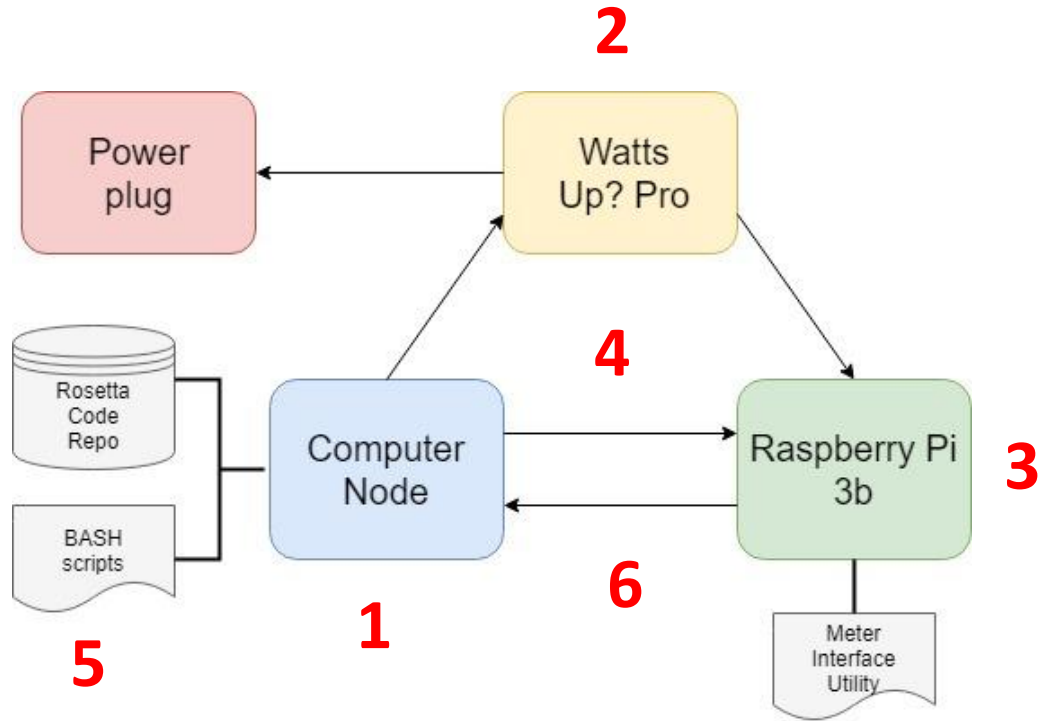
Data Set

Categories	Tasks
Arithmetic	<i>exponentiation-operator and numerical-integration</i>
Compression	<i>huffman-coding and lzw-compression</i>
Concurrent	<i>concurrency-computing and synchronous-concurrency</i>
Data structures	<i>array-concatenation and json</i>
File handling	<i>file-input-output</i>
Recursion	<i>Factorial, ackermann-function and palindrome-detection</i>
Regular Expression	<i>regular expression</i>
Sorting algorithms	<i>selection, insertion, merge, bubble, and quick</i>
String manipulation	<i>url-encoding/decoding</i>
Object-Oriented	<i>inheritance single/multiple, class, and call-an-object-method</i>
Functional	<i>function-composition</i>

Experimental Platform



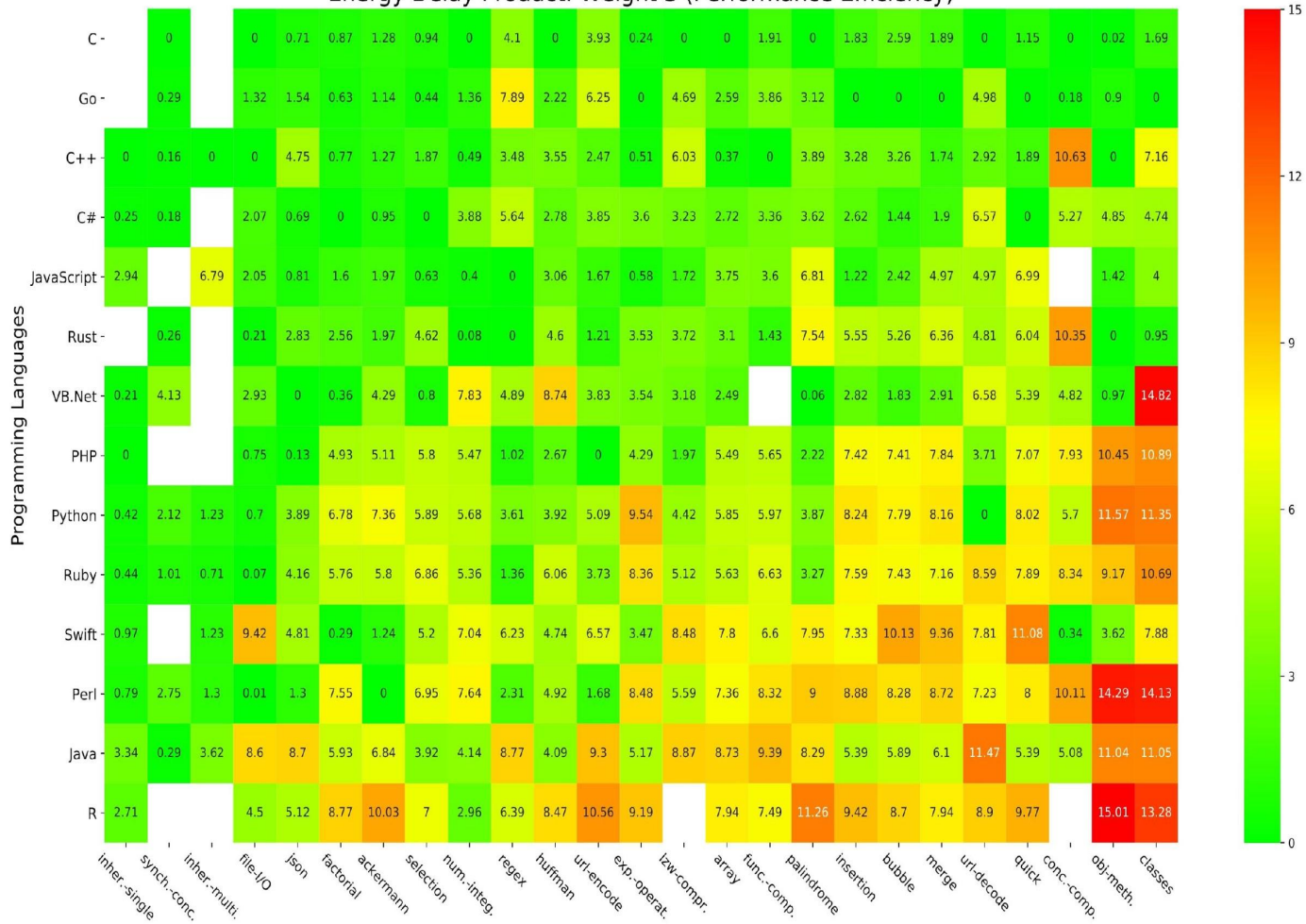
Execution Process

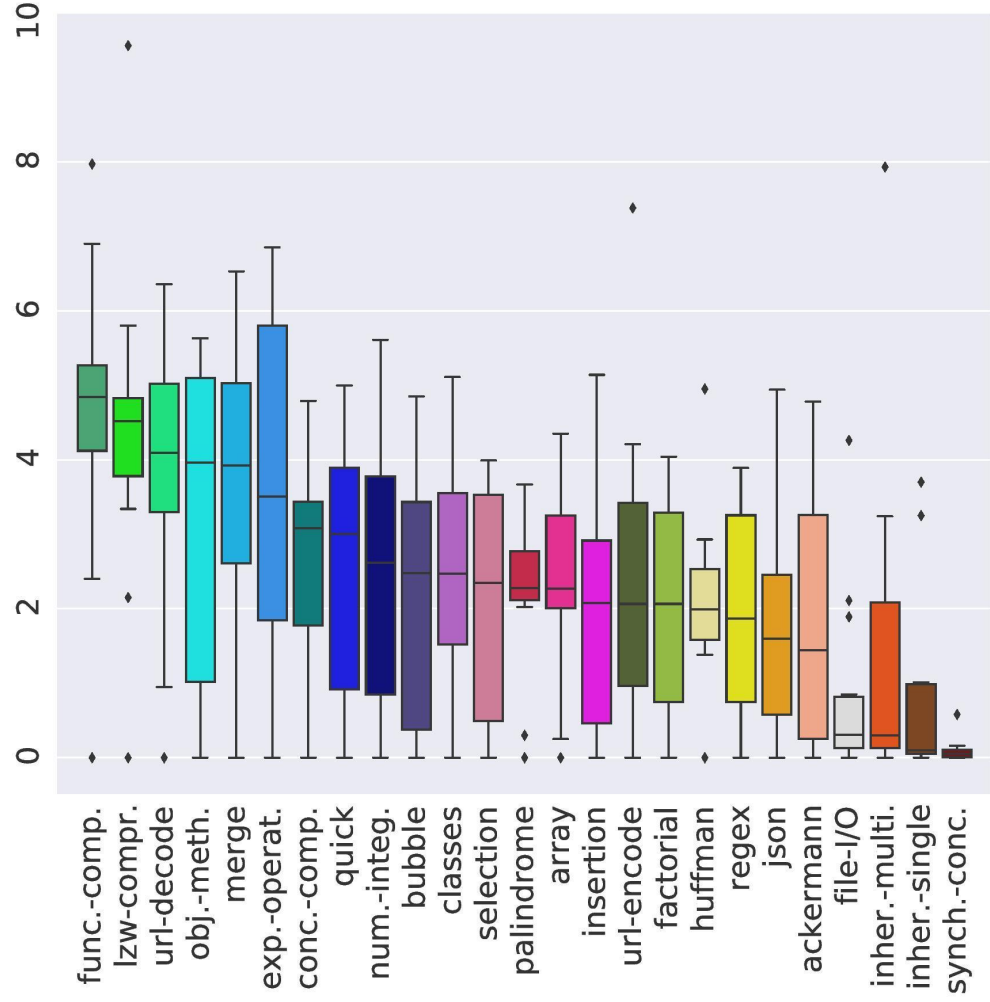


Questions

1. Drawback of using a hardware or software-based energy profiler?
2. Which factors can affect our experiment setup?

Energy Delay Product: Weight 3 (Performance Efficiency)





RQ1. Which programming languages are the most EDP efficient and inefficient for particular tasks?

