

The Sense of Logging in the Linux Kernel

Keyur Patel · João Faccin · Abdelwahab
Hamou-Lhadj · Ingrid Nunes

Received: date / Accepted: date

Abstract Logging plays a crucial role in software engineering because it is key to perform various tasks including debugging, performance analysis, and detection of anomalies. Despite the importance of log data, the practice of logging still suffers from the lack of common guidelines and best practices. Recent studies investigated logging in C/C++ and Java open-source systems. In this paper, we complement these studies by conducting the first empirical study on logging practices in the Linux kernel, one of the most elaborate open-source development projects in the computer industry. We analyze 22 Linux releases with a focus on three main aspects: the pervasiveness of logging in Linux, the types of changes made to logging statements, and the rationale behind these changes. Our findings show that logging code accounts for 3.73% of the total source code in the Linux kernel, distributed across 72.36% of Linux files. We also found that the distribution of logging statements across Linux subsystems and their components vary significantly with no apparent reasons, suggesting that developers use different criteria when logging. In addition, we observed a slow decrease in the use of logging—reduction of 9.27% between versions v4.3 and v5.3. The majority of changes in logging code are made to fix language issues, modify log levels, and upgrade logging code to use new logging libraries, with the overall goal of improving the precision and consistency of the log output. Many recommendations are derived from our findings such as the use of static analysis tools to detect log-related issues, the adoption of common writing styles to improve the quality of log messages, the development of conventions to guide developers when selecting log levels, the establishment of review sessions to review logging code, and so on. Our recommendations can serve as a basis for developing logging guidelines as well as better logging processes, tools, and techniques.

Keywords Linux Kernel · Software Logging · Empirical Studies · Software Engineering

K. Patel and A. Hamou-Lhadj
Concordia University, Montreal, CA
E-mail: {keyur.patel,wahab.hamou-lhadj}@concordia.ca

J. Faccin and I. Nunes
Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, BR
E-mail: {jgfaccin,ingridnunes}@inf.ufrgs.br

1 Introduction

Software logging is a practice that has long been used by developers to record valuable runtime information (e.g., events and data) about a running system (Yuan et al. 2012b; Pecchia et al. 2015; Miranskyy et al. 2016). These records are in the form of log messages, which are generated by logging statements in the source code. Studies have shown that the analysis of log data can help with many tasks, including debugging and diagnosis of system failures (Khatuya et al. 2018; El-Masri et al. 2020), detection of security vulnerabilities (Yen et al. 2013; Zhou et al. 2020), profiling of distributed systems (Pi et al. 2018), troubleshooting cloud computing environments (Kc and Gu 2011; Miranskyy et al. 2016), detection of anomalies (Bertero et al. 2017; Islam et al. 2018; Oliner et al. 2008), and reliability evaluation of applications (Tian et al. 2004).

Despite the importance of log analysis, the practice of logging remains largely *ad-hoc* with no recognized guidelines and systematic processes (Yuan et al. 2012a; Fu et al. 2014; Pecchia et al. 2015). Software developers continue to insert logging statements in the source code without clear and sufficient guidance. The decisions on how and where to log are often left to the discretion of the developers, resulting in inconsistencies even among developers working on the same projects. In a study performed at Microsoft (Zhu et al. 2015), the authors showed that 68% of the participants (mainly Microsoft developers) found it challenging to make decisions on how and where to log. The lack of systematic and automated approaches for logging raises serious questions as to the validity of the generated log data as well as the efficacy of existing log analytic tools, which may impede failure diagnosis and other log-related tasks, threatening the stability of deployed software systems.

There exist studies that investigate the practice of logging in software engineering (e.g., (Zhou et al. 2020; Yuan et al. 2012a; Chen and Jiang 2017b; Zhu et al. 2015; Zeng et al. 2019)). Yuan et al. (2012a) made the first characterization of logging practices in open-source projects. The authors studied logging practices in four server-side C/C++ projects and provided many findings with respect to the pervasiveness of logging code, how often logging code is changed, and what kind of changes are made to it. Chen and Jiang (2017b) conducted an extensive replication study of this work with a focus on open-source projects written in Java. These studies revealed many issues related to the practice of logging. For example, they showed that software developers use logging in an inconsistent way. In addition, they showed that developers often have to change logging code in subsequent versions of the system as afterthoughts, adding to the maintenance burden of software systems.

In this paper, we build on these studies by conducting the first empirical study to understand the practice of logging in the Linux kernel. Linux is perhaps one of the greatest collaborative efforts in the computer industry with more than a thousand experienced developers contributing to its growth on a regular basis. Many studies have examined the structure and evolution of the Linux kernel from different perspectives (Bagherzadeh et al. 2018; Israeli and Feitelson 2010; Lotufo et al. 2010; Passos et al. 2012; Lu et al. 2014; Fadel 2011), but little is known about the logging practices followed by Linux kernel developers. Existing Linux development guidelines do not include guidelines for making logging decisions.

Similarly to previous studies (Yuan et al. 2012a; Chen and Jiang 2017b), we explore three main aspects related to the practice of logging: (1) the pervasiveness of logging in Linux, (2) the types of changes made to logging statements over several releases, and (3) the rationale underlying changes to logging code. From this per-

spective, this is a replication study with a focus on the Linux kernel. Because Linux developers are considered as experienced developers, understanding how they log their code provides useful insights into the practice of logging in software engineering, further contributing to the corpus of knowledge in this area. Our findings show that there is a logging statement for every 27 lines of code in Linux v5.3 (the latest version when we collected data for this study). The pervasiveness of logging, however, varies from one kernel subsystem to another, with *filesystem* and *drivers* being the most logged subsystems. After investigating changes made to logging statements across 22 releases of Linux, we found that developers tend to modify logging code to improve the quality of the logging output by enhancing precision, conciseness, and consistency. By means of a qualitative analysis of 900 commits that target fixes and improvements of logging code, we found that the reasons behind these changes are similar to those reported by other studies (Hassani et al. 2018; Yuan et al. 2012a; Chen and Jiang 2019, 2017a), such as ambiguous log messages, redundant information and logging the wrong variables. We also found many issues that pertain to the Linux kernel such as revealing vulnerable data and problems related to early logging. Based on these results, we derive many recommendations that are not only applicable to Linux, but can be readily generalized to other software systems. Our long-term goal is to help design better processes, techniques, and tools for effective logging, and set the groundwork for establishing logging standards.

In summary, the main contributions of our work are listed as follows.

- We study the pervasiveness of logging in the Linux kernel. To our knowledge, this is the first study that focuses on the practice of logging in Linux.
- We present a study that shows the evolution of logging statements in 22 releases of the Linux kernel.
- We study the rationale behind changes made to logging statements across multiple releases of the kernel.
- Based on the analysis of afterthought changes to logging code in the Linux kernel, we derive several guidelines that can help the maintenance of logging code in Linux. These guidelines can possibly be applied to other systems as well.

This study can benefit researchers and practitioners interested in understanding and improving the practice of logging in software engineering. The ultimate objective is to better define guidelines and processes for logging software systems. Additionally, Linux developers can directly refer to the findings of this paper to develop better logging practices for Linux.

The remainder of this paper is structured as follows. We discuss existing work related to logging in Section 2. Next, we present our study design in Section 3. The results obtained with the analysis of logging practices in the Linux kernel are presented in Section 4. In Section 5, we discuss our findings and implications as well as threats to validity. Finally, we conclude the paper in Section 6.

2 Related Work

In this section, we review existing studies on the practice of logging (i.e., how to log). We also review related studies on what and where to log. We discuss how these studies differ from ours at the end of the section.

2.1 Research Studies on the Practice of Logging

In an attempt to understand the characteristics of software logging in open-source projects, Yuan et al. (2012a) conducted one of the first studies that examines the practice of software logging by mining the evolution history of four C/C++ projects (Apache httpd, OpenSSH, PostgreSQL, and Squid). The authors found that logging statements account for 3.30% of the source code. They also reported that developers often find it challenging to get logging statements right at the first try. They showed that developers changed 36% of all log messages at least once as afterthoughts. The changes are categorized as changes to *verbosity level*, *static content*, and *variables*. They observed that 75% of log modifications of the static content of the logs consist of fixing inconsistent and confusing log messages. Additionally, the authors showed that developers often struggle to use the right verbosity level. To address this problem, they designed a simple verbosity level checker. Chen and Jiang (2017b) conducted a replication study of the work of Yuan et al. (2012a) with the objective to understand the practice of logging in Java systems. Their findings differ in many aspects from the original study. For example, they found that consistency updates account for 41% of all the logging code updates. These updates are much more present in C/C++ project and represent 67% of the logging code updates.

Pecchia et al. (2015) investigated the practice of logging on the development process of *critical software systems* in industry. They identified three main reasons for using logging code: state dump, execution tracing, and event reporting. They found that the logging practice lacks standardization over key-value representations and missing contextual information, which hinders the adoption of automated log analysis tools.

Shang et al. (2014) conducted a study on one large open-source and one industrial software system to explore the evolution of logging statements (referred to as communicated information) by mining the execution logs of these systems. They found that logging statements change across versions, impacting log processing applications. They also showed that the use of advanced log analysis techniques can reduce the impact of these changes on log preprocessing tools.

Zeng et al. (2019) focused on understanding the practice of logging in *Android apps*. Their findings suggest that there are differences between logging practices observed in mobile apps when compared to desktop and server applications. They provided evidence that logging code in mobile apps is less pervasive and not maintained as actively as in desktop and server applications.

The analysis of two open-source software systems conducted by Shang et al. (2015) showed a relationship between post-release defects and some log-related metrics. This study suggests that more maintenance effort should be devoted to files with logging code as they are more susceptible to contain problems.

The quality of logging code was investigated by Hassani et al. (2018). The authors characterized log-related issues using the number of files involved in an issue, the time required to fix it, and who provides the fix. In addition, they manually examined log-related issues to identify common problems associated with logging code, which served as the basis for developing a tool to automatically detect problems such as spelling errors and empty catch blocks.

Cinque et al. (2010) conducted a quantitative study in which they assessed the effectiveness of logging mechanisms to generate logs when a fault is triggered. They injected faults in three open-source systems and observed whether faults that lead

to failures generate logs or not. The authors found that the percentage of logged failures ranges between 35.6% and 42.1% among the three systems they used in the case study. Their experimental results are then exploited to suggest places where to place additional logs based on the analysis of memory dumps.

Recently, Li et al. (2020) conducted a qualitative study to understand the benefits of software logging and the costs associated with it by surveying 66 developers and examining 223 logging-related issue reports. They found that developers do not have a systematic strategy for reconciling the benefit and cost of logging. The authors argued that the lack of clear strategies may lead to logging practices with a negative impact.

Li et al. (2017) proposed a predictive model that suggests the most appropriate log levels for new logging statements based on features such as the average log level and log churn. This is important in light of the findings of (Yuan et al. 2012a; Chen and Jiang 2017b; Hassani et al. 2018) who showed that developers have difficulties determining the correct log level to be used.

Chen and Jiang (2017a) analyzed many code-independent log updates from three open-source Java projects and identified five logging anti-patterns. A tool, called *LCAnalyzer*, was developed to allow the automatic identification of these poor logging practices.

Zhou et al. (2020) investigated the impact of logging practices on data leakage with a focus on Android mobile applications. They studied logging code in 5,000 mobile apps and showed using taint analysis that out of 276 apps with taint flow paths, 200 (72%) leaked sensitive data due to poor logging practices such as forgetting to remove debug logging statements before deployment. They categorized the types of data leakage that are related to logging practices into four types: network, account, location, database. They manually analyzed source-sink paths of log-related taint flow paths of the 200 apps and found that 49% of the data that is leaked is network-related, followed by database sensitive data (45%).

2.2 Studies on where and what to log

Fu et al. (2014) performed an empirical investigation on enterprise applications written in C# in order to understand where developers insert logging statements. The authors found that logging statements can be grouped based on where they are inserted in the code using these categories: *assertion-check logging*, *return-value-check logging*, *exception logging*, *logic-branch logging*, and *observing-point logging*. They also found that 39%–53% of logging statements are placed to capture information when the software fails while 47%–61% of logging statements are used to trace the normal execution flow. To complement their findings, the authors trained a decision tree model that suggests where to log based on contextual keywords extracted from code snippets. Their recommendation tool achieved a precision of 81.1%–90.2%, and recall of 80.8%–90.4%. A similar tool, called *LogAdvisor*, was proposed by Zhu et al. (2015). The tool aims to help developers by providing suggestions on where to log with an accuracy of 84.6%–93.4%.

In their position paper, Zhao et al. (2017) proposed a strategy for automatically placing logs in the source code using information theory concepts by measuring the software's uncertainty using Shannon Entropy. Their approach respects a predefined performance overhead (less than 2%). The idea is to place logs in execution paths

that reduce uncertainty. The approach was tested on a small example. The authors stated that the preliminary results show that their approach holds real promise for automating the placement of logging statements. The problem with this technique, however, is that it is completely agnostic to the semantic of the code blocks that are being logged. It simply adds logs in paths that would most likely be executed. The authors did not discuss the quality of the generated logging statements when used for debugging and other tasks.

Yang et al. (2018) proposed NanoLog, a logging mechanism that is 1–2 orders of magnitude faster than existing logging systems and that achieves a high throughput (80 million log messages per second), while exposing the traditional printf-like statements. NanoLog modifies the log messages at compile-time, and uses compaction techniques to output logs in a compact binary format. The authors showed that their approach can improve the performance of log analytic tools.

Because there is no standard way to know what information should be included in logs, developers insert logging statements into the code in an *ad-hoc* manner. Sometimes, the information from these logs may be insufficient to diagnose software failures. To handle this issue, Yuan et al. (2012b) designed *LogEnhancer* a tool that can infer *what* information could be helpful to narrow down the root cause of failures and to include that information into existing logging code automatically. To help developers decide which variables should be included in the logging statements, a recent study by Liu et al. (2019) proposed a recurrent neural network-based model. The work of He et al. (2018) also focuses on what to log, with the objective to understand the goal of the static text used in these statements. Their results suggest that static text is used essentially for describing program operation, error conditions, and high-level code semantics. They also proposed an information retrieval approach for generating logging messages automatically.

Li et al. (2018) used topic modelling to understand whether there is any relationship between *topics* of source code and having logging statements. Their findings suggest that some source code *topics* are more likely to include logging statements than others.

The single attempt to address the challenge of logging in Linux was made by Tschudin et al. (2015). The authors proposed a machine learning-based approach to suggest the most appropriate logging functions, using the evolution history of the Linux kernel. However, the approach consists of a conceptual description of a framework that still needs to be implemented and validated.

