SMOKE2.0 Whitebox Anonymization of Intellectual Property in Simulink with Structure Preservation

Alexander Boll, Manuel Ohrndorf and Timo Kehrer

Software Engineering Group, University of Bern, Switzerland.

*Corresponding author(s). E-mail(s): alexander.boll@unibe.ch;
Contributing authors: manuel.ohrndorf@unibe.ch;
timo.kehrer@unibe.ch;

Abstract

Simulink is widely used across various industries to model and simulate cyber-physical systems. Most industry-built models contain sensitive information, which prevents companies from sharing models with interested third parties, such as researchers or collaborating companies. However, advancing model-based engineering research requires access to such models – either to derive empirical insights or to evaluate new tools. While initiatives to replace industry-built models with open-source alternatives exist, they offer only a limited remedy. In this work, we present a novel approach with SMOKE, a Simulink anonymization tool designed to selectively remove sensitive information within models. This allows companies to share relevant parts of their models with researchers or other third parties while safeguarding all sensitive information. SMOKE's white-box design preserves the model's original format and structure, ensuring that meaningful insights remain accessible. We evaluated the tool on an extensive set of open-source models and found it successfully removes sensitive information, while preserving model structure. A video demonstration of SMOKE is available online at youtu.be/0i42BzgJAUA.

Keywords: Data Masking, Sanitization, Obfuscation, Anonymization, Abstraction, Filtering, Open Science, Simulink

1 Introduction

Across various industries, Simulink is a widely-used tool to design, implement and simulate cyber-physical systems [1–4]. The popularity of Simulink also gives rise to a considerable research interest, e.g., into better understanding Simulink models and their evolution [5, 6]. However, within the research community, it is well known that companies often refrain from sharing their Simulink models to protect any sensitive information within them [7–9]. In some cases, industry partners share their models under severe limitations, such as non-disclosure agreements, which strictly prohibit the publication of model files or any visual representation of the models. Companies may also restrict access to models to company computers and locations. This practice creates significant challenges for the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [10] in Simulink research and often leaves researchers without useful study subjects [6]. The modeling research community has begun addressing this issue by developing corpora of open-source Simulink models [1, 6, 11] and managing these corpora [12, 13]. These open-source models can serve as substitutes for proprietary models in research. In general, however, open-source models are much smaller than industry models, and are often no adequate substitutes [1, 7].

In this work, we present SMOKE: **S**imulink **M**odel **O**bfuscator **K**eeping structur**E**, an extensible, open-source tool for selective anonymization of Simulink models through removal of sensitive information. Our goal is to anonymize models by removing sensitive information, i.e., intellectual property or any other information that must not be shared with others, while preserving their general usefulness, especially for research. Thus, a model being anonymized using SMOKE is still a valid Simulink model whose structure is isomorphic to the original one, yet exposing a different visual appearance and behavior, depending on the desired degree of anonymization. The modifications of visual appearance and behavior – the anonymizations – are both implemented through unidirectional model transformations [14], i.e., they are non-reversible without logging or domain knowledge. The first class of modifications are layout obfuscations, which preserve behavior but reduce understandability. The second class comprises sanitization techniques, which remove sensitive data or typically by breaking or altering model behavior. This is a concept we adapt from database sanitization, originally developed for relational databases [15]. The concrete transformations to be applied can be selected by the user, depending on the desired degree of anonymization. SMOKE supports both interactive and non-interactive selections of transformations, which are then composed to an executable transformation workflow.

SMOKE addresses two use cases which are of interest for researchers. First, it simplifies obtaining approval for publishing visual representations of Simulink models by selectively removing layout information. A precursor to SMOKE was already used to obfuscate industry models, supporting the publication of screenshots in scientific articles [16, 17]. Second, companies may permit further study or use of sanitized models where sensitive data or behavior is altered or removed. As opposed to traditional obfuscators concealing a program's [18] or model's [19] entire behavior within an uninspectable and immutable virtual black box [20], our white-box anonymization yields a native Simulink model open to further inspection. This means that the model can be opened with the standard MATLAB/Simulink editor or other tools working with

Simulink models. A particular feature of SMOKE is that it preserves the structure of the original models. Even if highly anonymized, the obtained 'structure-only' models still remain valuable study subjects for many research interests. To better understand various aspects of a model or its evolution, empirical research on Simulink models often focuses on metrics related to model structure [1, 11, 21, 22], or relies on third-party analysis tools such as clone detectors, differencing tools, slicers, or variant and information flow analyzers that require an intact model structure to function [23–27].

In addition to addressing researchers' interest, SMOKE can also be employed by industry partners themselves. In software value chains, models are developed in co-engineering fashion, and such collaborations only work if the parties have access to the models [28]. However, companies may still hesitate to grant full access to others or may need to comply with various competition laws [29]. Moreover, (partially) anonymized models can be indexed by model search engines, as they can search models by basic metric structure [12, 30, 31]. This way, interested parties may find models according to basic information, and can get into contact with the owners to agree on the terms of full access.

We give an overview of SMOKE's anonymization capabilities, and showcase the effect of interactively applying a subset of those on a realistic example. Moreover, we report about our evaluation applying SMOKE's full anonymization capabilities on thousands of open-source models taken from the SLNET dataset [11], with a particular focus on evaluating its general applicability and functional correctness. While SMOKE focuses on Simulink models, we argue that other modeling ecosystems – such as UML, Modelica, or SysML – could also benefit from anonymization, as demonstrated by Munk et al. [32], who obfuscate SysML models by Bosch.

Our tool SMOKE, written in MATLAB, along with its documentation, and all artifacts from our experimental evaluation, is open-source and available at github.com/lanpirot/SMOKE. A brief video demonstration of SMOKE is available online at youtu.be/0i42BzgJAUA.

This paper is a substantial extension of the short paper "SMOKE: Simulink Model Obfuscator Keeping Structure" by Boll et al. [33], demonstrated at the MODELS'24 demo track. We have built upon our prior work through the following improvements:

- We updated SMOKE, extending it with new anonymization functionality. Notably, it is now applicable to library models, and anonymizations can be performed on user-selected scopes of the model.
- We repeated our entire evaluation, incorporating the aforementioned new functionality and now including library models. We greatly improved the evaluation of the behavior (non-)preservation of SMOKE.
- We further added discussions on SMOKE's coverage and threats to validity.
- Leveraging the expanded space available compared to the original short paper, we have meticulously rewritten and significantly enhanced all sections, notably Sections 2 to 5, from scratch. This revision includes extensive additional detail and thorough explanations that were not feasible to incorporate previously.

2 Background on Simulink and Obfuscation Techniques

In this section, we first lay the foundational background knowledge of Simulink that is needed for this paper. We then elaborate on obfuscation and sanitization techniques, and how they can be applied in Simulink. Note that we do not discuss an exhaustive list of obfuscation and sanitization techniques, but focus on those that are relevant to our context, i.e., the ones that are applicable to Simulink models, while keeping structural integrity.

2.1 Simulink

Simulink [34] is a modeling language and versatile integrated development environment developed by Mathworks.[1] It is based on and integrated into the MATLAB IDE and can be used for abstract modeling, implementation of functionality, simulation, and code generation. As its graphical modeling language is intuitive to use and understand, it is often used as a low-code development platform [35] in non-programmer domains like automotive, aerospace, medical and other technical industry domains [2, 36–40].

A Simulink model consists of *Blocks*, which can be connected by *Lines* – both placed on a modeling canvas, cf. Figure 1.[2] Lines transport values from Block OutPort(s) to Block InPort(s). Blocks transform their input from their InPort(s) to an output which they emit via their OutPort(s). To handle growing and complex models, users have the choice of multiple partitioning mechanisms, such as the *Subsystem* Block. Subsystem Blocks may nest Blocks and Lines, allowing developers to hierarchically structure their models. Subsystem Blocks can also nest other Subsystem Blocks and with this, developers can hierarchically structure their models.

The Simulink IDE offers different model *views* via the Subsystem Blocks. In a view, only the direct content of the currently selected Subsystem is visible, while Subsystems hide their nested implementation details from the outer view. While Figure 1 depicts the root view of a model, the Subsystems' inner views of the Subsystems 'Kelvin to Celsius' (lower right of Figure 1) and 'Heat Sources' (bottom middle of Figure 1) is given in other views, shown in Figure 2. The root Subsystem, i.e., the model itself, may also have InPorts and OutPorts, which act as inputs and outputs of the model. Other inputs and outputs may be read from or written to the file system, workspace variables, or user input. In many models, InPorts are driven by physical sensors and outputs drive actuators. Additionally, there are also models without input or output, e.g., library models that contain useful sets of custom Blocks, or functionality used in other models where they can be imported and reused. While Simulink already offers an extensive set of different Block types, users can use this mechanism to define their own (complex and reusable) Block behaviors. In addition, model elements come with their own set of *Parameters*[3] that can be adjusted individually for each instance of an

[1] mathworks.com (visited on 2025-11-7)

[2] Example Simulink model from mathworks.com/help/simulink/slref/simulating-automatic-climate-control-systems.html (visited on 2025-11-7)

[3] In EMF models, the counterpart of Parameters are called Attributes [41]. The Parameters of a Block define how it exactly behaves during simulation, and are not external output/input like parameters in textual programming languages.

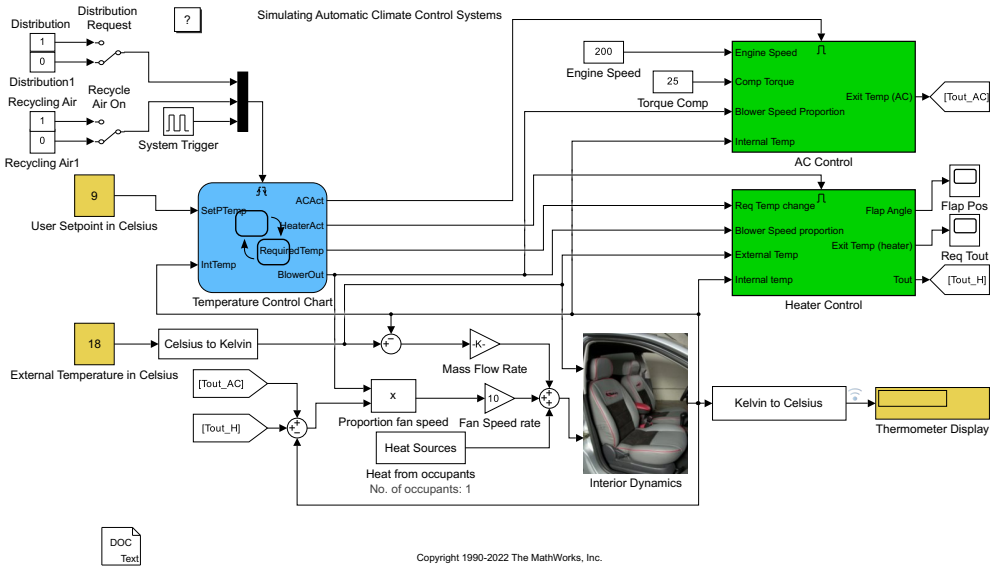
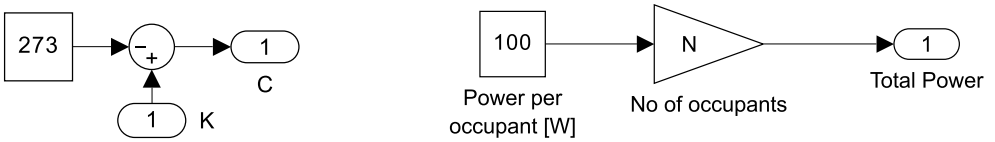


Figure 1: An exemplary Simulink model showing various kinds of Blocks, connected by Lines.



