Feature-Interaction Aware Configuration Prioritization for Configurable Code

Son Nguyen*, Hoan Nguyen†, Ngoc Tran*, Hieu Tran*, and Tien N. Nguyen*
*Computer Science Department, The University of Texas at Dallas, USA,
†Amazon Corporation, USA, Email: nguyenanhhoan@gmail.com

Abstract—Unexpected interactions among features induce most bugs in a configurable software system. Exhaustively analyzing all the exponential number of possible configurations is prohibitively costly. Thus, various sampling techniques have been proposed to systematically narrow down the exponential number of legal configurations to be analyzed. Since analyzing all selected configurations can require a huge amount of effort, fault-based configuration prioritization, that helps detect faults earlier, can yield practical benefits in quality assurance. In this paper, we propose CoPro, a novel formulation of feature-interaction bugs via common program entities enabled/disabled by the features. Leveraging from that, we develop an efficient feature-interaction-aware configuration prioritization technique for a configurable system by ranking the configurations according to their total number of potential bugs. We conducted several experiments to evaluate CoPro on the ability to detect configuration-related bugs in a public benchmark. We found that CoPro outperforms the state-of-the-art configuration prioritization techniques when we add them on advanced sampling algorithms. In 78% of the cases, CoPro ranks the buggy configurations at the top 3 positions in the resulting list. Interestingly, CoPro is able to detect 17 not-yet-discovered feature-interaction bugs.

Keywords—Configurable Code, Feature Interaction; Configuration Prioritization; Software Product Lines;

I. INTRODUCTION

Several software systems enable developers to configure to different environments and requirements. In practice, a highly-configurable system can tailor its functional and non-functional properties to the needs and requirements of users. It does so via a very large number of configuration options [9], [10] that are used to control different features [5], [29]. For example, Linux Kernel supports more than 12,000 compile-time configuration options, that can be configured to generate specific kernel variants for billions of scenarios.

In a configurable system, features can interact with one another in a non-trivial manner. As a consequence, such interaction could inadvertently modify or influence the functionality of one another [59]. Unexpected interactions might induce bugs. In fact, most configuration-related bugs are caused by interactions among features [2], [25], [42], [46], [56]. Unfortunately, traditional methods cannot be directly applied to work on configurable code since they focus on detecting bugs in a particular variant. Furthermore, exhaustively analyzing the systems is infeasible due to the exponential number of all possible configurations. In practice, configuration testing is often performed in a manual and ad-hoc manner by unsystematically selecting common variants for analysis [26], [39].

To systematically perform quality assurance (QA) for a highly-configurable system (Figure 1), researchers have proposed several techniques to narrow the configuration space by eliminating invalid configurations that violate the feature model of the system, which defines the feasible configurations via the constraints among the features [17], [18], [27], [30], [29], [50]. However, the number of configurations that need to be tested is still exponential. To address this explosion problem, researchers introduce various configuration selection strategies. The popular strategies include the sampling algorithms which achieve feature interaction coverage such as combinatorial interaction testing [48], [47], [28], [40], one-enabled [42], one-disabled [2], most-enabled-disabled [52], statement-coverage [53], to reduce the number of configurations to be analyzed. Still, those algorithms assume the chances of detecting interaction bugs are the same for all those combinations. Thus, interaction faults might be discovered only after the last variants in such samples is tested. Thus, after configuration selection, the selected set of configurations need to be prioritized for QA activities [3]. Note that configuration prioritization is different from test case prioritization because after configuration prioritization, any QA activities can be applied on the ranked list of prioritized configurations, including test generation and testing, static bug detection, or manual code review (Figure 1).

To motivate configuration prioritization, let us take an example of Linux Kernel. In Linux, the number of different configuration options is over 12,000, leading to $+2^{120}$ different configurations. After applying all the constraints on various combinations of options, the number of valid configurations for QA is an exponential number. For configuration selection, by using six-wise sampling algorithm, the number is still extremely large, up to $+500K$ configurations [42]. Hence, without configuration prioritization, many bugs that are dependent on configurations might still be hidden due to this large configuration space, especially when the resources for QA (e.g., time and developers’ efforts) are limited.

In practice, developers even do not perform QA activities on a particular configuration until it was reported to have defects by the users. In this case, users have already suffered the consequences of those defects. Due to the large number of configurations after selection for QA, even compile-time
errors and flaws cannot be quickly detected by a compiler or a bug detection in the appropriate configuration. Indeed, in the Variability Bugs Database (VBDb) [2], a public database of real-world configuration-related bugs reported for the Linux kernel, there are 42 out of 98 bugs and flaws that are compile-time: 25 declaration errors, 10 type errors, and 7 cases of dead code.

Despite the importance of configuration prioritization, the state-of-the-art methods for such prioritization are still limited in detecting feature-interaction bugs. The similarity-based prioritization method (SP) [3] is based on the idea that dissimilar test sets are likely to detect more defects than similar ones [3]. In SP, the configuration with the maximum number of features is selected to be the first one under test. The next configuration under test is the configuration with the lowest feature similarity compared to the previously chosen one. Despite its success, there are two key problems with SP. First, SP aims to cover as many features different from the previous ones. The different features to be considered next might not be the ones that potentially causes violations. SP does not examine the interaction between features, which is the key aspect causing interaction bugs in a variant. Second, in SP, the quality of the resulted prioritization order strongly depends on the selection of the first configuration.

In this paper, we propose CoPRO, a novel configuration prioritization approach for configurable systems by analyzing their code to detect feature-interaction bugs. Our key idea in CoPRO is as follows. In a configurable system, features are implemented as blocks of code, which are expressed via the program instructions/operations (e.g., declarations, references, assignments, etc.) on the data structures/program entities (e.g., variables, functions, etc.). Features interaction occurs when the operations on the program entities enabled shared between the features have impacts on each other.

Those operations, when the features are enabled or disabled, potentially create a violation(s) that makes the program not-compileable or having a run-time error. Detecting feature interactions via operations would help identify potential feature-interaction bugs. An example of a violation is that a feature disables the only initialization of a variable while another enables one of its dereferences (the violation of “dereferencing an un-initialized variable”). This violation could lead to a NULL pointer exception. It is clear that the configuration in which the former feature is disabled and the latter is enabled, is more suspicious than the one where both of them are either enabled or disabled. The suspiciousness of a configuration is indicated via the potential feature-interaction violations. Hence, a higher number of potential violations makes the configuration more suspicious. The suspiciousness levels are used to rank the configurations, which helps testing, bug detection, or other QA activities more efficient.

