



---

## KernelHaven – An Open Infrastructure for Product Line Analysis

Christian Kröher, Sascha El-Sharkawy, Klaus Schmid  
Software Systems Engineering, Institute of Computer Science,  
University of Hildesheim, Germany  
{kroehler, elscha, schmid}@sse.uni-hildesheim.de

---

### Please cite this publication as follows:

Christian Kröher, Sascha El-Sharkawy, and Klaus Schmid. “KernelHaven – An Open Infrastructure for Product Line Analysis”. In: *Proceedings of the 22nd International Systems and Software Product Line Conference — Volume 2*. SPLC '18. ACM, 2018, pp. 5–10. DOI: 10.1145/3236405.3236410.

### The corresponding BibTeX-entry is:

```
@INPROCEEDINGS{KroehlerEl-SharkawySchmid18,  
  author    = {Christian Kr{o}her and Sascha El-Sharkawy and Klaus Schmid},  
  title     = {KernelHaven -- An Open Infrastructure for Product Line Analysis},  
  booktitle = {Proceedings of the 22nd International Systems and Software Product Line  
              Conference --- Volume 2},  
  series    = {SPLC '18},  
  year      = {2018},  
  publisher = {ACM},  
  pages     = {5--10},  
  keywords  = {Software product line analysis; Variability extraction; Static analysis;  
              Empirical software engineering},  
  doi       = {10.1145/3236405.3236410},  
}
```

ACM, 2018. This is the authors version of the work. It is posted here by permission of the ACM for your personal use. Not for redistribution. The definitive version was published in *Proceedings of the 22nd International Systems and Software Product Line Conference — Volume 2*, DOI: 10.1145/3236405.3236410.

# KernelHaven—An Open Infrastructure for Product Line Analysis

Christian Kröher, Sascha El-Sharkawy, Klaus Schmid

University of Hildesheim, Institute of Computer Science

Hildesheim, Germany

{kroher,elscha,schmid}@sse.uni-hildesheim.de

## ABSTRACT

KernelHaven is an open infrastructure for Software Product Line (SPL) analysis. It is intended both as a production-quality analysis tool set as well as a research support tool, e.g., to support researchers in systematically exploring research hypothesis. For flexibility and ease of experimentation KernelHaven components are plug-ins for extracting certain information from SPL artifacts and processing this information, e.g., to check the correctness and consistency of variability information or to apply metrics. A configuration-based setup along with automatic documentation functionality allows different experiments and supports their easy reproduction.

Here, we describe KernelHaven as a product line analysis research tool and highlight its basic approach as well as its fundamental capabilities. In particular, we describe available information extraction and processing plug-ins and how to combine them. On this basis, researchers and interested professional users can rapidly conduct a first set of experiments. Further, we describe the concepts for extending KernelHaven by new plug-ins, which reduces development effort when realizing new experiments.

## CCS CONCEPTS

• **General and reference** → **Empirical studies; Experimentation**; • **Software and its engineering** → **Software product lines**;

## KEYWORDS

Software product line analysis, variability extraction, static analysis, empirical software engineering

### ACM Reference Format:

Christian Kröher, Sascha El-Sharkawy, Klaus Schmid. 2018. KernelHaven—An Open Infrastructure for Product Line Analysis. In *22nd International Systems and Software Product Line Conference - Volume B (SPLC '18)*, September 10–14, 2018, Gothenburg, Sweden. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3236405.3236410>

## 1 INTRODUCTION

KernelHaven is a tool for supporting professional research, in particular, in the area of Software Product Line (SPL) analysis and verification. In order to provide optimal research support, it rests on the underlying metaphor of an *experimentation workbench*: it is explicitly designed to support rapid evaluation of various research

hypotheses, easy creation of different setups in the exploration stage, reuse of research results, and support of reproducibility and documentation as described in [14]. It thus aims to support researchers in conducting experiments in a similarly broad way as a development workbench like Eclipse supports a developer.

While KernelHaven explicitly focuses on product line analysis and verification, it aims to support a wide range of tasks in this domain and due to its openness is closer to an open ecosystem than to a product line. As we discuss in Section 2, it currently supports (up to) three different input pipelines, which are fed into a processing step typically addressing analysis or verification activities. All individual steps in extraction and processing can be realized by different, independent plug-ins, creating a broad range of possible applications. There exist already a significant number of publicly available plug-ins. Some are actually wrapped components of well-known research tools like Undertaker [18] or TypeChef [11], while others have been developed as part of the KernelHaven-project<sup>1</sup>. Due to the integration framework that KernelHaven offers, it is easy to set up an experiment combining code analysis from Undertaker with a variability analysis using KConfigReader [12]. And it is also a matter of minutes to determine whether results differ, if as an alternative for code-analysis the Code Block extraction plug-in is used. As a practical consideration it is sometimes also important that as a result of this clear separation into plug-ins, the various tools may have different licenses and can still be used in combination as long as the user is allowed to use each individual part.