(a) (b)

Figure 2: The inner view of two Simulink Subsystems of the model in Figure 1: the Subsystem in Figure 2a transforms temperature from Kelvin to Celsius: $C = K - 273$, the Subsystem in Figure 2b computes the total heat generation of N occupants: $TotalPower = 100 \cdot N$.

element, cf. Figure 8a. The variety of Block types and their individual Parameters offer a comprehensive range of design options supporting both static modeling (memory-less systems), and dynamic modeling (time-evolving systems with states).

We give a highly simplified metamodel of Simulink depicted in Figure 3. More detailed, but also unofficial metamodels are available in Sánchez et al. [41], or from the Massif group.[4] We included the core classes and attributes and left out details such as Busses or special Block or Port types, as they are not necessary to grasp the main ideas of this work. A SimulinkModel serves as the root container, comprising Blocks, which can be nested via Subsystems to form complex hierarchies. Each Block exposes Ports – specifically InPort and OutPort – to define its interfaces. An OutPort can

[4] https://github.com/viatra/massif (visited on 2025-12-01)

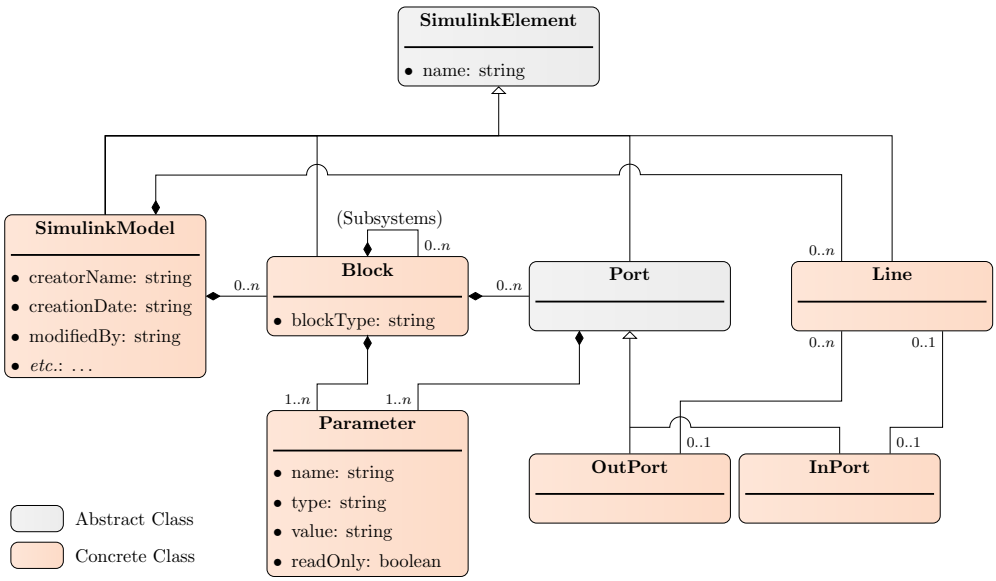


Figure 3: Simplified metamodel of Simulink models.

connect to multiple Lines, which enables a one-to-many flow from the OutPort of a Block to other Blocks. Meanwhile, an InPort connects maximally one Line, enforcing a unique source and direction of the incoming dataflow of a Block. Both Blocks and Ports are configurable via Parameters, which encode their semantic properties (e.g., Block behavior or Port attributes).

In this work, we define the model's *structure* as the typed graph of Blocks connected by Lines at their Ports (cf. Figure 3).[5] When we speak of *structure-preserving* anonymizations or transformations, we are thus looking for obfuscations and sanitizations that do not alter this underlying graph. This means that two models have the same structure if their graphs of Blocks, Ports, and Lines are isomorphic, and only Parameters or class attributes are altered.

2.2 Obfuscation Techniques

Obfuscations are transformations that change aspects of a program while preserving its behavior – that is, the numerical outputs, signal values, and execution results of the model remain identical. Obfuscating a program – in our case 'a program' is to be understood as 'a model' – increases the difficulty of understanding or accessing its purpose or logic [42]. Using obfuscations, one can thus remove sensitive information of a program and make it harder to reuse or repurpose it.

Collberg et al. [18] identify three basic classes of obfuscation: (1) *layout obfuscation*, (2) *data obfuscation*, and (3) *control obfuscation*. (1) Layout obfuscations are simple

[5] Simulink's metamodel contains additional constructs beyond our definition, such as invisible connections between blocks like Goto/From. Although SMOKE can obfuscate or break these, we exclude them from our notion of *structure*.

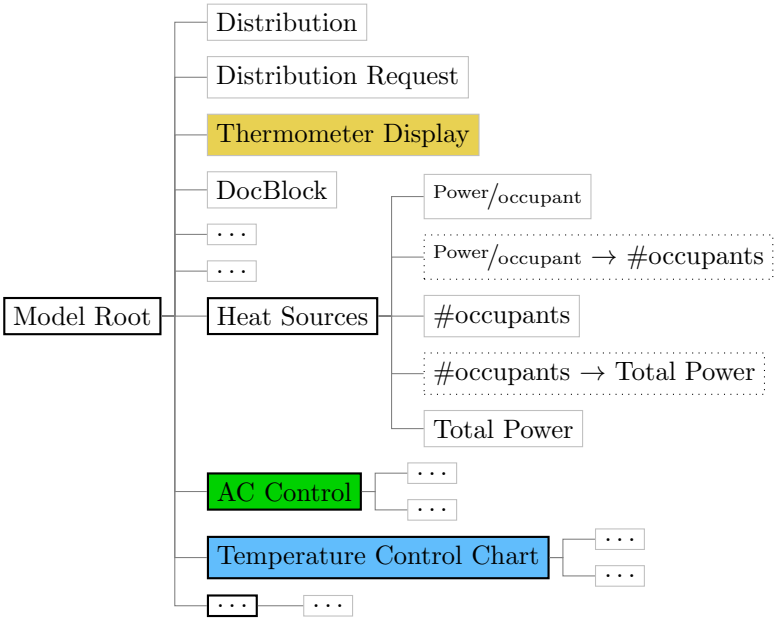


Figure 4: The partial structure of the model of Figure 1. All structure elements are part of this tree, hierarchically organized by the Subsystems. The root Subsystem is on the very left. Each Subsystem's child is a Block or Line within it. Only the content of the Subsystem 'Heat Sources' (cf. Figure 2b) is fully shown, all other content is only hinted at.

transformations that adapt documentation, names, or formatting, usually by removing them, resetting them to default values, or replacing them by arbitrary values. As Simulink is a graphical language, it offers many more layout obfuscations than classical textual languages, e.g., Blocks and Lines can be repositioned, resized, or recolored, fonts can be changed, and media elements can be used. (2) Data obfuscations alter the storage, encoding, or ordering of data. A different encoding, or even encryption, makes it difficult to interpret the data that is being processed in the program. Additional data abstraction transformations can be splitting up related data, or bunching up data that is unrelated. Simulink offers multiple data types, like `int8, double, boolean`,[6] with which one could obfuscate data encoding. Additionally, transformations on data abstraction are possible by using BUS-Blocks, that bundle up data from multiple Lines into one. (3) Control obfuscation manipulates the abstraction layers, adds useless conditions, or dead code. Obfuscating the abstraction layers is similar to the obfuscation of data abstraction: the removal of meaningful hierarchy layers or the addition of extraneous ones. In Simulink this can be achieved, e.g., by resolving Subsystems or adding new arbitrary ones into the model.

[6] It is further possible to create new, custom data types.

2.3 Sanitization Techniques

While obfuscation does not change the behavior of a program, *sanitization* does. In this work, we generalize the concept of *data sanitization* [15, 43] to our purpose of *model sanitization*: data sanitization is applied on sensitive data in relational databases by selectively removing parts of the database, while valuable insights into the rest of the data still remain possible. Data sanitization is performed so that the protected database can be shared with others, without risking the sensitive data being leaked.

In our case, we want to preserve the structure of a model, so that at least basic insights into the model are preserved, while the model behavior or data may be removed, changed or even completely broken. This means, the model should still be syntactically correct after all sanitization transformations, and still be loadable and inspectable in the IDE. On the other hand, a completely sanitized model should not show the same behavior, which means it either does not compile anymore, the simulation crashes or stops prematurely, or, given the same input, the output of the model changes. All of these conditions ensure that the original model can not be reverse engineered by simply observing the model's input/output behavior. Sanitizing a certain model element either removes or breaks all model behavior that depends on it. Similarly, the data sanitization transformations remove sensitive data or reset it to default values.

As the structure of the model shall remain stable, SMOKE's sanitizations change Parameters (see Figure 3). Such changes either result in altered behavior of Block calculations and thus overall model behavior, or misconfigurations that lead to non-compilability or simulation crashes.

3 Related Work

3.1 Simulink-specific Obfuscation

While obfuscation in traditional text-based programming languages is a well-established discipline with decades of research [18, 42], we found only limited prior work on the obfuscation of Simulink models.

Simulink itself features the Protected Model Creator,[7] a versatile tool that transforms Simulink models into black boxes of another file format, or white boxes that are not editable. However, these immutable white boxes do not anonymize at all. Other built-in options to create black box versions are so-called S-Functions or static libraries.[8]

A subset of SMOKE's layout obfuscations could be found in a tool by Ohashi[9] and the Obfuscate-Model tool by Jaskolka et al.[10] The former, however, is unmaintained and broken since at least 2017, according to the tool's reviews on MATLAB FileExchange and confirmed by us with MATLABR2025b. The latter is more mature and has been utilized for obfuscation in an industry-research partnership [16, 17]. Though it

[7] https://www.mathworks.com/help/rtw/ref/protectedmodelcreator.html (visited on 2025-11-7)
[8] https://www.mathworks.com/matlabcentral/answers/91537-how-do-i-protect-the-ip-of-my-simulink-model-when-sharing-it-with-others-who-may-include-it-in-thei (visited on 2025-11-7)
[9] https://www.mathworks.com/matlabcentral/fileexchange/54359-model-obfuscation-tool (visited on 2025-11-7)
[10] https://github.com/McSCert/Obfuscate-Model (visited on 2025-11-7)

supports only layout obfuscations, we built SMOKE based on a fork of the Obfuscate-Model tool, extending it with new capabilities, refined prior capabilities, and fixed bugs. For a detailed list of all changes and improvements, see Section 4.2.1. Moreover, we evaluated SMOKE's functionality on a large model set.