Task categories	Most efficient/inefficient
Arithmetic	C/R, VB.NET
Compression	C/VB.NET, Java
Concurrent	C/VB.NET, Perl
File Handling	Rust/VB.NET
Regular Expressions	JavaScript/Java
Sorting	Go/Swift, R
Functional	C++/Swift, Perl

RQ2. Which types of programming languages are, on average, more EDP efficient and inefficient for each of the selected platforms?

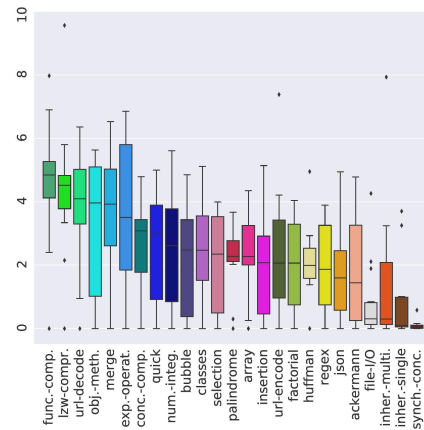
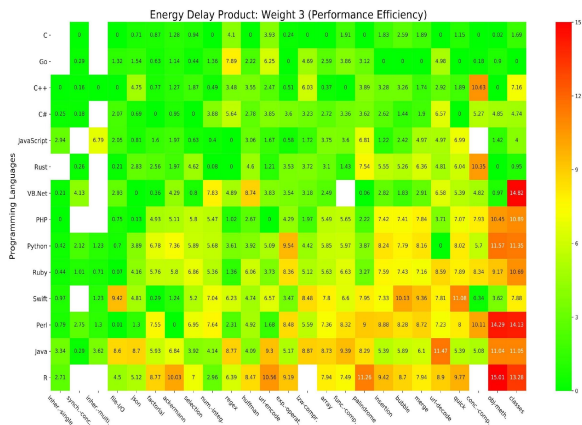
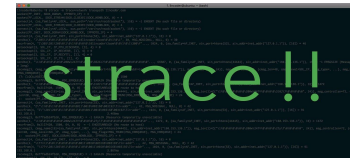
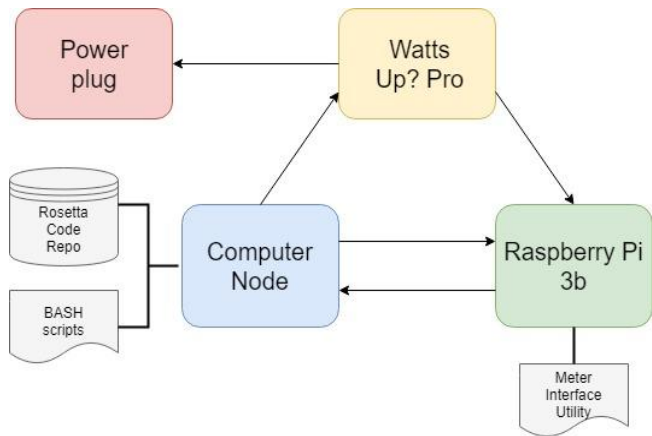
Rank	Embedded	Laptop	Server
1	C	C	C
2	C++	Go	Go
3	Go	C++	C++
4	Rust	JavaScript	C#
5	JavaScript	Rust	JavaScript
6	C#	C#	Rust
7	VB.NET	VB.NET	VB.NET
8	PHP	PHP	PHP
9	Ruby	Ruby	Python
10	Python	Swift	Ruby
11	Perl	Python	Swift
12	Java	Perl	Perl
13	Swift	Java	Java
14	R	R	R

RQ3. How much does the EDP of each programming language differ among the selected platforms?

- *Hypothesis H0: A programming language's average EDP, does not have a statistically important difference between the measurement platforms.*

In some cases, there is a significant difference between the average EDP in embedded and laptop platforms.

Takeaways



Energy-Delay Investigation of Remote Inter-Process Communication (IPC) Technologies

Energy-Delay Investigation of Remote Inter-Process Communication Technologies

Stefanos Georgiou^{*}, Diomidis Spinellis^{*}

^{*}Department of Management Science and Technology, Athens University of Economics and Business, Patission 76, Athens 10434

ARTICLE INFO

Keywords:
Energy Efficiency
Programming Languages
Remote Inter-Process Communication
System Calls

ABSTRACT

Most modern information technology devices use the Internet for creating, reading, updating, and deleting shared data through remote inter-process communication (IPC). To evaluate the energy consumption of IPC technologies and the corresponding run-time performance implications, we performed an empirical study on popular IPC systems implemented in Go, Java, JavaScript, Python, PHP, Ruby, and C#. We performed our experiments on computer platforms equipped with Intel and ARM processors. We observed that JavaScript and Go implementations of gRPC offer the lowest energy consumption and execution time. Furthermore, by analysing their system call traces, we found that inefficient use of system calls can contribute to increased energy consumption and poor execution time.

1. Introduction

The energy consumption,¹ for the IT-related products, is an evergrowing matter that has caught the attention of academic researchers and industry. This is primarily due to the increasing costs, as IT-related energy consumption is estimated to reach 15% of the world's total by 2020 [42]. Environmental impact is another major concern, as IT's total greenhouse gas emissions are expected to reach 2.3% by the same year [11]. Energy consumption of IT systems is particularly important in two areas. First, the data centres, one of the vital contributors of IT sector's global energy consumption and greenhouse gas emissions. These are housing large number of server nodes communicating with clients through energy-intensive remote inter-process communication (IPC) technologies. Second, the blossoming field of IoT, where low energy performance is critical, has multiple embedded devices connected with hyper-physical systems to exchange, share, and transmit data. To this end, providing sustainable solutions, by reducing energy consumption, to ensure data centres' and IoT infrastructures environmental sustainability and business growth is of paramount importance.

Researchers have carried out studies on different aspects and granularity of software artifacts to investigate the energy consumption of data structures [32, 12, 27, 28, 44], different programming languages [26, 3, 35, 18], multi-threaded applications [31, 29, 30], and coding practices [41, 19, 37, 39, 33]. In terms of remote IPC technologies, prior work [13, 8, 7, 22, 23] focused on investigating the energy consumption and run-time performance of smart phones and embedded systems on Java implementations for remote IPC such as RPC, REST, SOAP, and WebSockets. However, IPC

technologies have not been investigated in terms of energy consumption and run-time performance for different programming language implementations.

In this work, we research computer platforms equipped with Intel and ARM processors using three different IPC technologies available in Java, JavaScript, Go, Python, PHP, Ruby, and C#. We try to identify which programming language and IPC technology implementations offer the best energy and run-time performance when invoking remote procedures. Furthermore, we focus on pointing out the reasons behind our results to help software developers, specifically those concerned with IPC library development, build more energy and run-time performance-efficient implementations.

To accomplish this, we perform an empirical study on the selected computer systems on seven popular programming languages that offer implementations of three well known remote IPC technologies and investigate their energy and run-time performance cost. Our results highlight the efficiency of different implementations and libraries. We also examine whether the energy consumption of IPC technologies is proportional to the run-time performance or the systems' resource usage.

Results reveal that JavaScript and Go implementations of gRPC offer the most energy-efficient and best run-time performance implementations among the considered IPC technologies, while Ruby, PHP, Python, Java, and C# perform most inefficiently, for the most cases. We also found that the energy consumption and run-time performance is not proportional for all the examined IPC technology implementations. Besides, from the extracted system call traces, we were able to name certain misuse cases of computer resources that can contribute to higher energy demands and lower run-time performance.

This work is organised as follows. Section 3 presents our

^{*}stgeorgi@aubueh.gr (S. Georgiou); dspinellis@aubueh.gr (D. Spinellis)
ORCID(s):

Motivation

The Internet in Real Time

Like 267

Share

Save



By the time you finish reading this sentence, there will have been 219,000 new Facebook posts, 22,800 new tweets, 7,000 apps downloaded, and about \$9,000 worth of items sold on Amazon... depending on your reading speed, of course. Now that the Internet is widely available, just one second of global online activity is jam-packed full of events, from communication with others to data storage to entertainment options galore.

For example, in the amount of time you've been on this page, this is how much data has already passed through the Internet.

4,104,000

GIGABYTES OF DATA

Research Questions



RQ1. Which IPC technology implementation offers the most energy and run-time performance efficient results?