We conducted several experiments to evaluate CoPRO in two complementary settings. First, in a benchmark setting, we ran CoPRO on the Variability Bugs Database (VBDb) [2]. We compared CoPRO with the two state-of-the-art approaches in random prioritization and similarity-based prioritization (SP) [3], when we added each of the compared techniques on top of several state-of-the-art sampling algorithms [42]. We found that CoPRO significantly outperforms the other techniques. In 78.0% of the cases, CoPRO ranks the buggy configurations at the top-3 positions in the list, while the SP and Random approaches rank them at the top-3 positions for only 41.3% and 26.1% of the cases. Interestingly, CoPRO was able to detect 17 feature-interaction bugs that were not yet discovered in VBDb including high-degree interaction bugs, memory leaking bugs, etc. In the second setting, we connect CoPRO with a compiler to run on large, open-source configurable systems, and CoPRO can detect 4 newly discovered bugs and programming flaws.

In summary, in this paper, our main contributions include:
- A formulation of feature-interaction bugs using common program entities enabled/disabled by the features;
- CoPRO: an efficient feature-interaction-aware configuration prioritization technique for configurable systems;
- An extensive experimental evaluation showing the effectiveness of CoPRO over the state-of-the-art approaches.

II. MOTIVATING EXAMPLE

In this section, we illustrate the challenges of configuration prioritization and motivate our solution via an example.
A. Examples of Bugs in Configurable Code

Let us consider the simplified version of the real buggy code in the Linux kernel [2] at the commit 40410715715 of Linus-stable at https://git.kernel.org (shown in Figure 2). This version has more than 5,200 compile-time configuration options and about 30,000 files. The code in Figure 2 contains two feature-interaction bugs that were discovered in certain configurations:

- A compile-time error occurs (use the undeclared function of_platform_populate on line 24) in the variants where CONFIG_TWL4030_CORE, CONFIG_OF_DEVICE, and CONFIG_SPARC are enabled.
- A run-time error occurs (dereferencing the NULL pointer ops on line 12) in the configurations where CONFIG_TWL4030_CORE, CONFIG_OF_DEVICE are enabled, and CONFIG_SPARC and CONFIG_OF_IRQ are disabled.

For this example in Linux kernel, brute-force testing of all possible variants to discover these interaction bugs faces the problem of combinatorial explosion in the exponential configuration space (up to $2^{25,200}$ possible configurations). With a huge number of configurations and without an assessment of the potential buggy level of those configurations, the QA process (e.g., debugging) will require a great deal of effort from developers. To deal with such large number of configurations, first, one will eliminate the invalid configurations that violate the constraints among the features in the system [17], [18], [27], [29], [30], [50]. However, the number of configurations after this step is still exponential. To balance between bug detection rate and the number of configurations to be examined, the configuration selection process is applied. An example of selection algorithms is the k-way combinatorial approach [28], [40], [47], [48], which considers the system under test as a blackbox and selects a subset with at most k features. However, even with a small value of k, e.g., k = 6, inspecting a very large number of selected configurations without prioritizing the variants most likely having defects is still inefficient. Therefore, one would need a prioritization strategy to rank the configurations to be examined.

The current state-of-the-art configuration prioritization algorithm is the similarity-based configuration prioritization (SP) [3]. Unfortunately, SP is still ineffective in detecting feature-interaction bugs. Let us illustrate this via our example. Table I shows the partial set of configurations chosen by 4-wise sampling algorithm and prioritized by SP [3]. The variant, where TWL4030_CORE, IRQ_DOMAIN, OF_IRQ, and OFDEVICE are enabled, and SPARC is disabled, with the maximum number of features is selected to be examined first by the SP algorithm. For the next configuration, the configuration that has the minimum number of similar features compared to the previously selected configuration is picked (i.e., the one in which TWL4030_CORE, IRQ_DOMAIN, OF_IRQ, and OF_DEVICE are disabled, and SPARC is enabled). Although this second configuration is most dissimilar to the first one, it does not contain the features whose interactions cause violations, and there is no bug revealed by the second configuration. As a result, by SP’s strategy, the result is not an efficient order for inspection because the aforementioned compile-time and run-time errors are not detected until the 4th and 6th configurations are inspected respectively. The configuration with both interaction bugs would only be discovered via the 7th configuration. In our experiment (will be presented in Section V), 36.2% of the feature-interaction bugs in the public benchmark, the Variability Bugs Database (VBDb) [2], cannot be revealed until at least 10 configurations are inspected in the resulting list ranked by the SP approach.

B. Observations

Let us consider the code in Figure 2 with the two following feature interactions that can cause the violations of program semantics: 1) the declaration of the function of_platform_populate in feature L (line 5) and its use in Z (line 24), and 2) the assignment of ops in feature Y (line 19) and its reference in K (line 22). There are two potential bugs: 1) the use of the function of_platform_populate without its declaration; and 2) the reference to the variable ops without its initiation. The configuration that enables Z and disables L (CONFIG_OFDEVICE=T, CONFIG_SPARC=T) and enables K and disables Y (CONFIG_TWL4030_CORE=T, CONFIG_OF_IRQ=F) should be inspected earlier to detect the two bugs. Based on this observation, those interactions between features should be comprehended to quickly discover these above interaction bugs. That motivates us to propose an approach that first analyzes the source code to more precisely detect the potential interactions among features, and then assesses the probabilities to be faulty of the configurations to prioritize to inspect/test them in a more efficient order.

C. CoPRO Overview

Detecting all interactions among features in a sound and complete manner requires an analysis on all combinations of configuration options. That is prohibitively expensive and impractical. To deal with this problem, we statically analyze the source code to locally and heuristically identify the interactions between features via the shared program entities and the operations on them. For example, T shares the function of_platform_populate with Z (which is declared on line 5 and used on line 24) and K interacts with Y via the variable ops (which is assigned on line 19 and referred to on line 22). Importantly, the operations such as declaration, assignment, or references on the shared entities could become invalid.
when certain features (via configuration options) are enabled or disabled. As a consequence, that could lead to a violation.

For instance, a violation occurs when `CONFIG_TWBL4030CORE`, `CONFIG_OF_DEVICE`, and `CONFIG_SPARC` are enabled because the function `platform_populate` would be used (K and Z are enabled) while its declaration is turned off (L is disabled). The other violation occurs in the case that `CONFIG_TWBL4030CORE` is true, that enables K, while Y is disabled as `CONFIG_OF_IRQ` is disabled. This would induce the bug of dereferencing the NULL pointer on variable `ops` (line 12). With our strategy, the 7th variant in Table I is more suspicious than the 4th, 6th, and any other ones. Generally, the suspiciousness of a variant is determined by the number of violations that it might induce. Finally, a configuration can be ranked according to its suspiciousness score, thus, we could create a prioritization order of variants that maximizes fault detection rate.

III. FORMULATION

Let us formulate the problem of feature-interaction-aware configuration prioritization.

A. Program Entities and Operations

In a program, we are interested in the program entities and the operations performed on them.

Definition 1. (Program Entity). A program entity is a program element that is uniquely identified with a name and a scope. The scope and the name of an entity are used together as the identity of the entity.