2.3 Discussion

Our study focuses on understanding the practice of logging adopted by Linux developers. To our knowledge, this is the first study that focuses on the Linux kernel, which makes our study unique. To achieve our goal, we address the same research questions on characterizing the practice of logging as the ones discussed in related work (Yuan et al. 2012a; Chen and Jiang 2017b). In this sense, this study can be considered as a replication study with a focus on the Linux kernel. It complements the studies that aim to better understand of the practice of logging in software development. We found that Linux developers tend to modify logging code to improve the quality of the logging output by enhancing precision, conciseness, and consistency. We also found that the reasons behind afterthought changes to logging statements

include to the ones reported in other studies (Hassani et al. 2018; Yuan et al. 2012a; Chen and Jiang 2019, 2017a). This confirms the need for better guidelines and processes for logging in software engineering. We derive many recommendations from our study that can improve the logging practices in Linux and other systems. Many of these can be implemented in tools such as the ones discussed in the previous subsection to make suggestions to developers as they write logging code.

3 Study Design

3.1 Research questions

Our goal is to characterize how logging is used in the Linux kernel so that we can understand the adopted practices and identify the challenges. To achieve this goal, we focus on answering the following research questions.

RQ1 *What is the pervasiveness of logging in the Linux kernel?*

RQ2 *How does the logging code in the Linux kernel evolve?*

RQ3 *What are the causes other than keeping consistency in the code that lead developers to change the logging code?*

By answering RQ1, we aim to understand whether logging is a practice consistently adopted by Linux kernel developers. This involves the analysis of the location of logging statements to understand the rationale of why they are placed in specific locations. In RQ2, we focus on investigating the lifecycle of logging statements, i.e. when they are created, updated, or removed. By making such analyses, we are able to understand the state of practice of the Linux kernel, that is, how much logging exists in the code, how it is distributed, and its lifestyle. The goal is to observe discrepancies in how logging is adopted so that we can raise questions related to the adopted practices. Such questions can be the basis for future studies to better understand logging decisions or to improve how logging is added and maintained in the code.

These two research questions are similar to those answered in previous studies (Yuan et al. 2012a; Chen and Jiang 2017b). This would allow us not only to understand these aspects in a project with distinguished characteristics but also to contrast obtained results. These previous studies also further analyzed logging changes to logging code, classifying them as consistent changes and afterthought changes. The former refers to changes that are due to changes in other parts of the code that requires the logging code to be updated to keep consistency, for example, when a variable referred to in a logging statement has its name changed. Afterthought changes, in turn, are changes made to logging statements to fix or improve the logging statements for other purposes. In this paper, we investigate such purposes in RQ3 to provide insights with respect to the effort required to maintain logging code. We aim to identify the reasons behind afterthought changes, which we hope can help researchers and practitioners develop techniques to automatically prevent, detect and automate such changes.

3.2 Subject project

The Linux kernel, our subject project, is a free and open-source software distributed under the GPLv2¹ license. It is responsible for managing the interactions between hardware components and higher-level programs that use these components. Used across a wide range of computer systems, from mobile devices and server systems to supercomputers, the kernel can be considered as one of the most important software projects in the computing industry. In fact, we are surrounded by the Linux kernel in one way or another. Being developed by roughly fifteen thousand developers throughout its history, as of 2020, it contains around 18M lines of C code. This project is continuously evolving to meet both hardware manufacturers' requirements and end-user expectations.

For RQ1, we study the pervasiveness of logging in version v5.3 (the last version when doing this study). This version received 14,605 commits from 1,881 developers, adding 837,732 and removing 253,255 lines of code, which represents an increase of 584,477 lines of code when compared to its previous version.²

To understand changes to logging code across various Linux releases and hence answer RQ2 and RQ3, we need to select a number of releases to include in the study. Our first attempt was to cover ten years of development going back starting from v5.3 (the last version when we conducted this study), but the amount of data that was generated was simply too high to conduct a viable analysis, especially that there is a lot of manual work to answer RQ3 and part of RQ2. We therefore decided to focus on a smaller dataset without loss of generality by going back 20 releases starting from v5.3. We settled for v4.3³ to v5.3⁴ (22 releases to be exact). This corresponds to an interval of four years of software development, from November 2015 to September 2019. All releases used in this study are available in the official Linux kernel repository.⁵ We believe that this dataset is representative since our objective is not to uncover all the problems related to logging in Linux but rather to provide insight into the practice of logging in software development by looking at how Linux developers (even in a narrower scope) use logging.

3.2.1 Linux subsystems

In our analysis, we frequently refer to Linux subsystems. The Linux kernel consists of five major subsystems;⁶ each covering particular aspects of the project. Table 1 shows the system decomposition. The *core* subsystem is in charge of memory management, inter-process communication, management of I/O operations, among others. The *filesystem* subsystem is responsible for providing the file system interface as well as individual file system implementations. The *drivers* subsystem provides device and sound drivers as well as the implementation of cryptography algorithms used within the system and other security-related code. Finally, the *net* and *arch* subsystems are responsible for networking and architecture-specific code, respectively.

¹ <https://www.gnu.org/licenses/old-licenses/gpl-2.0.html>

² <https://github.com/gregkh/kernel-history/>

³ <https://github.com/torvalds/linux/commit/6a13feb9>

⁴ <https://github.com/torvalds/linux/commit/4d856f72>

⁵ <https://github.com/torvalds/linux>

⁶ <https://github.com/gregkh/kernel-history/blob/master/scripts/genstat.pl>

Table 1: Linux kernel architectural decomposition.

Subsystem	Top level directories
<i>core</i>	init, block, ipc, kernel, lib, mm, virt
<i>filesystem</i>	fs
<i>drivers</i>	crypto, drivers, sound, security
<i>net</i>	net
<i>arch</i>	arch

Table 2: The eight possible log levels defined in the Linux kernel.

Log level	Description
KERN_EMERG	System is unusable
KERN_ALERT	An action must be taken immediately
KERN_CRIT	Critical conditions
KERN_ERR	Error conditions
KERN_WARNING	Warning conditions (default)
KERN_NOTICE	Normal but significant condition
KERN_INFO	Informational
KERN_DEBUG	Debug-level messages

3.2.2 Logging in the Linux kernel

Logging statements are used to record relevant information about a running system. A logging statement usually comprises three elements: (i) the *log level* of the event being recorded; (ii) a *static message* that describes that event; and, optionally, (iii) values of *variables* related to the logged event. A log message is typically generated by the execution of a function specifically created for that purpose. These functions can be either project-specific or provided by external logging libraries and frameworks.

The Linux kernel provides developers with its own set of logging functions. One of the simplest ways to write a message to the kernel log buffer is by using the `printk()` function. It is the kernel's equivalent of `printf()`, with the difference that it allows developers to specify the log level of the event being recorded (Corbet et al. 2005). An example of a logging statement is as follows:

```
printk(KERN_ERR "Device initialized with return code %d\n", code);
```

where `KERN_ERR` corresponds to the log level, `"Device initialized with return code %d\n"` is the static message, and `code` is the corresponding variable. There are eight log levels defined in the Linux kernel, representing different degrees of severity (see Table 2). `KERN_WARNING` is the default log level and is assigned to a message in case no log level is specified when calling `printk()`.

Additional sets of logging functions were introduced in the Linux kernel v1.3.983 with the aim of making logging statements more concise. These functions incorporate log levels in their names. Therefore, in order to log a debug or informational message, instead of using the `printk()` function with the `KERN_DEBUG` and `KERN_INFO` levels as parameters, developers can use the `pr_debug()` and `pr_info()` functions,

Table 3: New sets of logging functions.

Log level	pr_*() function	dev_*() function
KERN_EMERG	pr_emerg()	dev_emerg()
KERN_ALERT	pr_alert()	dev_alert()
KERN_CRIT	pr_crit()	dev_crit()
KERN_ERR	pr_err()	dev_err()
KERN_WARNING	pr_warn()	dev_warn()
KERN_NOTICE	pr_notice()	dev_notice()
KERN_INFO	pr_info()	dev_info()
KERN_DEBUG	pr_debug()	dev_dbg()

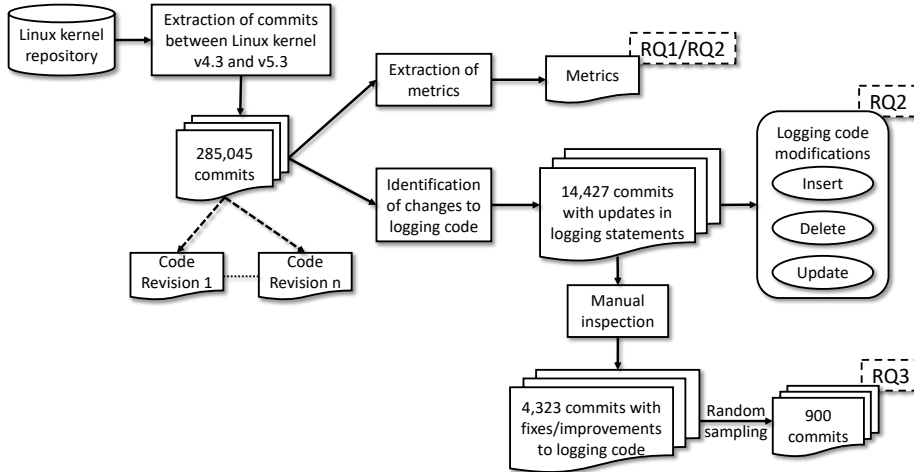


Fig. 1: Overview of the study procedure.

respectively. Another family of functions specifically designed for device drivers, e.g. `dev_dbg()` and `dev_info()`, automatically include device names in their outputs, making it easier to identify the origin of log messages. Table 3 lists both sets of functions and their corresponding log levels. Currently, some kernel components present their own logging functions, which are able to generate messages with service-specific information. Examples include network device and TI `wl1251` drivers, which provide the `netdev_*` and `wl1251_*` families of logging functions, respectively.

3.3 Procedure

In this section, we detail how we collected logging data from the Linux kernel. Figure 1 summarizes the tasks that were carried out. In the following subsections, we describe how we (1) extracted logging statements from the target project; (2) calculated metrics; (3) identified changes made to logging code between different versions; and (4) selected a subset of changes in logging statements to be analyzed.

3.3.1 Identification of logging functions

Traditional approaches to automatically locate logging statements into the source code rely on logging functions that are known in advance (e.g., Log4j in Java) or use regular expressions looking for variations of the term “log”. Although these methods have been successfully adopted in previous studies (Shang et al. 2015; Zhu et al. 2015; Chen and Jiang 2017b; Yuan et al. 2012a), they are not applicable in the context of the Linux kernel, which uses a wide variety of functions and macros for logging purposes. It is common to find situations in which customized calls to these macros are implemented by particular services (e.g., `_enter`⁷, `pr_pic_unimpl`⁸). Listing 1 shows an example. A macro named `adsp_dbg()` is implemented with the aim of ensuring the invocation of `dev_dbg()` with a particular set of parameters and formatting (lines 2–3) so that there is no need to specify such elements in further calls (line 5). This type of logging statements cannot be detected unless we know that `adsp_dbg()` is a logging function.

```

1 [...]
2 #define adsp_dbg(_dsp, fmt, ...) \
3     dev_dbg(_dsp->dev, "%s: " fmt, _dsp->name, ##__VA_ARGS__)
4 [...]
5 adsp_dbg(dsp, "Wrote %zu bytes to %x\n", len, reg);
6 [...]
```

Listing 1: Logging macro defined in `/sound/soc/codecs/wm_adsp.c`.

To identify Linux logging functions and macros in our study, we take advantage of the pattern-based method presented by Tschudin et al. (2015). The authors proposed an identification method that consists of three semantic patterns. These patterns, although describing general properties when taken individually, are able to correctly characterize logging functions when combined, yielding a low number of false positives. The first pattern states that logging functions should be called at least once inside an `if` block that ends with a `return` statement. The second pattern indicates that logging functions have at least one string argument, representing a log message. Finally, the third pattern requires logging functions to have a variable number of arguments. Listing 2 shows how macro `adsp_dbg()` satisfies all three patterns and thus can be considered as a logging function.

```

1 [...]
2 if (val == 0) {
3     adsp_dbg(dsp, "Aked control ACKED at poll %u\n", i);
4     return 0;
5 }
6 [...]
7 adsp_dbg(dsp, "Wrote %zu bytes to %x\n", len, reg);
```

Listing 2: A logging macro defined in `/sound/soc/codecs/wm_adsp.c` that satisfies the identification patterns.

⁷ <https://github.com/torvalds/linux/blob/v5.3/fs/afs/internal.h#L1449>

⁸ <https://github.com/torvalds/linux/blob/v5.3/arch/x86/kvm/i8259.c#L37>

Although this approach yields good results, it might miss logging functions with a fixed number of arguments (e.g., `PRINTK_2`⁹ and `PRINTK_4`¹⁰). These functions do not satisfy the third pattern. We therefore decided to relax the third pattern by including logging functions that contain a fixed number of arguments (i.e., the ones that are not overloaded).

We have carefully reviewed the list of logging functions extracted with this approach. Although we believe that this is a very comprehensive list of functions, we cannot guarantee 100% coverage, which is a threat to internal validity of this study. To mitigate this threat, we selected randomly 20 files and checked manually their content to see if we missed any logging functions. We found that our list covered all the logging functions invoked in these files. Further, we make the list of logging functions (and the data used for RQ1, RQ2, and RQ3) available to the community.¹¹ Once we know which functions/macros are used for logging, we write a script to detect all calls to these functions/macros to retrieve the specific logging statements.

Another alternative to handle the problem of the many forms of log printing functions caused by macros would be to expand macro definitions using a compiler preprocessor (e.g., `gcc -E`). However, Tschudin et al. (2015) found that this approach may result in false positives because of the possibility of having non-logging functions that contain calls to `printk`-like functions.

3.3.2 Logging code metrics

To assess the pervasiveness of logging in the Linux kernel, we consider four metrics, summarized in Table 4. Similarly to previous studies (e.g., Yuan et al. (2012a)), we extract the number of *lines of source code* (SLOC) and the number of *lines of logging code* (LLOC). We only included `.c` files in our analysis, that is, header files were not taken into account. Comments and empty lines are discarded when measuring SLOC. These metrics are collected considering different *levels* of granularity: (i) overall system and subsystems; (ii) files; and (iii) program constructs. The latter comprises functions as well as blocks of control statements, namely `do-while`, `if`, `else`, `else-if`, `for`, `switch`, and `while`. We also compute the *Log Density* (Yuan et al. 2012a; Chen and Jiang 2017b; Zeng et al. 2019) and *Log Ratio* metrics (see Equations 1 and 2). The Log Density measures the number of lines of source code per logging statement, whereas the Log Ratio measures the number of logging statements per line of source code. For example, a log density of 10 would mean that for every ten lines of code, there is one logging statement. A log ratio of 0.2 means that 20% of the code consists of logging statements.

$$\text{Log Density} = \frac{SLOC}{LLOC} \quad (1) \qquad \text{Log Ratio} = \frac{LLOC}{SLOC} \quad (2)$$

⁹ <https://github.com/torvalds/linux/blob/v5.3/drivers/char/mwave/mwavedd.h#L79>

¹⁰ <https://github.com/torvalds/linux/blob/v5.3/drivers/char/mwave/mwavedd.h#L89>

¹¹ <https://github.com/iamkeyur/linux-logging-2021>

Table 4: Extracted metrics related to logging code.