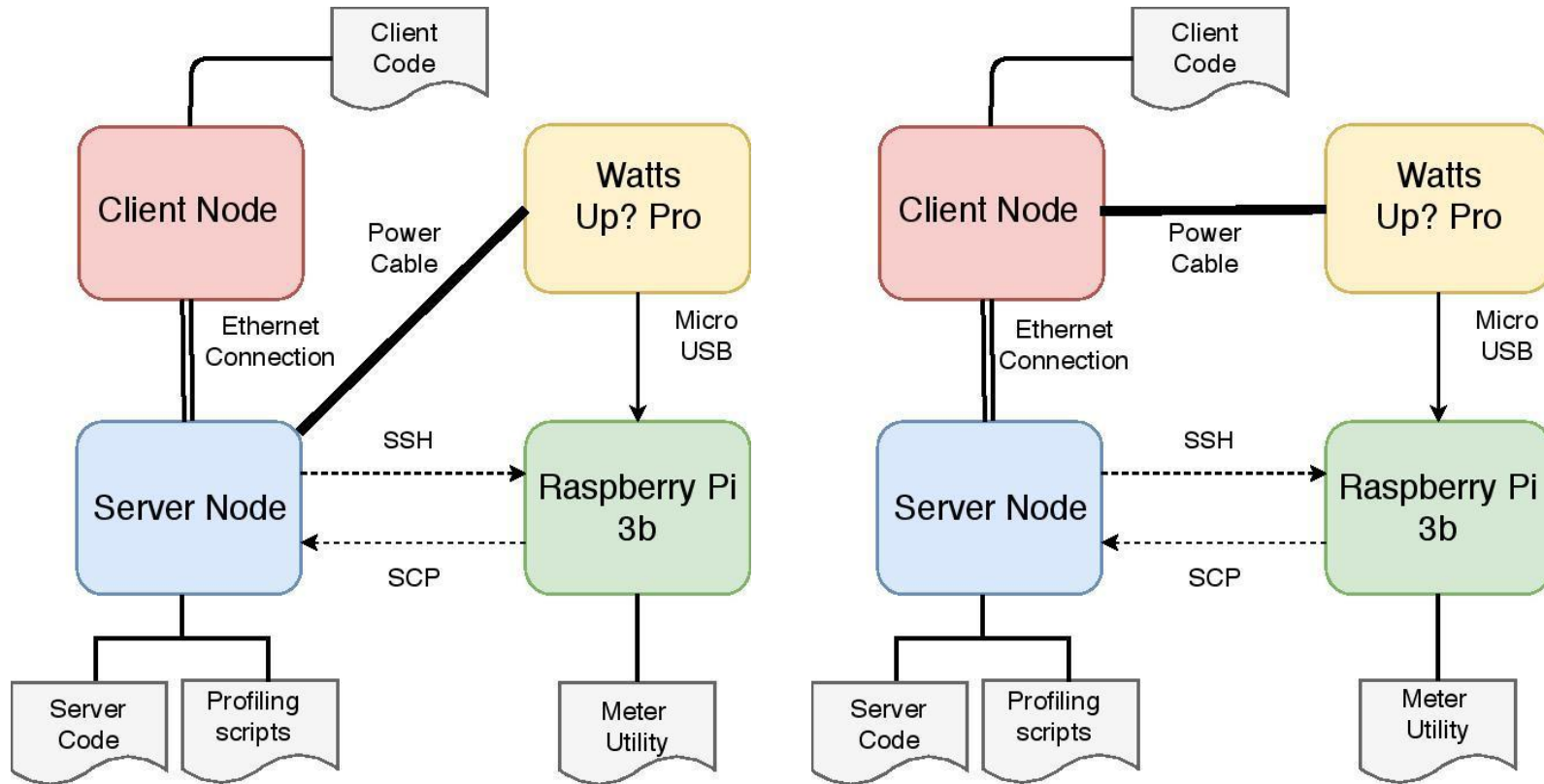
RQ2. What are the reasons that make certain IPC technologies more energy and run-time performance efficient?

RQ3. Is the energy consumption of the IPC technologies proportional to their run-time performance or resource usage?

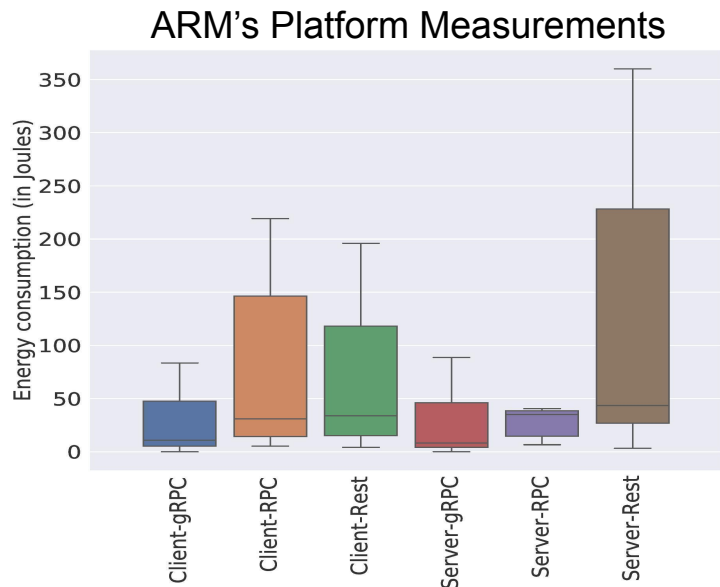
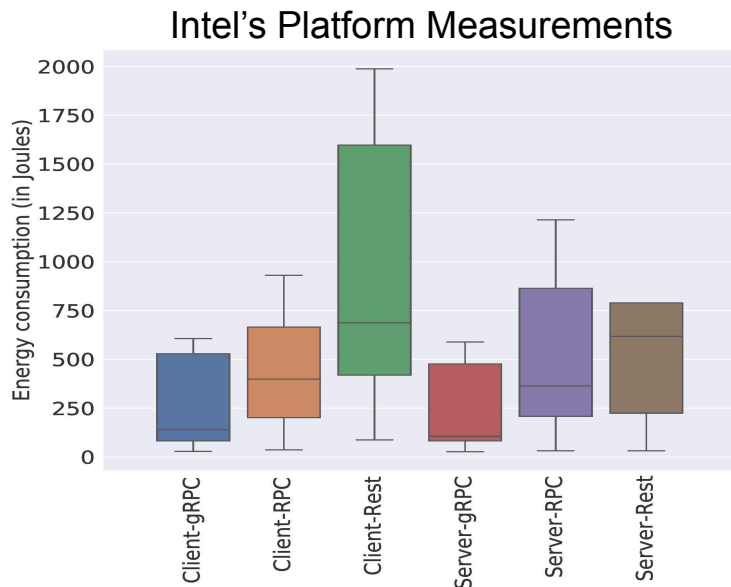
Subject Systems

Categories	Programming Languages	Compiler and Interpreter Versions	
		ARM Processor	Intel Processor
Compiled	Go	1.9.4	1.9.4
Semi-Compiled	Java	1.8.0	1.8.0
	C#	4.8.0	4.8.0
Interpreted	JavaScript	10.4.0	10.4.0
	Python	2.7.14	2.7.14
	PHP	7.2.12	7.2.12
	Ruby	2.5.3p	2.5.3p

Execution Process

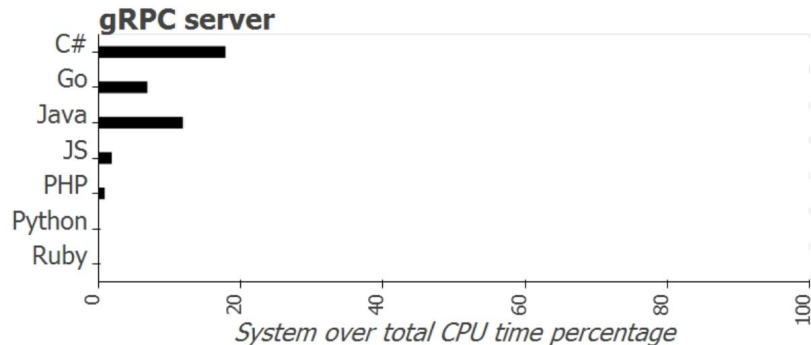
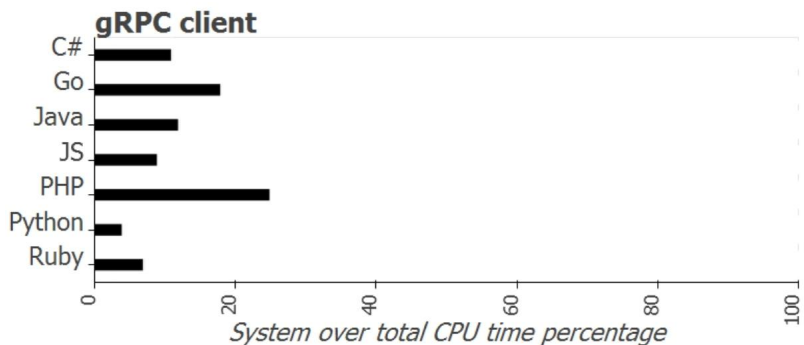


RQ1. Which IPC technology implementation offers the most energy and run-time performance-efficient results?