In our formulation, we are interested in two types of program entities: variable and function. An entity is represented in the form of `{scope,ent_name}`, where `scope` and `ent_name` are the scope and the name of the program entity respectively. For example, the code in Figure 2 contains the variables `GLOBAL.irq_domain_simple_ops`, `twl_probe.ops`, the function `GLOBAL.twl_probe`, etc.

We define 4 types of operations on variables and functions.

Definition 2. (Program Operation). We define four types of operations on variables: declare, assign, use, and destruct. Let `OP` be the set of program operations, `OP` = `{declare, assign, use, destruct}`. Of all the four operations are applicable to variables, while declare and use are only applicable on functions.

For variables, the assign operation is used to assign a non-null value to a variable. A NULL assignment to a variable is treated as a special case of an assignment. In Figure 2, function `GLOBAL.of_platform_populate_probes` is declared at line 5, and `used` at line 24. `twl_probe.ops` is declared (line 17), assigned a value (line 19), and then used/referred to (line 22).

B. Configurations and Features

A configurable system contains several segments of code that are present in any variant that implements its basic functionality. Those segment form the core of the system.

In practice, a configurable system usually provides a large number of configuration options to configure several optional segments of code to be present or absent, in addition to the core of the system. Those optional segments of code are aimed to realize the optional features of the system. For example, in the Linux Kernel, the configuration options have the prefix of `CONFIG_`, and they can have different values. Without loss of generality, we assume that the value of a configuration option is either true(T) or false(F). (We can consider the entire conditional expressions of non-boolean options as boolean ones, e.g., `CONFIG.A=10` as `CONFIG.A=10=T/F`).

Definition 3. (Configuration Option). A configuration option (option for short) is an element that is used to configure the source code of a configurable system, such that the option’s value determines the presence or absence of one or more segments of code.

In a configurable system, the presence or absence of code segments is dependent on the values of multiple options. In Figure 2, the lines 19 and 20 are presented only when both `CONFIG_TWBL4030CORE` and `CONFIG_OF_IRQ` are `true`. Thus, at line 19, `irq_domain_simple_ops` is potentially used to assign as a value to the variable `ops` when both of those options are `true`.

Definition 4. (Selection Functions). In a configurable system, we define selection functions as the functions from `O × V → 2P`, where `O` is the set of configuration options, `V` = `{T, F}`, and `P` is the set of program entities used in the code of the configurable system. We define four selection functions:

• `α : O × V → 2P`, where `α(o,v) = D`, where `o ∈ O, v ∈ {T, F}`, and `D` is the set of entities potentially declared if `o = v`.
• `β : O × V → 2P`, where `β(o,v) = D`, where `o ∈ O, v ∈ {T, F}`, and `D` is the set of entities potentially assigned if `o = v`.
• `γ : O × V → 2P`, where `γ(o,v) = D`, where `o ∈ O, v ∈ {T, F}`, and `D` is the set of entities potentially used if `o = v`.
• `δ : O × V → 2P`, where `δ(o,v) = D`, where `o ∈ O, v ∈ {T, F}`, and `D` is the set of entities potentially destructed if `o = v`.

For example, in Figure 2:

• `α(CONFIG_SPARC, F)={GLOBAL.of_platform_populate, of_platform_populate.node}`
• `β(CONFIG_OF_IRQ, T)={twl_probe.ops}`
• `γ(CONFIG_OF_IRQ, T)={GLOBAL.irq_domain_simple_ops}`

Definition 5. (Configuration). Given a configurable system, a configuration is a specific selection of configuration options, which defines a variant of the system.

Configuration options are used to control the features that are represented by certain segments of code. For example, in Figure 2, the feature represented by the segment of code `X` (feature `X`) is enabled if the value of the configuration option `CONFIG_OF_IRQ_DOMA IN` is true, whereas feature `Y` is enabled if both `CONFIG_OF_IRQ` and `CONFIG_TWBL4030CORE` are true.

Definition 6. (Feature). In a configurable system, a feature `f` is implemented by applying program operations on a set of program entities, whose presence/absence is controlled by certain configuration options. We denote it by `f ∼ OP`
where OP is the set of program operations and ρ is the set of program entities.

A special case of features is that \( f \) is the core feature (F), \( A \cup B \cup \Gamma \cup \Delta = \rho \), where \( A, B, \Gamma, \Delta \) are the sets of program entities that are declared, assigned, used and destructed in the core system. \( F \) is not controlled by any configuration option.

C. Feature Interactions

In a configurable system, a feature may influence or modify (often called interact with) the functions offered by other features through shared program entities that are used to implement the features. For example, features \( X, K \) and \( Z \) interact with one another via the variables \( \text{GLOBAL.irq_domain.simple.ops} \) and \( \text{twl.probe.temp} \). The manners the features interacting with each other depend on how the shared entities are operated. For example, feature \( Y \) assigns \( \&\text{irq_domain.simple.ops} \) to \( \text{ops} \) and feature \( K \) uses that variable (line 22). If no assignment was done in \( Y \), dereferencing in \( K \) would be invalid, causing a NULL pointer exception.

Multi-way feature-interaction. We present only on the interactions between pairs of features because the interactions between more than two features can be modeled as the operations on the shared variables between pairs of features. Let us provide a sketch of the proof for this statement. We assume that there exists an interaction among \( m \) features \((m > 2)\). For simplicity, we consider the case of \( m = 3 \), and the interaction among \( f_1 \sim OP \times p_1, f_2 \sim OP \times p_2 \) and \( f_3 \sim OP \times p_3 \). There are two cases of this interaction. First, there exists an entity that shared by all 3 features, \( p_1 \cap p_2 \cap p_3 = \omega \). Since \( p_1 \cap p_2 \supset \omega \) and \( p_2 \cap p_3 \supset \omega \), identifying interactions between pairs directly captures the interaction among 3 features. The second case is that \( p_1 \cap p_2 = \omega_1, p_2 \cap p_3 = \omega_2 \) and \( \omega_1 \cap \omega_2 = \emptyset \). Meanwhile, \( f_3 \) is influenced by \( f_1 \) (because the roles of \( f_1 \) and \( f_3 \) features in this case are equal). This leads to that there exist entities: \( e_1 \in \omega_1, e_2 \in \omega_2 \), such that \( e_2 = p(e_1) \), where \( p \) is a propagation function. This means the value of \( e_1 \) is propagated to \( e_2 \), and that influences \( f_3 \). Hence, the interaction among 3 features is still captured by determining interactions between pairs of features.

For instance, the interaction among features \( X, K \) and \( Z \) can be broken down into the shared program entities between two pairs of features as follows: \((X, K)\) via the variable \( \text{GLOBAL.irq_domain.simple.ops} \), and \((K, Z)\) via the variable \( \text{twl.probe.temp} \). Thus, our solution can still model the interactions with more than two features via the operations on their shared program entities. From now on, we refer to a feature interaction as an interaction determined via the shared program entities between a pair of features.