3.2 Obfuscating Other Model Types

Obfuscation approaches are also suggested in various other domains, such as conceptual models, business process models, Functional Mockup Interfaces, or Machine Learning models.

Fill [44] discusses obfuscating conceptual models with the intent of making them shareable, but his transformation breaks the structure by splitting models into multiple parts. Other transformations he suggested break the model's syntax. Nacer et al. [45] also explore business process model obfuscation by fragmenting models.

A further concept of model sharing in a co-engineering scenario with multiple roles (see Martinez et al. [46]) is explored in the work of Weber et al. [47]. In a first step, the model is split into sub-modules and their interfaces. Each sub-module is encrypted and only decryptable by authorized parties – effectively fragmenting the model. In our tool, selective aspects of the model (e.g., whole sub-modules or only sensitive properties within the model or some sub-modules) are removed.

Sihler et al. [28] build a sanitizer tool for IRIS, a graphical modeling tool for technology road maps (TRM). Their tool can perform various one-way transformations with the intent of preserving model behavior, in addition to the model still being in the same file format. During their transformations, the model structure becomes completely removed, though. Their transformations provably uphold criteria, such as self-containment of the model parts that are left; the sanitized model should reveal as little information as possible about removed elements; and preserving behavior. This makes their tool similar to SMOKE. However, SMOKE also serves the use case of preserving model structure and selectively sanitizing behavior. Sihler et al. further identify Functional Mockup Interfaces [48] as a similar approach, as they enable black box functionality without exposing the implementation details.

Less related are the ideas of Gupta et al. [49], who protect CAD models by inserting sabotaging elements into the model that hamper reproducing physical instances of the model. One may interpret this as the model structure largely being preserved, while its behavior is changed.

Finally, the *ModelObfuscator* by Zhou et al. [50] is intended for Deep Learning models and uses various obfuscation techniques, some of them structure-preserving. However, many of them intentionally alter the model's structure, such as extra layer injection.

4 Approach and Tool

4.1 Approach

The main goal of SMOKE is to support users with the selective removal of sensitive information from their models via obfuscation, or sanitization, while keeping the model

structurally intact and loadable. We thus create one-way and resilient [51] (aka. unidirectional [14]) obfuscation and sanitization transformations in Simulink that fulfill these criteria for SMOKE.

To come up with a list of possible transformations, we first went over the metamodel approximation of Simulink from the Massif group[11] and identified model elements that could be altered to obfuscate or sanitize. In a second step, we consulted the Simulink Documentation for model elements that are not listed in this metamodel, e.g., 'Model Information' which we added as attributes to the 'SimulinkModel' class in our metamodel in Figure 3. In a last step (see Section 5.5), we inspected the raw model files for possibly sensitive user information in any of the model Parameters or attributes, that were still present in models after we applied transformations.

As Simulink is a vast ecosystem,[12] we cannot provide a complete obfuscation or sanitization option for every type of model element. Many model elements interact with each other in complex ways, which is impossible to deal with in a research prototype. Instead, we chose to offer options for those transformations that we deem the most intuitive ones: transformations on elements which are used often, are prominently displayed to users, or are used to hold sensitive or model-wide information. In Section 4.2 we will see though, that SMOKE can be easily extended to obfuscate or sanitize any model elements users may wish to alter.

Our extensive list of selectable elements to be obfuscated and/or sanitized can be seen in the main menu of SMOKE in Figure 5: in the lower right block 'Optical' is the list of obfuscations, in the lower left block 'Functional' is the list of sanitizations. In the following, we will shortly elaborate on a selection of these shown transformations, and our rationale behind them.

4.1.1 Obfuscations

In Simulink, a model's primary central information, such as `CreatorName, CreationDate, and ModifiedBy`, are stored in its 'Model Information'. SMOKE can remove this information, i.e., reset it to default values, and thus break the model's traceability. Next, Simulink offers various ways of commenting and documenting within models, such as 'DocBlocks,' 'Annotations,' and 'Descriptions'. Note that these documentation options are directly embedded into the model file, while additional model 'Notes' exist [52]. By removing such documentation SMOKE can impair the model's understandability. Additionally, Simulink offers a feature called 'Subsystem Content Preview,' which allows users to preview the contents of a Subsystem, such as the blue Subsystem shown on the left of Figure 1. Disabling this preview ensures that no external information is visible in the current model view, meaning screenshots will only display information from the currently opened Subsystem. Furthermore, all other obfuscation options shown in Figure 5 either remove custom names or custom design elements from the model, both of which impair the model's understandability.

[11] https://github.com/viatra/massif (visited at 2025-11-7)

[12] In addition to the hundreds types of blocks Simulink already ships with, various toolboxes offer additional Block types. Each Block type has its own specific Parameters in addition to the common ones. We observed more than 100 Block types and more than 10,000 Parameters. Finally, Simulink comes with its own sub-languages: StateFlow, and SimScape. This results in a multitude of corner cases, that need to be dealt with, if one strives for an exhaustive tool.

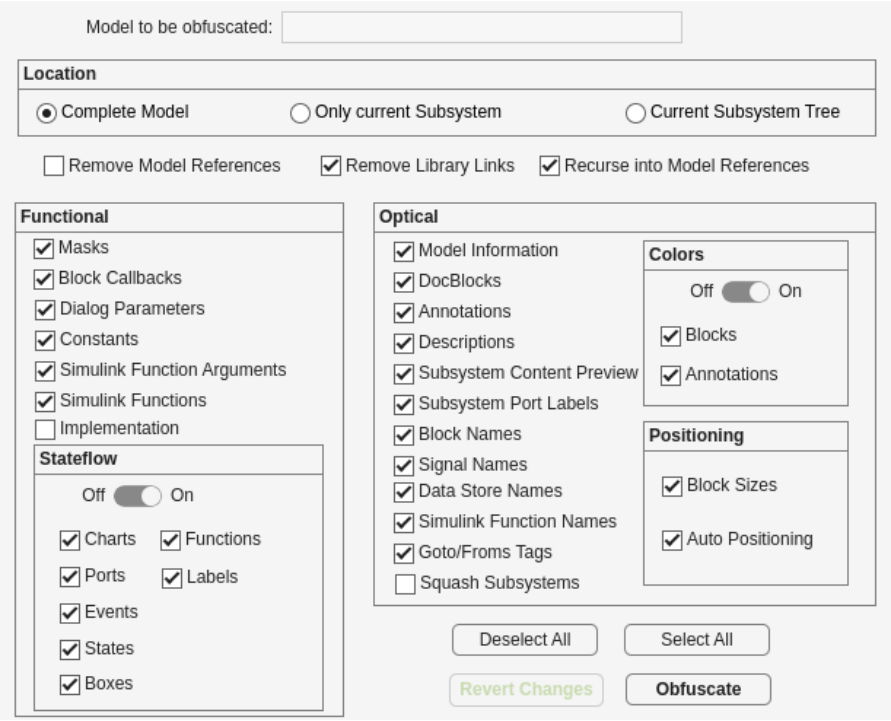


Figure 5: The main menu of SMOKE listing all transformations that can be performed as menu elements. The menu's elements are described in Section 4.3.1.

4.1.2 Sanitizations

In Simulink 'Masks' are used to hide and protect the internals of a Subsystem. This can be a simple picture icon visually (e.g., the car seat picture on the Subsystem in the lower bottom of Figure 1), but can also offer complex custom Parameters. SMOKE removes such Masks, and the bare Subsystems remain under them. 'Block Callbacks' are custom MATLAB scripts that are executed on certain events, like loading the model or changing a view. They are powerful and can, e.g., check for the presence of certain model elements, and depending on the result of the check, prevent the saving of the model. 'Dialog Parameters' are probably the most important sanitizer, as all kinds of Block functionality and behavior can be changed here for each Block individually (cf. Figure 8). SMOKE will reset Parameters to Block default values, whenever possible. 'Constants' are special Blocks, that drive a constant input into the model. Constant Blocks' values are reset to either default number or string values, depending on their type. More interestingly, users can embed a whole MATLAB script into a Function Block, which are sanitized, i.e., their body is removed by activating the checkbox 'Simulink Functions'. We also address various parts of the Simulink sub-language 'Stateflow' which offers to embed state machines or stateflows into Blocks, with more fine-grained options.

4.1.3 Structural Transformations

As explained in Section 2.1, almost all of our obfuscations and sanitizations only alter class attributes or Parameters to preserve the model structure. The options 'Squash Subsystems' and 'Implementation' (the only unchecked options in the bottom of the panels 'Functional' and 'Optical' in Figure 5) do in fact change the model structure, though. We still incorporated them into SMOKE, as they were recommended to us by users. 'Squash Subsystems' removes all Subsystems, i.e., they get resolved and a completely flattened model hierarchy emerges. When checking 'Implementation' all Blocks and Lines apart from InPorts and OutPorts of the currently selected Subsystem get removed. This option is useful when certain parts of the model can be shared, while other parts are so sensitive that even their structure needs to be removed.

Strictly speaking, the removal of DocBlocks – such as the one visible in the bottom-left corner of Figure 1 – does introduce a minor structural alteration to the model. However, this change is negligible in practice: DocBlocks are, by design, isolated components with no functional connections to other Blocks in Simulink. They are never connected to other Blocks meaning their presence or absence does not impact the model's computational behavior or simulation semantics. Thus, while the model's structure is technically modified, its behavior remains preserved. Additionally the connected elements of a model remain unaltered.

4.1.4 Reversing Transformations

In general, obfuscation and sanitization transformations should be hard or impossible to reverse; otherwise, they need not to be applied. We thus implement (see Section 4.2) all our transformations to actually remove or reset, and not to just hide model information. In particular, we remove model elements, whenever possible. Elements, whose removal would result in a changed structure, or broken model, are reset to default values. As such custom information is removed, it can only be reconstructed using additional domain knowledge or by extrapolating other model information that was not selected to be removed by the user.

Additionally, we made sure that model information gets permanently removed once a transformation was applied by the user. There appear to be only two possibilities for reversing any deletion: either the deleted information, while not visible to the user in the Simulink IDE, may still be part of the raw model file (1), or the Simulink IDE may offer an "undo action" functionality to restore the information (2).
Regarding (1): we compared pairs of models before and after transformations, manually and using scripts, to make sure the deleted information is indeed removed from the raw model file. We found our transformations to completely remove the selected information from the raw file. As the model information is saved into and recreated from this file, this path of transformation reversal is thus impossible.
Regarding (2): Simulink does offer to undo IDE actions from its current session. Once a model is saved and closed in the IDE, the edit history dissipates and is not recoverable from the saved file.[13] Furthermore, the transformations SMOKE calls are not reversible in the first place, as they are not IDE edits, but edits executed by scripted

[13]This applies to MATLAB versions through R2025a and is expected to remain valid in future versions.