Metric	Description
SLOC	Number of lines of source code excluding comments and empty lines
LLOC	Number of lines of logging code
Log Density	The number of lines of source code per logging statement
Log Ratio	The percentage of number of lines of logging code per source line of code

3.3.3 Detection of changes in logging statements

To understand how logging code evolves in the Linux kernel, we study three types of modifications made to logging code—log insertion, log deletion, and log update—similarly to previous work (Yuan et al. 2012a; Chen and Jiang 2017b; Zeng et al. 2019). We further divide insertion and deletion of logging statements into two categories: logging statements that are added or deleted along with the addition or deletion of files. To retrieve changes to logging statements, we wrote a script using the `git rev-list` command with `--no-merges` option to retrieve the commits. Note that the `--no-merges` option omits commits with more than one parent. This is necessary in order to avoid duplicate commits. For each commit, we extract two adjacent versions of each file that are changed in the commit. We only consider `.c` files. Then, to generate an edit script representing syntactic modifications by inferring changes at the level of abstract syntax tree, we use GumTree, which is the state-of-the-art AST differencing tool (Falleri et al. 2014). As result, we found that 14,427 commits modified at least one logging statement, which are those used to answer RQ2.

3.3.4 Selection of changes in logging statements

Considering the 14,427 commits used to answer RQ2, not all aim to provide fixes or improvements to logging code, which is the type of change in logging code that we are interested in to answer RQ3. To identify changes that were made targeting logging code, previous studies (Hassani et al. 2018; Mazuera-Rozo et al. 2020) rely on a keyword-based approach by searching for commits that contain in their message variations of the word “log”. This strategy may miss many commits in Linux such as the one described above. To address this, the first author of this study manually examined all 14,427 commits by reading the commit titles and messages. He identified 4,323 commits out of 14,427 commits that explicitly discuss changes to logging statements. This task took several weeks to complete and required multiple iterations. He then selected randomly 900 commits out of 4,323 for a qualitative analysis. This subset is statistically representative with a confidence level of 99% and a margin of error of $\pm 4\%$ (Boslaugh 2012).

4 Results and Analysis

We followed the previously described study procedure and obtained the results detailed in this section. We present and analyze the results by research question.

Table 5: Summary of metrics collected from the Linux kernel and its subsystems.

Subsystem	Component	SLOC	LLOC	Log Density	Log Ratio
core	lib	106,577	2,296	46	2.15%
	kernel	208,000	4,396	47	2.11%
	mm	86,921	1,940	45	2.23%
	block	31,846	713	45	2.24%
	ipc	6,468	13	498	0.20%
	init	2,935	192	15	6.54%
	virt	16,839	184	92	1.09%
		459,586	9,734	47	2.12%
filesystem	fs	840,527	35,780	23	4.26%
drivers	drivers	9,358,913	384,269	24	4.11%
	sound	765,988	19,673	39	2.57%
	security	57,974	1,192	49	2.06%
	crypto	52,888	835	63	1.58%
		10,235,763	405,969	25	3.97%
net	net	746,263	14,490	52	1.94%
arch	arch	1,107,992	32,924	34	2.97%
Total		13,390,131	498,897	27	3.73%

4.1 RQ1: Pervasiveness of logging in Linux kernel

To examine the pervasiveness of logging in the Linux kernel, we consider its latest available version when we collected data for this study (v5.3). Based on the introduced metrics, we quantitatively assess how logging statements are spread in the code, considering four granularity levels: system (i.e., the entire Linux kernel), subsystems, files, and programming constructs.

4.1.1 System and subsystem level

We first report the values obtained for each metric (SLOC, LLOC, Log Density and Log Ratio) considering the system and subsystems levels. Table 5 presents the results. We found that from a total of 13,390,104 lines of source code, 498,897 (i.e., 3.73%) are lines of logging code. It is equivalent to a logging density of 27, which means that for every 27 lines of source code in Linux there is one logging statement. This finding is in line with the observations made by Yuan et al. (2012a) who reported an average log density of 30 in the four C/C++ applications they examined. This result, however, differs from the study by Chen and Jiang (2017b) who reported an average log density of 50 in the 21 Java applications they studied. Although these results are not conclusive, they suggest that the language in which the project is written as well as the application domain may affect the prevalence of logging statements in the source code. In addition, this gives evidence of the relevance of logging code. Logging statements do not provide additional application functionality and they, still, account for almost 4% of the written code in our target system. Consequently, logging code should be maintained with the same practices of code associated with other concerns. For example, developers should carefully review logging code in code review.

When looking at each subsystem individually, we obtain heterogeneous results. The log ratio ranges from 1.94% to 4.26%. A deeper look at the components of the Linux subsystems shows important discrepancies. For example, we observe that the *init* component of *core*, which consists only of 2,935 SLOC, contains 192 lines of logging code (Log Ratio = 6.54%). This is considerably higher than the other components. This may be explained by the fact the *init* component is responsible for the initialization of the console and other key kernel services such as the security framework, scheduler, memory allocation (Bootlin 2020). For these services, it is important to log all possible errors in order to quickly debug potential failures. However, the idea that critical components are logged the most does not always hold. Take for example the *ipc* component from the *core* subsystem with the highest log density (Log Density = 498). This component, which contains only six files, is responsible for setting up the inter-process communication mechanisms on which the kernel processes rely. Yet, despite being critical, this component is the least logged.

Without further studies, we can only attribute these variations to the fact that different groups of developers are maintaining each subsystem, and there are no recognized (or common) guidelines on how to log. This calls for future studies to understand the rationale for adding a large amount of logging statements in particular modules of the code considering different goals, such as debugging or auditing. It is important to understand when it is critical to have detailed logged information, e.g. when a module have a specific role or is the target of frequent changes. Moreover, given that logging can have a performance impact, certain components may have a reduced number of logging statements due to that reason.

4.1.2 File level

Similarly to the study of Lal et al. (2015), we also calculate the log ratio at the file level. Note that we do not use log density here due to the fact that many files do not contain logging statements, and therefore this metric cannot be calculated because SLOC cannot be divided by zero. The log ratio provides a more accurate measurement by assessing the pervasiveness of logging across Linux files. Figure 2 shows the collected data. We found that 94.34% of the total number of files in Linux v5.3 have a log ratio between 0% and 10%. More precisely, 7,134 files (27.64%) do not contain any logging statements, while a significant number of files (66.70%) have a log ratio greater than 0% and less than or equal to 10%. Files with a log ratio greater than 10% account for only 5.66% of the total number of files. Figure 2 thus highlights that the number of logging statements by file is a skewed distribution. Having the majority of files with a few logging statements is expected because many files may include code statements that can result in errors, and these are typically logged. However, it is interesting to observe that there are files with a substantial amount of logging statements. As discussed above, a qualitative study focusing on these highly logged files could give directions why this occurs and is needed. We, in particular, manually inspected 6 files with a log ratio close to 90%. We found that they contain mostly debugging routines. For example, `drivers/scsi/qla4xxx/ql4_dbg.c` contains functions to dump relevant information about the Linux Host Adapter structure.

The analysis of this distribution by the Linux subsystems can be seen in Figure 3, which shows the distribution of files by subsystem according to their log ratio.

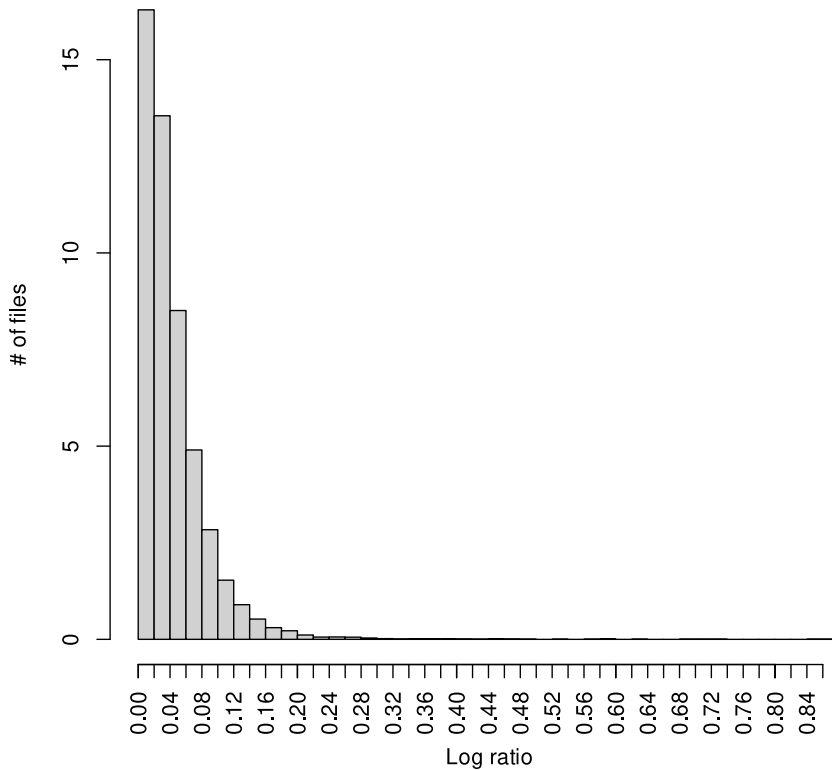


Fig. 2: Value of log ratio at the file level (Total #files = 25,814).

We can observe that 22%–44% of the files have no logging statements in the subsystem files. This indicates that in some systems—*core*, *net*, and mainly *arch*—the logging is concentrated in fewer files. Except *filesystem*, only a small percentage of files (3%–5%) has a high log ratio, that is, having many logging statements is an exception. This is not the case of *filesystem*, which has 11% of the files with more than 10% of their code being logging statements. In all subsystems, the majority (sometimes nearly half) of the files have at most 10% of their code as logging statements, but contain at least one. Similarly to what was discussed above, these differences cannot be explained without future studies.

4.1.3 Programming construct level

In this section, we measure the number of logging statements in program constructs of the Linux kernel including functions, `do-while`, `if`, `else`, `else-if`, `for`, `switch`, and `while`. Determining where logging statements are located can help understand

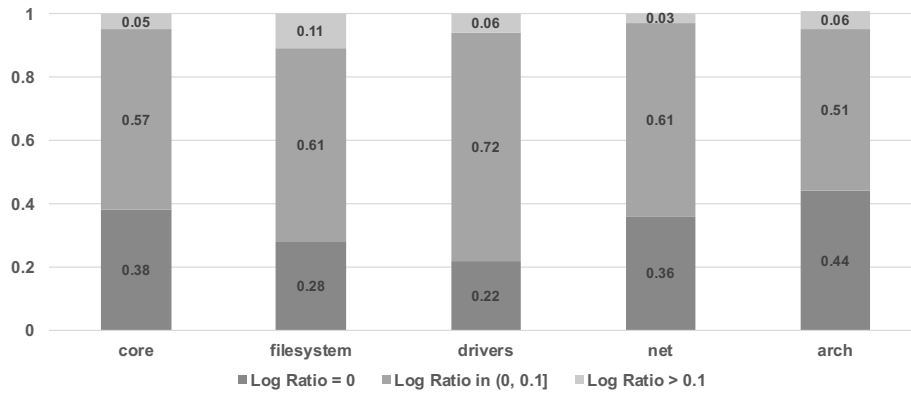


Fig. 3: Distribution of files by subsystem based on the log ratio.

Table 6: Number of functions that have a certain amount of logging statements.

#Logging Statements	#Functions
0	352,045 (73.88%)
1	64,553 (13.55%)
2	25,823 (5.42%)
3	12,432 (2.61%)
4	7,075 (1.48%)
5	4,432 (0.93%)
6	2,752 (0.58%)
7	1,902 (0.40%)
8	1,282 (0.27%)
≥ 9	4,226 (0.89%)
Total	476,522 (100.00%)

the purpose they serve. There exist studies that examine the location of logging statements in the source code. In particular, the work by Pecchia et al. (2015) measured the number of logging statements in the same program constructs as those listed above. Their work targeted two large C/C++ industrial applications. We followed the same approach and compared our results to theirs.

Table 6 shows the distribution of logging statements in the kernel functions. We found that from a total of 476,522 functions, 352,045 (73.88%) do not have any logging statements. For the remaining functions (26.12%), 88.28% of these have the number of logging statements in the range of one to five. Less than 1% of the functions have more than eight logging statements. Here, again, the distribution of logging statements is skewed. But differently, we observe that the vast majority of functions do not contain any logging statements at all. This shows that a relatively low number of functions need to have their information recorded in logs.

Complementing this information, Figure 4 shows the distribution of logging statements in different program constructs. We found that 55.66% (168,539 out of 302,799) of logging statements are used inside the `if` block and 4,585 (1.51%) are within `else-`

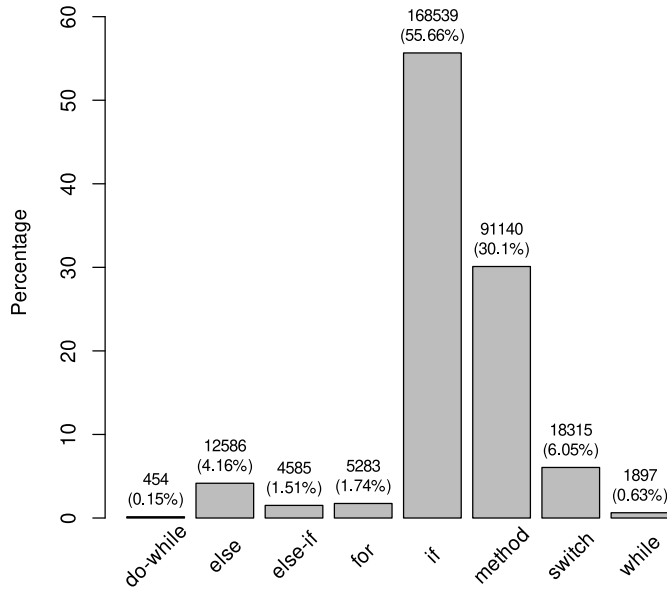


Fig. 4: Distribution of logging statements across different program constructs.