In CoPRO, we focus on the feature interaction through the shared program entities. The feature interactions when the variables are associated with the external data such as when they interfere with each other’s behaviors on files or databases are beyond the scope of our static analysis-based solution. Similarly, we will not detect the interactions through pointers or arrays in this work. As a consequence, if both features use (refer to) a program entity, they will not change the program’s state. Thus, there is no interaction between two features if they only use shared functions and variables.

With the above design focuses, in CoPRO, the interactions between two features \( f_1 \sim OP \times p_1 \) and \( f_2 \sim OP \times p_2 \) with \( p_1 \cap p_2 \neq \emptyset \), can be categorized into nine kinds of interactions that are displayed in Table II (the use-use case is eliminated as explained).

D. Feature Interaction Detection

In a configurable system, the features (except the core features of the system) are controlled by certain configuration options. Thus, if there exists an interaction among the features, the interaction will be one of the following:

- **declare-Declare**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \alpha(o_1, v_1) \cap \beta(o_2, v_2) \neq \emptyset \)
- **declare-Assign**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \alpha(o_1, v_1) \cap \beta(o_2, v_2) \neq \emptyset \)
- **Declare-Use**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \alpha(o_1, v_1) \cap \beta(o_2, v_2) \neq \emptyset \)
- **Assign-Declare**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \beta(o_1, v_1) \cap \alpha(o_2, v_2) \neq \emptyset \)
- **Assign-Assign**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \beta(o_1, v_1) \cap \beta(o_2, v_2) \neq \emptyset \)
- **Assign-Use**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \beta(o_1, v_1) \cap \gamma(o_2, v_2) \neq \emptyset \)
- **Use-Declare**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \beta(o_1, v_1) \cap \gamma(o_2, v_2) \neq \emptyset \)
- **Use-Assign**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \beta(o_1, v_1) \cap \delta(o_2, v_2) \neq \emptyset \)
- **Use-Use**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \beta(o_1, v_1) \cap \delta(o_2, v_2) \neq \emptyset \)
- **Destruct-Destruct**, there exist two option \( o_1, o_2 \) and their selected values \( v_1, v_2 \), such that \( \delta(o_1, v_1) \cap \delta(o_2, v_2) \neq \emptyset \)

### Table II

<table>
<thead>
<tr>
<th>Kind of Interaction</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 declare-declare</td>
<td>( \exists e \in p_1 \cap p_2, e ) is declared in both ( f_1 ) and ( f_2 )</td>
</tr>
<tr>
<td>2 declare-assign</td>
<td>( \exists e \in p_1 \cap p_2, e ) is declared in ( f_1 ) and then assigned in ( f_2 )</td>
</tr>
<tr>
<td>3 declare-use</td>
<td>( \exists e \in p_1 \cap p_2, e ) is declared in ( f_1 ) and used in ( f_2 )</td>
</tr>
<tr>
<td>4 declare-destructor</td>
<td>( \exists e \in p_1 \cap p_2, e ) is declared in ( f_1 ) and destructed in ( f_2 )</td>
</tr>
<tr>
<td>5 assign-assign</td>
<td>( \exists e \in p_1 \cap p_2, e ) is assigned in both ( f_1 ) and ( f_2 )</td>
</tr>
<tr>
<td>6 assign-use</td>
<td>( \exists e \in p_1 \cap p_2, e ) is assigned in ( f_1 ) and used in ( f_2 )</td>
</tr>
<tr>
<td>7 assign-destructor</td>
<td>( \exists e \in p_1 \cap p_2, e ) is assign in ( f_1 ) and destructed in ( f_2 )</td>
</tr>
<tr>
<td>8 use-destructor</td>
<td>( \exists e \in p_1 \cap p_2, e ) is used in ( f_1 ) and destructed in ( f_2 )</td>
</tr>
<tr>
<td>9 destruct-destruct</td>
<td>( \exists e \in p_1 \cap p_2, e ) is used in ( f_1 ) and destructed in ( f_2 )</td>
</tr>
</tbody>
</table>

The entities: \( \text{\texttt{GLOBAL.irq_domain.simple.ops}} \) and \( \text{\texttt{twl.probe.temp}} \).
Based on the above rules, our feature-interaction detection algorithm statically analyzes the code and configuration options, and then computes the sets α, β, γ, and δ for any two options o1 and o2. For example, we can detect a declare-declare interaction between 2 features if there exists 2 options o1 and o2, such that α(o1, v1) ∩ α(o2, v2) ≠ ∅, where v1, v2 are their selected values. Other detection rules are similarly derived. For example, because β(CONFIG_0F_IRQ, T) ∩ γ(API, T) = {ops}, there is a potential assign-use interaction among features. Thus, in this case, the actual assign-use interaction among Y and K exists.

For the core feature, if F and other features interact with one another, depending on the kinds of the interaction, there exists a selection v of an option o, such that α(o, v), β(o, v), γ(∅), δ(∅) intersect with A, B, Γ, Δ, i.e., intersecting with the entities in the core. Interactions among core features and others are similarly identified.

In this version of CoPRO, we formulate feature interaction statically through the completed set of operations on the entities that are shared between features. More sophisticated interactions relevant to pointers and external data such as files or databases can be detected by using different data structures in the same principle and using other types of analysis.

IV. CONFIGURATION PRIORITIZATION

A. Overview

In general, to prioritize a given set of configurations under test, our algorithm assigns a suspiciousness score to each configuration. The suspiciousness score is determined via the number of the potential feature-interaction bugs in different kinds that the variant corresponding to that configuration might potentially have.

Feature-interaction bugs can be induced by any kinds of interaction. Table III shows 10 different kinds of feature-interaction bugs that are potentially caused by the respective kinds of interactions listed in Table II of Section III. The interactions in Table III are called sensitive interactions with their suspicious selection of options. A configuration containing a suspicious selection potentially has the corresponding violation. For example, at line 6, if β(o1, v1) ∩ γ(o2, v2) ≠ ∅, there is an assignment-use potential interaction between f1 and f2. When o1 = v1′, o2 = v2, where v1′ ≠ v1, f1 might be disabled while f2 is enabled, which poses a violation of use without assignment. In Figure 2, because α(CONFIG, F) ∩ γ(CONFIG_0F_DEVICE, T) = {GLOBAL_of_platform_populate}, the program might not be compiled if CONFIG = T and CONFIG_0F_DEVICE = T (use without declaration).

B. Detailed Algorithm

The listing 1 shows the pseudo-code of CoPRO, our feature-interaction aware configuration prioritization algorithm. Given a configurable system, we first extract the set of options used in the system. Then, for each selection v of each option o, the sets α(o, v), β(o, v), γ(∅), and δ(∅) are computed via the function CollectProgramEntities (lines 4–5). After that, for each pair of option selections, it detects the potential interactions among the features and checks whether the interactions are sensitive as described in Table III. Sensitive interactions are used to specify suspicious selections. This information is used to compute the suspiciousness score for each configuration after configuration selection (line 16). This score is the number of suspicious selections contained by a configuration, and equal to the number of potential bugs that the corresponding variant might have. Finally, the configurations are ranked descendingly by their suspiciousness scores.