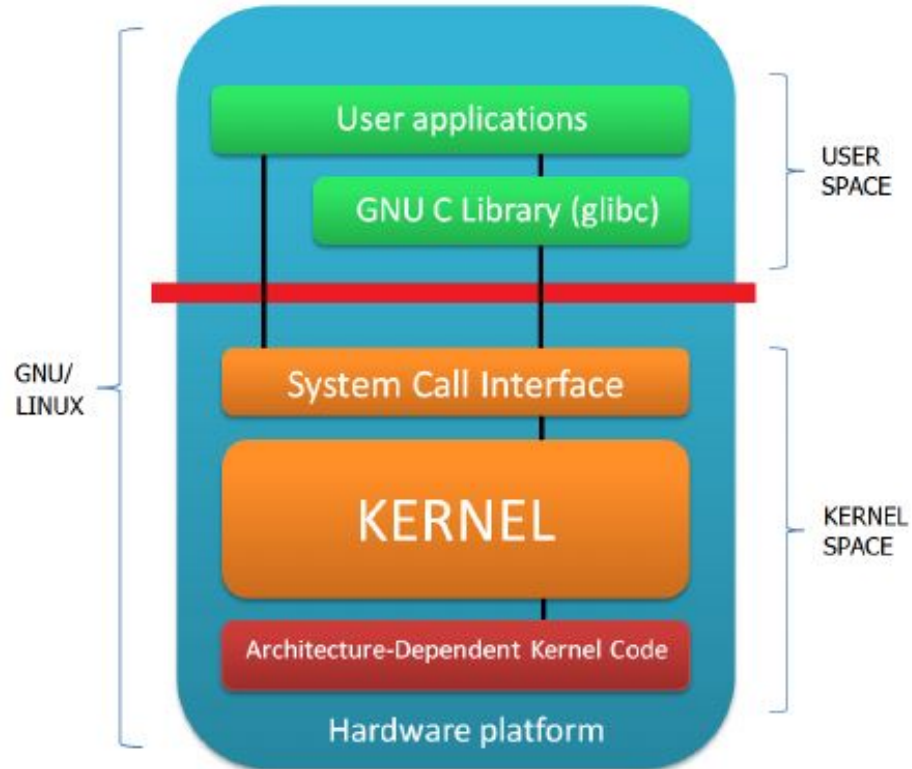
***JavaScript** and **Go** are the programming languages offering the most energy and run-time performance efficient library implementations for the Intel and ARM platforms. In addition, for almost all programming language implementations, we found that **gRPC** is the IPC technology having the most efficient results.*

RQ2. What are the reasons that make certain IPC technologies more energy and run-time performance-efficient?



Our analysis shows the frugal opening, connecting, closing, accepting, and shutting down connections can impact the energy consumption and run-time performance of the IPC technologies. The usage of `writew` system call appears in the most efficient implementations.

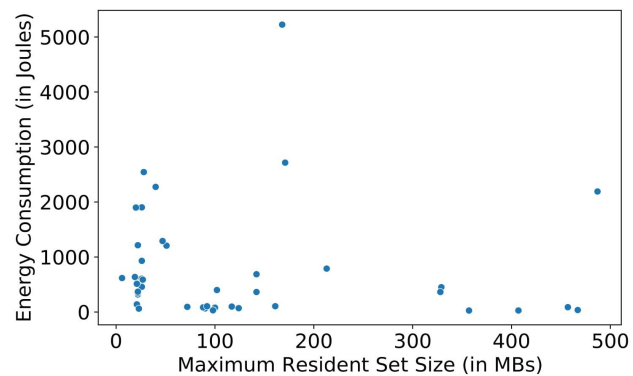
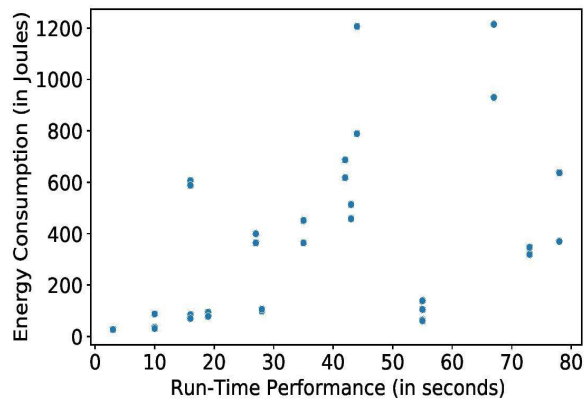
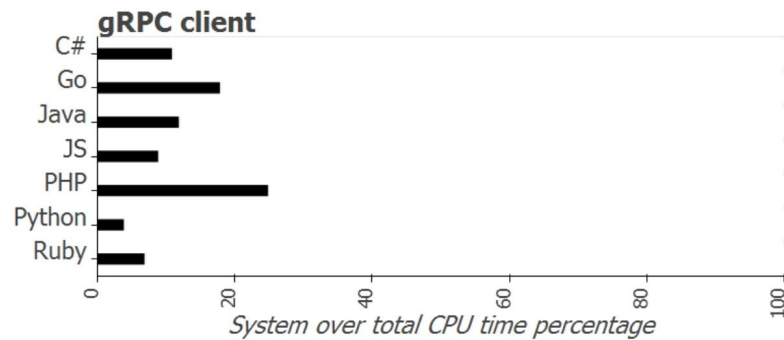
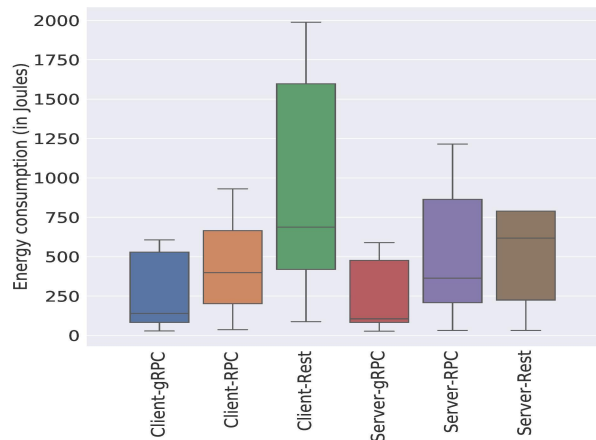
What are the system calls?



RQ3. Is the energy consumption of the IPC technologies proportional to their run-time performance or resource usage?

We found that there is a positive moderate and very strong monotonic correlation between the energy consumption and run-time performance of the Intel and ARM platforms, respectively. Also, we found a weak and very weak monotonic relationship between our energy measurements and resource usage. Therefore, none of the collected resource usage measurements can be used to justify the energy consumption results in terms of IPC technologies.

Takeaways





Energy-Efficient Computing in a Secure Environment

111

Energy-Efficient Computing in a Secure Environment

STEFANOS GEORGIU, Athens University of Economics and Business, Greece
DIMITRIS MITROPOULOS, Athens University of Economics and Business, Greece
DIOMIDIS SPINELLIS, Athens University of Economics and Business, Greece

An operating system (os) typically includes numerous security mechanisms that protect the confidentiality, integrity, and availability of its data and services. However, such safeguards may impact energy consumption and encumber the run-time performance of applications running on a system. We present a large scale study, where we investigate how the various security mechanisms affect energy and run-time performance cost at the os-level. We focus on well-known mechanisms including encrypted communication protocols, memory zeroing, gcc safeguards, and CPU vulnerability patches against critical vulnerabilities, such as Meltdown. To do so, we utilise 128 benchmarks of different application types found under the well-established Phoronix test suite. Our findings suggest that security mechanisms lead to an increased energy consumption and significantly degrade the run-time performance of various applications including web servers, databases systems, kernel operations, and disk usage. Notably, when we disabled the various security mechanisms, real-world applications such as Apache and Redis, indicated important energy (from 18% to 41%) and run-time performance (from 23% to 45%) gains. Additionally, we examined the correlation between energy consumption and performance. Our findings showed that the two are not always related. Overall, our results suggest that administrators should consider disabling such security mechanisms when a computer system runs inside a secure environment to benefit from energy and run-time performance gains.

CCS Concepts: • Hardware → Power and energy; • Security and privacy → Systems security.

Additional Key Words and Phrases: energy consumption, CPU vulnerability patches, secure environment, security mechanisms, gcc safeguards, encrypted network communications, memory zeroing, Spectre, Meltdown, mds.

ACM Reference Format:

Stefanos Georgiou, Dimitris Mitropoulos, and Diomidis Spinellis. 2020. Energy-Efficient Computing in a Secure Environment. *ACM Trans. Comput. Syst.* 37, 4, Article 111 (June 2020), 29 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnn>

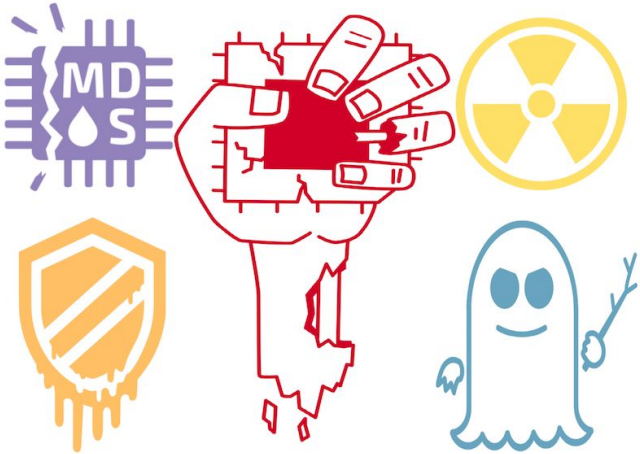
1 INTRODUCTION

The high computational demands and services, offered by the different IT-related products, increase the energy consumption of the computer systems. The study of energy consumption in computer systems has gained vast popularity in recent years due to the advent of mobile applications and data centers that are responsible for more than 10% of the world's [18, 25, 62]. Therefore, reducing the energy footprint of IT products is of increased importance.