`if`. Together, they represent 57.17% of the use of logging. These results are similar to those of the study by Pecchia et al. (2015), who reported that around 60% of the logging statements are used inside the `if` blocks. These logs are typically used for logging errors after checking the return value of function calls. The logging statements in the `else` block accounts for 4.16% (12,586 out of total 302,799 logging statements). The `switch` block, which is another control statement available in C, accounts for 6.05% of the total logging statements. The logging statements used directly inside loop controls represent only 2.52% of the logging statements. The logging statements used directly within functions (i.e., not in any of the program constructs) account for 30.10% of the total number of logging statements. This suggests that the cyclomatic complexity of a function may be related to the presence of logging statements. Given that developers may need to understand when a function follows a particular path during the program execution, they add logging to conditional constructs. However, this does not occur within loops. This can be explained by the fact that adding a log statement to a loop may generate a high number of log messages and slow down the function performance.

4.1.4 RQ1: Summary

RQ1 - Findings. The log density in Linux v5.3 is 27, i.e., on average there is one logging statement for every 27 lines of code. This is similar to the study of Yuan et al.

(2012a) on C/C++ systems, but different from the result obtained by Chen and Jiang (2017b) when working on Java projects. The logging code represents 3.73% of the entire source code, with *filesystem* and *drivers* being the most logged subsystems with a log ratio of 4.26% and 3.97% and a log density of 23 and 25, respectively. We found that 72.36% of the total number of files in the Linux kernel have at least one logging statement. However, only 26.12% of the functions in the Linux kernel have one or more logging statements. Further studies, such as the work from Li et al. (2018), are needed in order to understand which topics of source code are more likely to be logged. 57.17% of the logging statements are used within an `if` and `else-if` blocks, while 30.10% are used inside code blocks in no particular program constructs.

RQ1 - Implications. The fact that logging code accounts for almost 4% of the entire code base is a strong reason for developers to carefully review logging code with the same rigour they use to review the other code. The discrepancies in the log density of various Linux subsystems calls for future studies (with a protocol similar to that by Fu et al. (2014)) to understand the goals of logging (e.g., debugging, auditing, security, etc.) in the different parts of the kernel. This study should be extended to the file and function levels where the distribution of logging statements is also heterogeneous with no apparent reason. Another implication of these results is the need for common guidelines for logging to ensure that best logging practices are followed. Finally, we suggest to study the correlation between code quality (using metrics such as cyclomatic complexity) and the need for logging, which may explain the fact the majority of logging code is found in if-else blocks.

4.2 RQ2: Logging code evolution

As shown in the previous section, logging is widespread in Linux—for every 27 lines of code, there is one line of logging code. Now, we focus on understanding how the logging code evolves across various releases of the kernel in terms of size as well as the type and the nature of changes.

4.2.1 Evolution of the size of logging code

We start by analyzing the logging code size by looking at two perspectives. First, we observe how the proportion of logging code evolved, measured by the log ratio metric. Second, we compare the evolution of SLOC and LLOC. The evolution of the log ratio metric for the Linux system from v4.3 to v5.3 can be seen in Figure 5a. We observe that it has been decreasing over the years. This can be due to the addition of code that has fewer logging statements than the average or the deletion of code that has more logging statements than the average. To better understand this observed evolution, we complement this analysis with the data shown in Figure 5b, which indicates how LLOC and SLOC, individually, increased across the different versions. SLOC and LLOC are normalized using min-max normalization to fall in the $[0, 1]$ interval.

The SLOC and LLOC curves shown in Figure 5b generally have a similar behavior, except for versions v4.11 to v4.18. By going through the Linux kernel changelogs, we found that these inconsistencies seem to be the result of the addition/removal of drivers or filesystems, which contained a large amount of logging code. The Linux

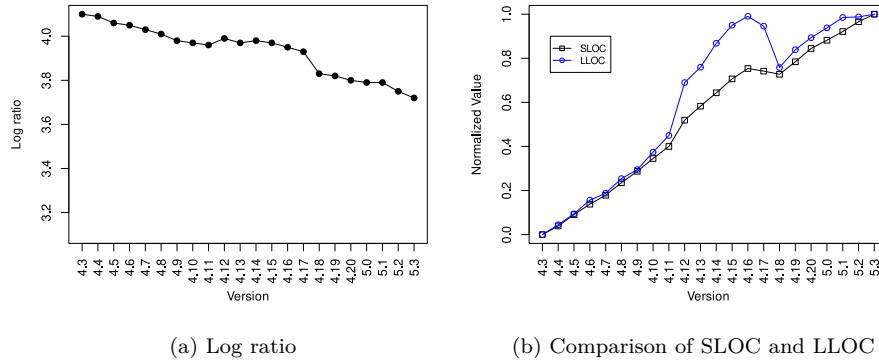


Fig. 5: Evolution of the use of logging code between Linux v4.3 and v5.3.

kernel v4.12 added Intel *atomisp* camera drivers (commit [a49d253](#)) and *rtl8723bs* sdio wifi driver (commit [554c0a3a](#)). These two changes increased the number of logging statements by 6,103, which contributed to a sharp increase in the number of lines of logging code. Similarly, we can see a sudden decrease in SLOC and LLOC between v4.16 and v4.18 releases. After inspecting the Linux kernel v4.18 changelogs, we found that SLOC of the Linux kernel v4.18 was smaller than its previous release, and that occurred just three times in the history of the Linux kernel before the release of kernel v4.18. The reason for this can be attributed to removal of the *lustre* filesystem (commit [be65f9e](#)) and the *atomisp* driver (commit [51b8dc51](#)), which earlier contributed to 10,442 logging statements. As discussed earlier, the amount of logging statements within modules and files varies. Consequently, the removal of particular parts of the system can largely affect the amount of logging code when compared to the rest of the code.

Considering both the log ratio and the number of lines of logging code, it is thus possible to observe that, though the number of lines of code increased as the Linux kernel evolved, the log ratio decreased. This can be due to a number of reasons, such as the types of changes made in the analyzed Linux versions—as it was discussed, different subsystems have varying amounts of logging. This may also be explained by the reliance on debugging and tracing tools, which can be used to diagnose problems as shown by Corbet (2016) and Edge (2019). These authors noticed an increase in the use of tracepoints rather than simple `printk()` in recent versions of Linux. The term tracing is used here to show the flow of execution of specific program constructs, e.g., traces of routine calls, and system calls (Hamou-Lhadj and Lethbridge 2002, 2004)). A discussion with one Linux expert (see Section 5.2) confirms this finding. We need to conduct further studies including user studies with Linux developers to understand the real causes behind this decline of logging code (see Section 5.3 on the implications for future research for more discussion), including the identification of the impact of the use of new tools that capture runtime information automatically without the need for logging statements added manually. It is relevant to understand what types of statements are removed and what types of statements remain in the code.

Table 7: Distribution of modifications made to logging statements.

Subsystem	Insertion	Deletion	Update
arch	3,987	4,620	2,962
core	1,863	647	1,700
driver	87,125	53,235	42,748
fs	2,776	1,626	3,502
net	1,490	1,677	1,479
Total	97,241	61,805	52,391
Percentage	45.99%	29.23%	24.78%

Similarly to previous studies (Yuan et al. 2012a; Chen and Jiang 2017b; Zeng et al. 2019; Li et al. 2019a), we calculate the ratio of the number of commits that involve modifications to logging code (i.e., additions, deletions, or updates to logging statements) to the total number of commits so that it is possible to understand how often a source code revision involves a modification to logging code. Out a total of 285,045 commits, we found that 39,351 (14%) commits involve modifications to logging code. This result compares to that of the study of Yuan et al. (2012a) on C/C++ systems, who found that 18% of the commits they studied involve log modifications. It is also similar to the results reported by Chen and Jiang (2017b) who found that there are around 21% of such commits in the studied Java systems. Some changes are expected because logging statements should be kept consistent with the corresponding code. However, it is relevant to understand why the logging statements change when this is not the case. This is in fact investigated in our RQ3.

4.2.2 Types of modifications in logging statements

Given that logging code changes overtime, we now look at the types of modifications that are made. Table 7 reports the number of logging statements added, deleted, and updated across Linux versions v4.3 to v5.3. There are 211,437 logging statement modifications, out of which 24.78% are log updates, 45.99% are log insertions, and log deletions account for 29.23%. The *drivers* subsystem alone represents 86.60% of the total modifications made to logging code, followed by *arch* (5.47%), and *filesystem* (3.74%). This is somewhat expected because the *drivers* subsystem is considerably larger (over 10 million SLOC) than the other subsystems (see Table 5). In addition, drivers are typically dynamic in operating systems. As new devices are released or become obsolete, the operating system must be evolved accordingly. Consequently, some of these modifications to the logging statements may be not due to changes specifically done on these statements, but to the addition or deletion of drivers. Moreover, as discussed in our RQ1, drivers typically include much logging code. This justifies why the *drivers* subsystem is associated with the vast majority of modification on logging code.

The percentage of log additions in Linux is similar to that reported by Chen and Jiang (2017b), who showed that log additions contribute to 18%–41% of the total log modifications. Yuan et al. (2012a) did not report the percentage of log additions in their examination of C/C++ systems.

Table 8: Impact of adding or deleting files on logging code.

Subsystem	Additions in new files	Deletions in a deleted files
arch	1,289	3,038
core	614	84
driver	51,178	30,452
fs	833	409
net	320	218
Total	54,234	34,201
	55.77% of the additions	55.34% of the deletions

We found that the number of log deletions represents 29.23% of the total number of modifications made to logging statements. This result differs significantly from the work of Yuan et al. (2012a) who reported that the number of log deletions is only 2% of the total number of modifications. Our result is similar to that of Chen and Jiang (2017b) who reported that log deletion contributes to 26% of the total modifications in Java systems. Such evolution of logging code may adversely impact the bug triaging process as developers rely on the logs contained in the bug reports, as noted by Ran (2019). The authors found that it is not possible to rebuild the execution paths for bug reproduction from bug report logs in 34% cases. They also argued that the continuous evolution of system logs can have an effect on the accuracy of log processing tools and machine learning models deployed for identifying anomalous activities, as models need to be retrained whenever logging statements are changed.

To drill down into the data, we grouped log additions and deletions into two categories: added along with the addition of a new file and deleted along with the deletion of an existing file. Table 8 shows a detailed breakdown. We found that 55.77% of all log insertions were made along with the addition of new files. Similarly, we found that 55.34% of the deleted logging statements were deleted along with the deletion of existing files. These data corroborate with our hypothesis of why the **drivers** subsystem accounts for many modification on logging code. These type of changes—that is, addition and deletion of files—have in fact a large impact in the log ratio, as shown in Figure 5.

4.2.3 Updates in logging statements

We now proceed to a deeper analysis of logging changes, analyzing how log statements are updated. We classify log updates into three categories depending on which part of the logging statement has been modified: (i) the *logging function* (or macro), including the log level, (ii) the *static content* representing the log message, and (iii) the *dynamic content*, i.e. the variables and function calls. Table 9 presents the results. Note that one update may consist of one or more changes to the same logging statement. For example, if the log function and the static content of a logging statement has changed, they appear as two updates. This explains why the total number of log updates is higher than the total number of the updated logging statements. We next analyze the different categories of updates.

Table 9: Distribution of updates made to logging statements.

Subsystem	Logging function	Static content	Dynamic content
arch	1,226	2,014	1,284
core	606	979	721
driver	15,685	23,127	28,501
fs	1,344	2,422	1,863
net	435	809	952
Total	19,296	29,351	33,321
% of updates	36.83%	56.02%	63.60%

Changes made to the logging function. Of the 52,391 updated logging statements, 19,296 (36.83%) include modifications to the used logging functions. A wide range of changes are possible, as shown in Table 10. In this table, we show only the most frequent changes in logging function, i.e. those with a frequency higher than 100.

The analysis of the changes to the logging functions revealed that 6,512 (33.75%) of these updates are changes between `printk`, `pr_<*>`, and `dev_<*>` macros. For example, in commit 26a0a10, a developer updated logging statements from `printk` to using device-aware `dev_err()/dev_info()` logging functions for improving the precision of the resulting logs by including device-specific information. This is further cemented by an observation that there has been a steady decrease in the usage of the `printk()` function, with a usage reduction of 29.32% between versions 4.3 and 5.3 of the Linux kernel. However, we found that the use of `pr_*(*)` and `dev_*(*)` functions has increased to replace `printk()` call sites (Corbet 2012). This gives evidence of evolving the importance of logging code in Linux, that is, developers have been modifying logging functions, possibly with the aim of creating standardised functions.

We also found an increasing use of the so-called “rate limited” logging functions such as the `<*>[_once/_ratelimited]` family of macros, which can be seen in commit 527aa2a, where all calls to `pr_info` were converted to `pr_info_ratelimited`. The objective of this type of functions is to prevent overloading the log buffers by controlling the amount of logs generated in a given period of time. Consequently, while writing logging code, it is important to consider not only the information being logged, but also the amount of information being generated and its impact on the application execution.

Moreover, we observed that many changes to logging functions are triggered by the need to make them more concise. For example, in commit 466414a, a developer introduced `btc_alg_dbg` and `btc_iface_dbg` logging macros, and converted all calls to `BTC_PRINTK` to the new functions (`btc_<*>_dbg`). The benefit is that software developers do not have to specify `btc_msg_type`, resulting in more concise logging statements. Ten months later, in commit 10468c3, all calls to `btc_<*>_dbg` were again changed to another logging function, named `RT_TRACE`, to be consistent with the use of this function in other drivers.

We thus conclude from the above observations that changes to logging functions aimed to improve the quality of the logging output by either enhancing precision, conciseness, or consistency. However, after analyzing many commits related to log updates, we could not find any evidence that there was a Linux-wide strategy, which

Table 10: Frequency of changes made to logging functions (frequency > 100).

Old logging function	New logging function	Frequency
printk	pr_err	950
printk	pr_info	881
BTC_PRINT	btc_alg_dbg	663
btc_alg_dbg	RT_TRACE	621
printk	pr_cont	606
printk	pr_warn	591
PDEBUG	gspca_dbg	477
pr_err	dev_err	440
pr_info	ioc_info	418
PDBG	pr_debug	392
dev_err	DRM_DEV_ERROR	387
pr_debug	dev_dbg	369
test_msg	test_err	345
brcmf_err	bphy_err	315
RT_TRACE	pr_err	275
pr_warning	pr_warn	262
pr_err	ioc_err	231
printk	pr_debug	226
BT_ERR	bt_dev_err	205
DTRACE	dml_print	173
PRINT_ER	netdev_err	169
dev_dbg	musb_dbg	163
pr_info	dev_info	161
SSI_LOG_DEBUG	dev_dbg	156
BUGMSG	arc_printk	153
SSI_LOG_ERR	dev_err	152
dev_info	dev_dbg	151
dev_err	dev_dbg	146
DRM_ERROR	DRM_DEBUG	145
dev_info	pci_info	139
PERR	gspca_err	129
btc_iface_dbg	RT_TRACE	119
BTC_PRINT	btc_iface_dbg	119
pr_info	pr_debug	119
gvt_err	gvt_vgpu_err	118
DRM_ERROR	DRM_DEV_ERROR	104
pr_warn	dev_warn	102
pr_info	pr_info_ratelimited	102
pr_err	pr_debug	100

suggests that these updates are a result of established guidelines. It appears that the decision on how to log is left to the discretion of the developers.