C. Static Analysis

In this version of CoPRO, to compute α, β, γ, and δ for the value v of an option o in CollectProgramEntities, CoPRO analyzes the code by using TypeChef, a variability-aware parser [33]. For a given configurable code, TypeChef is used to analyze and generate the corresponding variability-aware control-flow graph. In a variability-aware control-flow graph, the nodes refer to statements and the edges, which are annotated with the corresponding presence conditions, refer to the possible successor statements (conditional statements). For the example in Figure 2, the successor of the statement at line 22 is the conditional statement at line 24 if CONFIG_0F_DEVICE is on, otherwise the statement at line 26 is the direct successor of the statement at line 22. After that, CoPRO analyzes every conditional statements in the generated control-flow graph to identify the entities that are either declared, defined, used, or destructed in the statement and compute α, β, γ, and δ for the options and its values in the corresponding presence conditions. For the statement at line 24 in Figure 2, if the value of CONFIG_0F_DEVICE is T, the variable status is defined by using PLATFORM_populate and n. This leads to that the variable status is in β(CONFIG_0F_DEVICE,T), and γ(CONFIG_0F_DEVICE,T) contains the function PLATFORM_populate and the variable n.
Table III
DIFFERENT KINDS OF FEATURE-INTERACTION DEFECTS

<table>
<thead>
<tr>
<th>Kind of interaction</th>
<th>Detection condition</th>
<th>Suspicious selection</th>
<th>Potential violation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 declare-declare</td>
<td>(\alpha (n_1, n_2) \cap \beta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Declaration duplication</td>
</tr>
<tr>
<td>2 declare-use</td>
<td>(\alpha (n_1, n_2) \cap \gamma (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Use without declaration</td>
</tr>
<tr>
<td>3 declare-use</td>
<td>(\alpha (n_1, n_2) \cap \gamma (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Unused variables/functions</td>
</tr>
<tr>
<td>4 declare-destruct</td>
<td>(\alpha (n_1, n_2) \cap \delta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Destruction without declaration</td>
</tr>
<tr>
<td>5 declare-assign</td>
<td>(\gamma (n_1, n_2) \cap \beta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Assignment without declaration</td>
</tr>
<tr>
<td>6 assign-use</td>
<td>(\gamma (n_1, n_2) \cap \delta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Use without assignment</td>
</tr>
<tr>
<td>7 assign-destruct</td>
<td>(\gamma (n_1, n_2) \cap \delta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Destruction without definition</td>
</tr>
<tr>
<td>8 destruct-assign</td>
<td>(\beta (n_1, n_2) \cap \delta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Memory leak</td>
</tr>
<tr>
<td>9 destruct-destruct</td>
<td>(\beta (n_1, n_2) \cap \delta (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Destruction duplication</td>
</tr>
<tr>
<td>10 destruct-use</td>
<td>(\delta (n_1, n_2) \cap \gamma (n_2, n_3) \neq \emptyset)</td>
<td>(v_1 = v_2, v_2 = v_3)</td>
<td>Use after destruction</td>
</tr>
</tbody>
</table>

Table IV
CONFIGURATION OPTIONS AND THE VALUES OF 4 SELECTION FUNCTIONS \(\alpha, \beta, \gamma, \text{AND} \delta\) FOR THE EXAMPLE IN FIGURE 2

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>(\gamma)</th>
<th>(\delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OF_IRQ</td>
<td>T</td>
<td>twl_probe.ops</td>
<td>GLOBAL.irq_domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>simple_ops,</td>
<td>domain_add.ops</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLOBAL.irq_domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.add,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.add,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.add</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.ops</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRQ_DOMAIN</td>
<td>F</td>
<td>GLOBAL.of.platform</td>
<td>populate,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.ops,</td>
<td>.node,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.ops,</td>
<td>.pattern.t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OF_DEVICE</td>
<td>T</td>
<td>GLOBAL.twl_probe</td>
<td>.n,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.temp,</td>
<td>.status,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPARC</td>
<td>T</td>
<td>.temp,</td>
<td>.status,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWL4030_CORE</td>
<td>T</td>
<td>.probe.node,</td>
<td>.probe.temp,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.probe.status,</td>
<td>.probe.ops</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table V
TOP-3 CONFIGURATIONS RANKED BY COPRO FOR FIGURE 2

<table>
<thead>
<tr>
<th>Rank by CoPro</th>
<th>Rank by SP</th>
<th>OF_IRQ</th>
<th>IRQ_DOMAIN</th>
<th>OF_x DEVICE</th>
<th>SPARC</th>
<th>TWL4030_CORE</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>2</td>
</tr>
</tbody>
</table>

D. Running Example

Let us illustrate our algorithm via the example shown in Figure 2. COPro computes the sets of the selection functions for each option, and the result is shown in Table IV. Based on the description on Table III, the suspicious selections include:
- \(\text{CONFIG}_{-}\text{OF}_{-}\text{IRQ}=\text{F}, \text{CONFIG}_{-}\text{TWL4030}_{-}\text{CORE} = \text{T}\)
- \(\text{CONFIG}_{-}\text{SPARC}=\text{T}, \text{CONFIG}_{-}\text{TWL4030}_{-}\text{CORE} = \text{T}\)
- \(\text{CONFIG}_{-}\text{SPARC}=\text{T}, \text{CONFIG}_{-}\text{OF}_{-}\text{DEVICE} = \text{T}\)
- \(\text{CONFIG}_{-}\text{IRQ}_{-}\text{DOMAIN}=\text{F}, \text{CONFIG}_{-}\text{OF}_{-}\text{IRQ} = \text{T}\)
- \(\text{CONFIG}_{-}\text{IRQ}_{-}\text{DOMAIN}=\text{F}, \text{CONFIG}_{-}\text{TWL4030}_{-}\text{CORE} = \text{T}\)

Based on the suspicious selections, COPro assigns the suspiciousness scores and ranks all the configurations accordingly. Table V shows the ranked configurations for our example with their corresponding scores. The top-ranked configuration by COPro is the \(7^{th}\) configuration in the order generated by the ACTS tool [58], a combinatorial test generation tool (see Table I). The configuration covers both interaction bugs. Thus, after inspecting/testing the first configuration, those two bugs will be detected. In other words, COPro effectively ranks higher the potential buggy variant than the SP algorithm.

V. EMPIRICAL EVALUATION

To evaluate our configuration prioritization approach, we sought to answer the following:

RQ1 [Performance against a benchmark]. How does COPro perform on Variability Bugs Database (VBDb) [2], a public dataset of bugs in configurable code?

RQ2 [Comparison]. How does COPro improve over the baseline random prioritization and similarity-based prioritization [3] approaches when we add each of them on top of advanced sampling configuration selection algorithms?

RQ3 [Performance in the wild]. How does it perform on not-yet discovered interaction bugs in configurable systems?