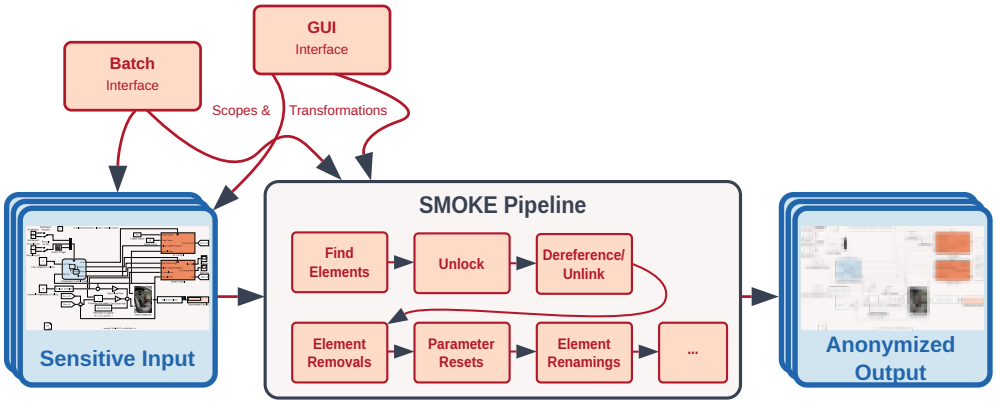


Figure 6: SMOKE's architecture: its pipeline of model transformations can be invoked from the GUI or in batch mode.

MATLAB transformations, as shown in Section 4.2. MATLAB can currently only undo IDE edits, though.
In conclusion: we are certain that sharing a SMOKE-transformed file will remove all the user-selected information permanently and irreversibly.

4.1.5 SMOKEing other Modeling Languages

Most ideas we present in this paper are similarly applicable to obfuscate or sanitize models created in other modeling languages. One simply has to choose what kind of information shall be preserved in the model (in our case, the model structure and ability to parse it with the IDE), and what types of information can or should be removed. For modeling languages with a known metamodel and fewer Block types and Parametrical corner cases than Simulink, it may also be possible to automatically derive obfuscation/sanitization transformations directly from the metamodel. Mathworks, however, never published Simulink's metamodel, and only the unofficial approximations by the Massif group and Sánchez et al. [41] exist.[14]

4.2 Implementation and Extensibility

We implemented SMOKE[15] in the MATLAB scripting language as a standalone application. Every transformation uses the Simulink model API of MATLAB, to alter the model via functions such as `delete(<modelElement>)` or `set_parameter(<blockID>,<Parameter>,<newValue>)`. We did not directly alter the model's raw XML files, as this risks breaking the model, i.e., making it impossible to load or edit the model.

After starting the SMOKE app, the user chooses a Simulink model as input, and then selects transformations (cf. Section 4.3.1) to be applied on the model. SMOKE transforms a model through a pipeline architecture, cf. Figure 6. If the user chooses

[14] https://github.com/viatra/massif (visited at 2025-11-7)
[15] Available at https://github.com/lanpirot/SMOKE.

multiple transformations, they will be applied sequentially. Before any transformation takes place, the model and all its Blocks are unlocked so they are editable in the further process. Next, the model references are resolved, and links to library Blocks or Models removed, as edits to them are not allowed by Simulink, otherwise. Finally, all selected transformations are applied sequentially.

Each transformation iterates over all the elements it pertains to and transforms them one by one, e.g., the DocBlock removal iterates over all DocBlocks. To ensure that SMOKE preserves the model structure, every transformation that is called removes exactly one model element or resets exactly one Parameter at a time. We only ever remove Annotations and DocBlocks (they are not structurally relevant) and check whether a Parameter change affects the model structure before attempting to change it. We thus guarantee structure-preservation by construction. For most transformations, our implementation is trivial, with just a few necessary interventions to keep the model intact, e.g., names of Blocks have to be unique in each view, or references to names have to be updated appropriately. For the resetting of names, we choose a trivial naming scheme, e.g., every Block will get the name `<block type of Block><Number>`.

One of the more interesting transformations is our sub-script `removeDialogParameters.m`. Here, all Blocks' Parameters with custom values shall be reset. Simulink does not offer a direct functionality for such a reset, and also does not give default values to which to reset them to. We thus create and insert a dummy Block of the same type into the model, and read out, and copy its values for Parameters. Sometimes, dummy Blocks do not have all Dialog Parameters of the actual Block (c.f. also Figure 8, where the Parameter list changes after transformation). In this case, we reset Parameters that are numeric to `0` and string type to the empty word `''`. Further, we preserve some Parameters of Blocks, that would alter the model structure: we found, e.g., 7 Parameters that change the number of ports of a Block. Resetting them could remove the incoming or outgoing Signal Lines of the Block and thus break the model structure. After the Block's Parameters are reset, we clean up the model by removing the temporary dummy Block again. To reset a Block's size, we use the same technique and copy a dummy Block's default dimensions.

Our implementation ensures linear scaling by iterating over the elements of a model exactly once per transformation. Most transformations only reset a single local Parameter, or remove the element itself – both operations execute in constant time. The most computationally intensive transformation is again the `removeDialogParameters.m` sub-script, as it is bounded by the Block type with the highest Parameter count in the Simulink model. Finally, the structure-changing 'Squash Subsystems' transformation is implemented top-down: it begins by squashing Subsystems closest to the root, ensuring that any element moved to a higher-level Subsystem ascends only one level per operation, which is achieved in constant time per element. In summary, each transformation performs a fixed number of operations per element, resulting in SMOKE's linear runtime.

Our transformations' implementation is *optimistic* [53], i.e., our transformations may try to change or remove model elements that could break the model. Breaking transformations are caught by Simulink's error detection via exceptions. SMOKE

catches the exceptions, and we simply skip the transformation for elements causing such errors, thus leaving the model syntactically correct. Some examples that get caught are: certain Blocks are not resizable; some model elements cannot be removed as Callback Functions are dependent on them, etc. This means that some non-removable information remains in the model. However, we are not aware of Callback Functions preventing the removal of user-relevant sensitive information. To guarantee complete removal success, users can select 'Block Callbacks' in their transformation list. SMOKE removes the Callbacks first, and can then safely remove the otherwise unremovable elements. All other non-applicable transformations we witnessed are related to the syntactical correctness of the model.

As each transformation takes and leaves a structurally intact model, most transformation orderings are interchangeable. For best performance, SMOKE still works according to a partial ordering. After unlocking the model and removing model Callbacks it is best to start with transformations that remove model elements, like Blocks. This ensures that no transformation is performed on elements that are later removed, anyway. The removal of sizes and positions should be performed last. Resetting the shapes and sizes of Blocks, in the penultimate step, ensures that potential text can still be completely displayed. The removal of positions should be the last transformation: here, all Blocks and their sizes are taken into account to achieve visually pleasing diagrams without any overlap of model elements.

When we debugged SMOKE in its development phase, we found it best to perform the renaming transformations last, as it is much easier to quickly match the original model and its transformed counterpart this way. For an easier comparison of the model pair, it's also best to deselect size and position removal, as then all elements 'stay in their place' and can be matched visually at a glance.

To extend SMOKE with new transformations, users can either: (1) add a transformation to their suitable transformation list within 'element removals', 'parameter resets', and 'element renamings' (the lower middle boxes in Figure 6); (2) append it in the '...' box. The sole requirement for any new transformation is that it must preserve syntactic correctness – that is, it should accept a valid Simulink model as input and produce a valid Simulink model as output, consistent with all prior transformations.

4.2.1 Comparison of SMOKE with Prior Versions

Our application is based on a predecessor app called `Obfuscate Model`[16] by Jaskolka et al. Here, we first compare `Obfuscate Model` to SMOKE in the state of the publication of [33], i.e., SMOKE1.0. Next, we compare SMOKE1.0 to SMOKE in this paper's state, i.e., SMOKE2.0.

Jaskolka et al. built a Simulink model obfuscation tool written in MATLAB. Besides some bug fixes (e.g., crashes on broken array indices), adding exception handling, and a slight redesign of the UI, in SMOKE1.0 we:

- added various functional transformations. `Obfuscate Model` was restricted to graphical obfuscations only.
- added the graphical obfuscations for resetting block position and block sizes.

[16] https://github.com/McSCert/Obfuscate-Model (visited on 2025-11-7)

- enabled deep obfuscations, where SMOKE1.0 also descends into masked, protected, variant or linked Blocks. These were not transformed before.
- added a 'Revert changes' option to quickly reset the transformation process and recover the original model.

Since our first publication [33], we fixed some further bugs and sped up SMOKE2.0's execution, as well as:

- The user can choose the location or scope where to apply transformations. This way, some parts of the model can be kept in their original state or be transformed differently than others.
- SMOKE2.0 can be applied to locked (library) models, which makes it applicable to all types of Simulink models, now.
- We added the most radical transformation: removing the complete implementation of a Subsystem. The whole local Subsystem tree is thus pruned.
- We added an obfuscation that removes the Subsystem hierarchy, and thus "flattens" the model, as all model elements are put into the same hierarchy level.
- We added a sanitization that removes Simulink Function bodies.

4.3 SMOKE in Action

4.3.1 Menu and User Interaction

The menu screen of SMOKE is presented to the user right at startup. At startup, the boxes are pre-checked as shown in Figure 5. In the upper line, the model selected for transformation is shown. For this, SMOKE automatically chooses the last model that is opened by the user. All chosen transformations will later be performed on the selected model. Next, in the second line, the user can select the scope of all chosen transformations. Either the complete model gets transformed, only the currently selected Subsystem and its direct elements, or a whole Subsystem Tree, i.e., the current Subsystem and its descendants. After that, users choose how to handle model references and library links. In the lower left box, all options that remove or alter model behavior, i.e., sanitizations, are listed. In the lower right box, all options pertaining to obfuscations are listed, which leave model's behavior intact. On the bottom right are the action buttons. In the upper row are two buttons for convenience that check all or none of the boxes. The 'Obfuscate' button triggers the obfuscations and sanitizations the user chose. The 'Revert changes' button reverts the model to its original, i.e., last saved state.

We envision two main user journeys using the GUI: (1) apply all pre-selected transformations as shown in Figure 5, which removes everything but the main model structure, (2) start with few transformations and selectively apply more and more of them, perhaps in different scopes, until all sensitive information is removed. Once the user is satisfied with the model's state and that its sensitive information is removed, they can save it, and it is ready to be shared.