The second reason for logging function modification is changing the severity of logging statements by specifying log levels. The `printk` function allows one of the eight log levels defined in `/include/linux/kern_levels.h`. For example,

```
printk(KERN_ERR "GCT Node MAGIC incorrect - GCT invalid\n");
```

However, the new logging API introduced in Linux 1.3.98¹² embedded log levels into the function names, such as `pr_debug` and `pr_info`. This change makes the logging API richer, as developers have dedicated functions for specific log levels.

¹² <https://repo.or.cz/davej-history.git?a=commit;h=aa66269c>

Table 11: Changes in the level of logging statements between Linux v4.3 and v5.3.

New Old	EMERG	ALERT	CRIT	ERR	WARN	NOTICE	INFO	DEBUG
EMERG	0	1	10	51	30	0	1	0
ALERT	0	0	6	8	10	0	0	0
CRIT	1	0	0	8	3	0	0	0
ERR	1	0	5	0	242	19	195	720
WARN	22	2	14	343	0	14	336	346
NOTICE	0	0	0	10	5	0	21	22
INFO	1	0	1	188	62	23	0	562
DEBUG	0	0	0	616	77	6	145	0

Table 11 provides the distribution of changes made to the severity of logging statements. If the developer updates a logging statement to use a new logging function, while keeping the same log level, we do not consider this change as a log level change. This avoids including in the analysis refactoring in the logging code that preserved the log level. As stated in previous work (Yuan et al. 2012a), developers often fail at determining how critical an error is in the first attempt. This observation is confirmed by the finding from our study that, out a total of 4,127 log level changes, approximately one third of the total log level changes were between ERR and DEBUG log levels. Specifically, 720 (17.45%) logging statements lowered the severity of the log message from ERR to DEBUG, and 616 (14.93%) from DEBUG to ERR. We found a total of 1,522 (36.88%) instances where developers increased the severity of a logging statement to increase their visibility. In addition, logging debugging messages at the ERROR level would result in log flooding, making it difficult to diagnose the real problems. We found that a total of 2,605 (63.12%) instances where developers reduced the severity of a logging statement in order to prevent log flooding. In fact, of these 4,127 log level modifications, 3,832 (92.85%) changes are between ERR, WARNING, INFO, and DEBUG log levels. This suggests, as stated before, a lack of standards when logging, in particular, when choosing a log level. Moreover, it indicates that it is by checking generated log messages generated by already written logging statements that developers are better able to reason about appropriate log levels.

Changes made to the static content. 56.02% of the 52,391 changed logging statements, i.e. the majority, include modifications to the static content (i.e. the log message). A similar result has also been observed by Chen and Jiang (2017b) and Yuan et al. (2012a), who reported a ratio of 14%–65% and 18%–56%, respectively. Prior studies list *fixing inconsistency*, *clarification*, and *spelling/grammar mistakes* as the major causes of this type of modifications (Chen and Jiang 2017b; Yuan et al. 2012a; Chen and Jiang 2019). Because static content consists of qualitative data, we are not able to make further claims about these changes in RQ2. However, this is further investigated in RQ3, with a qualitative analysis of logging modifications.

Changes made to the dynamic content. Of the 52,391 updated logging statements, 33,321 (63.60%) include modifications to the dynamic content (i.e. variables and function calls). Yuan et al. (2012a) found that developers often add variables into existing logging statements as afterthoughts, which can aid in the failure diagnosis

process. However, changes to dynamic content represent only 27% of all log modifications in their study. Our finding is, however, in line with the study of Li et al. (2019a) on 12 C/C++ open-source projects, where the authors found that 69.1% updates to logging statement made modifications to the dynamic content. This high number of changes to dynamic content has also been observed by Chen and Jiang (2017b) and Zeng et al. (2019), in which they studied server/desktop and Android applications written in Java, respectively. That is, despite Yuan et al. (2012a) already claims that the addition of variables to logging statement is frequent, in various projects—including Linux—it is more than twice more frequent than they observed in their target systems.

One possible explanation for this high number of changes of the logging statement dynamic content in Linux can be seen in commit 6be9005, where the developer switched to `DRM_DEV_DEBUG_<*>` instead of `DRM_DEBUG`. `DRM_DEV_DEBUG_<*>` are device-aware logging macros and they require `struct *device` as an argument to include device name in the log output. In order to help developers decide which variables should be included in the logging statement, a recent study by Liu et al. (2019) proposed a deep learning-based approach that achieved an average MAP score of 0.84 on nine open-source Java projects. Such approaches tailored to Linux kernel domain could be helpful to alleviate this problem of what information should be included in a logging statement.

4.2.4 RQ2: Summary

RQ2 - Findings. The log ratio (LLOC/SLOC) has been decreasing slowly from v4.3 to v5.3, with a log ratio of 4.10% and 3.72%, respectively. This represents a reduction of 9.27%. SLOC and LLOC are closely correlated across most versions of Linux kernel between v4.3 and v5.3. This seems to be caused by the increase use of tracing mechanism as a substitute to logging. We found that 14% of the commits made between the Linux kernel versions v4.3 and v5.3 involve modifications to logging code. Out of the 211,437 logging statements modifications, 24.78% are log updates, 45.99% are log insertions, and 29.23% are log deletions. The majority of changes to logging code (86.60%) are made in the *drivers* subsystem. The changes to logging functions are triggered by the need to improve the quality of the logging output by either enhancing precision, conciseness, or consistency. We found that 92.85% of the changes of log levels are between ERR, WARNING, INFO, and DEBUG log levels, suggesting that it is difficult for Linux developers to decide on which log level to use. We found that 63.60% of the updated logging statements include modifications to the dynamic content, while 56.02% are changes to the static content.

RQ2 - Implications. A study should be conducted to dig deeper into the reasons behind the decline of the log ratio over the years. The study should explore the correlation between logging and tracing as it appears that tracing is being used as a substitute to logging in more recent versions of Linux. We recommend to examine the benefits of logging, tracing, and other debugging mechanisms when used individually and together. This may lead to better ways to optimize the use of these mechanisms. For example, tracing has the advantage of being less dependent on the developer’s input, and hence requires less maintenance. However, it comes with an

added overhead due to the presence of trace points, which may limit its overall deployment. Another implication of these results is related to the lack of guidelines and best practices, which is the main trigger behind the many changes to logging statements. This is also evidenced in the high number of changes to log levels, which clearly show that developers face challenges when deciding on the log levels to use. These issues should be addressed by developing a set of guidelines and best practices that do not only consider the information being logged (i.e., what to log), but also the amount of information that is generated and its impact on the application execution. Finally, the high number of changes to the dynamic variables is problematic since many errors can occur, leading to faults and crashes. To address this issue, we suggest to use tools that would automatically recommend updates to the dynamic variables of logging statements when these variables have been modified in others parts of the code.

4.3 RQ3: Afterthought changes in logging code

For this research question, we look into *afterthought changes*, which are those that explicitly address a bug caused by or is used to enhance the logging code (Yuan et al. 2012a; Chen and Jiang 2017b). It is essential to study afterthought changes because they add to the overall maintenance effort. Having a large number of afterthought log changes may defeat the very purpose of logging, which is to reduce the maintenance effort by facilitating debugging and other failure diagnosis tasks.

An example of a commit addressing a problem associated with logging code can be seen in Listing 3. In this code snippet, there is an error message displayed by the *thinkpad_acpi* driver when brightness interfaces are not supported, encouraging the user to contact IBM for this problem. However, according to the developer that handled this commit, back-light interfaces on newer devices are supported by the *i915* driver. The developer decided to change the log level from “*error*” to “*info*” to reduce the visibility of the log message. Another example consists of a commit in which a developer decided to enhance the existing logging statements by including `qp_num(qp)` in debug messages to improve debugging tasks (commit `e404f94`).

```
--- a/drivers/platform/x86/thinkpad_acpi.c
+++ b/drivers/platform/x86/thinkpad_acpi.c
@@ -6459,8 +6459,7 @@ static void __init
     ↪ tpacpi_detect_brightness_capabilities(void)
         pr_info("detected a 8-level brightness capable ThinkPad\n");
         break;
     default:
-         pr_err("Unsupported brightness interface, "
-              "please contact %s\n", TPACPI_MAIL);
+         pr_info("Unsupported brightness interface\n");
+         tp_features.bright_unkfw = 1;
+         bright_maxlvl = b - 1;
     }
```

Listing 3: Commit `d618651` - *thinkpad_acpi*: Don't yell on unsupported brightness interfaces.

Table 12: Characterization of fixes to the logging code.

Group	Logging Fix	#Commits
Lack of Information	LF01: Missing variables and statements	92
	LF02: Imprecise logging messages	62
Issues in variable usage	LF03: Null pointer dereference	13
	LF04: Uninitialized variables	9
	LF05: Logging wrong variables	28
Inadequate logging configuration	LF06: Fixing incorrect log levels	156
	LF07: Fixing inconsistencies	163
	LF08: Deleting redundant information	31
Writing issues	LF09: Language mistakes	171
	LF10: Fixing format specifiers	125
	LF11: Message mistakes	6
	LF12: Formatting issues	71
Disclosure of Sensitive data	LF13: Revealing kernel pointers	16
Total		900

To understand the nature of afterthought changes made to logging code, we conducted a qualitative analysis, detailed in Section 3, in which we manually examined the corresponding fixes provided by the Linux developers. The ultimate goal is to gain deep insight into the reasons behind these changes, which can help developers and researchers design new approaches and tools to prevent these problems. To analyze the resulting 900 commits, for each commit, the first author reviewed the commit message, the commit diff, related artefacts such as bug reports, and discussions on the Linux kernel mailing lists, if available. He classified the reasons behind these afterthought changes into 13 categories, which were reviewed and validated by the other authors. These categories are shown in Table 12 and are discussed in more detail in the subsequent sections, split into five groups.

4.3.1 Lack of information

LF01: Missing variables and statements. We found that 10.22% of the studied commits are about improving debuggability by adding information to logging code with the aim of reducing the time needed to diagnose program failures. This is in line with the finding of Yuan et al. (2012a), who concluded that developers often add information to the existing logging statements to narrow down the root causes of the underlying problems. For example, in commit `d0de579`, a developer mentioned: “*Identify Namespace failures are logged as a warning but there is not an indication of the cause for the failure. Update the log message to include the error status.*” Another example is commit `077c066`, which added to the logging statement a local variable representing a section index (`idx`), as shown in Example 1 of Table 13. The reason for this change is given by the developer as: “*While debugging a bpf ELF loading issue, I needed to correlate the ELF section number with the failed relocation section reference. Thus, add section numbers/index to the `pr_debug`.*” We also found many cases where developers added logging statements to record additional runtime

Table 13: Examples of lack of information.

Example 1	tools/libbpf: improve the pr_debug statements to contain section numbers linux@077c066
Original	<code>pr_warning("failed to alloc name for prog under ↳ section %s\n", section_name);</code>
Updated	<code>pr_warning("failed to alloc name for prog under ↳ section(%d) %s\n", idx, section_name);</code>
Example 2	ASoC: tas6424: Print full register name in error message linux@919869214
Original	<code>dev_err(dev, "failed to read FAULT1 register: %d\n", ↳ ret);</code>
Updated	<code>dev_err(dev, "failed to read GLOB_FAULT1 register: ↳ %d\n", ret);</code>
Example 3	mmc: dw_mmc: fix misleading error print if failing to do DMA transfer linux@d12d0cb
Original	<code>/* We can't do DMA */ dev_err(host->dev, "%s: failed to start DMA.\n", ↳ __func__);</code>
Updated	<code>/* We can't do DMA, try PIO for this one */ dev_dbg(host->dev, ↳ "%s: fall back to PIO mode for current transfer\n" ↳ , __func__);</code>
Example 4	x86/fault: Reword initial BUG message for unhandled page faults linux@f28b11a
Original	<code>pr_alert("BUG: unable to handle kernel %s at %px\n", ↳ address < PAGE_SIZE ? "NULL pointer ↳ dereference" : "paging request", ↳ (void*)address);</code>
Updated	<code>if (address < PAGE_SIZE && !user_mode(regs)) pr_alert("BUG: kernel NULL pointer dereference, ↳ address = %px\n", (void *)address); else pr_alert("BUG: unable to handle page fault for ↳ address = %px\n", (void *)address);</code>

information. This is exemplified by commit 9ef8690, where a developer inserted few additional logging points in order to make errors more visible: “*The NCSI driver is mostly silent which becomes a headache when trying to determine what has occurred on the NCSI connection. This adds additional logging in a few key areas such as ...*”

LF02: Imprecise logging messages. Ambiguity in an error message can delay the process of diagnosis as it does not allow end users to easily uncover the part of the programs that failed as in, for instance, Example 2 of Table 13. In this example, a

developer chose to use the full register name in the error message because a short version of the register name may be ambiguous when diagnosing a fault. A similar example can be found in commit [2e5d04dad](#), where the *iwlagm* driver uses exactly the same error message in three different functions. Therefore, the developer chose to add the name of the function to the error string to disambiguate from where the error originated.

Another common problem is the ambiguity of a log message, which can be misleading during the analysis phase. Example 3 of Table 13 shows an example where the original log message “failed to start DMA” together with an error log level might mislead users to think that a fatal error occurred. In fact, although DMA could not be used, the transfer could still be completed by PIO mode. The updated message thus makes this clear as well as has a debug log level.

Another example of this type of logging fix is when developers reword the logging message to make the logs more informative and facilitate analysis. This is shown in Example 4 of Table 13, where developers decided to reword NULL pointer dereference message. The logging message was modified to drop “unable to handle” from the message, because it might imply that in some cases the kernel actually handles NULL pointer dereference, which is not valid. A similar example is found in commit [135e535](#), where the developer clarified an error message to avoid user confusion. He reported that: “Some user who install SIGBUS handler that does longjmp out therefore keeping the process alive is confused by the error message [188988.765862] Memory failure: 0x1840200: Killing cellsrv:33395 due to hardware memory corruption. Slightly modify the error message to improve clarity.” In conclusion, poorly worded log messages may lead to user confusion and fixing these logs takes up maintenance time and effort.

Using the same static text in a single file is also a case that contributes to ambiguity in logging statements. This practice of duplication in logging code was rightfully reported by Li et al. (2019b) as a logging *code smell*. One example of this practice is depicted in commit [a15e824](#) and commit [90cc7f1](#), where the developer added additional information so that log messages can be uniquely identified using search techniques.

