On the Verification of SCOOP Programs

Georgiana Caltais[a], Bertrand Meyer[a,b]

[a] *Department of Computer Science, ETH Zürich, Switzerland*
[b] *Software Engineering Lab, Innopolis University, Russia*

Abstract

In this paper we focus on the development of a unifying framework for the formal modelling of an object oriented-programming language, its underlying concurrency model and their associated analysis tools. More precisely, we target SCOOP – an elegant concurrency model, recently formalized based on Rewriting Logic (RL) and Maude. SCOOP is implemented in Eiffel and its applicability is demonstrated also from a practical perspective, in the area of robotics programming. Our contribution consists in devising and integrating an alias analyzer and a Coffman deadlock detector under the roof of the same RL-based semantic framework of SCOOP. This enables using the Maude rewriting engine and its LTL model-checker "for free", in order to perform the analyses of interest. We discuss the limitations of our approach for model-checking deadlocks and provide possible workarounds for the state explosion problem. On the aliasing side, we propose an extension of a previously introduced alias calculus based on program expressions, to the setting of unbounded program executions such as infinite loops and recursive calls. Moreover, we devise a corresponding executable specification easily implementable on top of the SCOOP formalization. An important property of our extension is that, in non-concurrent settings, the corresponding alias expressions can be over-approximated in terms of a notion of regular expressions. This further enables us to derive an algorithm that always stops and provides a sound over-approximation of the "may aliasing" information, where soundness stands for the lack of false negatives.

Keywords: concurrency, SCOOP, operational semantics, alias analysis, deadlock detection, Maude, rewriting logic

1. Introduction

In light of the widespread deployment and complexity of concurrent systems, the development of corresponding frameworks for rigorous design and analysis has been a great challenge. In this paper we focus on the development of a unifying framework for the formal modelling of an object oriented-programming language, its underlying concurrency model and their associated analysis tools.

We are targeting SCOOP [1], a simple object-oriented programming model for concurrency. Two main characteristics make SCOOP simple: 1) just one keyword programmers have to learn and use in order to enable concurrent executions, and 2) the burden of orchestrating concurrent executions is handled within the model, therefore reducing the risk of correctness issues. The reference implementation is Eiffel [2], but implementations have also been built on top of Java-like languages [3]. The success of SCOOP is demonstrated not only from a research perspective, but also from a practical perspective, with applications appearing, for instance, in the area of robotics programming [4].

The basis of a framework for the design and analysis of the SCOOP model has already been set. In this respect, we refer to the recent formalization of SCOOP in [1] based on Rewriting Logic (RL) [5], which is executable and straightforwardly implementable in the programming language Maude [6]. In [1] these

Email addresses: gcaltais@gmail.com (Georgiana Caltais), bertrand.meyer@inf.ethz.ch (Bertrand Meyer)

capabilities have been successfully exploited in order to reason on the original SCOOP model and to identify a number of design flaws.

Moreover, an executable semantics can be exploited in order to formalize and "run" analysis tools for SCOOP programs as well. This facilitates the extension of the aforementioned SCOOP formalization to the level of a unifying executable semantic framework for the design and analysis of both the model and its concurrent applications. In this paper we focus on the *development of a RL-based toolbox for the analysis of SCOOP programs* on top of the formalization in [1]. We are interested in constructing an alias analyzer and a deadlock detector.

Alias analysis has been an interesting research direction for the verification and optimization of programs. One of the challenges along this line of research has been the undecidability of determining whether two expressions in a program *may* reference the same object. A rich suite of approaches aiming at providing a satisfactory balance between scalability and precision has already been developed in this regard. Examples include: (i) intra-procedural frameworks [7, 8] that handle isolated functions only, and their inter-procedural counterparts [8, 9, 10] that consider the interactions between function calls; (ii) type-based techniques [11]; (iii) flow-based techniques [12, 13] that establish aliases depending on the control-flow information of a procedure; (iv) context-(in)sensitive approaches [14, 15] that depend on whether the calling context of a function is taken into account or not; (v) field-(in)sensitive approaches [16, 17] that depend on whether the individual fields of objects in a program are traced or not.