In its current form, especially given the currently available plug-ins, KernelHaven is well equipped for Linux-related analysis. However, we emphasize, that there is no architectural restriction in this regard. Rather, it can be used for any type of product line and any type of static analysis. The architecture is indeed so open that it can also be used itself as some form of post- or pre-processing in other scientific analysis. We are currently investigating such applications.

Besides the support for research, the well-defined packaging of the various plug-ins allows to apply these analyses more easily and flexibly in industry. We experienced this in some of our industrial cooperations, where the framework greatly enhanced the usability also for the industrial partners [8].

The remainder of this paper is structured as follows: Section 2 provides a short overview of the basic architecture of the tool. Section 3 then looks more in-depth into the current capabilities, in particular, the different available plug-ins. The intention is to provide a better understanding of the out-of-the-box capabilities of the tool as it is available right now. Section 4 reviews the approach for extensibility of the framework. This is particularly relevant to anyone aiming to integrate fundamentally new analysis approaches. Finally, in Section 5, we conclude.

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

SPLC '18, September 10–14, 2018, Gothenburg, Sweden

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5945-0/18/09.

<https://doi.org/10.1145/3236405.3236410>

<sup>1</sup><https://github.com/KernelHaven>

## 2 OVERVIEW

KernelHaven is designed to support a variety of similar experiments in the domain of static SPL analysis and verification [14]. This design enables rapid prototyping as well as detailed analyses by offering an open plug-in infrastructure implemented in Java. The core components of this infrastructure are three extraction pipelines, the data processing, and a pipeline configurator as illustrated in Figure 1 and described in this section.

Each **pipeline** extracts and provides information of a particular type of artifact typically considered in variability-based analyses:

- The code pipeline in the upper part of Figure 1 extracts information from code files resulting in a set of element trees. An element tree represents a single code file and provides information about the available code elements on different levels of abstraction. We provide more details on this data model in Section 3.
- The build pipeline in the middle of Figure 1 extracts and provides information from build files. The result of this extraction is a map of files and their presence conditions (PC in Figure 1) [6, 16]. These conditions define constraints, which must be satisfied to compile and link or, in general, build a specific (set of) file(s).
- The (variability) model pipeline in the lower part of Figure 1 translates information from variability model files into a list of features and propositional formulas. They represent the features and constraints, which define the planned products of the SPL.

The quality of the extracted information depends on the specific extractors defined for the individual pipelines. Extractors can be easily exchanged by adapting the configuration file described later in this section. This exchangeability allows, e.g., to use a fast but imprecise extractor for early experiments, which serve the improvement of the experiments core processes and algorithms, and to switch to a rather slow but detailed extractor for final results.

The information provided by one or multiple pipelines is input to the core **processes and algorithms** of the conducted experiment. As experiments can be of different nature, the particular processing of the input information depends on the desired results. Hence, this processing can be exchanged similar to the extractors. We discuss the already available processing variants in Section 3, ranging from static analysis and verification tasks to software metrics. A particular feature of KernelHaven processing plug-ins is that they can call other processing plug-ins to consume their results, like the configuration mismatch analysis described in Section 3. Hence, there can be a complete sub-pipeline for processing, too.

The **configuration file** in the lower left part of Figure 1 defines the particular setup of an experiment and, hence, a specific instance of KernelHaven. It consists of a set of parameters for preparing the infrastructure (input and output locations, etc.) as well as defining the required pipelines, the desired extractors, and processing plug-ins. The pipeline configurator reads these parameters to configure KernelHaven prior to its execution.