In addition to SMOKE's interactive mode, it can also be called via MATLAB's scripting API. In this way, a whole project's models can be anonymized in batch mode,

using pre-selected transformations. We used this mode in our evaluation in Section 5 to handle thousands of models at a time.

4.3.2 Exemplary Obfuscation

To give an intuitive overview of the capabilities and impact of SMOKE, we give an exemplary user journey of a model's obfuscation (Figure 7) and sanitization (Figure 8) of the model from Figure 1. In a first step, a user chooses to remove all documentation elements, labels and names, which leaves the model's elements completely nameless (hidden names are reset to generic default names) in Figure 7a. The model's layout otherwise is completely unchanged, and a visual mapping from the original is trivial. Next, the user decides to further flatten the model's hierarchy and all Subsystems in the model get resolved. This affects some Subsystems on the left and lower side of the model, which get replaced by their inner Blocks in Figure 7b. To further obfuscate the model, the user removes the Blocks' colors and resets their sizes in Figure 7c. In a final step, the user obtains a clean model layout by using SMOKE's autopositioning feature, with which one can cycle through semi-random model layout arrangements. The final version is now completely obfuscated, while preserving the model structure and behavior. The original model from Figure 1 is hardly recognizable in its final form in Figure 7c.

If the user decides to (perhaps additionally to the obfuscation) sanitize the model, they can choose to apply various transformations affecting behavior (cf. lower left options in Figure 5). Most of these effects are not immediately visible in the IDE view, in contrast to the obvious obfuscation transformations. However, a Block's Parameters and their values can be inspected in popup menus like the one given in Figure 8a. If the user decides to reset the 'Dialog Parameters' some of the Block's Parameters are reset to their default values. For this Block's reset, the conditional Parameter 'Sample time' is revealed in Figure 8b. Changing a Block's Parameters can have a dramatic behavioral impact, as can be seen for the Pulse Generator Block from Figure 8, whose behavior changed from the one shown in Figure 9a to the one in Figure 9b, after the Parameter reset shown in Figure 8.

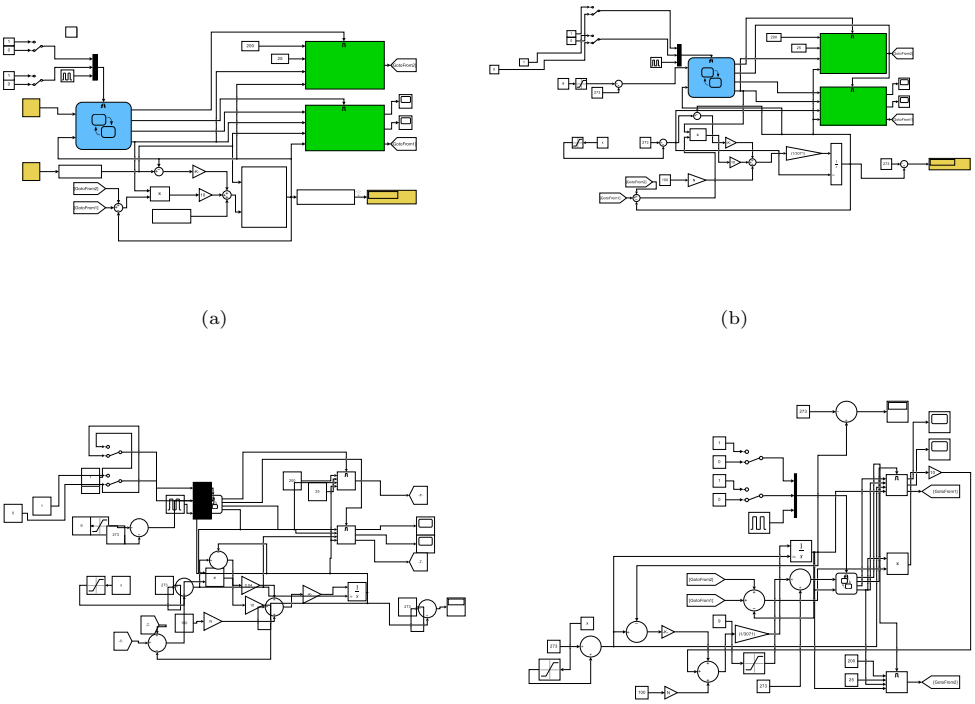
5 Evaluation

To ensure that SMOKE works as intended and designed, we test it in multiple ways: we test, whether SMOKE's obfuscation preserves structural integrity for all its transformations in Section 5.2, whether SMOKE's obfuscation does not alter a model's behavior in Section 5.3, and whether SMOKE's sanitization *does alter* model behavior in Section 5.4. We further check SMOKE's coverage in Section 5.5. For all our experiments, we use a diverse set of models, which we briefly introduce in Section 5.1.

5.1 Experimental Design

5.1.1 Subjects and Setup

As described earlier, Simulink is a vast and diverse ecosystem. To ensure that the various kinds of models and use cases are covered by SMOKE, we apply SMOKE on the

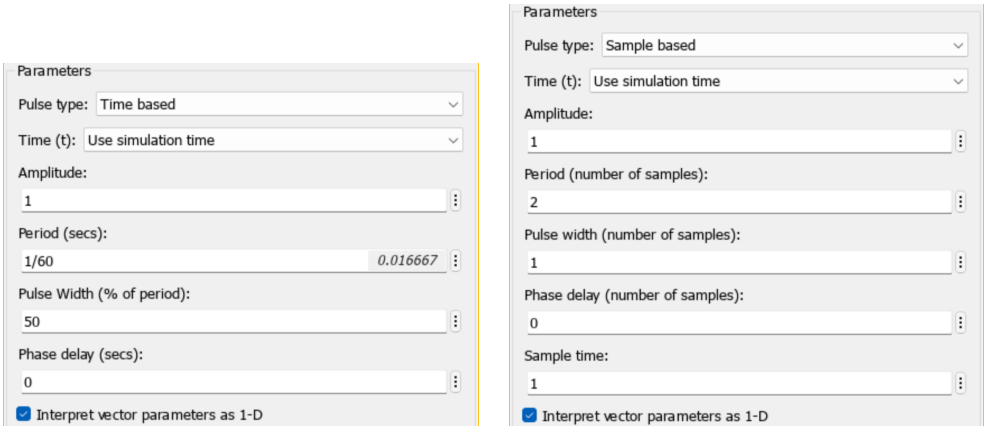


(a) (b)

(c) (d)

Figure 7: Step-wise obfuscation of the model from Figure 1. First, Annotations, Docblocks, labels and names are removed in Figure 7a. The Subsystem hierarchy is flattened in Figure 7b. Then, the Block colors and sizes are reset in Figure 7c. Finally, in Figure 7d the Block and Line positions are reset.

model collection SLNET [11]. This is a comprehensive set of 9,105 open-source models covering multiple domains like electronics, aerospace, robotics, or medicine. The collection encompasses small and large models for various purposes like toy projects, industrial application, etc. SLNET was previously used as a benchmark set in other empirical studies on Simulink models [52, 54–56]. 8,814 of the models were loadable error-free in our setup of Simulink with MATLAB R2024b on our laptop with Windows11, 96GB RAM, and Intel i9-13980HX processor. Models that were not loadable were not necessarily broken, but typically had some libraries missing that were needed in Callbacks of Blocks. Five more models were excluded from our evaluation, as they caused MATLAB crashes during model simulation or model backup – both of these functions are called in our evaluation pipeline. This left 8,809 models for our evaluation. In these 8,809 models, we found 160 unique Block types, and more than 10,000



(a) (b)

Figure 8: Figure 8a gives a snippet from the popup menu of the Parameters of the Pulse Generator Block (located in the upper left in Figure 1, called 'System Trigger'). Users can alter various Block specific Parameters in this menu. Figure 8b gives the Block's Parameters after sanitization: the Parameters 'Pulse type', 'Period', and 'Pulse Width' are affected by the resetting of Parameters. Due to the sanitization, other default Block Parameters may become accessible, like 'Sample time' in this case.

unique types of Block Parameters. This richness enables various, complex ways of interaction in the models. Thus, SLNET presents a large set of numerous challenging corner cases for SMOKE.

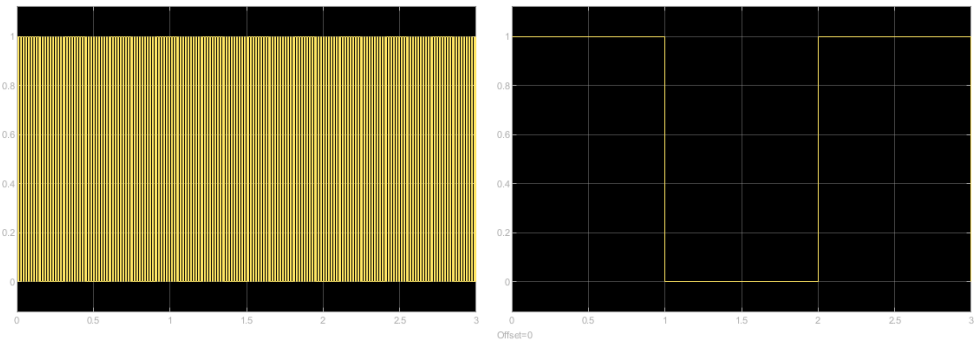
In our evaluation, we applied SMOKE sequentially on each SLNET model to obtain a pair of an original model and a transformed model. We then analyzed each pair further as described in the next sections.

5.1.2 Measuring Robustness, Performance, and Structural Integrity

In a first step, we measured the robustness of SMOKE, i.e., how often SMOKE was applicable to our models without error. We also recorded SMOKE's runtime performance, e.g., transformed Blocks per second, when applying the complete list of obfuscations and sanitizations that are shown as checked in Figure 5.

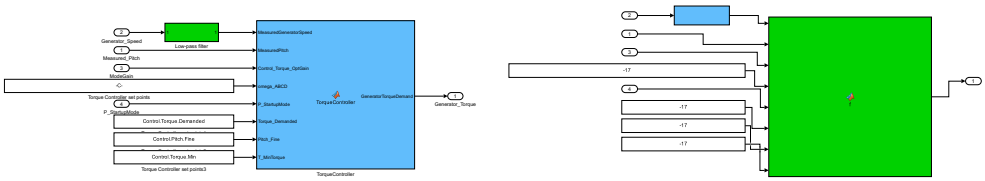
One of our goals of SMOKE is that it does not alter a model's structure, for both sanitization and obfuscation. Our first approach of validating the structural integrity was to employ an existing model comparison tool. However, the built-in tool of Simulink[17] was not able to accurately match pairs of original and transformed models. This was surprising, as it even failed to match easy cases like the one shown in Figure 10. There, the Model Comparison tool erroneously matched the green colored and the blue colored blocks of this small Subsystem. In our second attempt, we tried the clone

[17] mathworks.com/help/simulink/model-comparison.html (visited on 2025-11-7)



(a) (b)

Figure 9: The exemplary effect of resetting Dialog Parameters: the original Block (Figure 8a) produces rapid pulses (Figure 9a), while the sanitized version (Figure 8b) has a much longer cycle length (Figure 9b).