4.3.2 Issues in variable usage

LF03: Null pointer dereference. In 1.44% of the cases, a developer attempted to dereference a pointer that may have a NULL value or an empty variable. This is something that can occur in Example 5 of Table 14, which illustrates a fix made in commit [95d2a32](#).¹³ In this commit, the original logging statement included the dereferencing of the *skb* pointer, which can be NULL. A NULL pointer dereferencing causes a runtime crash. To prevent this type of problems, developers should incorporate tools such as Coccinelle¹⁴ in their workflow for detecting dereferences of NULL pointers.

LF04: Uninitialized variables. Out of 900 commits, we found 9 cases in which log messages refer to device names before they are registered. These logs contain mes-

¹³ Note that the change made to the logging statement is not semantically equivalent to the original statement, but with the bug fixed. The goal in the change was to fix the NULL pointer and the change was accepted to fix this issue.

¹⁴ <http://coccinelle.lip6.fr/rules/#null>

Table 14: Examples of issues in variable usage.

Example 5	staging: ks7010: don't print skb->dev->name if skb is null linux@95d2a32
Original	<code>printk(KERN_WARNING "%s: Memory squeeze, dropping ↪ packet.\n", skb->dev->name);</code>
Updated	<code>printk(KERN_WARNING "ks_wlan: Memory squeeze, ↪ dropping packet.\n");</code>
Example 6	staging: fsl-dpaa2/eth: Don't use netdev_err too early linux@0f4c295
Original	<code>netdev_err(net_dev, "Failed to configure hashing\n");</code>
Updated	<code>dev_err(dev, "Failed to configure hashing\n");</code>
Example 7	btrfs: tree-log.c: Wrong printk information about namelen linux@286b92f
Original	<code>btrfs_crit(fs_info, "invalid dir item name len: %u", ↪ (unsigned)btrfs_dir_data_len(leaf, dir_item));</code>
Updated	<code>btrfs_crit(fs_info, "invalid dir item name len: %u", ↪ (unsigned)btrfs_dir_name_len(leaf, dir_item));</code>

sages referring to “unnamed net device” or “uninitialized”, which lead to logs that may confuse end users. Example 6 of Table 14 shows a fix to one of these problems. The logging statements consists now of a call to the function `dev_err`, which considers the possibility that the first parameter is null. In this case, it will add “NULL net_device” to the log message.

LF05: Logging wrong variables. Developers specified the wrong variables as arguments of the logging function calls in 28 cases. The root causes of these issues are simple copy/paste mistakes or typographical errors.¹⁵ This is illustrated in Example 7 of Table 14, in which the original and updated logging functions have similar names, but the former calls the function `btrfs_dir_data_len`, reporting `data_len` in the error message. However, the correct function is `btrfs_dir_name_len`, which reports `name_len` in the message. Such errors are difficult to detect using static analysis tools, because both `btrfs_dir_name_len` and `btrfs_dir_data_len` have the same return type.

4.3.3 Inadequate logging configuration

LF06: Fixing incorrect log levels. We found that in 17.33% of the cases there are changes in the log levels. This includes cases where developers failed to make a distinction between fatal errors and errors that are recoverable. Moreover, using a non-error log level for logging error conditions would make it hard to diagnose such errors as the corresponding error messages would go unnoticed. Likewise, logging debugging messages as errors will also result in a flood of log messages making it

¹⁵ <https://cwe.mitre.org/data/definitions/688.html>

Table 15: Examples of inadequate logging configuration.

Example 8	i2c: imx: notify about real errors on dma i2c_imx_dma_request linux@5b3a23a
Original	<code>dev_dbg(dev, "can't configure rx channel\n");</code>
Updated	<code>dev_err(dev, "can't configure rx channel (%d)\n", ret);</code>
Example 9	mei: bus: remove redundant uuid string in debug messages linux@2bcfdc2
Original	<code>dev_dbg(&cldev->dev, "running hook %s on %pU1\n", ↪ __func__, mei_me_cl_uuid(cldev->me_cl));</code>
Updated	<code>dev_dbg(&cldev->dev, "running hook %s\n", __func__);</code>

hard to concentrate on the real problems. In example of such an issue is shown in Example 8 (Table 15), where the developer mentioned that “... *In contrast real problems that were only emitted at debug level before should be described at a higher level to be better visible and so understandable.*” Thus, the developer decided to change the log level from DEBUG to ERROR.

Developers need to carry out logging activities taking into account performance constraints (Ding et al. 2015; Sigelman et al. 2010). One of the most frequent issues is log spamming, which often leads to degradation of system performance. This occurred in commit 7f20d83, where the developer downgraded the log message to DEBUG level to suppress frequent “VTU miss violations” messages, as “VTU miss violations” are rather common.

LF07: Fixing inconsistencies. We found 163 commits (18.11%) that were the result of ongoing modernization of logging code. Usually, such improvements do not address severe problems associated with the logging code; instead, they arise from attempts to improve the consistency of logging code across various parts of the Linux kernel. One such change is the use of device-managed logging macros in order to simplify error handling, reduce source code size, improve readability, and/or reduce the risks of bugs. Another improvement is the use of `__func__` and `pr_fmt` rather than hard-coded function names and module names in error messages. For example, in commit a8ab042, a developer mentioned: “*Instead of having the function name hard-coded (it might change and we forgot to update them in the debug output) we can use __func__ instead and also shorter the line so we do not need to break it.*” As Yuan et al. (2012a) pointed out, inconsistency in the function names referred in the log message is one of the main reasons for changes made to logging code.

LF08: Deleting redundant information. We found 31 cases in which developers report information that is redundant or not needed. One illustration of this case found in commit 2bcfdc2 is shown in Example 9 (Table 15), where the developer removed `uuid` from the debug messages in `bus-fixup.c` as this was already part of the device name. Another common pattern that we observed is the removal of `__func__` from `dev_dbg()` calls. The reason for this change is given in commit b814735 as “*Dynamic debug can be instructed to add the function name to the debug output using the +f switch, so there is no need for the nfit module to do it again. If a user decides to add the*”

Table 16: Examples of writing issues.

Example 10	usbip: vhci: fix spelling mistake: “synchronuously” → “synchronously” linux@cb48326
Original	<pre>dev_dbg(&urb->dev->dev, "urb seq# %u was unlinked ↳ %ssynchronuously\n", seqnum, status == -ENOENT ? ↳ "":"a");</pre>
Updated	<pre>dev_dbg(&urb->dev->dev, "urb seq# %u was unlinked ↳ %ssynchronously\n", seqnum, status == -ENOENT ? ↳ "":"a");</pre>
Example 11	ti_sci: Use %pS printk format for direct addresses linux@595f3a9
Original	<pre>dev_err(dev, "Mbox timedout in resp(caller: %pF)\n", ↳ (void *)_RET_IP_);</pre>
Updated	<pre>dev_err(dev, "Mbox timedout in resp(caller: %pS)\n", ↳ (void *)_RET_IP_);</pre>
Example 12	greybus: gpb: Fix print mistakes linux@b908dec
Original	<pre>if (gb_usb_protocol_init()) { pr_err("error initializing usb protocol\n"); goto error_usb; } if (gb_i2c_protocol_init()) { pr_err("error initializing usb protocol\n"); goto error_i2c; } }</pre>
Updated	<pre>if (gb_usb_protocol_init()) { pr_err("error initializing usb protocol\n"); goto error_usb; } if (gb_i2c_protocol_init()) { pr_err("error initializing i2c protocol\n"); goto error_i2c; } }</pre>

+f switch for nfit’s dynamic debug this results in double prints of the function name Thus remove the stray __func__ printing.” In addition, removing `__func__` from `dev_*()` callsites helps reduce the Linux kernel size as pointed out by Wolfram Sang, the current maintainer of the Linux I2C subsystem.¹⁶

4.3.4 Writing issues

LF09: Language mistakes. 19% of the 900 studied log updates are caused by spelling or grammar mistakes. Example 10 (Table 16) shows an example where the word “*synchronously*” was misspelt as “*synchronuously*” in the static text. Another example can be seen in commit 748ac56. To correct a grammatical mistake, the static text was changed from “Failed to registered ssb SPROM handler” to “Failed to register ssb SPROM handler”. We also noticed that there is no standardization over the use of capitalization, grammatical style, punctuation, etc.

LF10: Fixing format specifiers. We found 13.89% of the 900 studied revisions are the result of using improper format specifiers in logging statements.¹⁷ A typical issue related to `printf` format specifier is shown in Example 11 (Table 16). In the example, the developer decided to use the `%pS` `printf` format specifier for printing symbols from direct addresses. Moreover, the explanation for this change was given as “*This is important for the ia64, ppc64 and parisc64 architectures, while on other architectures there is no difference between %pS and %pF. Fix it for consistency across the kernel.*” It appears that most of this type of changes are motivated by the intention to fix build warnings, which developers should have addressed at commit-time. We found that developers often get around using proper format specifiers by resorting to unnecessary casts.¹⁸ This practice was rightfully reported by Chen and Jiang (2017a) as a logging *anti-pattern*.

LF11: Message mistakes. We identified six instances where developers made mistakes in the static messages. Similarly to LF05 (*Logging wrong variables*), it seems that developers aimed at save time by copying and pasting a few snippets of code. However, some of the necessary modifications are left behind. This happens in Example 12 (Table 16), where the developer entered an incorrect logging message in the `gpbriidge_init` method, likely caused by a copy/paste mistake. It seems that the developer copied the line `if (gb_usb_protocol_init())` and corrected it. However, the debugging statements were not updated accordingly, referring to the *i2c* protocol, as in the updated statement.

LF12: Formatting issues. Developers fixed formatting of log messages in 71 cases. Poorly formatted messages make it difficult to search for matching text in the log file. We can see this issue in commit a790634, where certain messages were appearing on separate lines resulting in a strange output. This was fixed using line continuations where necessary. Another issue that falls under this category is that developers frequently break log message strings over several lines to meet the checkpatch’s 80 characters per line restriction. According to commit 4bd69e7b, this is no longer considered a good practice, because it makes it more difficult to `grep` for strings in the source code.

Table 17: Example of sensitive data.

Example 13	drm/exynos: Print kernel pointers in a restricted form linux@9cdf0ed
Original	<pre>dev_dbg(dev, "< xfer %p: tx len %u, done %u, rx len ↳ %u, done %u\n", xfer, length, xfer->tx_done, ↳ xfer->rx_len, xfer->rx_done);</pre>
Updated	<pre>dev_dbg(dev, "< xfer %pK: tx len %u, done %u, rx len ↳ %u, done %u\n", xfer, length, xfer->tx_done, ↳ xfer->rx_len, xfer->rx_done);</pre>

4.3.5 Disclosure of sensitive data

LF13: Revealing kernel pointers. Insufficient information in log messages can delay the diagnosis process. However, revealing sensitive information such as cryptographic keys or kernel addresses can lead to information leaks.¹⁹ We found 16 commits, which mentioned these cases. An example can be seen in commit 9cdf0ed shown in Example 13 (Table 17). It states in the commit message that “*Printing raw kernel pointers might reveal information which sometimes we try to hide (e.g. with Kernel Address Space Layout Randomization). Use the “%pK” format so these pointers will be hidden for unprivileged users.*” Other examples of such cases can be seen at CVE-2018-5995²⁰ and CVE-2018-7273²¹, where developers print kernel addresses into logs, which can allow an attacker to extract sensitive information. The problem of accidental data leakage through the misuse of logs has recently been examined by Zhou et al. (2020). The authors showed that logs could reveal sensitive information in Android apps. Similar studies should be conducted for larger systems such as the Linux kernel to understand the extent of this serious problem.

4.3.6 RQ3: Summary

RQ3 - Findings. After manually analyzing 900 commits randomly selected from those that include fixes or improvements in logging code, we identified 13 types of logging fix, categorized into five groups: lack of information, issues in variable usage, inadequate logging configuration, writing issues, and disclosure of sensitive data. The most common type of logging fix consists of language mistakes, appearing in 19% of the commits, followed by fixing inconsistencies (18%), log levels (17%), and format specifiers (13%). The last three types of logging fix indicate a lack of logging standards. Another type of fix that appeared with a certain frequency is missing variables and statements (10%), which suggests that developers should further consider all information that is needed for debugging when adding logging information.

¹⁶ <https://git.kernel.org/pub/scm/linux/kernel/git/wsa/linux.git/commit/?h=strings/rtc-no-func&id=762c5af234c5b816b7da3687a3e703cf8cdc2214>

¹⁷ <https://www.kernel.org/doc/Documentation/printk-formats.txt>

¹⁸ <https://github.com/torvalds/linux/commit/3fcb3c836ef413d3fc848288b308eb655e08d853>

¹⁹ <https://cwe.mitre.org/data/definitions/200.html>

²⁰ <https://nvd.nist.gov/vuln/detail/CVE-2018-5995>

²¹ <https://nvd.nist.gov/vuln/detail/CVE-2018-7273>

RQ3 - Implications. The findings show that there are many errors in the logging code, which may have a significant impact on log analysis tasks. Based on these findings, we derived a list of guidelines and recommendations to help address each of the issues raised in this section. The proposed guidelines can be found in Section 5.4.

5 Discussion

In this section, we discuss insights derived from our study. First, we focus on lessons learned associated with the pervasiveness of logging, followed by comparison of our findings to previous similar studies. Then, we present guidelines and point out directions for automated support for logging. Next, we detail the feedback provided by a Linux expert regarding the findings of our study. Finally, we discuss threats to the validity of our study.

5.1 On the Pervasiveness of Logging

Logging code represents 3.69% of the Linux kernel, being pervasive in its source code. However, when we evaluate the pervasiveness of logging at the file and method levels, we found that the distribution of logging code is highly skewed, suggesting that certain aspects of the Linux kernel are more likely to have logging code than others. It would be interesting to assess the correlation between the log ratio of a method and its cyclomatic complexity in future studies.

Although logging code corresponds to only less than 4% of the Linux kernel code, this type of code is actively maintained. We found that 14% of commits related to Linux kernel versions 4.3 to 5.3 involves modifications to logging code resulting in a total of 211,437 logging statements modifications, which represents 66.19% of the total number of logging statements present in Linux kernel v5.3. Moreover, we identified that logging code deletion accounts for 29.23% of the total modifications to logging code, in contrary to findings reported by Yuan et al. (2012a). Despite the majority of logging code deletion happens when a file is deleted, there is a non-negligible amount (44.66%) of deletion of logging code occurring as afterthoughts. A possible explanation for this might be observed in commit `a8d5dad`, where a developer deleted two logging statements that report memory allocation failures. There is a rule in `checkpatch.pl` to check for possible unnecessary *out of memory* message. However, it seems that developers do not care to fix this issue at commit time, because there is a large number of commits removing *out of memory* error messages as afterthoughts. Logging statements are also deleted in cases in which they are redundant. For instance, in commit `c99a23e55`, a developer removed an error message present within error handling code when `i2c_mux_add_adapter` fails because `i2c_mux_add_adapter` already prints an error message when it fails.