RQ4 [Time Complexity] What is COPro’s running time?

To answer RQ1 and RQ2, we conducted an experiment to evaluate COPro in a controlled environment with the VBDb public benchmark of configuration-related bugs [2]. Answering RQ2 helps evaluate how much improvement COPro gains over the random prioritization and the state-of-the-art similarity-based prioritization [3], when adding COPro on top of the advanced configuration selection techniques [42]. We answer RQ3 to evaluate COPro in the real-world setting. While the bug detection tools cannot directly work on configurable code, with COPro, we run them on the list of suspicious configurations ranked by COPro.

A. SUBJECT SYSTEMS

To evaluate COPro, we used two datasets in two different experiments. To answer RQ1 and RQ2, we used the...
Variability Bugs Database (VBDb) [2] as a benchmark. This publicly available bug database has 98 manually verified configuration-related bugs in different versions of highly-configurable systems: the Linux kernel [38], BusyBox [13], Marlin [41], and Apache HTTPD [4]. Because the VBDb contains configuration-related bugs other than feature-interaction ones, we kept only 46 feature-interaction bugs in those systems. Table VI shows their information including the minimum and maximum numbers of configuration options (MinOpt, MaxOpt), the minimum and maximum numbers of files (MinFile, MaxFile), and the number of feature-interaction bugs (Bugs).

For the second experiment of RQ3, we selected an open-source configurable system with a long history: libpng [36] v0.89 with 40KLOC in 19 files and 80 options, and xterm [57] v2.24 with 50KLOC in 50 files, and 501 configuration options.

### B. Experimental Procedure

For each known buggy version of a subject system, we chose to include the maximum number of files of 100 and the maximum number of inclusion level of 3 (due to the limitation of the TypeChef tool [32] that we used for variability-aware parsing). We first applied a configuration selection process. That is, to produce the sampled sets of configurations for each buggy version, we ran sampling algorithms to select a subset of configurations. For each buggy system version and a particular sampling algorithm, we ran CoPro on the set of configurations selected by a sampling algorithm. For comparison, we ran the random prioritization and similarity-based prioritization techniques [3] on the same configurations.

To evaluate CoPro on detecting not-yet reported interaction bugs in VBDb, we first ran it on a subject system to achieve the ranked list of the configurations. We also collected and analyzed the sensitive interactions and potential suspicious selections reported by our tool to detect unknown bugs. For the top-ranked configurations in the list with the reported potential suspicious interactions, we used a compiler to detect bugs.

### C. Evaluation Metric

For evaluation, we adopted the Average Percentage Faults Detected (APFD) [51], a widely-used metric in evaluating test prioritization techniques. APFD is originally applied for evaluating the average percentage bugs detected by a test suite. In this work, since we used CoPro with a bug detection tool, we used APFD to measure prioritization effectiveness in terms of the rate of bug detection of a configuration set, which is defined by the following formula:

\[
APFD = 1 - \frac{\sum_{i=1}^{m} CF_i}{n \times m} + \frac{1}{2 \times n}
\]

where \( n \) and \( m \) denote the number of configurations and the number of bugs, respectively. \( CF_i \) is the smallest number of configurations in the list, which is needed to be inspected to detect the \( i^{th} \) bug. The APFD score is from 0 to 1. For the fixed numbers of faults and configurations, the higher APFD, the higher fault-detection rate and the better ranking order.

### D. Effectiveness and Comparison (RQ1 and RQ2)

1) **Comparative Results**: Table VII shows the comparative results in term of the average APFD and average rank (AVG Rank) between CoPro and the state-of-the-art prioritization methods, when we ran all of them on the results of the advanced sampling techniques [42]. As seen, CoPro achieves 2-32% higher APFD (14.9% on average) compared to SP and 5-28% higher (17.8% on average) compared to Random approach. CoPro also achieves much better ranking compared to SP and Random. For example, using CoPro with One-disabled, which is recommended by the authors of VBDb [2], the interaction bugs are revealed after more than 15 configurations on average in the resulting ranked list by CoPro are inspected, instead of more than 37 configurations in the lists prioritized by SP and Random. Especially, in 78.0% of the cases, CoPro ranks the buggy configurations at the top-3 positions in the list, while the SP and Random approaches rank them at the top-3 positions for only 41.3% and 26.1% of the cases.

We can also see that CoPro outperforms the SP and Random prioritization techniques consistently on the resulting configurations selected by various advanced sampling algorithms. That is, if one uses CoPro to rank the configurations selected by advanced algorithms, the inspection order by CoPro is better than those of the SP and Random prioritization. Note that in the case of Most-enabled-disabled [42], for each buggy system, there are only two configurations selected by the sampling algorithm, and 23 out of 46 bugs cannot be revealed by the selected set of configurations. That makes all three prioritization approaches not perform well in this case and achieve nearly equal average APFDs and ranks. In brief, CoPro is able to rank the buggy configuration in a much higher rank than SP and Random approaches. In other words, if we add CoPro as the prioritization technique on top of the most advanced sampling algorithms, we would achieve a more

<table>
<thead>
<tr>
<th>Systems</th>
<th>MinOpt</th>
<th>MaxOpt</th>
<th>MinFile</th>
<th>MaxFile</th>
<th>#Bugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux</td>
<td>3463</td>
<td>5504</td>
<td>18886</td>
<td>34012</td>
<td>43</td>
</tr>
<tr>
<td>BusyBox</td>
<td>439</td>
<td>1449</td>
<td>236</td>
<td>799</td>
<td>18</td>
</tr>
<tr>
<td>HTTPD</td>
<td>602</td>
<td>791</td>
<td>254</td>
<td>426</td>
<td>23</td>
</tr>
<tr>
<td>Marlin</td>
<td>233</td>
<td>713</td>
<td>38</td>
<td>135</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Systems</th>
<th>MinOpt</th>
<th>MaxOpt</th>
<th>MinFile</th>
<th>MaxFile</th>
<th>#Bugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux</td>
<td>3463</td>
<td>5504</td>
<td>18886</td>
<td>34012</td>
<td>43</td>
</tr>
<tr>
<td>BusyBox</td>
<td>439</td>
<td>1449</td>
<td>236</td>
<td>799</td>
<td>18</td>
</tr>
<tr>
<td>HTTPD</td>
<td>602</td>
<td>791</td>
<td>254</td>
<td>426</td>
<td>23</td>
</tr>
<tr>
<td>Marlin</td>
<td>233</td>
<td>713</td>
<td>38</td>
<td>135</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Systems</th>
<th>MinOpt</th>
<th>MaxOpt</th>
<th>MinFile</th>
<th>MaxFile</th>
<th>#Bugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux</td>
<td>3463</td>
<td>5504</td>
<td>18886</td>
<td>34012</td>
<td>43</td>
</tr>
<tr>
<td>BusyBox</td>
<td>439</td>
<td>1449</td>
<td>236</td>
<td>799</td>
<td>18</td>
</tr>
<tr>
<td>HTTPD</td>
<td>602</td>
<td>791</td>
<td>254</td>
<td>426</td>
<td>23</td>
</tr>
<tr>
<td>Marlin</td>
<td>233</td>
<td>713</td>
<td>38</td>
<td>135</td>
<td>14</td>
</tr>
</tbody>
</table>