(a) (b)

Figure 10: The Simulink-internal Model Comparison tool struggles to match even easy cases: here the green and blue Blocks from the original (Figure 10a) and the obfuscated model (Figure 10b) are erroneously matched.

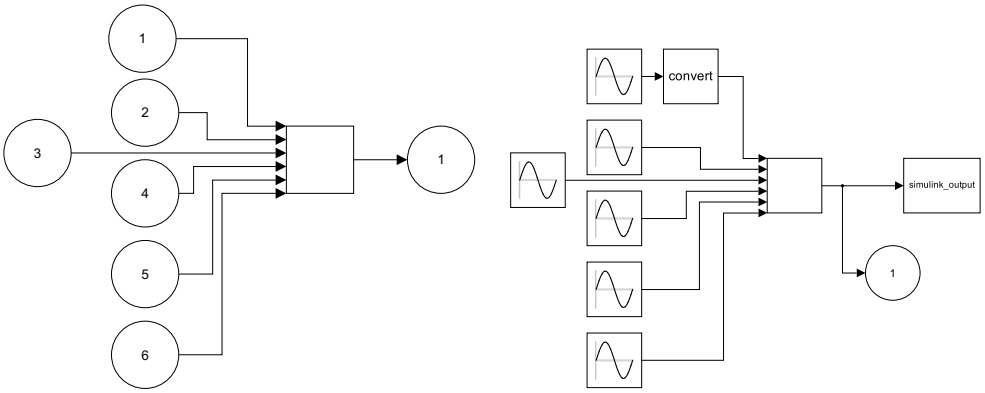
detection tools Conqat and SIMONE. However, we did not get Conqat to run, and SIMONE is already outdated, as it is only able to handle `.mdl` models.[18]

We thus employed a signature-based model comparison [57] as a proxy, using model metrics for which we developed evaluation scripts specifically for this assessment. Our signature uses simple model metrics like the number of Blocks, Lines, Subsystems, unique Block types of a model, and its cyclomatic complexity. A model's signature is formed by combining these model metrics as a tuple. All the metrics were either used previously in the literature [1, 6], or are given as relevant by Simulink researchers themselves [12]. We use an unchanged model signature as a proxy for structural integrity, though we discuss limitations of this assumption in Section 5.6.

5.1.3 Behaviorial (Non-)Integrity of Transformations

In regard to the preservation of model behavior, our goal with SMOKE is two-fold: all obfuscation transformations shall *preserve* the original model's behavior, while the

[18] Simulink uses `.slx` models by default since 2012.



(a) (b)

Figure 11: Our black box wrapping setup: all model InPorts in Figure 11a are replaced by signal generators in Figure 11b. The first port of this example model is of type boolean and a boolean converter is thus added. Similarly, all output signals are tapped into with a To Workspace Block to record the model output.

sanitization transformations shall *alter* a model's behavior. However, an altered model behavior only indicates that (at least) *a part* of the model's behavior is sanitized. It does not guarantee a complete sanitization of all sensitive information in the model. To automatically test whether a model's behavior is preserved or altered after transformation, we treat models as black boxes into which we feed the same inputs and observe their outputs. If a transformed model shows a different output, given the same input, we classify it as altered behavior. We optimistically classify all obfuscated models that show the same output as behavior-preserving. This is optimistic in the sense of our setup possibly failing to try a 'deciding' input that produces diverging outputs. All sanitized models that have the same output are inspected, and classified manually, in a second step.

To evaluate a model's behavior, we record its input and output behavior and our evaluation scripts construct a wrapper for each model. The wrapper replaces the model InPorts with signal generators, and if necessary type converters, and additionally records the values of the outgoing signals, as shown in Figure 10. We then simulate both models of a pair and compare their outputs.

Note that we only construct wrappers for models that are actually compilable, and thus could demonstrate any behavior. However, we observed that often (see Table 1) the status of compilability is changed from the sanitization. Each compilability status change is also counted as breaking the model's behaviorial integrity, because only one of the models can demonstrate behavior, while the other cannot. Although library models cannot show behavior due to their lack of inputs and outputs, we still included them in our experimental setup, i.e. to ensure SMOKE is applicable on all model types.

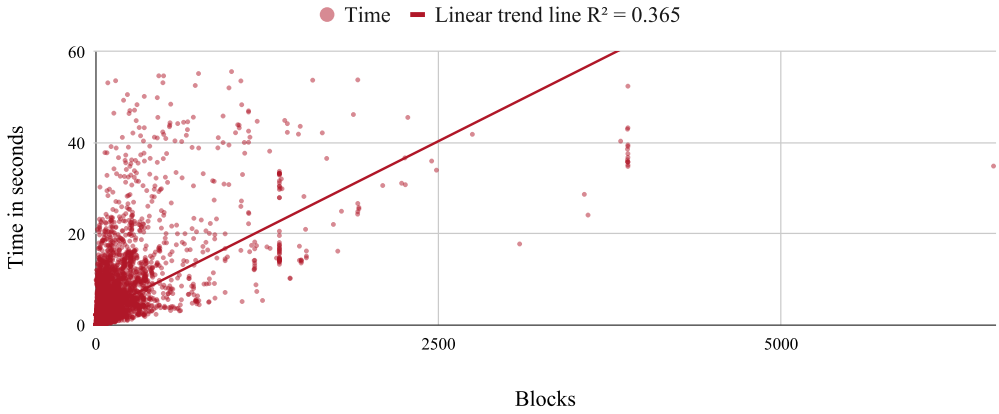


Figure 12: Scatter plot of execution time per model size in Blocks.

5.2 Results: Robustness, Performance, Structural Integrity

SMOKE was able to apply all obfuscations and all sanitizations on all 8,809 models successfully. SMOKE worked at a speed of 25.4 Blocks per second, 33.8 Signals per second, 3.5 Subsystems per second, and 0.62 cyclomatic complexity units per second, when all transformations shown in Figure 5 are applied. A scatter plot of SMOKE's execution time for the model set is given in Figure 12. The clustered 'vertical lines' of points at model sizes of around 1,300 and 4,000 Blocks are the result of anonymizing nearly-identical models. While SMOKE scales linearly in the worst case (see Section 4.1.5), the shown linear trend line has a coefficient of determination R^2 of only 0.365. This suggests a moderate level of prediction from Block size to execution time. The number of Subsystems showed the highest correlation to SMOKE execution time of +0.521, while cyclomatic complexity had the lowest correlation of +0.140.

In all models, the structural integrity was preserved, i.e., no Blocks, Lines, Subsystems were ever added or removed or types of Blocks changed. We found that the cyclomatic complexity changed in 1,954 sanitized models (never in the obfuscated models). This, however stemmed from Simulink's method of calculating cyclomatic complexity: only a compilable model's cyclomatic complexity can be calculated. The compilability of sanitized models, however, switched in many model pairs, as we can see in Table 1. We thus ignored cyclomatic complexity for our analysis, and find SMOKE to preserve structural integrity for all transformations, as desired.

5.3 Behaviorial Integrity Results: Obfuscation

As expected, we found all models' behavior to be preserved by SMOKE's obfuscation transformations.

5.4 Behaviorial (Non-)Integrity Results: Sanitization

In a first step of behavior analysis, we compile all pairs of original and sanitized models – a necessary precursor to their simulation in the next step. Table 1 gives a result of

Table 1: Compilation status of models before and after sanitization.

		After Sanitization	
		Compilation Fails	**Compilable**
Before Sanitization	**Compilation Fails**	5,933	199
	Compilable	1,638	1,039

the compilation. We can see that most original models are not compilable, already before the sanitization, i.e., $6,132 = 5,933 + 199$ models fail to compile from the start. This is because many of them serve some other purpose, like library models, and are not intended (or impossible) to be compiled, or run. We view non-compilable models as not showing any behavior that could be altered or evaluated automatically, and thus do not evaluate them further in the upcoming steps. The sanitization breaks the compilation of 1,638 models, and interestingly, 199 models become compilable *after* sanitization. This is due to models being in some kind of broken state (in respect to compilation), which sometimes gets fixed by resetting Parameters with Smoke. We view a change in compilability in a model pair as a behaviorial alteration, as only one of the two models can demonstrate any behavior, while the other cannot. Only 1,039 models are compilable before and after transformation, and these are the models we simulated next.

Table 2: Output status of models before and after sanitization.

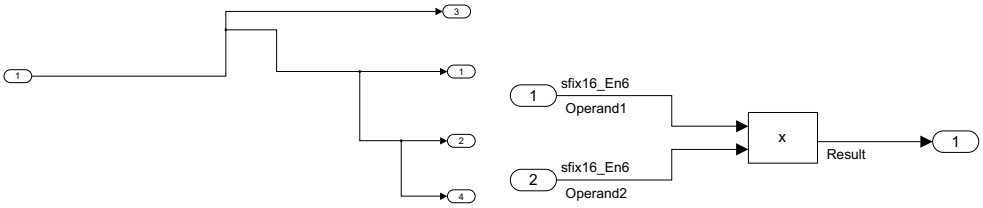
		After Sanitization	
		Crash/no output	**Output**
Before Sanitization	**Crash/no output**	823	27
	Output	31	158

We give the results of the simulation in Table 2. Most models produce no output, i.e., they crash in their execution right away, or there was no output to harness in our test setup (e.g., library models). Similar to the compilation, we see 31 models producing no output after the sanitization, and more notably, 27 sanitized models to start producing output. Only 158 models ran their execution crash-free for both model versions. Of these, 16 models produced the exact same output. From our initial set of 8,809 models, we thus automatically classified $8,809 - 16 = 8,793$ models as behaviorially altered, or without behavior.

The models that became compilable (199), or executable (27) via the sanitization deserve a closer look. We observed that the sanitization removed custom (but broken) model parts, such as configurations, or it reset data types of Parameter values, removed faulty Callbacks, etc. The other way around of sanitized models becoming broken (for compiling 1,638, or running 31) is more obvious: after the sanitization, needed variables, intra-model dependencies, or data types are missing, or Block Parameters become inconsistent with each other.

Table 3: Metric comparison of SLNET models vs. their subset of behaviorially stable models.

		Blocks	Block Types	Signal Lines	Subsystems
mean	**SLNET models**	133.5	10.8	177.8	18.6
	stable models ⊂ SLNET	7.6	3.4	8.7	0.3
median	**SLNET models**	30	9	36	4
	stable models ⊂ SLNET	5	3	3	0



(a) (b)

Figure 13: Two exemplary models unaffected by sanitizing. They are too small, and or too simple, and SMOKE did not change any behavior-affecting Parameters in each.