We found that developers often face difficulties to specify the right log level in the first attempt, because there is no such a static verification of whether a log level is correct. For example, in commit `3b364c659`, a developer downgraded the logging message from `WARN` to `INFO` log level, and reason for this change was: “*On an embedded system it is quite possible for the bootloader to avoid configuring PCIe*

Table 18: Lack of consistency in the text of the error messages

<pre>spear_thermal = thermal_zone_device_register(...); if (IS_ERR(spear_thermal)) { dev_err(&pdev->dev, "thermal zone device is NULL\n"); [...] }</pre>	drivers/thermal/spear_thermal.c
<hr/>	
<pre>priv->zone = thermal_zone_device_register(...); if (IS_ERR(priv->zone)) { dev_err(dev, "can't register thermal zone\n"); [...] }</pre>	drivers/thermal/rcar_thermal.c
<hr/>	
<pre>sensor->thermal_dev = thermal_zone_device_register(...); if (IS_ERR(sensor->thermal_dev)) { dev_err(dev, "failed to register thermal zone device\n"); [...] }</pre>	drivers/thermal/st/st_thermal.c

devices if they are not needed." It is surprising that it took the developers 26 months to notice this problem.

The majority of the logging statements (55.66%) are used within an `if` block. Moreover, most of them is used for logging errors after checking the return value of function calls. As the decision regarding logging is left to developers, we noticed many inconsistencies in the text of an error message, logging function used, and information included in an error message. Table 18 shows an example of this situation. A call to `thermal_zone_device_register()` returns a pointer to the newly created struct `thermal_zone_device`, and in case of error it returns an `ERR_PTR`.²² All these three drivers perform a registration of a new thermal zone device and check the return value using `IS_ERR()` macros and, in case of an error, log an error using `dev_err()`. Even though all three drivers are performing the same action, there is no consistency among the error messages. This can prevent automated post-mortem analysis.

We also observed inconsistencies in the information included when logging an error message. For example, there are many locations that report an error when there is a failure when calling the function `devm_request_irq`. However, plenty of the callers of this function do not include the `irq` requested and the returned error code.²³ A possible solution to avoid such inconsistencies would be centralizing error reporting rather than leaving the decision to the developers. One such a change made to get more consistent error reporting can be seen in commit 7723f4c, the reason

²² https://github.com/torvalds/linux/blob/v5.3/drivers/thermal/thermal_core.c#L1211

²³ <https://github.com/torvalds/linux/blob/v5.3/drivers/usb/dwc2/gadget.c#L4846>

for this change is that “A *grep* of the kernel shows that many drivers print an error message if they fail to get the *irq* they’re looking for. Furthermore, those drivers all decide to print the device name, or not, and the *irq* they were requesting, or not, etc. Let’s consolidate all these error messages into the API itself, allowing us to get rid of the error messages in each driver.” This centralization of error reporting helps reduce the size of the Linux kernel code. This also decreases the need for commits made as afterthoughts to add additional information or fixing typographical mistakes.

5.2 Comparison to Existing Studies

Our study replicates existing studies that had as target systems C/C++ and Java projects. We, in turn, analyzed the Linux kernel, which is one of the greatest collaborative efforts in the computer industry. Based on these previous studies and their results, we show in Table 19 a comparison of the results obtained in our study and those from Yuan et al. (2012a) and Chen and Jiang (2017b), which explore similar aspects related to the practice of logging.

5.3 Feedback from a Linux Expert

We reached out to a Linux development expert, who has also been very active in discussing the use of logging libraries in Linux, to obtain feedback on this study and discuss possible improvements. We also gave him an early version of the paper to review. In general, the expert agreed with the fact that logging is not a practice that is well governed in Linux, which explains the fact that logging varies significantly from one component of the system to another, depending on the development teams of the components. He added that most of the findings related the change of the logging code are caused by a lack of guidelines across all the development teams in Linux. He added that he had himself committed many updates to fix issues related to logging code to improve the quality of the logging statements.

Further, the Linux expert made three suggestions that do not only provide insights into the results, but can also be used to drive future studies. The first suggestion is to separate code written specifically for Linux from that ported to Linux such as some external drivers. This would perhaps provide a better picture of the logging landscape of Linux-specific code. Having said, our answer to RQ1 shows that the most logged components are “filesystem” and “core”, which are specific to Linux. In other words, it is not clear whether excluding the drivers would have an important effect on the results of RQ2 and RQ3. Besides drivers are an integral part of Linux, removing them may weaken the study. The second suggestion is related to the use of various tools (codespell, integrated spelling tests in checkpatch, coccinelle, etc) that have recently been made available to Linux developers. These tools are now recommended to improve the quality of code. On one hand, this comment confirms our findings with respect to RQ3 where we concluded that many errors in the logging code could have been avoided using this type of tools. On the other hand, the point raised by the expert warrants a study on how these tools are used to check logging code and what the impact on the quality of logging statements would be. Finally, the expert explained that in recent years, there has been an increase in the use of tracing tools such as *ftrace* to log the entry/exit of function, which justifies

Table 19: Comparison of results from our study and previous work.

Research Question	Our study	Yuan et al. (2012a)	Chen and Jiang (2017b)
<i>RQ1: Pervasiveness of logging in Linux kernel</i>	On average, every 27 lines of code contained one line of logging code in the Linux kernel.	On average, every 30 lines of code contained one line of logging code in four open-source C/C++ projects.	On average, every 51 lines of code contained one line of logging code in 21 Java projects.
<i>RQ2: Logging code evolution</i>	Logging code is modified in 14% of all analysed commits.	Logging code is modified in 18% of all committed revisions.	Logging code is modified in 16% of all committed revisions.
	Logging deletions account for 29% of the number of modifications made to logging code.	Logging deletions account for 2% of the number of modifications made to logging code.	Logging deletions account for 26% of the number of modifications made to logging code.
<i>RQ3: Afterthought changes in logging code</i>	30% of analyzed commits are related to afterthought updates.	33% of updates to the log printing code are afterthought updates.	59% of updates to the log printing code are afterthought updates.
	Fixing incorrect log levels accounts for 17% of analyzed commits.	Fixing incorrect log levels accounts for 26% of afterthought updates.	Fixing incorrect log levels accounts for 21% of afterthought updates.
	Almost 16% of analyzed commits are related to issues in variable usage or missing variables and statements.	27% of the afterthought updates are related to variable logging.	32% of afterthought updates are related to variable logging.
	Fixing language mistakes is the most frequent update to static text.	Fixing misleading information is the most frequent updates to the static text.	Fixing misleading information is the most frequent updates to the static text.

the decrease in the use of logging. This comment is inline with our finding in RQ2 where we showed that despite the increase of the number of LOCs across versions, the number of logging statements have decreased. We attributed this to the emergence of tracing and other debugging mechanisms as substitutes to logging on the premise of the studies of Corbet (2016) and Edge (2019). The Linux expert’s opinion converges with this finding. The input provided by the Linux expert encourages us to conduct a formal user study to gain more insights into the way logging is used in Linux. This study should build on the results obtained in this paper, which are an important first step to understand the logging landscape in Linux in particular and in software engineering in general.

5.4 Logging Guidelines

Based on the analysis of the afterthoughts changes in the logging statements in the Linux kernel source code, we derived the following guidelines that can help the maintenance of logging code. For each listed guideline, we also indicate the identified types of logging fix that serves as a foundation for it and a description. We also indicate opportunities for automation when applicable.

A. Be precise, concise and consistent in logging statements. (*LF02, LF07, LF08, LF10, LF12*) In order for logs to be useful, they should be precise (not leaving room for ambiguity), concise (to not create huge amounts of logs with redundant information) and consistent (to facilitate post-mortem analysis). An example of specific recommendations is to *not use hard-coded module or function names* and refer to *complete variable names*.

Developers should strive for writing concise logging statements to prevent redundant information. Automatic tools can be developed to detect redundancies. However, some redundancies may require domain knowledge. These can be detected by developers when reviewing the code. There are static analysis tools such as *smatch*²⁴ and *sparse*²⁵, that are useful at detecting incorrect format specifiers. Developers should consider using these tools to detect this type of problems at commit time rather than fixing the corresponding code as afterthought.

It is recommended to use a *spell checker*, as provided by many IDEs. This is an initial step to prevent the need for corrections in language issues. Additional tools for grammar checking could be embedded in current IDEs. Guidelines for a common writing style should be developed and promoted. Tools such as *kernelscan*²⁶ should be incorporated into build pipelines.

B. Specify (in advance) and follow logging conventions. (*LF06, LF08*) Even though there are existing descriptions of how the different log levels should be used in the Linux kernel, a suggestion is to further detail logging conventions. For instance, the error level should only be used in situations in which components may fail and compromise system operation. Moreover, conventions can be used to avoid redundant or missing information, and use similar static messages, because for the same situation, such as an error, different messages and variables are logged in distinct code locations. This can be aided by dedicated logging functions.

²⁴ <https://repo.or.cz/w/smatch.git>

²⁵ <https://sparse.wiki.kernel.org/>

²⁶ <https://github.com/ColinIanKing/kernelscan>

C. Program for debuggability. (*LF01, LF12*) To prevent posterior logging changes, we recommend that developers provide beforehand detailed information on where the failures are located and any other information relevant to failure that can facilitate debugging. For example, including error code in the logging statement, if available, is always a good idea. In addition, developers should avoid breaking log message lines to facilitate post-mortem analysis of the logs.

D. Reason about possible variable values. (*LF03, LF04*) NULL pointer dereferencing causes a runtime crash. Consequently, it is crucial to reason if a referred pointer can be NULL or a variable can be empty. A potential NULL pointer dereferencing can be avoided by adding automatic checks to logged variables, and function return values that are used as logging statements arguments. Furthermore, developers should only use the device or network-specific logging macros after checking that the devices were correctly initialized and registered. Static analysis tools can be used to check that all variables are initialized before they are used for logging.

E. Be careful with copy-paste. (*LF05, LF11*) Developers should avoid copying and pasting logging statements, as this is a source of mistakes. If the statements do not match the targeted subject, they may confuse and mislead developers when debugging and analyzing logged messages. In the case of a copy-paste of a logging statement, it is essential to check if (i) a called logging function has been updated; (ii) the static message has been adequately changed; and (iii) referred variables are correct.

F. Consider security issues. (*LF13*) Tests should be put in place to verify that logs do not accidentally cause security and data privacy breaches. This effort should adhere to the broader task of ensuring the security of the Linux kernel.

G. Review logging code. (*LF02, LF09*) To avoid many of the raised issues, such as imprecise, redundant, and inconsistent logging statements as well as copy-paste mistakes, can be detected at commit time during a code review process. Therefore, we recommend logging code to be carefully reviewed. This task can be done before accepting the commits.

H. Use recommender systems. (*LF01, LF05, LF06, LF07*) There is an ongoing research in developing recommendation tools to assist developers in logging. For example, Panthaplackel et al. (2020) proposed a deep-learning-based approach to automatically update the static text in the logging statements based on changes made to the surrounding source code. Fu et al. (2014) conducted a study at Microsoft where they showed that it is possible to use machine learning techniques to predict with a good accuracy where to log. Developers should consider including these tools, if shown effective in the context of the Linux kernel, to predict log levels, variables that need to be logged, static analysis of logging statements, etc.

5.5 Implications for Future Research

Our study allowed us to understand how logging is present in the Linux kernel and to identify issues that caused the evolution of logging statements. The analysis of collected data revealed facts related to logging in the Linux kernel, leaving room for future research work. We discuss directions for future studies as follows.

The analysis of the pervasiveness of the logging in the Linux kernel showed that the log ratio in the many Linux subsystems and components varies. This can be explained by various arguments, such as: (i) long-lived less-changed components may

be stable and are the cause of few bugs, so they do not need to be as logged as other components; (ii) logging can have an impact on system performance and, therefore, certain code parts may be less logged to prevent performance decay; and (iii) different parts of the code may be maintained by different teams of developers, with diverging logging practices. Based on the data collected for this study, it is not possible to reach a conclusion regarding the rationale for the current logging decisions and their variability across the different Linux subsystems and files. This calls for future studies that investigate this. Possible types of studies that are suitable for testing these hypotheses are:

- a user study in which developers think out loud, justifying the logging statements that they add, modify or delete;
- interviews with developers immediately after they commit a code change that involves logging changes; or
- a qualitative study of code review comments, in which there are requests for changes in logging statements with an accompanying rationale for the change.

With respect to our logging guidelines, a future study can be conducted to evaluate their impact on the evolution of logging practices. The presented guidelines were derived using the principles of grounded theory—the guidelines are grounded on the conducted qualitative analysis. Although there are supporting evidences for each derived guideline, it is important to assess if their adoption in the Linux kernel (or a software project) can have a positive impact on the standardization of the logging statements, reducing the need for constant modifications due to inadequate logging. An experimental study in which the intervention is the use of our derived guidelines is suitable for answering this question.

5.6 Threats to Validity

Internal Validity. Threats to internal validity are associated with factors that may impact our results. In this study, a source of bias is the automated data collection process. To identify logging statements, we rely on the semantic patterns specified by Tschudin et al. (2015). However, this approach may not identify statements that lack variability in their use. To mitigate this issue, we manually examined macros containing calls to basic functions, such as `printk()`, `pr_*()`, and `dev_*()`, in order to identify missed logging functions. The set of functions collected by combining both processes was thus manually reviewed with the aim of eliminating obvious false positives. In addition, although we believe that the list of logging functions extracted with this approach is a very comprehensive, we cannot guarantee 100% coverage. To mitigate this threat, we selected randomly 20 files and checked manually their content to see if we missed any logging functions. We found that our list covered all the logging functions invoked in these files. We also want to emphasize that we do not think that the missing of some logging functions affects much the conclusions of this study. For RQ1 (the pervasiveness of logging), we already established that logging is pervasive in Linux (on average for every 27 lines of code, there is one logging statement), so finding more logging statements can only confirm this fact. For RQ2—The evolution of logging code, we found that logging code is constantly maintained by Linux developers (14% of commits modify logging statements). Additional logging statements will eventually confirm this finding. The nature of changes are due to the

need to improve the quality of the logging output by either enhancing precision, conciseness, or consistency. RQ3 (analysis of afterthought log changes) is based on a sample of commits and is not affected by the number of logging statements.

In order to automatically identify and classify changes made to logging code, we use a pipeline that consists of GumTree (Falleri et al. 2014) and a script specifically developed for this study. To mitigate possible misclassifications, we manually inspected a sample of 100 randomly selected modifications, which showed an accuracy of 98% on classified modifications. Although 2% have been misclassified, given the amount of analyzed code, it can be considered noise in the data.