There is a huge literature on heap analysis for aliasing [18], but hardly any paper that presents a calculus allowing the derivation of alias relations as the result of applying various instructions of a programming language. Hence, of particular interest for the work in this paper is the untyped, flow-sensitive, field sensitive, inter-procedural and context-sensitive calculus for may aliasing, introduced in [19]. The calculus covers most of the aspects of a modern object-oriented language, namely: object creation and deletion, conditionals, assignments, loops and (possibly recursive) function calls. The approach in [19] abstracts the aliasing information in terms of explicit access paths [20] referred to as *alias expressions* straightforwardly computed in an equational fashion, based on the language constructs. As we shall see later on in this paper, the language-based expressions can be exploited in order to reason on "may aliasing" in a finite number of steps in non-concurrent settings and, moreover, can be easily incorporated in the semantic rules defining SCOOP in [1].

The work most similar to our contribution is the one in [21], where aliasing information is identified by exploiting regular behaviour of (non-concurrent) programs, in a RL setting. The main difference with the results in [21] consists in how the abstract memory addresses corresponding to pointer variables are represented. In [21] these range over a finite set of natural numbers. In this paper we consider alias expressions build according to the calculus in [19], based on program constructs.

In [22] the authors focus on alias analysis and type inference for PHP and sketch a hypothetical solution of their approach in Maude. The alias analysis in [22] is inter-procedural as well, and can handle function pointers, recursive calls and references. One of the major differences with our work is that in [22], in order to derive correct analysis results, the authors run aliasing and type inference analysis in "tandem" until a fixpoint is reached. We guarantee the termination our analysis by exploiting a sound over-approximation of aliasing via the so-called regular alias expressions.

Deadlock is one of the most serious problems in concurrent systems. It occurs when two or more executing threads are each waiting for the other to finish. Along time, the complexity of the problem determined various approaches to combat deadlocks [23]. Examples include: (i) deadlock prevention [24] which ensures that at least one of the deadlock conditions cannot hold, (ii) deadlock avoidance [25] that provides a priori information so that the system can predict and avoid deadlock situations, (iii) deadlock detection [26, 27] that detects and recovers from a deadlock state.

Our focus is on *deadlock detection* for SCOOP programs. We base our work on the fact that this type of analysis is in strict connection with the underlying model of interest. Consequently, as described in the corresponding subsequent sections, our approach consists in formalizing deadlocks in the context of the SCOOP concurrency model and enriching its semantics in [1] with the equivalent operational-based definition of deadlocks. This enables using the Maude rewriting capabilities "for free" in order to test SCOOP programs for deadlock. Nevertheless, the more ambitious goal of using the Maude LTL model-checker for deadlock detection is not straightforward. As discussed in more detail later on in this paper, verification of deadlocks was possible after reducing the SCOOP semantics in [1] and abstracting it based on aliasing information, and modifying a series of implementation aspects (such as indexed-based parameterizations) that determined state explosion issues.

The literature on using static analysis [8] and abstracting techniques for (related) concurrency models is considerable. We refer, for instance, to the recent work in [46] that introduces a framework for detecting deadlocks by identifying circular dependencies in the (finite state) model of so-called contracts that abstract methods in an OO-language. Nevertheless, the integration of a deadlock analyzer in SCOOP on top of Maude is an orthogonal approach that aims at constructing a RL-based toolbox for SCOOP programs laying over the same semantic framework.

In [47] SCOOP programs are verified for deadlocks and other behavioral properties using GROOVE [48]. The work in [47] proposes a redefinition of the most common features of the SCOOP semantics based on graph transformation systems (GPSs). This is a bottom-up approach, as it aims at redefining the SCOOP semantics from scratch via GPSs, orthogonal to our rather top-down strategy of narrowing the original semantics proposed in [1].

There are several RL-based approaches to deadlock detection in the literature. In [28, 29], for instance, the authors use the Maude LTL model checker to show deadlock-freedom in a dining philosophers implementation in C. In [30], the Maude LTL model checker is used to verify safety and liveness properties of Orc [31] programs such as a dining philosophers implementation and an online auction system. Nevertheless, none of the aforementioned results are used in the context of a unifying framework for specifying and clarifying a concurrency model [1], and analyzing its corresponding applications.

The results in this paper centre around SCOOP as tackling aliasing and deadlock related aspects is particularly challenging in the context of a concurrency model as complex as SCOOP. However, exploiting a unifying RL-based framework for both designing the concurrency model and analyzing its corresponding applications raises certain issues that were not obvious at the stage of just formalizing SCOOP in [1]. Some of the identified design and implementation decision affecting the verification of SCOOP applications in a negative way can serve as guidelines in the context of other RL-based models as well.