We further inspected the last 16 stable models to determine why the sanitizing process did not alter their behavior. Our first observation is easily recognizable from Table 3: the stable models are much smaller for the various size metrics than their counterparts of the complete SLNET set. Most of their implementations are also completely flat, i.e., they are devoid of any Subsystems. This is intuitive because, on average, smaller models exhibit less complex behavior that could be affected by our sanitizations.

A manual inspection further showed that these models are also simple, probably toy projects. We give two example models in Figure 13, where one can see them to be too simple and too small for the sanitization to have any effect. We argue they show no real behavior that needs to be sanitized, i.e., is sensitive for users, in the first place.

In conclusion, SMOKE's sanitization altered the behavior of most models, with only a few small and trivial models remaining stable. The more complex and larger models all showed behavior alteration. Although this observation does not guarantee that all sensitive information is sanitized, it shows that at least a part of the original model behavior is removed.

5.5 Coverage

In a final step to enhance SMOKE's capabilities, we examined the raw model files to improve its coverage, i.e., we checked if SMOKE leaves potentially sensitive model elements untouched. In the SLNET model files, we found 160 different Block types and 10,731 different Parameters. These Parameters are either model-wide Parameters

or for its various elements, such as Lines, Blocks, Annotation, etc. Many of these Parameters are for internal Simulink use, and users are not supposed to edit them.

Smoke does not support *all* of these Blocks and Parameters, as many of them are not supposed to be changed, or would need individual handling in our implementation, which is not feasible. When designing Smoke, we first came up with a number of obfuscation and sanitization candidates (compare Section 4.1). After anonymizing them, we next ran a script that finds model elements that Smoke has thus far left out, but could potentially hold sensitive information. These sensitive parts were then included into transformations of Smoke, and we started the process again. To identify possibly sensitive parts, our script gathered all unique values in all models for each Parameter. We manually went over this list of values and inspected the first, or the first couple of them, to see whether they might hold valuable information. If they did, we updated Smoke so that they are also anonymized. Our heuristic here was to give a closer look at the values of text or numeric type – especially if there were more than 10 different values for a Parameter. Our argument here is that Parameters holding few different values, or even always the same value, are less likely to be sensitive to disclosure. We further inspected a handful of raw transformed model files visually, to see whether any more possibly sensitive information might still be left. Our investigation process does not guarantee that *every possible* sensitive part of models will be removed by Smoke, but it does remove all those that we found.

5.6 Threats to Validity

5.6.1 Internal Threats

The biggest threat to the internal validity of our experimental results is in our test setup for running the models. Our harnessing to capture outputs (cf. Figure 11) is based on the heuristic that the model inputs and outputs are its outermost input and output Blocks and not hidden somewhere in the model but are on the outer layer. Additionally, some models may need inputs of certain types or values, which our test setup did not try. However, the size of our model pool made a non-model-specific and fast heuristic necessary. Furthermore, our test setup captured outputs for more than a fourth of the models that were compilable (cf. Table 2), of which many of the models were never intended to be compiled or run in the first place, i.e., design, toy, or library models, etc.
Another aspect of our model output analysis in Section 5.4 is that we only measured whether the output of a sanitized model differed *in any way* to the original model's output. We cannot sufficiently decide whether "enough" of a model or a model's behavior is altered or removed. Users have to make sure themselves whether a transformed model satisfies their need for information protection.

Regarding the ability to reverse-engineer information or Smoke's coverage (Section 5.5): we did not investigate, whether users (or Simulink) hides possibly valuable data in an encrypted way in the model (or raw files). During our investigation of the raw files, we found a few parts that were not intelligible. In our investigation, we found them to be Base64 encoded UTF16 strings holding simple MATLAB UI variables and values, though. We otherwise completely remove model elements or

map Parameter values to the same value. Mathematically, this means that reversing our transformations would require reconstructing the original elements or Parameter values, via a reverse transformation, that would have to 'guess' correctly [14]. Such guessing might be helped if enough contextual information is still present in the model. We thus caution users of SMOKE to anonymize their models enough, so that anonymized parts, cannot be inferred from the remaining information, possibly coupled with domain knowledge.

Lastly, our signature-based approach to check for structural integrity of our transformations (see Section 5.1) only reliably detects structural changes. If the signature of the model before and after transformation is identical, structural integrity is not necessarily implied. However, SMOKE's transformations were explicitly designed and implemented to preserve the model structure.

5.6.2 External Threats

Although the SLNET evaluation set is extensive and diverse, we have not yet tested SMOKE with actual corporate models, which are its intended target. However, prior studies have shown that a subset of models from SLNET is 'industry-like' [1]. Although industry models may be larger than the models in our dataset – e.g., industry models may be as large as 100k Blocks – SMOKE's linear computational complexity ensures that even large models can be transformed within an hour, see Figure 12. While potential industry partners may have additional requests of obfuscating or deleting elements that SMOKE currently ignores, such features are easily integrated, as we already demonstrated in SMOKE's evolution (compare Sections 4.2.1 and 5.5).

6 Conclusion

With SMOKE, we provide a versatile tool that allows users to share Simulink models while protecting their sensitive information. The tool enables selective and fine-grained obfuscation or sanitization of models, all while preserving their structure. Our hope is that SMOKE will enhance collaboration among researchers and industry partners by facilitating the selective protection of models. This will foster secure sharing within the research community and between companies.

We are currently integrating SMOKE as a filter step into a workflow of creating research data management containers [58, 59]. With SMOKE, we ensure sensitive model parts are excluded before they become part of an immutable container. In the future, we hope that Simulink will integrate features of SMOKE natively to make model anonymization even more accessible. SMOKE is open-source[19] and easily extendable, e.g., with additional transformations. We welcome any suggestions for additional features the community would like to see integrated.

References

[1] Boll, A., Brokhausen, F., Amorim, T., Kehrer, T., Vogelsang, A.: Characteristics, potentials, and limitations of open-source Simulink projects for empirical research.

[19] https://github.com/lanpirot/SMOKE

Software and Systems Modeling **20**(6), 2111–2130 (2021)

[2] Bertram, V., Maoz, S., Ringert, J.O., Rumpe, B., Wenckstern, M.: Component and Connector Views in Practice: An Experience Report. In: 20th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, MODELS, pp. 167–177. IEEE Computer Society, Austin, TX, USA (2017)

[3] Ro, J.W., Malik, A., Roop, P.: High Fidelity Simulation of Hybrid Systems using Higher Order Hybrid Automata. IEEE Transactions on Computers **71**(7), 1668–1680 (2022) https://doi.org/10.1109/TC.2021.3100746

[4] Xu, X., Wang, S., Zhan, B., Jin, X., Talpin, J.-P., Zhan, N.: Unified graphical co-modeling, analysis and verification of cyber-physical systems by combining AADL and Simulink/Stateflow. Theoretical Computer Science **903**, 1–25 (2022) https://doi.org/10.1016/j.tcs.2021.11.008

[5] Jaskolka, M., Pantelic, V., Wassyng, A., Lawford, M., Paige, R.: Repository Mining for Changes in Simulink Models. In: 2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS), pp. 46–57 (2021). https://doi.org/10.1109/MODELS50736.2021.00014

[6] Shrestha, S.L., Boll, A., Chowdhury, S.A., Kehrer, T., Csallner, C.: EvoSL: A Large Open-Source Corpus of Changes in Simulink Models & Projects. In: 2023 ACM/IEEE 26th International Conference on Model Driven Engineering Languages and Systems (MODELS), pp. 273–284 (2023)

[7] Boll, A., Vieregg, N., Kehrer, T.: Replicability of experimental tool evaluations in model-based software and systems engineering with MATLAB/Simulink. Innovations in Systems and Software Engineering, 1–16 (2022)

[8] Bertram, V., Maoz, S., Ringert, J.O., Rumpe, B., Wenckstern, M.: Component and Connector Views in Practice: An Experience Report. In: Proceedings of the ACM/IEEE 20th International Conference on Model Driven Engineering Languages and Systems. MODELS '17, pp. 167–177. IEEE Press, Austin, TX, USA (2017). https://doi.org/10.1109/MODELS.2017.29

[9] Tomita, T., Ishii, D., Murakami, T., Takeuchi, S., Aoki, T.: A Scalable Monte-Carlo Test-Case Generation Tool for Large and Complex Simulink Models. In: 2019 IEEE/ACM 11th International Workshop on Modelling in Software Engineering (MiSE), pp. 39–46 (2019). https://doi.org/10.1109/MiSE.2019.00014

[10] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L.B., Bourne, P.E., *et al.*: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data **3**(1), 1–9 (2016)

[11] Shrestha, S.L., Chowdhury, S.A., Csallner, C.: SLNET: a redistributable corpus of 3rd-party Simulink models. In: Proceedings of the 19th International Conference on Mining Software Repositories. MSR '22, pp. 237–241. Association for Computing Machinery, New York, NY, USA (2022)

[12] Shrestha, S.L., Boll, A., Kehrer, T., Csallner, C.: ScoutSL: An Open-Source Simulink Search Engine. In: 2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C), pp. 70–74 (2023)

[13] Indamutsa, A., Di Rocco, J., Almonte, L., Di Ruscio, D., Pierantonio, A.: Advanced discovery mechanisms in model repositories. Software: Practice and Experience **54**(11), 2214–2248 (2024) https://doi.org/10.1002/spe.3332 https://onlinelibrary.wiley.com/doi/pdf/10.1002/spe.3332

[14] Czarnecki, K., Helsen, S.: Feature-based survey of model transformation approaches. IBM systems journal **45**(3), 621–645 (2006)

[15] Oliveira, S.R.M., Zaiane, O.R.: Protecting sensitive knowledge by data sanitization. In: Third IEEE International Conference on Data Mining, pp. 613–616 (2003)

[16] Pantelic, V., Postma, S., Lawford, M., Jaskolka, M., Mackenzie, B., Korobkine, A., Bender, M., Ong, J., Marks, G., Wassyng, A.: Software engineering practices and Simulink: bridging the gap. International Journal on Software Tools for Technology Transfer **20**, 95–117 (2018)

[17] Jaskolka, M., Pantelic, V., Wassyng, A., Lawford, M.: A Comparison of Componentization Constructs for Supporting Modularity in Simulink. SAE Technical Paper (2020-01-1290) (2020)

[18] Collberg, C., Thomborson, C., Low, D.: A taxonomy of obfuscating transformations. Technical report, Department of Computer Science, The University of Auckland, New Zealand (1997)

[19] Tevajärvi, J.: Protecting Intellectual Property in Multi-Supplier Ship Powertrain Co-Simulation. Master's thesis, Aalto University, Otaniemi (December 2 2023)

[20] Barak, B., Goldreich, O., Impagliazzo, R., Rudich, S., Sahai, A., Vadhan, S.P., Yang, K.: On the (Im)possibility of Obfuscating Programs. In: Kilian, J. (ed.) Advances in Cryptology - CRYPTO 2001, 21st Annual International Cryptology Conference. Lecture Notes in Computer Science, vol. 2139, pp. 1–18. Springer, Berlin/Heidelberg, Germany (2001)