Besides these core components and the resulting flexibility in setting up different experiments, KernelHaven also offers features, which particularly address common research challenges [3, 4, 7]: comprehensive documentation and reproduction of experiments. KernelHaven supports automatic archiving of all relevant plug-ins, input, output, and intermediate data along with its core infrastructure. Further, the resulting archive contains the configuration file,

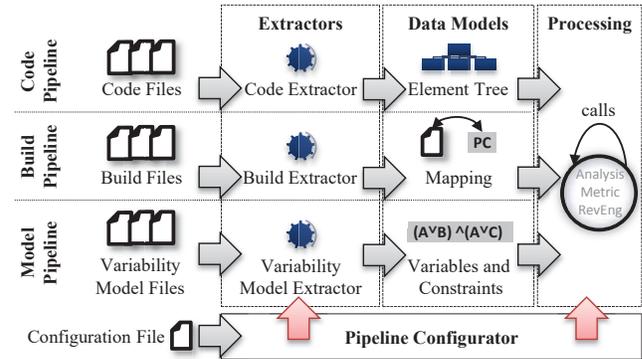


Figure 1: KernelHaven architecture

which defines the experimental setup and, hence enables reproduction of an entire experiment on a new machine. This only requires the installation of Java on an operating system, which the extractors of the experiment support. In the following section, we explicitly include these dependencies on specific operating systems as part of our discussion of existing KernelHaven capabilities.

## 3 CAPABILITIES

The flexibility in combining and exchanging extractors and processing plug-ins is one of KernelHaven’s core capabilities, which, in particular, aims at supporting different users to conduct their desired analysis. As a starting point, different plug-ins are publicly available, which can be used out of the box. In this section, we introduce them by, first, discussing available extractors, their purpose, and extracted information and, second, describing realized processing plug-ins and their possible combinations with the extractors.

KernelHaven provides seven different extractors as summarized in Table 1. Most of these extractors are **code extractors**. They rely on the code model illustrated in a simplified form in Figure 2 to provide code information of different granularity: ranging from code blocks to entire Abstract Syntax Trees (ASTs). This range of expressiveness is required to support a variety of external tools wrapped as KernelHaven plug-ins, like **Undertaker** [18], **srcML** [5], and **TypeChef** [11]. Except for Undertaker, which is typically available as native Linux binary, these extractors are executable on any operating system. Due to this limitation of Undertaker, we also provide a reimplementation of its algorithms in Java resulting in the operating-system-independent and faster **Code Block** extractor. Hence, both extractors extract conditional code blocks (right side of Figure 2) from C-code files defined as input in Table 1. In contrast, srcML and TypeChef provide ASTs of their input files, which results in a set of possibly nested `SyntaxElements` as illustrated on the left side of Figure 2. While the AST of the srcML extractor currently only provides a granularity of single line statements<sup>2</sup>, the TypeChef extractor enables variability-aware parsing including macro-expansion<sup>3</sup>, which results in significantly more details about code files.

The **KBuildMiner** extractor is the only publicly available **build extractor** so far. It wraps the corresponding tool [2], which

<sup>2</sup>This is due to the ongoing work on this extractor; the original srcML tool provides more details.

<sup>3</sup>TypeChef automatically considers header files (\*.h) without an explicit definition of those files as input.

**Table 1: Available KernelHaven extractors (April 2018)**

Name	Type	Input	OS
Undertaker <sup>1</sup>	Code	*.c/*.h/*S	Linux
Code Block <sup>2</sup>	Code	*.c/*.h/*S	Any
srcML <sup>1</sup>	Code	*.c/*.h	Any
TypeChef <sup>1</sup>	Code	*.c	Any
KBuildMiner <sup>1</sup>	Build	KBuild/Make	Any
KConfigReader <sup>1</sup>	VM	Kconfig	Linux
DIMACSVariabilityModel <sup>2</sup>	VM	DIMACS	Any

VM = Variability Model

<sup>1</sup> This extractor wraps an external tool not developed by the authors

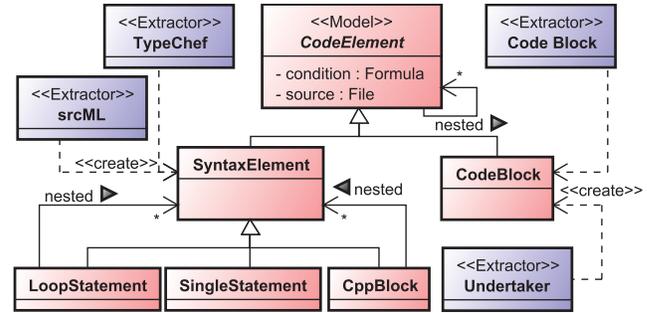
<sup>2</sup> This extractor is solely developed by the authors

takes KBuild- and Make-files as input and provides the mapping of files and presence conditions as described in Section 2. In particular, KBuild-files stem from the Linux kernel build system [13] and its specific variability realization technique, which uses `CONFIG_*` symbols in code and build files to reference configuration options defined in the variability model. Hence, this extractor explicitly searches for such references to build the mapping. Build extractors for other variability realization techniques can easily be implemented by exploiting the extensibility features described in Section 4.