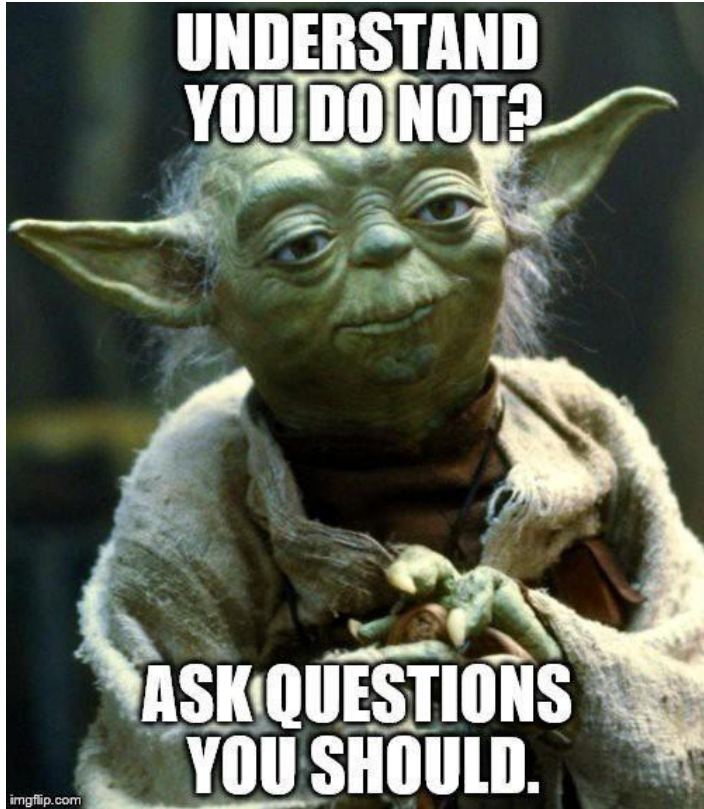
Unix-like os's are equipped with various services, background processes, and daemons [59]. Among such services are the security mechanisms that protect computer systems from attackers trying to exploit potential vulnerabilities. Such security mechanisms can affect the energy consumption and the run-time performance of a

Authors' addresses: Stefanos Georgiou, sgeorgiou@aub.gr, Athens University of Economics and Business, Patission 76, Athens, Greece, 10434; Dimitris Mitropoulos, dimitro@aub.gr, Athens University of Economics and Business, Patission 76, Athens, Greece, 10434; Diomidis Spinellis, dds@aub.gr, Athens University of Economics and Business, Patission 76, Athens, Greece, 10434.

Motivation



Research Questions



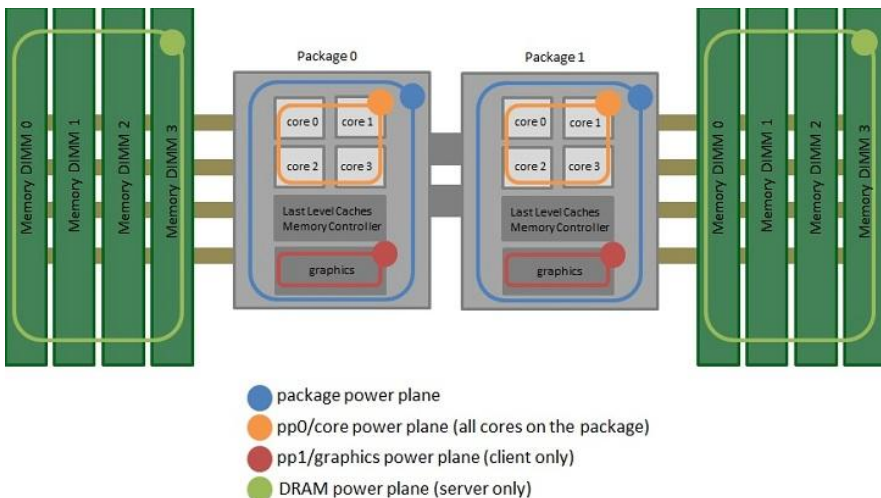
RQ1. What are the energy and run-time performance implications of the investigated security mechanisms on a computer system?

RQ2. Is the energy consumption of the examined security mechanisms proportional to their run-time performance?

RQ3. How do security mechanisms affect the energy consumption and the run-time performance of diverse applications and utilities?

Subject Systems

System	Kaby Lake (x86)
Microarchitecture	Kaby Lake
Processor/Soc	Core i7-7700
Cores × threads	4 × 2
Base Frequency	3.6 GHz
Max Frequency	4.2 GHz
Cache line size	64 B
L1-D/L1-I Cache	4 x 32 KiB 8-way
L2 cache	4 x 256 KiB 4-way
L3 cache	8 MB 16-way
I-TBL	4-KByte pages, 8-way
D-TBL	1-GB pages, 4-way
L2-TBI	1-MB, 4-way
RAM	16 GB UDIMM, DDR4 2400



Scenarios

CPU Vulnerability Patches

GCC Security Flags

HTTP/HTTPS

Stock

Stock

HTTP

Meltdown

Stack Protector

HTTPS

Spectre

FORTIFY_SOURCE

MDS

PIC/PIE

AllOff

RELRO

AllOff

Data-set

Category	Benchmark Suites
Audio Encoding	encode-mp3, encode-flac
Video Encode/Decode	dav1d, svt-av1, svt-hevc, svt-vp9, vpxenc, x264, x265, ffmpeg
Code Compilation	build-php, build-linux-kernel, build-gcc, build-gdb, build-llvm, build2
File Compression	compress-p7zip, compress-bzip2, compress-zstd, compress-xz, lzbench
Database Suite	sqlite, redis, rocksdb, cassandra, mcperf, pymongo-insert
CPU Massive	aircrack-ng, apache, blogbench, brl-cad, byte, cloverleaf, cpp-perf-bench, crafty, dacapo-bench, ebizzy, embree, fhourstones, glibc-bench, gmpbench, himeno, hint, hmma, hpcg, javascimark2, m-queens, minion, nero2d, nginx, node-express-loadtest, numenta-nab, phpbench, primesieve, pybench, pyperformance, rodinia, rust-prime, scimark2, stockfish, swet, sysbench, sudokut, tensorflow, xsbench, sunflow, bork, java-jmh, renaissance, tiobench, openssl, blake2s, john-the-ripper, botan, octave-bench, oidn
Disk Suite	fs-mark, iohome, dbench, postmark, aio-stress
Kernel	schbench, ctx-clock, stress-ng, osbench
Machine Learning	rbenchmark, numpy, scikit-learn, mkl-dnn
Memory Suite	ramspeed, stream, t-test1, cachebench, tinymembench, mbw
Networking Suite	iperf, network-loopback
Imaging	graphics-magick, inkscape, rawtherapee, tjbench, dcraw, darktable, rsvg, gegl
Renderers	tungsten, ospray, aobench, c-ray, povray, smallpt, ttsiod-renderer, indigobench, rays1bench, j2bench, qgears, jxrendermark
Desktop Graphics	xonotic, openarena, tesseract, paraview, unigine-valley, unigine-heaven, nexuiz, glmark2

```
while (get_time() - start_t < TIME) {  
    create_files();  
}
```

```
for (int i = 0; i < 1000; i++) {  
    create_files();  
}
```

Results

1--3.3%	3.4--6.6%
6.7--9.9%	10% >

Tasks	Stock		Meltdown		Spectre		MDS		AllOff		Difference (%)	
	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time
cassandra_read	127	15	125	15	124	15	125	15	123	15	3.6	0
cassandra_write	115	13	112	13	112	13	111	13	111	13	3.6	0
mcperf_add	804	30	760	28	665	25	777	29	590	21	26.6	30
mcperf_append	752	29	710	27	632	24	712	27	555	20	26	31
mcperf_delete	550	18	521	17	441	14	508	17	387	12	29.6	33.3
mcperf_get	546	18	503	16	434	14	500	17	386	12	29.2	33.3
mcperf_prepend	748	29	710	27	631	23	711	27	555	20	25.8	31.1
mcperf_replace	747	30	710	28	630	25	720	29	553	20	25.8	31
mcperf_set	802	30	762	28	666	25	766	29	594	21	25.9	30
pymongo	883	38	890	39	866	37	870	38	867	38	1.8	0
redis_get	1972	66	1631	53	1761	58	1685	55	1166	36	40.8	45.4
redis_lpop	1985	65	1629	52	1756	58	1672	54	1167	36	41.1	45.4
redis_lpush	1988	66	1640	53	1763	58	1665	54	1166	36	41.3	45.4
redis_sadd	1981	66	1642	52	1764	58	1680	54	1169	36	40.9	45.5
redis_set	1972	66	1622	53	1761	58	1680	54	1167	36	40.8	44.6
rocksdb_fillrand	1281	31	1271	31	1243	31	1250	31	1212	31	5.3	0
rocksdb_fillseq	976	21	969	21	951	21	960	21	936	20	4.1	4.7
sqlitebench	944	134	962	136	878	125	930	134	908	130	3.8	3

Impact on Kernel operations

Tasks	Stock		Meltdown		Spectre		MDS		Alloff		Difference (%)	
	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time
ctx_clock	6192	318	3523	174	6178	318	3165	169	1121	56	81.8	82.3
osb_files	659	29	631	27	593	26	622	27	542	23	17.6	20.6
osb_processes	887	22	852	20	889	22	870	21	803	19	9.4	13.6
osb_threads	588	19	549	17	574	19	550	18	494	15	15.9	21
osb_mem_alloc	514	23	503	22	508	23	501	22	480	21	6.6	8.6
osb_programs	584	11	553	10	576	11	546	10	511	9	12.4	18.1
stress-ng_fork	1232	24	1185	23	1222	24	1170	22	1108	21	10	12.5
stress-ng_matrix	799	14	759	13	775	13	786	14	752	13	5.9	7.1
stress-ng_msg	1024	22	862	18	988	21	636	13	519	10	49.2	54.5
stress-ng_sem	911	20	919	21	888	20	782	17	812	18	9.3	10
stress-ng_sock	1719	30	1392	28	1591	28	1367	28	1309	26	23.8	13.3
stress-ng_switch	1431	26	1411	25	1329	23	1204	21	1125	19	21.3	26.9
stress-ng_vec	1446	29	914	18	913	18	910	18	911	18	36.9	37.9