This paper is an extended version of [32] where we proposed:

1. a translation of the (finite) alias calculus in [19] to the setting of unbounded program executions such as infinite loops and recursive calls, together with a sound over-approximation technique based on (finitely representable) "regular alias expressions" capturing unbounded executions in non-concurrent settings;
2. a RL-based specification of the extended calculus suitable for integration within the SCOOP formalization in [1] (for this purpose we use a notation inspired by the $\mathbb{K}$-framework enabling compact and modular definitions);
3. an algorithm for "may aliasing" (exploiting the finiteness property in 1.) that always terminates in non-concurrent settings.

The current work adds to 1.–3. above:

4. the full RL-based specification in 2. and the complete formal proofs showing the soundness of the over-approximating technique based on "regular alias expressions";
5. examples of exploiting the algorithm in 3. and its implementation on top of the SCOOP formalization in Maude [1];
6. the formalization and integration of a deadlock detection mechanism on top of the SCOOP operational semantics [1], together with discussions on the limitations of our approach and associated workarounds.

Paper structure. The paper is organized as follows. In Section 2 we provide a brief overview of SCOOP. In Section 3 we introduce the extension of the alias calculus in [19] to unbounded executions. In Section 4 we provide the full RL-based executable specification of the calculus. The implementation in SCOOP and further applications are discussed in Section 5. Section 6 is dedicated to deadlocking in SCOOP. In Section 7 we draw the conclusions and provide pointers to future developments.

2. Brief introduction to SCOOP

As already stated, the purpose of the current work is the development of a toolbox for the analysis of SCOOP programs by exploiting the semantics proposed in [1]. SCOOP is particularly attractive due to its simplicity and elegance, as it allows the switch from sequential to concurrent programming in a rather straightforward fashion, by means of just one keyword, namely, *separate*. Transparent to the user, the key notion in SCOOP is the processor, or *handler* (that can be a CPU, or it can also be implemented in software, as a process or thread). Handlers are in charge of executing the routines of "separate" objects, in a concurrent fashion.

For an example, assume a processor p that performs a call $o.f(a_1, a_2, \ldots)$ on an object o. If o is declared as "separate", then p sends a request for executing $f(a_1, a_2, \ldots)$ to q – the handler of o (note that p and q can coincide). Meanwhile, p can continue. Moreover, assume that $a_1, a_2, \ldots$ are of "separate" types. According to the SCOOP semantics, the application of the call $f(\ldots)$ will *wait* until it has been able to *lock* all the separate objects associated to $a_1, a_2, \ldots$. This mechanism guarantees exclusive access to these objects. Given a processor p, by $W(p)$ we denote the set of processors p *waits* to release the resources p needs for its asynchronous execution. Orthogonally, by $H(p)$ we represent the set of resources (more precisely, resource handlers that) p already acquired.

The semantics of SCOOP in [1] is defined over tuples of shape

$$\langle p_1 :: St_1 \mid \ldots \mid p_n :: St_n, \sigma \rangle \tag{1}$$

where, p_i denotes a processor (for $i \in \{1, \ldots, n\}$), St_i is the call stack of p_i and σ is the *state* of the system. States hold information about the *heap* (which is a mapping of references to objects) and the *store* (which includes the binding of the formal parameters to actual arguments, local variables, *etc.*). Processors communicate via *channels*.

Roughly speaking, one could classify the operational rules formalizing SCOOP in [1] in: a) *language rules* that provide the semantics of language constructs such as "**if** ... **then** ... **else** ... **end**" or "**until** ... **loop** ... **end**", and b) *control rules* implementing mechanisms such as locking or scheduling.

For an example in category a) above, consider the rules specifying "**if**" instructions:

$$\frac{\text{a is fresh}}{\begin{array}{c}\langle p :: \textbf{if } e \textbf{ then } St_1 \textbf{ else } St_2 \textbf{ end}\,; St,\, \sigma \rangle \to \\ \langle p :: \mathrm{eval}(a, e); \mathrm{wait}(a); \mathit{provided\ a.data\ then\ } St_1 \mathit{\ else\ } St_2; St,\, \sigma \rangle\end{array}} \tag{2}$$

$$\frac{\cdot}{\langle p :: \mathit{provided\ true\ then\ } St_1 \mathit{\ else\ } St_2; St,\, \sigma \rangle \to \langle p :: St_1; St,\, \sigma \rangle} \tag{3}$$

$$\frac{\cdot}{\langle p :: \mathit{provided\ false\ then\ } St_1 \mathit{\ else\ } St_2; St,\, \sigma \rangle \to \langle p :: St_2; St,\, \sigma \rangle} \tag{4}$$

Intuitively, "eval(a, e)" evaluates e and puts the result on a fresh channel a and "wait(a)" enables processor p to use the evaluation result stored in *a.data*. It is straightforward to see that, according to (3), in case the condition e is evaluated to *true* then the "**if** branch" St_1 is placed on top of the call stack of p. Otherwise, based on (4), if e is evaluated to $false$, the "**else** branch" is executed.