The **KConfigReader** and the **DIMACSVariabilityModel** extractors represent two fundamentally different **variability model extractors**. While the former wraps the **KConfigReader** tool [12], which explicitly takes **Kconfig**-files [13] as input and uses build tools of the Linux kernel restricting its applicability to Linux only, the latter extracts propositional formulas in DIMACS format and converts them into the internal variability model representation of KernelHaven on any operating system. Similar to the **KBuildMiner** build extractor, **KConfigReader** is the extractor of choice if an experiment requires variability model information from the Linux kernel build system. The **DIMACSVariabilityModel** extractor can be used, for example, for any SPL using a feature model as this type of model can be easily translated into propositional formulas [1] and, hence, into DIMACS format.

KernelHaven allows any combination of the above extractors for a specific experiment. However, the actual processes and algorithms implemented in a processing plug-in may require a particular setup of the pipelines. Table 2 includes such dependencies for the available processing plug-ins as an example. The **UnDeadAnalysis** detects dead code blocks or missing features as described in [17]. Therefore, it requires the respective block information from code files, which multiple code extractors provide. Hence, one of them needs to be defined in the respective KernelHaven configuration file. Further, the analysis requires build information to determine whether the code file including the respective code block is part of the final product. The availability of a particular file and its inherent code blocks is typically controlled by configuration options defined in the variability model, which requires a corresponding extractor.

The **FeatureEffect** analysis requires a different setup of KernelHaven as it is a reverse engineering approach, which identifies variability constraints from code and build files [15]. Hence, it does not need a variability model extractor as well as the respective

**Figure 2: Simplified code model of KernelHaven**

pipeline. For this purpose, a user simply excludes the corresponding parameters from KernelHaven’s configuration file. Further, the **ConfigurationMismatches** analysis builds upon the same configuration as it uses the results of the feature effect analysis to check whether the implemented (effective) variability matches the variability model [9]. This concatenation of processing plug-ins results in processing or analysis pipelines, which we describe in detail in [14]. In essence, processing pipelines are defined by combining processing plug-ins in the order they consume their results in the configuration file without any further implementation effort.

A fundamentally different type of experiment is represented by **MetricHaven**. This processing plug-in provides currently more than 7,300 metric variations [10] to measure SPL artifact types in isolation as well as in combination. Similar to the **UnDeadAnalysis**, this plug-in requires all pipelines and, in particular, **srcML** as it provides the necessary implementation details.

The set of already available plug-ins shows the potential of KernelHaven to support various types of experiments ranging from verification and analyses to metrics. A user can conduct these experiments with different levels of details, e.g., code blocks or entire ASTs, depending on the configured extractors. Further, new plug-ins can be realized and combined with the existing ones to extend the current capabilities as described in the following section.

## 4 EXTENSIBILITY

The architecture of KernelHaven is designed to simplify the development of new extensions by means of separation of concerns and reuse. This allows researchers to focus on the realization of novel concepts while abstracting from technical details. We outline the three design concepts for achieving this simplicity in this section.

**Decoupled extractors and processing.** Independent data models decouple the extraction of relevant information and the processing of the extracted data. On the one side, this facilitates the reuse of already developed processing plug-ins for new artifact types as only a new extractor is required. On the other side, the realization of novel processing algorithms, e.g., a new code metric, may reuse existing extractors for already handled artifact types.

**Reuse of processing algorithms.** KernelHaven processing plug-ins do not differentiate between the output of extractors or other processing plug-ins. Thus, new processing plug-ins may reuse existing and mature processing plug-ins to form a processing or analysis pipeline, like the concatenation of the feature effect and

**Table 2: Available KernelHaven processing plug-ins and possible extractor combinations (July 2018)**

Name	Goal	Code Extractor	Build Extractor	VM Extractor
UnDeadAnalysis <sup>1</sup>	Detection of dead code blocks (never part of any product) or missing features	Undertaker, Code Block, srcML, or TypeChef	KBuildMiner	KConfigReader
ConfigurationMismatches <sup>2</sup>	Detection of inconsistencies between modeled and effective variability	<i>Processes the results of the FeatureEffect analysis and, hence, does not directly use a particular extractor</i>		
FeatureEffect <sup>1</sup>	Reverse engineering of variability constraints from code and build files to improve an existing or create a new VM	Undertaker, Code Block, srcML, or TypeChef	KBuildMiner	<i>not required</i>
MetricHaven <sup>1,2</sup>	More than 7,300 metrics and their variants for measuring SPL properties	srcML (TypeChef possible, but not realized yet)	KBuildMiner	KConfigReader