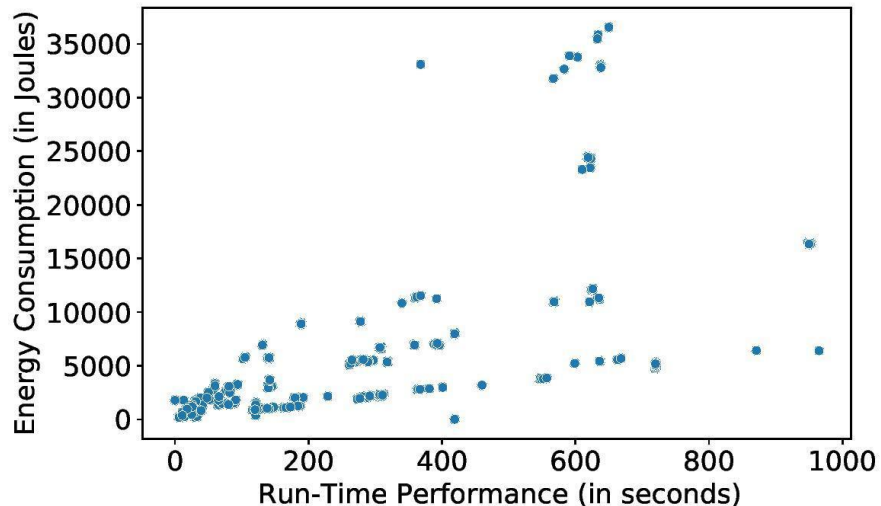
RQ1. What are the energy and run-time performance implications of the investigated security mechanisms on a computer system?

Tasks	Stock		Meltdown		Spectre		MDS		AllOff		Difference (%)	
	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time
build2	10419	196	10352	197	10329	197	10282	197	10247	195	1.6	0.5
gcc	17039	359	16830	353	16898	355	16796	350	16539	345	3	3.9
gdb	5791	134	5695	131	5729	132	5666	130	5587	128	3.5	4.4
kernel	81490	1555	73093	1382	73330	1393	72572	1371	72103	1374	11.5	11.6
llvm	45784	864	45569	861	45488	862	45302	858	45052	851	1.5	1.5
php	4403	99	4380	98	4366	98	4360	98	4318	97	1.9	2

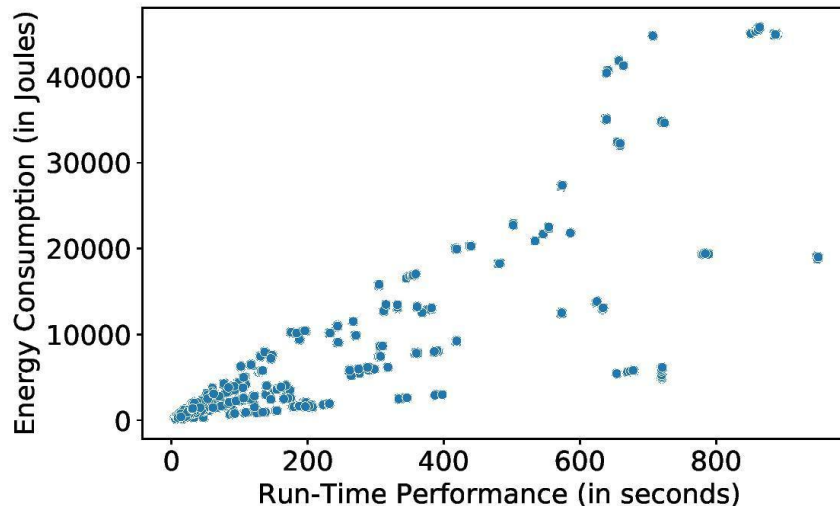
*CPU vulnerability patches can **impact the energy and run-time performance real-world applications from 18% up to 45%**, respectively. Similarly, GCC safeguards affect the energy and run-time performance of applications **up to 10%**. Similar results appear for the communications-related security mechanisms as well.*

RQ2. Is the energy consumption of the examined security mechanisms proportional to their run-time performance?

CPU Vulnerability Patches Measurements



GCC Security Flags Measurements



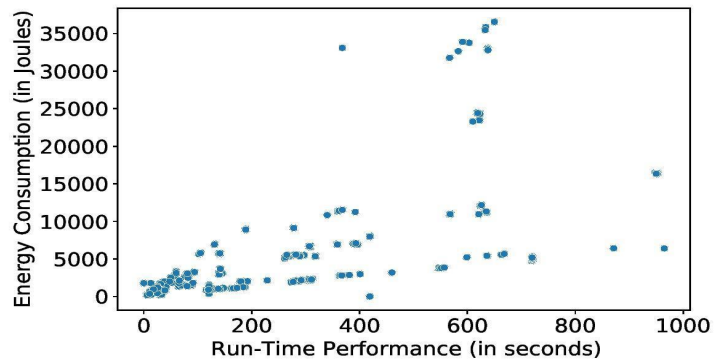
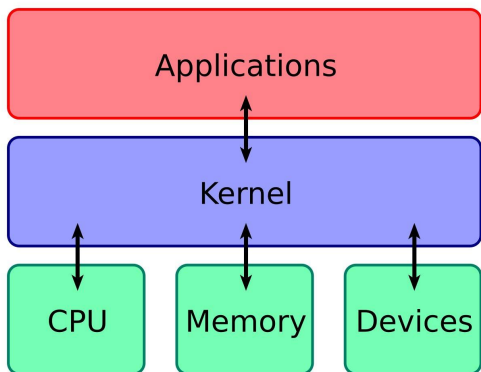
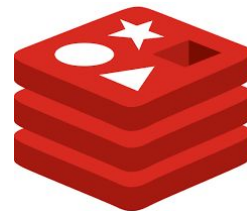
Our findings suggest that energy consumption and the run-time performance have a very strong monotonic correlation for the investigated benchmarks in the case of the CPU vulnerability patches and the GCC safeguards.

RQ3. How do security mechanisms affect the energy consumption and the run-time performance of diverse applications and utilities?

*Application types such as **database systems**, **code compilation**, **compute-intensive**, **kernel operations**, **disk usage** had the highest energy and run-time performance gains after disabling the CPU vulnerability patches. For the GCC safeguards, **compute-intensive**, **databases systems**, and **file compression** applications had the highest energy and run-time performance gains after disabling security flags.*

Takeaways

Tasks	Stock		Meltdown		Spectre		MDS		AllOff		Difference (%)	
	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time	Energy	Time
cassandra_read	127	15	125	15	124	15	125	15	123	15	3.6	0
cassandra_write	115	13	112	13	112	13	111	13	111	13	3.6	0
mcperf_add	804	30	760	28	665	25	777	29	590	21	26.6	30
mcperf_append	752	29	710	27	632	24	712	27	555	20	26	31
mcperf_delete	550	18	521	17	441	14	508	17	387	12	29.6	33.3
mcperf_get	546	18	503	16	434	14	500	17	386	12	29.2	33.3
mcperf_prepend	748	29	710	27	631	23	711	27	555	20	25.8	31.1
mcperf_replace	747	30	710	28	630	25	720	29	553	20	25.8	31
mcperf_set	802	30	762	28	666	25	766	29	594	21	25.9	30
pymongo	883	38	890	39	866	37	870	38	867	38	1.8	0
redis_get	1972	66	1631	53	1761	58	1685	55	1166	36	40.8	45.4
redis_lpop	1985	65	1629	52	1756	58	1672	54	1167	36	41.1	45.4
redis_lpush	1988	66	1640	53	1763	58	1665	54	1166	36	41.3	45.4
redis_sadd	1981	66	1642	52	1764	58	1680	54	1169	36	40.9	45.5
redis_set	1972	66	1622	53	1761	58	1680	54	1167	36	40.8	44.6
rocksdb_fillrand	1281	31	1271	31	1243	31	1250	31	1212	31	5.3	0
rocksdb_fillseq	976	21	969	21	951	21	960	21	936	20	4.1	4.7
sqlitebench	944	134	962	136	878	125	930	134	908	130	3.8	3



Green AI: Do Deep Learning Frameworks Have Different Costs?

Green AI: Do Deep Learning Frameworks Have Different Costs?

Stefanos Georgiou^{*}
Queen's University
stefanos.georgiou@queensu.ca

Maria Kechagia^{*}
University College London
m.kechagia@ucl.ac.uk

Tushar Sharma
Dalhousie University
tushar@dal.ca

Federica Sarro
University College London
f.sarro@ucl.ac.uk

Ying Zou
Queen's University
ying.zou@queensu.ca

ABSTRACT

The use of Artificial Intelligence (AI), and more specifically of Deep Learning (DL), in modern software systems, is nowadays widespread and continues to grow. At the same time, its usage is energy demanding and contributes to the increased CO₂ emissions, and has a great financial cost as well. Even though there are many studies that examine the capabilities of DL, only a few focus on its *green aspects*, such as energy consumption.

This paper aims at raising awareness of the costs incurred when using different DL frameworks. To this end, we perform a thorough empirical study to measure and compare the energy consumption and run-time performance of six different DL models written in the two most popular DL frameworks, namely PyTorch and TensorFlow. We use a well-known benchmark of DL models, DEEPLARNINGEXAMPLES, created by NVIDIA, to compare both the training and inference costs of DL. Finally, we manually investigate the functions of these frameworks that took most of the time to execute in our experiments.