As we shall see in Section 5, an operational view on the alias calculus in [19] exploiting the instructions of a programming language will enable a straightforward implementation on top of the "language rules" of SCOOP.

For the case b) above we refer to the locking rule:

$$\frac{\cdot}{\begin{array}{c}\langle p :: \mathit{lock}(\{q_1, \ldots\ q_m\}); St,\, \sigma \rangle \to \\ \langle p :: St,\, \sigma.\mathit{lock\_rqs}(p, \{q_1, \ldots\ q_m\}) \rangle\end{array}} \quad \forall q_i \in \{q_1, \ldots, q_m\} : \sigma.\mathit{rq\_locked}(q_i) = \mathit{false} \tag{5}$$

stating that a processor p can lock a set of handlers $\{q_1, \ldots, q_m\}$ by calling *lock_rqs* on the state σ whenever none of the handlers q_i has previously been acquired by other processors, *i.e.*, $\sigma.\mathit{rq\_locked}(q_i) = \mathit{false}$.

As it will become clear in Section 6, "control rules" pave the way to an immediate implementation of a corresponding "deadlock rule" on top of the Maude formalization of SCOOP in [1].

3. The alias calculus

The calculus for *may aliasing* introduced in [19] abstracts the aliasing information in terms of explicit access paths referred to as "alias expressions". Consider, for an example, the case of a linked list. We write x_i $(i \geq 0)$ to represent node i in the list, and use a setter to assign the next node of the list:

```
**create** $x_0$
**loop**
    $i := i + 1$
    **create** $x_i$
    $x_i.set\_next(x_{i-1})$
**end**
```
(6)

The result of the execution of the code above can be intuitively depicted as the infinite sequence:

$$x_0 \xleftarrow{next} x_1 \xleftarrow{next} \ldots x_{k-1} \xleftarrow{next} x_k \xleftarrow{next} x_{k+1} \ldots$$

Hence, x_0 becomes aliased to $x_1.next$, $x_2.next.next$, $x_3.next.next.next$, so on and so on. In short, the set of associated alias expressions can be equivalently written as:

$$\{[x_i,\, x_{i+k}.next^k] \mid i \geq 0 \wedge k \geq 1\}. \tag{7}$$

The sources of imprecision introduced by the calculus in [19] are limited to ignoring tests in conditionals, and to "cutting at length L" for the case of possibly infinite alias relation corresponding to unbounded executions as in (6). The cutting technique considers sequences longer than a given length L as aliased to all expressions.

In this section we define an extension of the calculus in [19], to unbounded program executions. Moreover, based on the idea behind the *pumping lemma for regular languages* [33], we devise a corresponding sound over-approximation of "may aliasing" in terms of regular expressions, applicable in sequential contexts. This paves the way to developing an algorithm for the aliasing problem, as presented in Section 4. The machinery for collecting aliases is defined using a notation inspired by the $\mathbb{K}$ framework [34], as its operational flavor enables a straightforward integration within the SCOOP formalization in [1].

We proceed by recalling the notion of *alias relation* and a series of associated notations and basic operations, as introduced in [19].

Definition 1 (Alias expressions [19])**.** *We call an* alias expression *a (possibly infinite) path of shape $x.y.z.\ldots$, where x is a local variable, class attribute or Current, and $y, z, \ldots$ are attributes. Here,* Current*, also known as* this *or* self*, stands for the current object. For an arbitrary alias expression e, it holds that $e.Current = Current.e = e$.*

Definition 2 (Alias relations [19])**.** *Let E represent the set of all expressions of a program. An* alias relation *is a symmetric and irreflexive binary relation $r \subseteq E \times E$. Let $[e_1, e_2] \in r$; we call $[e_1, e_2]$ an* alias pair *in r.*

Given an alias relation r and an expression e, we define

$$r/e = \{e\} \cup \{x\colon E \mid [x, e] \in r\}$$

denoting the set consisting of all elements in r which are aliased to e, plus e itself.

Let x be an expression; we write $r - x$ to represent r without the pairs with one element of shape $x.e$:

$$r - x = \{[e_1, e_2] \mid \forall e{:}E\ .\ [e_1, e_2] \in r \wedge e_1 \neq x.e \wedge e_2 \neq x.e\}.$$

By $a \in dom(t)$ we refer to an attribute in the class corresponding to the object referred by the alias expression associated to t. For instance, given a class NODE with a field next of type NODE, and a NODE object x, we say that next is in the domain of $t = x.next.next$.