VM = Variability Model; <sup>1</sup> This processing is a reimplementaion in KernelHaven of existing work; <sup>2</sup> This processing is developed by the authors

configuration mismatches analyses in Section 3. In this way, developers can realize complex processing algorithms incrementally. Individual KernelHaven plug-ins may represent these increments, which can be used both in isolation as well as in combination with other processing algorithms and, hence, avoid extra effort for re-implementing required processing capabilities.

**Abstraction from technical details.** The KernelHaven infrastructure separates conceptual work from technical implementation. This way, researchers do not need to consider aspects like multi-threading or saving of results as the infrastructure takes care of that. This means that KernelHaven executes by default all extraction pipelines (cf. Figure 1) and all processing plug-ins in parallel. For instance, the configuration mismatch analysis may start to validate that the variability model covers the detected code dependencies of the feature effect analysis, while this analysis still detects further code dependencies. However, developers can force all processing components (extractors or processing plug-ins) to complete before a new component starts. To save the results, a developer only needs to annotate data classes and KernelHaven takes care of storing the results. Thus, developers usually neither need to specify any output formats nor to take care of stream handling.

## 5 CONCLUSION

We presented KernelHaven as an experimentation workbench to support different types of users in their goal to conduct their individual experiments in the domain of static Software Product Line (SPL) analyses and verification. Users benefit from the available plug-ins, which wrap some of the most prominent analysis tools in this domain and, hence, simplify their application even in new experiments. KernelHaven addresses researchers by explicitly supporting the documentation and reproduction of experiments. Developers take advantage of the rather simple extensibility due to reusability of existing plug-ins and abstraction from technical details provided by the core infrastructure. We believe that KernelHaven is also applicable in teaching SPL engineering as it significantly reduces time and effort students need to setup a particular extraction, analysis, metric, etc. Rather, it is simply a matter of extracting an archive and executing KernelHaven with the desired configuration.

## ACKNOWLEDGMENTS

This work is partially supported by the ITEA3 project REVaMP<sup>2</sup>, funded by the BMBF (German Ministry of Research and Education) under grant 01IS16042H. Any opinions expressed herein are solely by the authors and not by the BMBF.

We thank the following contributors: Moritz Flöter, Adam Krafczyk, Alice Schwarz, Kevin Stahr, Johannes Ude, Manuel Nedde, Malek Boukhari, and Marvin Forstreuter.