The results of our empirical study reveal that there is a statistically significant difference between the cost incurred by the two DL frameworks in 94% of the cases studied. While TensorFlow achieves significantly better energy and run-time performance than PyTorch, and with large effect sizes in 100% of the cases for the training phase, PyTorch instead exhibits significantly better energy and run-time performance than TensorFlow in the inference phase for 66% of the cases, always, with large effect sizes. Such a large difference in performance costs does not, however, seem to affect the accuracy of the models produced, as both frameworks achieve comparable scores under the same configurations. Our manual analysis, of the documentation and source code of the functions examined, reveals that such a difference in performance costs is under-documented, in these frameworks. This suggests that developers need to improve the documentation of their DL frameworks, the source code of the functions used in these frameworks, as well as to enhance existing DL algorithms.

CCS CONCEPTS

• **Hardware** → **Power and energy**; • **Software and its engineering** → **Software libraries and repositories**; • **Computing methodologies** → **Machine learning**;

KEYWORDS

Energy consumption, run-time performance, deep learning, APIs

ACM Reference Format:

Stefanos Georgiou, Maria Kechagia, Tushar Sharma, Federica Sarro, and Ying Zou. 2022. Green AI: Do Deep Learning Frameworks Have Different Costs?. In *44th International Conference on Software Engineering (ICSE '22)*, May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Deep learning (DL) is a field of machine learning (ML) that has recently gained significant attention from researchers and practitioners. Along with the increase of computational power and availability of data, the use of deep learning has contributed to the improvement of several applications (e.g., in the medical, financial, transportation sectors) that, for instance, use speech and image recognition, machine translation, and natural language processing (NLP). The advancement in these areas would have not been possible without the great advancement in DL.

While the research community has spent a significant effort towards improving the accuracy of DL approaches, it has often overlooked their costs. As recently reported by Schwartz et al. [76], DL has been assisting in an increase in the computational costs of the state-of-the-art AI research as big as 3000,000x between 2012 and 2018. Such a dramatically increasing trend in resource consumption, dubbed as RED AI, is not just often prohibitively expensive for researchers and practitioners, but also environmentally unfriendly.

This has motivated the field of Green Software Engineering (SE) research, which aims to decrease software environmental footprints and supports *inter alia* GREEN AI [58]. Optimizing resource

Questions

- What programming languages are TensorFlow and PyTorch written in?
- Which computer component makes the biggest difference when training Deep Learning models?

Why is this important?

- The energy spent to train a model can vary significantly between frameworks, but how much?
- Investigate the energy and run-time performance for training and inferencing Deep Learning algorithms build in popular ML frameworks and suggest which to use in specific cases.
- What tuning parameters can affect the energy and run-time performance of the selected models?

Research Questions

RQ1: Which is the most energy and run-time performance-efficient Deep Learning framework for the models examined?

RQ2: How much accuracy do energy and run-time performance efficient Deep Learning frameworks sacrifice for the models under examination?

RQ3: What are the most energy and run-time performance inefficient APIs of Deep Learning frameworks for the models under examination?

HOW SHOULD I ANSWER



Public Repository

The screenshot shows the GitHub interface for the repository `NVIDIA / DeepLearningExamples`. The repository is public and has 252 watchers, 2.1k forks, and 7.5k stars. The main navigation bar includes links for Pull requests, Issues, Marketplace, and Explore. The repository's main navigation includes Code, Issues (116), Pull requests (14), Actions, Projects, Wiki, Security, and Insights. The repository is currently on the `master` branch, with 6 other branches and 0 tags. A recent commit by `nv-kkudrynski` is highlighted, titled "Merge: [nnUNet/PyT] Fix DALI inference pipeline", made 11 days ago with 1,066 commits. Below the commit list, three folders are shown: `.github/ISSUE_TEMPLATE` (updated 2 years ago), `CUDA-Optimized/FastSpeech` (updated 6 months ago), and `DGLPyTorch/DrugDiscovery/SE3Tra...` (updated 2 months ago). The right sidebar shows the repository's name, a Readme link, 7.5k stars, 252 watchers, and 2.1k forks.

Search or jump to... / Pull requests Issues Marketplace Explore

NVIDIA / **DeepLearningExamples** Public

Watch 252 Fork 2.1k Starred 7.5k

<> Code Issues 116 Pull requests 14 Actions Projects Wiki Security Insights

master 6 branches 0 tags

Go to file Add file Code

About
Deep Learning Examples

Readme
7.5k stars
252 watching
2.1k forks

nv-kkudrynski Merge: [nnUNet/PyT] Fix DALI inference pipeline 589604d 11 days ago 1,066 commits

.github/ISSUE_TEMPLATE	Update issue templates	2 years ago
CUDA-Optimized/FastSpeech	Adding links to performance benchmark page	6 months ago
DGLPyTorch/DrugDiscovery/SE3Tra...	Remove RoseTTAFold	2 months ago

Tasks

Categories	Models	Datasets
Recommender Systems	NCF	ML-20M
NLP	Transformer-XL	WikiText-103
	GNMT	WMT16 EN-DE
Computer Vision	ResNet-50	Coco 2014
	SSD	Coco 2017
	MaskRCNN	Coco 2017

Tools

Tue Aug 3 17:16:41 2021

```
+-----+
| NVIDIA-SMI 460.32.03    Driver Version: 460.32.03    CUDA Version: 11.2    |
+-----+-----+-----+-----+-----+
| GPU Name      Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|====+=====+====+=====+=====+=====+=====+=====+
|  0  Quadro P4000      On      | 00000000:2D:00:00 Off |          N/A         |
| 64%   82C   P0     63W / 105W |  7789MiB /  8110MiB |      98%    Default  |
|                              |                      |          N/A         |
+-----+-----+-----+-----+-----+

+-----+
| Processes: |
| GPU  GI  CI       PID   Type   Process name                      GPU Memory |
|  ID   ID   ID             |                 | Usage     |
+-----+-----+-----+-----+-----+
|  0   N/A N/A     2023    G    /usr/lib/xorg/Xorg                  9MiB      |
|  0   N/A N/A     2086    G    /usr/bin/gnome-shell                5MiB      |
|  0   N/A N/A    21488    C    python                             7769MiB   |
+-----+-----+-----+-----+-----+
```

cProfile

Python 3.X

cProfile and profile provide deterministic profiling of Python programs. A profile is a set of statistics that describes how often and for how long various parts of the program executed. These statistics can be formatted into reports via the pstats module. The Python standard library provides two different implementations of the same profiling interface:

 [More at Python 3 Documentation](#)

[Share Feedback](#)

```
Display all 240 possibilities? (y or n)
```

```
tushar@Enterprise:~/media/tushar/Cupboard/machine_learning_frameworks_analysis$ cat measurements/transformer_xl_PyTorch_perf_2.txt
```

```
Performance counter stats for 'system wide':
```

```
144,643.48 Joules power/energy-pkg/
15,078.98 Joules power/energy-ram/
```

```
1431.584266139 seconds time elapsed
```

```
tushar@Enterprise:~/media/tushar/Cupboard/machine_learning_frameworks_analysis$ █
```

Our platform

```
tushar@Enterprise: /media/tushar/Cupboard/machine_learning_frameworks_analysis/DeepLearningExamples/TensorFlow/Detection/SSD
File Edit View Search Terminal Tabs Help
tushar@Enterprise: /media/tushar/Cupboard/machine_le... x tushar@Enterprise: /media/tushar/Cupboard/machine_le... x tushar@Enterprise: /media/tushar/Cupboard/machine_lea... x tushar@Enterprise: /media/tushar/Cupboard/machine_lea... x
1 [|||] 3.9% 19 [ 0.6% 37 [||] 1.3% 55 [ 0.0%
2 [||] 2.6% 20 [ 0.0% 38 [||] 1.9% 56 [ 0.0%
3 [||] 1.3% 21 [ 0.0% 39 [||] 0.6% 57 [ 0.6%
4 [||] 2.0% 22 [||] 2.6% 40 [||] 0.6% 58 [ 0.6%
5 [ 0.0% 23 [||] 0.6% 41 [||] 0.6% 59 [ 0.6%
6 [||] 0.6% 24 [||] 0.6% 42 [||] 0.6% 60 [ 0.0%
7 [ 0.0% 25 [||] 0.6% 43 [ 0.0% 61 [||] 0.7%
8 [||] 0.6% 26 [||] 0.6% 44 [ 0.0% 62 [ 0.0%
9 [ 0.0% 27 [||] 0.6% 45 [ 0.0% 63 [ 0.0%
10 [||] 0.6% 28 [ 0.0% 46 [||] 1.3% 64 [ 0.0%
11 [||] 0.6% 29 [ 0.0% 47 [|||] 3.2% 65 [||] 1.3%
12 [||] 0.6% 30 [ 0.6% 48 [|||||] 44.7% 66 [ 0.0%
13 [||] 0.6% 31 [ 0.0% 49 [ 0.0% 67 [||] 0.6%
14 [||] 0.6% 32 [||] 1.9% 50 [||||] 6.5% 68 [|||] 4.5%
15 [||] 1.9% 33 [||] 1.9% 51 [||] 2.0% 69 [||] 1.3%
16 [ 0.0% 34 [ 0.0% 52 [||] 1.9% 70 [||] 1.3%
17 [||] 1.9% 35 [ 0.0% 53 [ 0.0% 71 [||] 3.2%
18 [||] 1.3% 36 [ 0.0% 54 [ 0.0% 72 [||] 0.6%
Mem[|||||] 4.81G/93.1G Tasks: 164, 1171 thr; 2 running
Swp[||] 1.83G/94.7G Load average: 2.40 2.53 2.27
Uptime: 42 days, 14:02:08
```