[21] Chowdhury, S.A., Varghese, L.S., Mohian, S., Johnson, T.T., Csallner, C.: A curated corpus of Simulink models for model-based empirical studies. In: Bures,

T., Fitzgerald, J.S., Schmerl, B.R., Weyns, D. (eds.) Proceedings of the 4th International Workshop on Software Engineering for Smart Cyber-Physical Systems, ICSE 2018, Gothenburg, Sweden, May 27, 2018, pp. 45–48. ACM, New York City, NY, USA (2018)

[22] Amorim, T., Boll, A., Bachman, F., Kehrer, T., Vogelsang, A., Pohlheim, H.: Simulink bus usage in practice: an empirical study. Journal of Object Technology **22**(2), 2–114 (2023)

[23] Stephan, M., Cordy, J.R.: Identification of Simulink model antipattern instances using model clone detection. In: Lethbridge, T., Cabot, J., Egyed, A. (eds.) 18th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, MoDELS 2015, Ottawa, ON, Canada, September 30 - October 2, 2015, pp. 276–285. IEEE Computer Society, New York City, NY, USA (2015)

[24] Schlie, A., Wille, D., Schulze, S., Cleophas, L., Schaefer, I.: Detecting Variability in MATLAB/Simulink Models: An Industry-Inspired Technique and its Evaluation. In: Cohen, M.B., Acher, M., Fuentes, L., Schall, D., Bosch, J., Capilla, R., Bagheri, E., Xiong, Y., Troya, J., Cortés, A.R., Benavides, D. (eds.) Proceedings of the 21st International Systems and Software Product Line Conference, SPLC 2017, Volume A, Sevilla, Spain, September 25-29, 2017, pp. 215–224. ACM, New York City, NY, USA (2017)

[25] Reicherdt, R., Glesner, S.: Slicing MATLAB Simulink models. In: Glinz, M., Murphy, G.C., Pezzè, M. (eds.) 34th International Conference on Software Engineering, ICSE 2012, June 2-9, 2012, Zurich, Switzerland, pp. 551–561. IEEE Computer Society, New York City, NY, USA (2012)

[26] Schlie, A., Schulze, S., Schaefer, I.: Comparing Multiple MATLAB/Simulink Models Using Static Connectivity Matrix Analysis. In: 2018 IEEE International Conference on Software Maintenance and Evolution, ICSME 2018, Madrid, Spain, September 23-29, 2018, pp. 160–171. IEEE Computer Society, New York City, NY, USA (2018)

[27] Mikulcak, M., Herber, P., Göthel, T., Glesner, S.: Information Flow Analysis of Combined Simulink/Stateflow Models. Inf. Technol. Control. **48**(2), 299–315 (2019)

[28] Sihler, F., Tichy, M., Pietron, J.: One-Way Model Transformations in the Context of the Technology-Roadmapping Tool IRIS. Journal of Object Technology **22**(2), 2–114 (2023). The 19th European Conference on Modelling Foundations and Applications (ECMFA 2023)

[29] Lobrano, G.: Making Sense of Competition Law Compliance For, A Practical Guide for SMEs. Business Europe, Brussels, Belgium (2017)

[30] López, J.A.H., Cuadrado, J.S.: An efficient and scalable search engine for models.

Software and Systems Modeling **21**(5), 1715–1737 (2022)

[31] Reza, S.M., Badreddin, O., Rahad, K.: ModelMine: a tool to facilitate mining models from open source repositories. In: Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings. MODELS '20. Association for Computing Machinery, New York, NY, USA (2020)

[32] Munk, P., Nordmann, A.: Model-based safety assessment with SysML and component fault trees: application and lessons learned. Software and Systems Modeling **19**(4), 889–910 (2020)

[33] Boll, A., Kehrer, T., Goedicke, M.: SMOKE: Simulink Model Obfuscator Keeping Structure. In: Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems. MODELS Companion '24, pp. 41–45. Association for Computing Machinery, New York, NY, USA (2024)

[34] Klee, H., Allen, R.: Simulation of Dynamic Systems with MATLAB and Simulink, 3rd edn. Crc Press, Boca Raton, FL, USA (2018)

[35] Di Ruscio, D., Kolovos, D., Lara, J., Pierantonio, A., Tisi, M., Wimmer, M.: Low-code development and model-driven engineering: Two sides of the same coin? Software and Systems Modeling **21**(2), 437–446 (2022)

[36] Liggesmeyer, P., Trapp, M.: Trends in Embedded Software Engineering. IEEE Software **26**(3), 19–25 (2009)

[37] Pajic, M., Jiang, Z., Lee, I., Sokolsky, O., Mangharam, R.: From Verification to Implementation: A Model Translation Tool and a Pacemaker Case Study. In: 2012 IEEE 18th Real Time and Embedded Technology and Applications Symposium, pp. 173–184 (2012)

[38] Badi, N., Khasim, S., Al-Ghamdi, S.A., Alatawi, A.S., Ignatiev, A.: Accurate modeling and simulation of solar photovoltaic panels with Simulink-MATLAB. Journal of Computational Electronics **20**(2), 974–983 (2021)

[39] Tomar, V., Chitra, A., Krishnachaitanya, D., Rao, N.R., Raziasultana, W., *et al.*: Design of powertrain model for an electric vehicle using MATLAB/Simulink. In: 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1–7 (2021). IEEE

[40] Faisal, M., Abbas, S., Ahmad, F., Abbas, Z.: Maximizing wind turbine efficiency using MATLAB SIMULINK with integrated PMSG and MPPT. Next Sustainability **6**, 100200 (2025) https://doi.org/10.1016/j.nxsust.2025.100200

[41] Sánchez, B., Zolotas, A., Rodriguez, H.H., Kolovos, D., Paige, R.: On-the-fly Translation and Execution of OCL-like Queries on Simulink Models. In:

2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS), pp. 205–215 (2019). IEEE

[42] Ceccato, M., Di Penta, M., Falcarin, P., Ricca, F., Torchiano, M., Tonella, P.: A family of experiments to assess the effectiveness and efficiency of source code obfuscation techniques. Empirical Software Engineering **19**, 1040–1074 (2014)

[43] Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., Verykios, V.: Disclosure limitation of sensitive rules. In: Proc. 1999 Workshop on Knowledge and Data Engineering Exchange (KDEX'99), pp. 45–52 (1999). IEEE

[44] Fill, H.-G.: Using Obfuscating Transformations for Supporting the Sharing and Analysis of Conceptual Models. In: Robra-Bissantz, S., Mattfeld, D. (eds.) Multikonferenz Wirtschaftsinformatik 2012 - Teilkonferenz Modellierung Betrieblicher Informationssysteme. GITO Verlag, Braunschweig (2012)

[45] Nacer, A.A., Goettelmann, E., Youcef, S., Tari, A., Godart, C.: Obfuscating a business process by splitting its logic with fake fragments for securing a multi-cloud deployment. In: 2016 IEEE World Congress on Services (SERVICES), pp. 18–25 (2016). IEEE

[46] Martínez, S., Gerard, S., Cabot, J.: On the Need for Intellectual Property Protection in Model-Driven Co-Engineering Processes. In: Reinhartz-Berger, I., Zdravkovic, J., Gulden, J., Schmidt, R. (eds.) Enterprise, Business-Process and Information Systems Modeling, pp. 169–177. Springer

[47] Weber, T., Weber, S.: Model Everything but with Intellectual Property Protection - The Deltachain Approach. In: Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems. MODELS '24, pp. 49–56. Association for Computing Machinery, New York, NY, USA (2024)

[48] Blochwitz, T., Otter, M., Åkesson, J., Arnold, M., Clauss, C., Elmqvist, H., Friedrich, M., Junghanns, A., Mauss, J., Neumerkel, D., *et al.*: Functional mockup interface 2.0: The standard for tool independent exchange of simulation models. In: 9th International Modelica Conference, pp. 173–184 (2012). The Modelica Association

[49] Gupta, N., Chen, F., Tsoutsos, N.G., Maniatakos, M.: ObfusCADe: Obfuscating Additive Manufacturing CAD Models Against Counterfeiting: Invited. DAC '17. Association for Computing Machinery, New York, NY, USA (2017)

[50] Zhou, M., Gao, X., Wu, J., Grundy, J., Chen, X., Chen, C., Li, L.: ModelObfuscator: Obfuscating Model Information to Protect Deployed ML-Based Systems. In: Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis. ISSTA 2023, pp. 1005–1017. Association for Computing Machinery

[51] Collberg, C., Thomborson, C., Low, D.: A Taxonomy of Obfuscating Transformations. Technical Report 148, University of Auckland, New Zealand (1997)

[52] Boll, A., Rani, P., Schultheiß, A., Kehrer, T.: Beyond code: Is there a difference between comments in visual and textual languages? Journal of Systems and Software **215**, 112087 (2024)

[53] Bubenik, R.G.: Optimistic computation. PhD thesis, Rice University (1990)

[54] Shrestha, S.L., Chowdhury, S.A., Csallner, C.: Replicability Study: Corpora For Understanding Simulink Models & Projects. In: 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 1–12 (2023)

[55] Yu, Z., Yang, Y., Su, Z., Wang, R., Tao, Y., Jiang, Y.: Knight: Optimizing Code Generation for Simulink Models With Loop Reshaping. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **44**(2), 444–457 (2025) https://doi.org/10.1109/TCAD.2024.3438691

[56] Zhang, J., Ghobari, D., Sabetzadeh, M., Nejati, S.: Simulink Mutation Testing using CodeBERT. In: 2025 IEEE/ACM International Conference on Automation of Software Test (AST), pp. 24–28 (2025). https://doi.org/10.1109/AST66626.2025.00009

[57] Selonen, P., Kettunen, M.: Metamodel-based inference of inter-model correspondence. In: 11th European Conference on Software Maintenance and Reengineering (CSMR'07), pp. 71–80 (2007). IEEE

[58] Goedicke, M., Lucke, U.: Research Data Management in Computer Science - NFDIxCS Approach. In: Demmler, D., Krupka, D., Federrath, H. (eds.) 52. Jahrestagung der Gesellschaft Für Informatik, INFORMATIK 2022, Informatik in Den Naturwissenschaften, 26. - 30. September 2022, Hamburg. LNI, vol. P-326, pp. 1317–1328. Gesellschaft für Informatik, Bonn, Bonn, Germany (2022)

[59] Laban, F.A., Bernoth, J., Goedicke, M., Lucke, U., Striewe, M., Wieder, P., Yahyapour, R.: Establishing the Research Data Management Container in NFDIxCS. In: Sure-Vetter, Y., Goble, C.A. (eds.) 1st Conference on Research Data Infrastructure - Connecting Communities, CoRDI 2023, Karlsruhe, Germany, September 12-14, 2023. TIB Open Publishing, New York City, NY, USA (2023)