## REFERENCES

- [1] Don Batory. 2005. Feature Models, Grammars, and Propositional Formulas. In *9th International Conference on Software Product Lines*. Springer-Verlag, Berlin, Heidelberg, 7–20.
- [2] Thorsten Berger and Christian Kästner. 2016. KBuildMiner. (2016). <https://github.com/ckaestne/KBuildMiner> Last visited: 18.04.2018.
- [3] Carl Boettiger. 2015. An Introduction to Docker for Reproducible Research. *Operating Systems Review – Special Issue on Repeatability and Sharing of Experimental Artifacts* 49, 1 (2015), 71–79.
- [4] Jürgen Cito and Harald C. Gall. 2016. Using Docker Containers to Improve Reproducibility in Software Engineering Research. In *38th International Conference on Software Engineering Companion*. ACM, New York, NY, USA, 906–907.
- [5] srcML Community. 2018. srcML. (2018). <http://www.srcml.org/> Last visited: 18.04.2018.
- [6] Christian Dietrich, Reinhard Tartler, Wolfgang Schröder-Preikschat, and Daniel Lohmann. 2012. A Robust Approach for Variability Extraction from the Linux Build System. In *16th International Software Product Line Conference*, Vol. 1. ACM, New York, NY, USA, 21–30.
- [7] Holger Eichelberger, Aike Sass, and Klaus Schmid. 2016. From Reproducibility Problems to Improvements: A journey. *Softwaretechnik-Trends* 36, 4 (2016), 43–45.
- [8] Sascha El-Sharkawy, Saura Jyoti Dhar, Adam Krafczyk, Slawomir Duszynski, Tobias Beichter, and Klaus Schmid. 2018. Reverse Engineering Variability in an Industrial Product Line: Observations and Lessons Learned. In *Proceedings of the 22nd International Systems and Software Product Line Conference (SPLC '18)*. ACM, New York, NY, USA, 11. accepted.
- [9] Sascha El-Sharkawy, Adam Krafczyk, and Klaus Schmid. 2017. An Empirical Study of Configuration Mismatches in Linux. In *21st International Systems and Software Product Line Conference*, Vol. A. ACM, New York, NY, USA, 19–28.
- [10] Sascha El-Sharkawy, Nozomi Yamagishi-Eichler, and Klaus Schmid. 2018. *Implementation Metrics for Software Product Lines – A Systematic Literature Review*. Technical Report 1/2018, SSE 1/18/E. Institute of Computer Science, University of Hildesheim. Available at <https://sse.uni-hildesheim.de/en/research/projects/revamp2/spl-metrics/>.
- [11] Christian Kästner. 2013. TypeChef. (2013). <https://ckaestne.github.io/TypeChef/> Last visited: 18.14.2018.
- [12] Christian Kästner. 2016. KConfigReader. (2016). <https://github.com/ckaestne/kconfigreader> Last visited: 18.04.2018.
- [13] KBuild. 2018. KBuild - The Linux Kernel Build System. (2018). <https://www.kernel.org/doc/Documentation/kbuild/> Last visited: 18.04.2018.
- [14] Christian Kröher, Sascha El-Sharkawy, and Klaus Schmid. 2018. KernelHaven - An Experimentation Workbench for Analyzing Software Product Lines. In *40th International Conference on Software Engineering: Companion Proceedings*. ACM, New York, NY, USA, 73–76.
- [15] Sarah Nadi, Thorsten Berger, Christian Kästner, and Krzysztof Czarnecki. 2015. Where do Configuration Constraints Stem From? An Extraction Approach and an Empirical Study. *IEEE Transactions on Software Engineering* 41, 8 (2015), 820–841.
- [16] Sarah Nadi and Ric Holt. 2012. Mining Kbuild to detect variability anomalies in Linux. In *16th European Conference on Software Maintenance and Reengineering*. IEEE Computer Society, Washington, DC, USA, 107–116.
- [17] Reinhard Tartler, Daniel Lohmann, Julio Sincero, and Wolfgang Schröder-Preikschat. 2011. Feature Consistency in Compile-time-configurable System Software: Facing the Linux 10,000 Feature Problem. In *6th Conference on Computer Systems*. ACM, New York, NY, USA, 47–60.
- [18] VAMOS/CADOS Team. 2015. Undertaker. (2015). <https://vamos.informatik.uni-erlangen.de/trac/undertaker> Last visited: 18.04.2018.

## A DEMONSTRATION AND CASE STUDIES

The KernelHaven demonstration starts with downloading a bundled release from the main infrastructure website<sup>4</sup>. A bundle contains everything needed to conduct available analyses out of the box. Figure 3 shows an example of the extracted content of such a bundle, which we briefly describe to understand the default structure of the tool and the purpose of the included files and directories.

Name	Type
cache	Folder
log	Folder
output	Folder
plugins	Folder
res	Folder
KernelHaven_withsource.jar	Archive
config_template.properties	Text
dead_code_in_Linux_w_CodeBlock.properties	Text
dead_code_in_Linux_w_TypeChef.properties	Text
single_metric.properties	Text

Figure 3: KernelHaven bundle content (example)

In Figure 3 the bundle contains the main infrastructure (KernelHaven\_withsource.jar), the extraction and processing plug-ins (plugins directory), and a set of configuration files (\*.properties) defining a particular setup of the tool. These configurations use the included set of empty directories, e.g., for saving cached information (cache), log files (log), and the final analysis results (output). Hence, we can start the tool, for example to collect metrics, on a command line prompt<sup>5</sup> by calling the main infrastructure with the metrics configuration file as single parameter:

```
java -jar KernelHaven_withsources.jar
↳ single_metric.properties
```