### Table VII

<table>
<thead>
<tr>
<th></th>
<th>APFD</th>
<th>AVG Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>SP</td>
<td>COPro</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>SP</td>
</tr>
<tr>
<td>Pairwise</td>
<td>0.88</td>
<td>0.75</td>
</tr>
<tr>
<td>Four-wise</td>
<td>0.88</td>
<td>0.94</td>
</tr>
<tr>
<td>Five-wise</td>
<td>0.89</td>
<td>0.93</td>
</tr>
<tr>
<td>One-enabled</td>
<td>0.64</td>
<td>0.69</td>
</tr>
<tr>
<td>One-disabled</td>
<td>0.60</td>
<td>0.56</td>
</tr>
<tr>
<td>Most-enabled-disabled</td>
<td>0.52</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Note that in the case of Most-enabled-disabled [42], for each buggy system, there are only two configurations selected by the sampling algorithm, and 23 out of 46 bugs cannot be revealed by the selected set of configurations. That makes all three prioritization approaches not perform well in this case and achieve nearly equal average APFDs and ranks. In brief, CoPro is able to rank the buggy configuration in a much higher rank than SP and Random approaches. In other words, if we add CoPro as the prioritization technique on top of the most advanced sampling algorithms, we would achieve a more
2) **Further Analysis:** We further studied the cases in which CoPRO correctly ranks the buggy configurations at the top positions. For the cases with correct ranking (1–3), we found that in 77% (30 out of 39) of these bugs, the features interact with one another via shared program entities. Thus, our rules in Section III are applicable to detect the majority of feature-interaction bugs in the public VBDb benchmark.

We also found an interesting scenario of **indirect feature-interactions** that CoPRO detected. In some of those 30 cases, CoPRO identifies sensitive interactions among features indirectly via entities. For example, variable x is initialized in the feature controlled by option A with A=T. x is assigned to y in the feature enabled if the option B is on. Then, y is referred to in another place that controlled by option C, C=T. In this case, if A=F, B=T, and C=T, a **null pointer exception** might be induced. In this case, since of the propagation of variables’ values, the interaction between two features controlled by A and C can be captured by CoPRO via the feature controlled by B. Thus, the buggy configurations are ranked on the top. This also indicates CoPRO’s capability in detecting configurations containing bugs relevant to more than two features.

3) **Examples on Feature-Interaction Bugs:** Let us present the configuration-related bugs involving high-degree feature interactions and the cases that CoPRO detected the feature-interaction bugs not-yet-discovered in the VBDb benchmark.

A bug involving 6 configuration options. Figure 3 shows a bug in Apache HTTPD at commit 2124ff4. The bug is in the file mod.cgid.c. In this example, the bug is observed when RLIMIT_CPU, RLIMIT_NPROC, RLIMIT_DATA, RLIMIT_VMEM, and RLIMIT_AS are disabled, while RLIMIT_NPROC is enabled. With the selections of the combinations of those options, the field limits of any variable of the type cgid_red_t (e.g. req) used in any features is not declared (line 3). Meanwhile, the filed limits is used in req.limits on line 12 when RLIMIT_NPROC is enabled. By identifying the suspicious interactions between the features controlled by the pairs of RLIMIT_NPROC and each of these 5 other options via the field req.limits, CoPRO specifies that the selection that RLIMIT_NPROC = T, RLIMIT_CPU = F, RLIMIT_NPROC = F, RLIMIT_DATA = F, RLIMIT_VMEM = F, and RLIMIT_AS = F is more suspicious than all other selections containing those six configuration options.

Not-yet discovered feature-interaction bugs in VBDb benchmark. Interestingly, while using VDb, we were able to use CoPRO detect the interaction bugs that were neither discovered and reported in those systems nor in VDb. In total, we found 17 such feature-interaction bugs including 12 using-without-declaration bugs, 2 memory-leak bugs, 2 declaration duplication bugs, and 1 dead code issue.

Figure 4 shows 2 not-yet-discovered bugs: a memory leak issue and an assignment without declaration bug at commit fac312d78bf (which also has use without declaration bug and destruction without declaration bug). The assignment without declaration bug occurs only if BB_FEATURE_LS_SORTFILES = F and CONFIG_FEATURE_LS_RECURSIVE = T. In this case, dndirs and dnd are not declared since lines 3–4 are not included, but they are used at lines 11 and 13. Moreover, dnd is destructed on line 15. This bug was fixed at commit ea224b6aah (in almost 6 years later). 3 years after that, a memory leak issue was reported and fixed at commit ffd4774ad25: as CONFIG_FEATURE_LS_RECURSIVE is disabled, the memory controlled by subdnp is initialized at line 9 and not released. With CoPRO, it would have been fixed earlier.

A run-time feature-interaction Bug in Busybox CoPRO is also able to detect run-time errors caused by feature interactions. Figure 5 shows a simplified bug in Busybox extracted from http://vbdb.itu.dk/#bug/busybox/061fd0a. In this case, a bug occurs when CONFIG_FEATURE_HDPARM.-HDIO.UNREGISTER_HWINF = T if c = ”U” and p = NULL. The execution goes to expected_hwinf.error. However, this label is visible only when CONFIG_FEATURE_HDPARM_HDIO_SCAN_HWINF = T. Otherwise, we would have a run-time error.
E. Effectiveness in Detecting Bugs in the Wild (RQ3)

To evaluate the effectiveness of CoPRO on the real-world, open-source projects, we ran it on the configurable systems libpng v0.89 and xterm v2.24 to detect interaction bugs. Interestingly, with CoPRO, we were able to detect 4 interaction bugs that have not been reported/discovered before. They have the same nature of using variables/functions without declarations. Let us discuss two of them in details. The other one can be found on our website [1].

In Figure 6, the code contains 2 bugs. The first one is observed when the option PNG_READ_INTERLACING_SUPPORTED or PNG_WRITE_INTERLACING_SUPPORTED is enabled (line 4) and PNG_INTERNAL is disabled (line 1). In this case, the function png_set_interlace_handling is declared (line 5), and PNG_INTERLACE (line 6) is used inside this function. Meanwhile, the constant PNG_INTERLACE (line 2) is declared only if PNG_INTERNAL is enabled. Thus, if PNG_INTERNAL is disabled, and either PNG_READ_INTERLACING_SUPPORTED or PNG_WRITE_INTERLACING_SUPPORTED is enabled, we will have a compiling error at line 6. The second bug occurs when both PNG_READ_INTERLACING_SUPPORTED and PNG_WRITE_INTERLACING_SUPPORTED are F. In this case, png_read_image uses an undeclared function (line 10).