Intuitively, we say that an alias relation is dot-complete*, whenever its elements are soundly extended via fields in their corresponding domains. The operation making a relation dot-complete is defined as follows:*

$$\textit{dot-complete}(r) = r \cup \{[t.a, s.a] \mid [t, s] \in r \wedge a \in \textit{dom}(t)\} \cup \{[s.a, v] \mid [t, s] \in r \wedge [t.a, v] \in r\}.$$

Let x be an alias expression and u be a set of expressions in E. We define the relation r augmented with pairs $[x, y]$, for $y \in u$, and made dot complete as

$$r[x = u] = \textit{dot-complete}(r') \text{ where } r' = r \cup \{[x, y] \mid y \in u\}.$$

3.1. Extension to unbounded executions

We further introduce an extension of the alias calculus in [19] to infinite alias relations corresponding to unbounded executions such as infinite loops or recursive calls. The main difference in our approach is reflected by the definition of loops, which now complies to the usual fixed-point denotational semantics.

The alias calculus is defined by a set of axioms "describing" how the execution a program affects the aliasing between expressions. As in [19], the calculus ignores tests in conditionals and loops.

Definition 3 (Program instructions [19])**.** *Let x, x' be variable names and s a valid sequence $x'.y_1.y_2 \dots y_n$ with $y_1 \in dom(x')$, $y_2 \in dom(x'.y_1)$, …, $y_n \in dom(x'.y_1 \dots y_{n-1})$. Let f be a routine name and l the actual parameters of a call of f. The* program instructions *are defined as follows:*

$$\begin{array}{rcl} p & ::= & p\,;p \mid \textbf{then}\ p\ \textbf{else}\ p\ \textbf{end} \mid \\ & & \textbf{create}\ x \mid \textbf{forget}\ x \mid x := s \mid \\ & & \textbf{loop}\ p\ \textbf{end} \mid \textbf{call}\, f(l) \mid x.\textbf{call}\, f(l). \end{array} \tag{8}$$

We write $r \gg p$ to represent the alias information obtained by executing p when starting with the initial alias relation r.

We further give the axioms of the extended alias calculus. The axiom for sequential composition is defined in the obvious way:

$$r \gg (p\,;q) = (r \gg p) \gg q. \tag{9}$$

Conditionals are handled by considering the union of the alias pairs resulted from the execution of the instructions corresponding to each of the two branches, when starting with the same initial relation:

$$r \gg (\textbf{then}\ p\ \textbf{else}\ q\ \textbf{end}) = r \gg p\ \cup\ r \gg q. \tag{10}$$

As previously mentioned, we define $r \gg \textbf{loop}\ p\ \textbf{end}$ according to its informal semantics : "execute p repeatedly any number of times, including zero". The corresponding rule is:

$$r \gg (\textbf{loop}\ p\ \textbf{end}) = \bigcup_{n \in \mathbb{N}} (r \gg p^n) \tag{11}$$

where p^n is the sequential composition of p with itself for n times and $\bigcup$ stands for the union of alias relations, as above. This way, our calculus is extended to infinite alias relations. This is the main difference with the approach in [19] that proposes a "cutting" technique restricting the model to a maximum length L. In [19], sequences longer than L are considered as aliased to all expressions. Orthogonally, for sequential settings, we provide finite representations of infinite alias relations based on over-approximating regular expressions, as we shall see in Section 3.2.

Both the creation and the deletion of an object x eliminate from the current alias relation all the pairs having one element prefixed by x:

$$\begin{array}{rcl} r \gg (\textbf{create}\ x) & = & r - x \\ r \gg (\textbf{forget}\ x) & = & r - x. \end{array} \tag{12}$$

The (qualified) function calls comply to their initial definitions in [19]:

$$
\begin{array}{rcl}
r \gg (\mathbf{call}\, f(l)) & = & (r[f^{\bullet}{:}l]) \gg \mid f \mid \\
r \gg (x.\mathbf{call}\, f(l)) & = & x.((x'.r) \gg \mathbf{call}\, f(x'.l)).
\end{array}
\tag{13}
$$

Here $f^{\bullet}$ and $\mid f \mid$ stand for the formal argument list and the body of f, respectively, whereas $r[u{:}v]$ is the relation r in which every element of the list v is replaced by its counterpart in u.