Listing 1 shows an excerpt of the single\_metric.properties file, which contains a set of information blocks that we explain next. The **directories** in Lines 2–6 specify the mandatory directories, like the location of the software product line to analyze (source\_tree), the location to store the results of the analysis (output\_dir), or the location of the extractors and analyses (plugins\_dir). While a bundle already contains those directories, the respective parameters can be used to specify custom ones. Lines 9–12 specify the amount of **logging** information (log.level), which the infrastructure prints to a file (log.file) in the specified directory (log.dir) and, optionally, to the console (log.console). The **code model parameters** in Lines 15–19 define the availability of the code pipeline in general and, in particular, the code extractor (code.extractor.class) via its fully qualified main class name. Further, the parameters define extracting information from all files in the source\_tree matching a given regular

```

1  ### Directories ###
2  resource_dir = res/
3  output_dir = output/
4  plugins_dir = plugins/
5  cache_dir = cache/
6  source_tree = <path/to/SPL>
7
8  ### Logging ###
9  log.dir = log/
10 log.console = false
11 log.file = true
12 log.level = INFO
13
14 ### Code Model Parameters ###
15 code.provider.timeout = 0
16 code.provider.cache.write = false
17 code.provider.cache.read = false
18 code.extractor.class = net.ssehub.kernel_haven.
19   ↳srcml.SrcMLExtractor
20 code.extractor.file_regex = .*\.c
21
22 ### Build Model Parameters ###
23 build.provider.timeout = 0
24 build.provider.cache.write = true
25 build.provider.cache.read = true
26 build.extractor.class = net.ssehub.kernel_haven
27   ↳kbuildminer.KbuildMinerExtractor
28
29 ### Variability Model Parameters ###
30 variability.provider.timeout = 0
31 variability.provider.cache.write = true
32 variability.provider.cache.read = true
33 variability.extractor.class = net.ssehub.
34   ↳kernel_haven.kconfigreader.
35   ↳KconfigReaderExtractor
36
37 ### Analysis Parameters ###
38 analysis.class = net.ssehub.kernel_haven.
39   ↳metric_haven.metric_components.
40   ↳MetricsRunner
41 analysis.metrics_runner.metrics_class = <class
42   ↳name of metric to execute>
43 metrics.round_results = true
44 analysis.output.type = xlsx

```

Listing 1: single\_metric.properties file (excerpt), place holder for local settings are marked as < >.

expression (code.extractor.file\_regex), which only identifies C-code class files. While a specific timeout for automatically terminating the extractor (code.provider.timeout) is not defined, caching intermediate code data (code.provider.cache.write and code.provider.cache.read) is disabled as it is currently not supported by the SrcMLExtractor. Similar to the code model parameters, Lines 22–25 define the **build model parameters**, like the build extractor (build.extractor.class), the timeout and the caching capabilities. The **variability model parameters** in Lines 28–31 specify the variability model extractor (variability.extractor.class) for the corresponding pipeline. Similar to the previous pipeline no timeout but caching is enabled. In order to exclude a pipeline from an experiment, only the respective block of parameters has to be deleted from the configuration.

<sup>4</sup><https://github.com/KernelHaven/KernelHaven#downloads>

<sup>5</sup>KernelHaven is a pure command line tool; it does not provide a graphical user interface

**Table 3: DLoC-metric results for Linux 4.15 (excerpt).**

Source	Line	Function	DLoC	LoF	PLoF
a20.c	100	enable_a20_kbc	7.0	0.0	0.0
build.c	248	efi_stub_defaults	3.0	3.0	1.0
build.c	282	efi_stub_defaults	0.0	0.0	0.0
cpu.c	37	show_cap_strs	21.0	20.0	0.95
cpu.c	75	validate_cpu	16.0	0.0	0.0
eboot.c	748	setup_e820	63.0	1.0	0.01
eboot.c	877	exit_boot_func	23.0	2.0	0.08
eboot.c	964	efi_main	116.0	0.0	0.0
misc.c	183	handle_relocations	33.0	14.0	0.42

The last block of information in Lines 34–37 of Listing 1 specifies the actual processing algorithm to use for this particular experiment by the fully qualified class name (`analysis.class`). In this example, the `MetricsRunner` is used, to load a metrics analysis and to execute all variations of the specified metric. This processing algorithm requires an additional parameter `analysis.metrics_runner.metrics_class` to select the particular metric algorithm to be executed. In the example, we choose `net.ssehub.kernel_haven.metric_haven.metric_components.DLoC` to compute three variations of the *Delivered Lines of Code*-metric for each code function of the analyzed product line. The last two parameters limit the output to two decimal places and specify that the results should be saved as a CSV-file. The KernelHaven bundles are shipped with a full documentation of available configuration settings (`config_template.properties`).