More challenges ahead



Low Disk Space

F [REDACTED]
YOU
NVIDIA



PrEngDL

Search or jump to... / Pull requests Issues Marketplace Explore

stefanos1316 / ICSE_2022_artifact Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code

stefanos1316 Add manual analysis data 45bd2c1 22 days ago 3 commits

configs	Add all tools and data	22 days ago
data	Add all tools and data	22 days ago
tools	Add all tools and data	22 days ago
.pre-commit-config.yaml	Add all tools and data	22 days ago
PyTorchVs.TensorFlow_manual_ana...	Add manual analysis data	22 days ago
README.md	Add all tools and data	22 days ago
install_and_configure.yml	Add all tools and data	22 days ago
run.sh	Add all tools and data	22 days ago

Training results

The  Recommender System, the  Natural Language Processing, and the  Computer Vision Category

Models	CPU & RAM Energy	<i>p</i> -value (A12)	GPU Energy	<i>p</i> -value (A12)	Run-Time	<i>p</i> -value (A12)
NCF	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)
Transformer-XL	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)
GNMT	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)
ResNet-50	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)
SSD	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)
Mask-RCNN	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)

Inference results

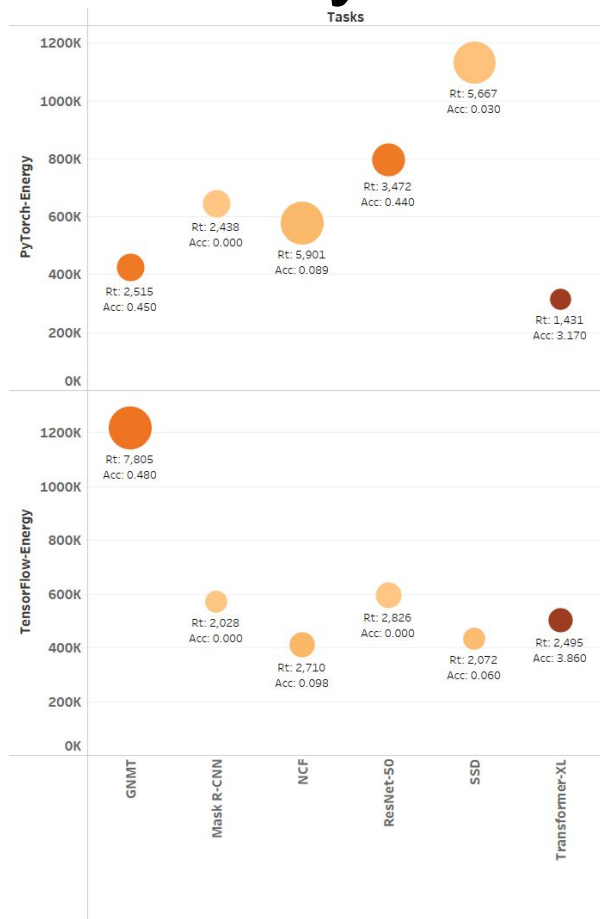
The  Recommender System, the  Natural Language Processing, and the  Computer Vision Category

Models	CPU & RAM Energy	<i>p</i> -value (A12)	GPU Energy	<i>p</i> -value (A12)	Run-Time	<i>p</i> -value (A12)
NCF	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)
Transformer-XL	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)
GNMT	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)
ResNet-50	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)	TensorFlow	< 0.001 (1)
SSD	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)
Mask-RCNN	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)	PyTorch	< 0.001 (1)

Answer to RQ1

- **TensorFlow** performs better for training **Recommender Systems** and **Computer Vision** tasks.
- **PyTorch** outperforms **TensorFlow** for **Natural Language Processing** tasks.
- For model inference, **TensorFlow** is more efficient for **Recommender Systems** and **ResNet-50**.
- Overall, **TensorFlow** is more energy and run-time performance efficient for **training models**, while **PyTorch** for models inference.

Accuracy Trade-Offs

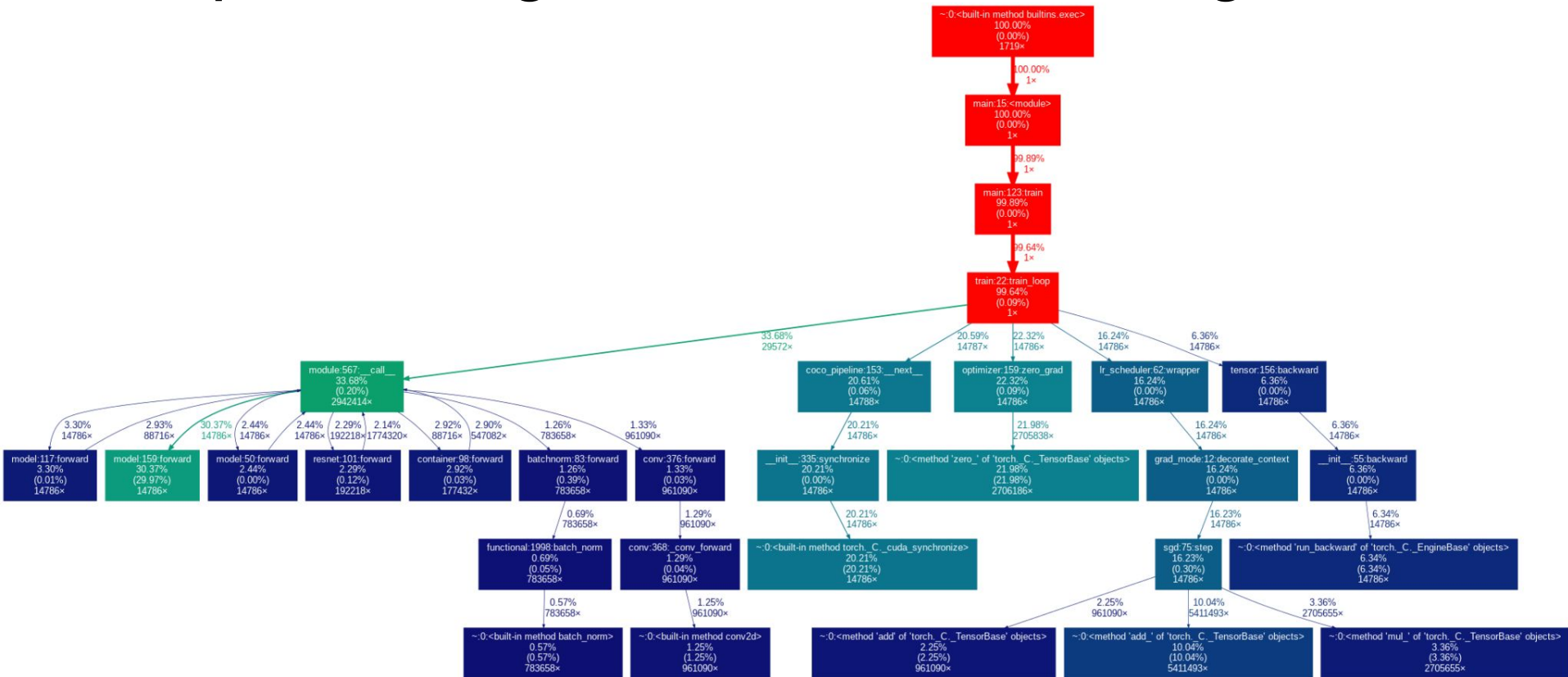


Answer to RQ2: The collected results suggest that better energy consumption and run-time performance—in most of the cases—yield better accuracy results as well. Overall, we find that TensorFlow has similar accuracy to PyTorch, under the configurations and parameter used in our study.

Identifying costly API calls – Spearman

Tuples	PyTorch	TensorFlow
PKG Energy–Run-Time	0.25	0.88
RAM Energy–Run-Time	0.88	0.94
GPU Energy–Run-Time	0.42	0.60

Deep Learning Frameworks Profiling



Symbols and their meanings

- Complex calculations
- Complex Implementation
- Large data
- ◇ Device dependency
- Unknown

PyTorch Inference

Model	Function Name	Ncalls	Run-Time	Cost	Type
Mask R-CNN	Torch.tensor	81,819	578	60.6%	◇
	Torch.nn.Conv2d	380,000	24	2.6%	■
	Torch.Tensor.nonzero	400,000	17	1.8%	—
	Torch.Tensor.float	1,505,005	16	1.7%	●
	Torch.Tensor.to	105,312	16	1.7%	□
	Torch.Tensor.type	9,735,395	14	1.5%	—

Takeaways



PyTorch

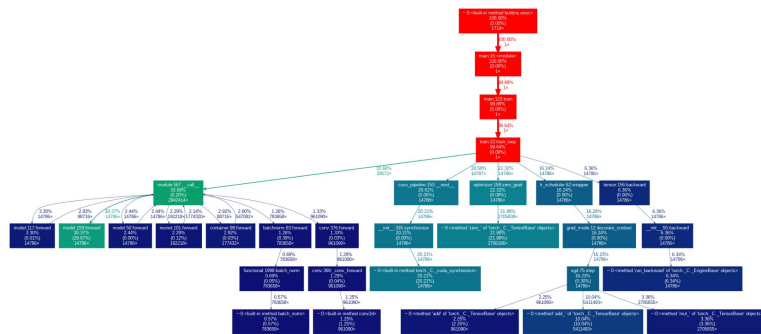
Table of Contents

PYTORCH DOCUMENTATION

PyTorch is an optimized tensor library for deep learning using GPUs and CPUs.

Features described in this documentation are classified by release status:

1. Test small programs energy consumption before running large scale experiments.
2. Use profiling approaches on small experiments to estimate resources to be used for large experiments afterwards.



What are we missing as a community?

- How to convince developers which programming language, module, or framework to choose in order to reduce energy consumption?
- How different components will affect the energy and the run-time performance of applications in the long term?
- An easy way to collect measurements.

Thank you for your attention!!!



Email: stefanos1316@gmail.com
GitHub: [stefanos1316](https://github.com/stefanos1316)

Backup slides