F. Time Complexity (RQ4)

We run our experiments on a computer with Intel Core i5 2.7GHz processor, 8GB RAM. The running time to analyze the most complex case that contains 43KLOC and 194 configuration options and rank 156 configurations is 211,020ms.

Figure 5. A Run-time Feature-Interaction Bug in Busybox

Figure 6. Two Not-yet-discovered Bugs in libpng

G. Limitations and Potential Solutions

For the cases that CoPRO did not rank well the buggy configurations, we found that the majority of them are not in the kinds of interaction-related defects listed in Section III. For example, a variable x is assigned a value v if option A is enabled, otherwise v=x. Then, x is referred to in a feature controlled by option B. In this case, CoPRO detects the interactions between those features. However, as a static technique, CoPRO could not conclude which option selections are more suspicious. To overcome such limitation, one could use a dynamic analysis approach for configurable code [45].

Figure 7 shows a simplified bug in HTTPD (commit 9327311d30f) that CoPRO did not rank well the buggy configurations. In Figure 7, a use without assignment is exposed when AP is enabled, otherwise x=x. Then, x is referred to in a feature controlled by option B. In this case, CoPRO did not work since there is no feature where rmm_lock is assigned. Consequently, no assign-use interaction exists.

H. Extension to CoPRO

Generally, to detect more kinds of bug such as in the above example, one can extend our set of conditions with the corresponding violations in Table III. One can define a new condition to detect this bug as follows: i) α(APU.Has.LDAP ) ∩ γ(APU.Has.SHARED_MEMORY, T) ∩ = {utilldap_cache_init.rmm_lock} and ii) there is no definition of rmm_lock in its scope, which is the function utilldap_cache_init.

Interestingly, note that for this buggy system, CoPRO ranked the configuration to reveal another flaw of unused variable (rmm_lock) when APU.Has.LDAP=T and APU.Has.SHARED_MEMORY=F.

In 14 cases out of 368 cases, the interactions that cause the interaction bugs are really detected, but the configurations that reveal the bugs are still ranked lower than others. The reason for these cases is that other configurations containing more suspicious selections that actually do not cause the bugs. To faster detect the bug in these situations, one can apply the Additional Prioritization strategy [22] to rank the set of configurations according to their numbers of potential bugs in an incremental manner. By this strategy, the next configuration to be selected is the one containing the largest number of
void <br> k <br> # on <br> rmm <br> configuration se- <br> showed <br> util <br> t <br> approaches to reduce the number of configurations to <br> printf(<br> did not rank well buggy configurations <br> ELA TED <br> similarity-based Prioritization <br> APR <br> is also related to the work on <br> is also related to <br> et al. <br> exception, one must rewrite such an analysis to <br> outperforms the other tech- <br> NULL <br> with an existing bug <br> most-enabled-disabled <br> configuration prioritization, thus, developers need to spend <br> files, and build-system information influence the performance <br> be rewritten in a variability-aware fashion. For example, to <br> detect NULL exception, one must rewrite such an analysis to <br> consider all different configurations in a configurable code. <br> In our experiment, we connect CoP with an existing bug <br> detection tool to work on configurable code. Variability-aware <br> execution [45], [43] explores multiple paths of execution at <br> the same time to detect feature-interaction bugs. However, it <br> suffers scalability issue. <br> Several approaches were proposed to detect feature inter- <br> actions [6], [25]. Verification [7] is also used to detect feature- <br> interaction bugs. Other prioritization approaches aim for state- <br> ment coverage [52], [53] via static checkers. The issue is that <br> computing an optimal solution for the coverage problem is NP- <br> hard, and including each block of optional code at least once <br> does not guarantee that all possible combinations of individual <br> blocks of optional code are considered [42]. To avoid finding <br> optimal coverage solution, the most-enabled-disabled [52] <br> algorithm checks two samples independently of the number <br> of configuration options. When there are no constraints among <br> configuration options, it enables all options and then it disables <br> all configuration options. One-(enabled/disabled) algorithm [2] <br> enables/disables one configuration option at a time. Despite <br> different levels of heuristics, they do not analyze the entities <br> in source code. <br> Several approaches are aimed for testing for configurable <br> systems [14], [20], [26], [39]. In product-line testing [49] and <br> framework testing [15] it is a common strategy to unit test <br> components or plug-ins in isolation, while integration tests are <br> often neglected or performed only for specific configurations. <br> Greiler et al. suggest shipping test cases with plug-ins and <br> running them in client systems [26]. In essence, this postpones <br> tests of configurations until the configuration is actually used. <br> Other approaches have been proposed for static analysis <br> of product lines [11], [12], [16], [19], [21], [31], [54], [55]. <br> Researchers explore to represent and reason about partial <br> but finite configuration spaces compactly with BDDs or SAT <br> solvers (as used in our variability contexts) [8], [31], [44], <br> choices of structures [23] and complex structures [24], [37]. <br> VI. RELATED WORK <br> CoP is most closely related to the work by Al-Hajjaji <br> et al. [3] on similarity-based Prioritization (SP). The key <br> idea of SP approach dissimilar test sets are likely to detect <br> more defects than similar ones [3]. In SP, the configuration <br> with the maximum number of features is selected to be the <br> first one under test and the next configuration must have the <br> minimum number of similar features as the previously selected <br> configuration. In comparison, SP does not analyze the nature <br> of feature interactions, while CoP does. This avoids the <br> problem in SP that the different features to be considered next <br> might not be the ones that potentially causes violations. <br> CoP is also related to the work on configuration selec- <br> tion approaches to reduce the number of configurations to <br> be tested [42]. They focus on the step before configuration <br> prioritization, therefore the resulting set of configurations is <br> not ranked as in CoP. The t-wise (i.e., k-way) sampling <br> algorithm covers all combinations of t options [28], [35], [46], <br> [48], while pair-wise checks all pairs of configuration options <br> [40], [47]. Recent study by Medeiros et al. [42] showed <br> that realistic constraints among options, global analysis, header <br> files, and build-system information influence the performance <br> of most sampling algorithms substantially; and several algo- <br> rithms are no longer feasible in practice. Importantly, they lack <br> configuration prioritization, thus, developers need to spend <br> efforts to perform QA on all the variants. <br> CoP is also related to Variability-aware (VA) analy- <br> sis [37]. VA analysis is a variation of a traditional analysis <br> that considers the variability in the configurable code. The variability-aware analysis techniques have been proposed for <br> type checking [16], [31], [37], [54], model checking [19], [34], <br> data-flow analysis [11], [12], [37], and other analyses [21] on <br> multiple compile-time configurations of a system at a time. <br> The main drawback of this approach is that it cannot reuse <br> existing static analysis tools, and each type of analysis must <br> be rewritten in a variability-aware fashion. For example, to <br> Figure 7. CoP did not rank well buggy configurations
REFERENCES


[57] xterm. [https://invisible-island.net/xterm/](https://invisible-island.net/xterm/).