Table 3 shows an excerpt of the analysis results for the three variations of the DLoC-metric. The first three columns list the measured item: path of the measured C-file (folder names were removed for the sake of brevity), the line number, and the name of the measured function. The last three columns show the measuring results for the DLoC, the Lines of Feature code (LoF), and the fraction of feature code ( $PLoF = \frac{LoF}{DLoC}$ ) [10]. The results show two separate implementations of the function `efi_stub_defaults` at the Lines 248 and 282 of the file `build.c`. The first implementation contains three statements completely surrounded by inner `#ifdef`-blocks (DLoC and LoF have the same value and, thus, the PLoF value is 100%). The second implementation is an empty function definition for the case that the feature “EFI\_STUB” was deselected.

Another supported analysis is the detection of code dependencies, which was also applied at the Bosch PS-EC product line [8]. For confidentiality reasons, we cannot show details specific for Bosch, but the general part of the analysis is publicly available and may also be applied to open source system like Linux. However, this analysis requires a different setup. The most important changes are presented in Listing 2. In Line 2 the code extractor is exchanged to a faster and more lightweight code block extractor. This extractor does not extract C-code statements, which are not required for the analysis. The extractor considers only the relevant C-preprocessor statements and their conditions. Further, it provides support to handle macro definitions, which is necessary to handle the variability information of Linux correctly (cf. Line 3). Line 4 specifies that unparsable conditions shall be treated as `true`. This allows the extractor to continue parsing a file if a condition cannot be handled rather than skipping the complete file.

```

1  ### Code Model Parameters ###
2  code_extractor.class = block_extractor.
   ↪ CodeBlockExtractor
3  code_extractor.handle_linux_macros = true
4  code_extractor.invalid_condition = true
5
6  ### Analysis Parameters ###
7  analysis.class = analysis.
   ↪ ConfiguredPipelineAnalysis
8  analysis.pipeline = \
9    fe_analysis.fes.FeatureEffectFinder( \
10     fe_analysis.pcs.PcFinder( \
11      cmComponent(), bmComponent() \
12     ) \
13    )
14 analysis.output.type = xlsx
15 analysis.output.intermediate_results = PcFinder
16 analysis.consider_vm_vars_only = true

```

**Listing 2: Modified properties file to analyze code dependencies (delta).**

Lines 7–16 specify the analysis. First, `net.ssehub.kernel_haven.analysis.ConfiguredPipelineAnalysis` specifies a wiring of a (complex) processing pipeline via an embedded DSL (cf. Lines 8–13). This pipeline consists of two processing algorithms. The first algorithm gathers for each feature all presence conditions in which it is used in (`PcFinder`). The second algorithm combines all presence conditions to compute a condition under which the selection of a feature has an effect on the product derivation (`FeatureEffectFinder`). By default, KernelHaven stores only the results of the last processing algorithm. Lines 14 and 15 specify that both results should be saved into separate sheets of an Excel workbook (\*.xlsx). The last line is used to limit the results only to features specified in the variability model, otherwise all constants used in C-preprocessor statements will be collected. `cmComponent`, `bmComponent`, and `vmComponent` (not used in Listing 2) are pre-defined elements of KernelHaven’s DSL to specify that a processing algorithm shall take the information of the code, build, or variability model extractor as input. Please note that we removed the prefix `net.ssehub.kernel_haven` from all class names in Listing 2 to improve readability.

Table 4 presents an excerpt of the results of the feature effect analysis. The first column lists all features found in code files or in Make-files. The second column shows the precondition, which must be fulfilled in order that the selection of the feature has an effect of the product derivation. For instance, the selection of `CONFIG_ACPI_DEBUG` and `CONFIG_ACPI_DEBUGGER` change the code only if also `CONFIG_ACPI` is selected. This may happen if both features are always nested below `CONFIG_ACPI`, while `CONFIG_ACPI` may also be used as top-level statement.

**Table 4: Feature Effect results for Linux 4.15 (excerpt).**

Feature	Relevancy Condition
<code>CONFIG_ACPI</code>	<code>TRUE</code>
<code>CONFIG_ACPI_DEBUG</code>	<code>CONFIG_ACPI</code>
<code>CONFIG_ACPI_DEBUGGER</code>	<code>CONFIG_ACPI</code>
<code>CONFIG_ACPI_I2C_OPREGION</code>	<code>CONFIG_ACPI</code> $\wedge$ ( <code>CONFIG_I2C_MODULE</code> $\vee$ <code>CONFIG_I2C</code> )