Consider a call $x.\mathbf{call}\, f(l)$ executed on behalf of a certain client object, which applies f to a supplier object referenced by x. In order to compute the aliasing induced by $x.\mathbf{call}\, f(l)$ when starting with an aliasing environment r, we need to compute the aliasing induced by the execution of f from r as seen by the supplier object x. This supplier view is $x'.r$ with both elements of every pair in r prefixed by the negative reference x'. Intuitively, x' can be seen as a back pointer giving access to the client. In accordance, the alias calculus evaluates $x'.x$ to *Current*. Intuitively, the negative variable x' is meant to transpose the context of the qualified call to the context of the caller.

For an example, consider a class C in an OO-language, and an associated procedure f that assigns a local variable y, defined as: $f(x)\,\{\ y := \ x\ \}$. Then, for instance, the aliasing for $a.\mathbf{call}\, f(a)$ computes as follows:

$$
\begin{array}{rc}
\emptyset \ \gg \ a.\mathbf{call}\, f(a) & = \\
a.(a'.\emptyset \ \gg \ y := a'.a) & = \\
a.(\emptyset \ \gg \ y := \mathit{Current}) & = \\
\mathit{dot\text{-}complete}(\{[a.y, a]\}). &
\end{array}
$$

Recursive function calls can lead to infinite alias relations if they do not terminate. In sequential settings, as for the case of loops, the mechanism exploiting sound regular over-approximations in order to derive finite representations of such relations is presented in the subsequent sections.

The axiom for assignment is as well in accordance with its original counterpart in [19]:

$$
\begin{array}{rcl}
r \gg (x := \ s) & = & (r_1 - x)[x \ = \ (r_1/s \ - \ x)] - ox \\
 & & \text{where } r_1 = r[ox = x]
\end{array}
\tag{14}
$$

where ox is a fresh variable (that stands for “old x”). Intuitively, the aliasing information w.r.t. the initial value of x is “saved” by associating x and ox in r and closing the new relation under dot-completeness, in r_1. Then, the initial x is “forgotten” by computing $r_1 - x$ and the new aliasing information is added in a consistent way. Namely, we add all pairs (x, s'), where s' ranges over $r_1/s \ - \ x$ representing all expressions already aliased with s in r_1, including s itself, but without x. Recall that alias relations are not reflexive, thus by eliminating x we make sure we do not include pairs of shape $[x, x]$. Then, we consider again the closure under dot-completeness and forget the aliasing information w.r.t. the initial value of x, by removing ox.

Remark 4. *It is worth discussing the reason behind* not *considering transitive alias relations. Assume the following program:*

$$\mathbf{then}\ x := \ y\ \mathbf{else}\ y := \ z\ \mathbf{end}$$

Based on the equations (10) and (14) handling conditionals and assignments, respectively, the calculus correctly identifies the alias set: $\{[x, y], [y, z]\}$. Including $[x, z]$ would be semantically equivalent to the execution of the two branches in the conditional at the same time, which is not what we want.

3.2. A sound over-approximation

In a non-concurrent or, sequential setting, the challenge of computing the alias information in the context of (infinite) loops and recursive calls reduces to evaluating their corresponding “unfoldings”, captured by expressions of shape

$$r \gg p^{\omega},$$

with ω ranging over naturals plus infinity, r an (initial) alias relation ($r = \emptyset$), and p a *basic block* defined by:

$$\begin{array}{lll} p & ::= & p\,;p \mid \textbf{then } p \textbf{ else } p \textbf{ end} \mid \\ & & \textbf{create } x \mid \textbf{forget } x \mid \\ & & x := s. \end{array} \tag{15}$$

The value $r \gg p^{\omega}$ refers to the alias relation obtained by recursively executing the control block p, and it is calculated in the expected way:

$$\begin{array}{rcl} r \gg p^0 & = & r \\ r \gg p^{k+1} & = & (r \gg p^k) \gg p. \end{array}$$

Consider again the code in (6):

```
create $x_0$
loop
   $i := i + 1$
   create $x_i$
   $x_i.set\_next(x_{i-1})$
end
```

Its execution generates an alias relation including an infinite number of pairs of shape:

$$[x_i, x_{i+1}.next],\ [x_i, x_{i+2}.next.next],\ [x_i, x_{i+3}.next.next.next] \ldots \quad . \tag{16}$$

In what follows we provide a way to compute finite representations of infinite alias relations in sequential settings. The key observation is that alias expressions corresponding to unbounded program executions grow in a regular fashion. See, for instance, the aliases in (16), which are pairs of type $[x_i, x_{i+k}.next^{k \geq 1}]$.