Another threat to this study is the manual classification of log-related commits. We manually examined all commits by their title and message whenever necessary. However, we cannot eliminate the possibility that errors may have occurred during this manual analysis. We thus do not claim that our dataset is complete. This threat is mitigated by the fact that our goal was to collectively study the nature of the problematic logging code rather than collecting every possible revision made to fix or improve logging. Because we perform a qualitative analysis in order to categorize changes made to logging code and the nature of problematic logging code, researcher bias also becomes a threat to the internal validity of our study. In this kind of analysis, results are associated with researcher interpretation of the data. Therefore, to mitigate it, every classification that raised questions were discussed by the authors until they reached an agreement. In addition, all the results were largely discussed and validated among all authors.

External Validity. Threats to external validity are related to which extent our results can be generalized. One threat to external validity is due to the fact that we only examined 22 releases (v4.3 - v5.3) of the Linux kernel to answer part of RQ2 where we measure the evolution of logging code in these releases. A larger dataset may provide different results. This said, we still think our dataset is representative for the study as a whole since our objective is not to uncover all the problems related to logging in Linux but rather to provide insight on the practice of logging in software development by looking at how Linux developers, even in a narrower scope, use logging. In addition, the fact that we focused solely on the Linux kernel project, which is known for having its own development culture, may be a threat to external validity. We believe that this threats is mitigate by the fact that Linux is a large scale, open-source project maintained by developers from different companies, which constitutes a representative sample of C/C++ projects.

6 Conclusion

In this paper, we presented an empirical study that aims to shed light into existing logging practices in the Linux kernel. Although the logging code accounts for 3.73% of the total source code in the Linux kernel, we observed that the distribution of logging code is skewed when evaluated at file and program construct level. Future studies may employ survey methods to understand the circumstances in which logging is needed and the rationales of logging activities. We also found that 32.37% of the log-level changes are between error and debug log levels, suggesting that it is also difficult for Linux developers to decide between fatal errors and errors that are recoverable. We observed that the major causes underlying fixes to problematic

logging code occurring in the Linux kernel are language mistakes, appearing in 19% of the commits. Other common causes for changes are fixing inconsistencies (18%), log levels (17%), and format specifiers (13%). Based on our results, we discussed insights associated with the pervasiveness of logging code in the Linux kernel. In addition, we derived eight practical guidelines that can help developers to maintain logging code, including the need for automated support for logging.

As future work, we aim to perform an in-depth qualitative analysis of logging code to further understand common patterns in logging statements. Our present study targeting the Linux kernel, complemented by this future qualitative analysis, will be the basis for the next generation of tools to provide automated support to develop and evolve logging code.

Acknowledgements

Abdelwahab Hamou-Lhadj and Keyur Patel would like to thank Ericsson Global Artificial Intelligence Accelerator (GAIA) Group in Montreal and MITACS for supporting this project (Grant Number: IT15986). Ingrid Nunes would like to for CNPq grants ref. 313357/2018-8 and ref. 428157/2018-1, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. João Faccin would like to acknowledge the support of the National Council for Scientific and Technological Development of Brazil (CNPq) (grant ref. 141840/2016-1), and the support of the Government of Canada through the Emerging Leaders in the Americas Program (ELAP) program.

References

- Bagherzadeh M, Kahani N, Bezemer CP, Hassan AE, Dingel J, Cordy JR (2018) Analyzing a decade of linux system calls. *Empirical Software Engineering* 23(3):1519–1551
- Bertero C, Roy M, Sauvanaud C, Trédan G (2017) Experience report: Log mining using natural language processing and application to anomaly detection. In: 2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE), IEEE, pp 351–360
- Bootlin (2020) Embedded Linux kernel and driver development training. CreateSpace Independent Publishing Platform, URL <https://github.com/bootlin/training-materials>
- Boslaugh S (2012) *Statistics in a nutshell: A desktop quick reference.* ” O’Reilly Media, Inc.”
- Chen B, Jiang ZM (2017a) Characterizing and detecting anti-patterns in the logging code. In: 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE), IEEE, pp 71–81
- Chen B, Jiang ZMJ (2017b) Characterizing logging practices in java-based open source software projects—a replication study in apache software foundation. *Empirical Software Engineering* 22(1):330–374
- Chen B, Jiang ZMJ (2019) Extracting and studying the logging-code-issue-introducing changes in java-based large-scale open source software systems. *Empirical Software Engineering* 24(4):2285–2322

- Cinque M, Cotroneo D, Natella R, Pecchia A (2010) Assessing and improving the effectiveness of logs for the analysis of software faults. In: IEEE/IFIP International Conference on Dependable Systems & Networks (DSN), IEEE, pp 457–466
- Corbet J (2012) The perils of `pr_info()`. LWN net
- Corbet J (2016) Tracepoint challenges. LWN net
- Corbet J, Rubini A, Kroah-Hartman G (2005) Linux Device Drivers: Where the Kernel Meets the Hardware. ” O’Reilly Media, Inc.”
- Ding R, Zhou H, Lou JG, Zhang H, Lin Q, Fu Q, Zhang D, Xie T (2015) Log2: A cost-aware logging mechanism for performance diagnosis. In: Proceedings of the 2015 USENIX Conference on Usenix Annual Technical Conference, USENIX Association, USA, USENIX ATC ’15, p 139–150
- Edge J (2019) Unifying kernel tracing. LWN net
- El-Masri D, Petrillo F, Guéhéneuc YG, Hamou-Lhadj A, Bouziane A (2020) A systematic literature review on automated log abstraction techniques. *Information and Software Technology* 122:106276
- Fadel W (2011) Techniques for the abstraction of system call traces to facilitate the understanding of the behavioural aspects of the linux kernel. In: Master’s Thesis, Concordia University
- Falleri J, Morandat F, Blanc X, Martinez M, Monperrus M (2014) Fine-grained and accurate source code differencing. In: ACM/IEEE International Conference on Automated Software Engineering, ASE ’14, Vasteras, Sweden - September 15 - 19, 2014, pp 313–324, DOI 10.1145/2642937.2642982, URL <http://doi.acm.org/10.1145/2642937.2642982>
- Fu Q, Zhu J, Hu W, Lou JG, Ding R, Lin Q, Zhang D, Xie T (2014) Where do developers log? an empirical study on logging practices in industry. In: Companion Proceedings of the 36th International Conference on Software Engineering, pp 24–33
- Hamou-Lhadj A, Lethbridge TC (2002) Compression techniques to simplify the analysis of large execution traces. In: Proceedings 10th International Workshop on Program Comprehension, IEEE, pp 159–168
- Hamou-Lhadj A, Lethbridge TC (2004) A survey of trace exploration tools and techniques. In: Proceedings of the 2004 conference of the Centre for Advanced Studies on Collaborative research, pp 42–55
- Hassani M, Shang W, Shihab E, Tsantalis N (2018) Studying and detecting log-related issues. *Empirical Software Engineering* 23(6):3248–3280
- He P, Chen Z, He S, Lyu MR (2018) Characterizing the natural language descriptions in software logging statements. In: Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ACM, New York, NY, USA, ASE 2018, pp 178–189, DOI 10.1145/3238147.3238193, URL <http://doi.acm.org/10.1145/3238147.3238193>
- Islam MS, Khreich W, Hamou-Lhadj A (2018) Anomaly detection techniques based on kappa-pruned ensembles. *IEEE Transactions on Reliability* 67(1):212–229
- Israeli A, Feitelson DG (2010) The linux kernel as a case study in software evolution. *Journal of Systems and Software* 83(3):485–501
- Kc K, Gu X (2011) Elt: Efficient log-based troubleshooting system for cloud computing infrastructures. In: 2011 IEEE 30th International Symposium on Reliable Distributed Systems, IEEE, pp 11–20
- Khatuya S, Ganguly N, Basak J, Bharde M, Mitra B (2018) Adele: Anomaly detection from event log empiricism. In: IEEE INFOCOM 2018-IEEE Conference on

- Computer Communications, IEEE, pp 2114–2122
- Lal S, Sardana N, Sureka A (2015) Two level empirical study of logging statements in open source java projects. *International Journal of Open Source Software and Processes (IJOSSP)* 6(1):49–73
- Li H, Shang W, Hassan AE (2017) Which log level should developers choose for a new logging statement? *Empirical Software Engineering* 22(4):1684–1716, DOI 10.1007/s10664-016-9456-2, URL <https://doi.org/10.1007/s10664-016-9456-2>
- Li H, Chen THP, Shang W, Hassan AE (2018) Studying software logging using topic models. *Empirical Softw Engg* 23(5):2655–2694, DOI 10.1007/s10664-018-9595-8, URL <https://doi.org/10.1007/s10664-018-9595-8>
- Li H, Shang W, Adams B, Sayagh M, Hassan AE (2020) A qualitative study of the benefits and costs of logging from developers’ perspectives. *IEEE Transactions on Software Engineering*
- Li S, Niu X, Jia Z, Liao X, Wang J, Li T (2019a) Guiding log revisions by learning from software evolution history. *Empirical Software Engineering* pp 1–39
- Li Z, Chen TH, Yang J, Shang W (2019b) Dfinder: Characterizing and detecting duplicate logging code smells. In: 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE), IEEE, pp 152–163
- Liu Z, Xia X, Lo D, Xing Z, Hassan AE, Li S (2019) Which variables should i log? *IEEE Transactions on Software Engineering* pp 1–1
- Lotufo R, She S, Berger T, Czarnecki K, Wasowski A (2010) Evolution of the linux kernel variability model. In: *International Conference on Software Product Lines*, Springer, pp 136–150
- Lu L, Arpaci-Dusseau AC, Arpaci-Dusseau RH, Lu S (2014) A Study of Linux File System Evolution. *ACM Transactions on Storage* 10(1):1–32, DOI 10.1145/2560012, URL <https://dl.acm.org/doi/10.1145/2560012>
- Mazuera-Rozo A, Trubiani C, Linares-Vásquez M, Bavota G (2020) Investigating types and survivability of performance bugs in mobile apps. *Empirical Software Engineering* pp 1–43
- Miranskyy A, Hamou-Lhadj A, Cialini E, Larsson A (2016) Operational-log analysis for big data systems: Challenges and solutions. *IEEE Software* 33(2):52–59
- Oliner AJ, Aiken A, Stearley J (2008) Alert detection in system logs. In: 2008 Eighth IEEE International Conference on Data Mining, IEEE, pp 959–964
- Panthaplackel S, Nie P, Gligoric M, Li JJ, Mooney RJ (2020) Learning to update natural language comments based on code changes. 2004.12169
- Passos L, Czarnecki K, Wasowski A (2012) Towards a catalog of variability evolution patterns: the Linux kernel case. In: *Proceedings of the 4th International Workshop on Feature-Oriented Software Development - FOSD '12*, ACM Press, Dresden, Germany, pp 62–69, DOI 10.1145/2377816.2377825, URL <http://dl.acm.org/citation.cfm?doid=2377816.2377825>
- Pecchia A, Cinque M, Carrozza G, Cotroneo D (2015) Industry practices and event logging: Assessment of a critical software development process. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, IEEE, vol 2, pp 169–178
- Pi A, Chen W, Zhou X (2018) Profiling distributed systems in lightweight virtualized environments with logs and resource metrics. In: *Proceedings of the 27th International Symposium on High-Performance Parallel and Distributed Computing*, Association for Computing Machinery, New York, NY, USA, HPDC '18, p 9–10, DOI 10.1145/3220192.3220197, URL <https://doi.org/10.1145/3220192.3220197>

- Ran CA (2019) Studying and leveraging user-provided logs in bug reports for debugging assistance, URL <https://spectrum.library.concordia.ca/985950/>
- Shang W, Jiang ZM, Adams B, Hassan AE, Godfrey MW, Nasser M, Flora P (2014) An exploratory study of the evolution of communicated information about the execution of large software systems. *Journal of Software: Evolution and Process* 26(1):3–26, DOI 10.1002/smr.1579, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.1579>, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/smr.1579>
- Shang W, Nagappan M, Hassan AE (2015) Studying the relationship between logging characteristics and the code quality of platform software. *Empirical Software Engineering* 20(1):1–27
- Sigelman BH, Barroso LA, Burrows M, Stephenson P, Plakal M, Beaver D, Jaspan S, Shanbhag C (2010) Dapper, a large-scale distributed systems tracing infrastructure. Tech. rep., Google, Inc., URL <https://research.google.com/archive/papers/dapper-2010-1.pdf>
- Tian J, Rudraraju S, Li Z (2004) Evaluating web software reliability based on workload and failure data extracted from server logs. *IEEE Transactions on Software Engineering* 30(11):754–769
- Tschudin PS, Lawall J, Muller G (2015) 3l: Learning linux logging. In: Belgian-Netherlands software eVOLUTION seminar (BENEVOL 2015)
- Yang S, Park SJ, Ousterhout J (2018) Nanolog: A nanosecond scale logging system. In: 2018 USENIX Annual Technical Conference (USENIX ATC 18), USENIX Association, pp 335–350
- Yen TF, Oprea A, Onarlioglu K, Leetham T, Robertson W, Juels A, Kirda E (2013) Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. In: Proceedings of the 29th Annual Computer Security Applications Conference, pp 199–208
- Yuan D, Park S, Zhou Y (2012a) Characterizing logging practices in open-source software. In: Proceedings of the 34th International Conference on Software Engineering, IEEE Press, pp 102–112
- Yuan D, Zheng J, Park S, Zhou Y, Savage S (2012b) Improving software diagnosability via log enhancement. *ACM Trans Comput Syst* 30(1):4:1–4:28, DOI 10.1145/2110356.2110360, URL <http://doi.acm.org/10.1145/2110356.2110360>
- Zeng Y, Chen J, Shang W, Chen THP (2019) Studying the characteristics of logging practices in mobile apps: a case study on f-droid. *Empirical Software Engineering* 24(6):3394–3434
- Zhao X, Rodrigues K, Luo Y, Stumm M, Yuan DY, Zhou Y (2017) The game of twenty questions: Do you know where to log? In: 16th Workshop on Hot Topics in Operating Systems (HotOS), pp 125–131
- Zhou R, Hamdaqa M, Cai H, Hamou-Lhadj A (2020) Mobilogleak: A preliminary study on data leakage caused by poor logging practices. In: 2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER), IEEE, pp 577–581
- Zhu J, He P, Fu Q, Zhang H, Lyu MR, Zhang D (2015) Learning to log: Helping developers make informed logging decisions. In: Proceedings of the 37th International Conference on Software Engineering - Volume 1, IEEE Press, Piscataway, NJ, USA, ICSE '15, pp 415–425, URL <http://dl.acm.org/citation.cfm?id=2818754.2818807>