Regular alias expressions are defined similarly to the regular languages over an alphabet. We say that an alias expression is *regular* if it is a local variable, class attribute or *Current.* Moreover, the concatenation $e_1 . e_2$ of two regular alias expressions e_1 and e_2 is also regular. Given a regular alias expression e, the expression e^* is also regular; here $(-)^*$ denotes the Kleene star. We call an alias relation *regular* if it consists of pairs of regular alias expressions.

Proposition 5. *Let r be a regular alias relation and*

$$p ::= \textbf{create } x \mid \textbf{forget } x \mid x := s.$$

In a sequential setting, it holds that $r \gg p$ is regular.

Proof. First, observe that the name of a variable x and the right-hand side of an assignment $x := s$ as introduced in Definition 3, in Section 3, are regular alias expressions. Hence, by Definition 2, it follows that r/s, $r - x$, *dot-complete*(r) and $r[x = s]$ preserve regularity of alias relations. Consequently, by (12) and (14) it holds that $r \gg \textbf{create } x$, $r \gg \textbf{forget } x$ and $r \gg x := s$ are regular alias relations. □

Lemma 6. *Let r be a regular alias relation and p a basic block as in (15). In a sequential setting, it holds that:*

(a) $r \gg p$ *is regular;*

(b) $r \gg p^n$*, with* $n \geq 1$*, is regular.*

Proof. The proof of (a) follows by induction on the structure of p.

Base case: $p ::= \textbf{create } x \mid \textbf{forget } x \mid x := s$. The result follows by Proposition 5.

Induction step. Assume $r' \gg p'$ is regular for all regular relations r' and all basic blocks p' with simpler structure than p.
Let $p = \textbf{then } p_1 \textbf{ else } p_2 \textbf{ end}$. By definition, $\overline{r} = r \gg \textbf{then } p_1 \textbf{ else } p_2 \textbf{ end} = (r \gg p_1) \cup (r \gg p_2)$. Hence, by the induction hypothesis, it follows that $\overline{r}$ is regular.
Let $p = p_1\,; p_2$. By definition, $\overline{r} = r \gg p_1\,; p_2 = (r \gg p_1) \gg p_2$. By the induction hypothesis, it holds that $\tilde{r} = (r \gg p_1)$ is regular. Hence, by the induction hypothesis, it follows that $\overline{r} = \tilde{r} \gg p_2$ is regular as well.

The proof of (b) is by induction on n. The base case, when $n = 1$, follows by Lemma 6(a). The induction step follows immediately by the induction hypothesis, as $r \gg p^{n+1} = (r \gg p^n) \gg p$. □

Lemma 7. *Let r be a regular alias relation and*

$$\begin{array}{lll} p & ::= & p\,;p \mid \textbf{then}\ p\ \textbf{else}\ p\ \textbf{end} \mid \\ & & \textbf{create}\ x \mid \textbf{forget}\ x \mid x := s \mid \\ & & \textbf{loop}\ p\ \textbf{end} \end{array} \tag{17}$$

In a sequential setting, it holds that:

(a) $r \gg p$ *is regular;*

(b) $r \gg p^n$*, with* $n \geq 1$*, is regular.*

Proof. The proof of (a) is by induction on n – the number of *nested* loops in p.

> *Base case:* $n = 0$. The result follows immediately by Lemma 6(a), for p a basic block as in (15). Let $p = p_1\,; \textbf{loop}\ q\ \textbf{end}\,; p_2$, with p_1, p_2 and q basic blocks as in (15). It holds that
>
> $$r' = r \gg p_1\,; \textbf{loop}\ q\ \textbf{end}\,; p_2 = \bigcup_{n \in \mathbb{N}} r \gg p_1 \gg q^n \gg p_2.$$
>
> Hence, by Lemma 6(b), r' is regular.
>
> *Induction step.* Assume $r' \gg p'$ is regular for all regular relations r' and p' instruction blocks with maximum n nested loops, built as in (17). Let $p = p_1\,; \textbf{loop}\ q\ \textbf{end}\,; p_2$, with p_1, p_2 and q instruction blocks with maximum n nested loops, built as in (17). It holds that $\overline{r} = r \gg p_1\,; \textbf{loop}\ q\ \textbf{end}\,; p_2 = \bigcup_{n \in \mathbb{N}} r \gg p_1 \gg q^n \gg p_2$. Hence, $\overline{r}$ is regular by the induction hypothesis.

The proof of (b) is by induction on n. The base case, when $n = 1$, follows by Lemma 7(a). The induction step follows immediately by the induction hypothesis, as $r \gg p^{n+1} = (r \gg p^n) \gg p$.

□

Lemma 8. *Let r be a regular alias relation and $f(x_1, \ldots, x_n) = \{p_1\,; \textbf{call}\, f(y_1, \ldots, y_n)\,; p_2\}$ a recursive function, with p_1, p_2 as in (17). In a sequential setting, it holds that:*

(a) $r \gg \textbf{call}\, f(a_1, \ldots, a_n)$ *is regular,*

(b) $r \gg \textbf{call}\, f(a_1, \ldots, a_n)^n$*, with* $n \geq 1$*, is regular,*

where $a_1, \ldots, a_n$ is a valid list of actual parameters of f.

Proof. For proving (a) we proceed as follows. First, for simplicity of notation, we write l_i, $i \geq 0$, in order to refer to the actual parameters of f corresponding to its i'th recursive call. Let $l_0 = a_1, \ldots, a_n$. Then:

$$\begin{array}{l} r \gg \textbf{call}\, f(l_0) \ = \ r[f^\bullet{:}l_0] \gg \ \mid f \mid \ = \ r[f^\bullet{:}l_0] \gg p_1 \gg \textbf{call}\, f(l_1) \gg p_2 \ = \\ (r[f^\bullet{:}l_0] \gg p_1)[f^\bullet{:}l_1] \gg p_1 \gg \textbf{call}\, f(l_2) \gg p_2 \gg p_2 \ = \ \ldots \ = \\ ((r[f^\bullet{:}l_0] \gg p_1)[f^\bullet{:}l_1] \gg p_1 \ldots)[f^\bullet{:}l_n] \gg p_1 \gg \textbf{call}\, f(l_n) \gg p_2^n \end{array}$$

Let $\overline{r}$ be the alias relation obtained from r by repeatedly replacing the actual arguments of f with their formal counterparts, in the reasoning above. Thus, also based on the results in Lemma 7, it follows that $\overline{r}$ is regular. Additionally, the aliasing information corresponding to $r \gg \textbf{call}\, f(l_0)$ can be over-approximated by $\bigcup_{n \in \mathbb{N}} \overline{r} \gg p_1^n \gg p_2^n$. Hence, as $\overline{r}$ is regular and p_1, p_2 are built as in (17), the statement in Lemma 8(a) is a consequence of Lemma 7(b).
The proof of Lemma 8(b) follows by induction on n and Lemma 8(a). □

Lemma 9. *Consider r a regular alias relation and a mutually recursive function $f(x_1, \ldots, x_n) = \{p_1\,; \textbf{call}\, g(y_1, \ldots, y_m)\,; p_2\}$ with $g(y_1, \ldots, y_m) = \{q_1\,; \textbf{call}\, f(x_1, \ldots, x_n)\,; q_2\}$, and p_1, p_2, q_1, q_2 as in (17). In a sequential setting, it holds that:*

(a) $r \gg \mathbf{call}\, f(a_1, \ldots, a_n)$ *is regular,*

(b) $r \gg \mathbf{call}\, f(a_1, \ldots, a_n)^n$*, with* $n \geq 1$*, is regular,*

where $a_1, \ldots, a_n$ *is a valid list of actual parameters of* f*.*

Proof Sketch. The proof of (a) follows a reasoning similar to the one in the proof of Lemma 8(a). The alias relation $r \gg \mathbf{call}\, f(a_1, \ldots, a_n)$ can be over-approximated by $\bigcup_{n \in \mathbb{N}} \overline{r} \gg (p_1;q_1)^n \gg (q_2;p_2)^n$, where $\overline{r}$ is obtained from r by repeatedly replacing the actual arguments of f and g with their formal counterparts. As $\overline{r}$ is regular and p_1, p_2, q_1, q_2 are built as in (17), the statement in Lemma 9(a) is a consequence of Lemma 7(b). The proof of (b) follows by induction on n and Lemma 9(a). □

Theorem 10. *Assume* p *a program built according to the rules in (8). Then, in a sequential setting, the relation* $\emptyset \gg p$ *is regular.*

Proof Sketch. The proof follows by induction on the structure of p and Lemmas 6–9. □

Inspired by the idea behind the *pumping lemma for regular languages* [33], we define a *lasso* property for alias relations, which identifies the repetitive patterns within the structure of the corresponding alias expressions. The intuition is that such patterns will occur for an infinite number of times due to the infinite execution of loops or recursive function calls. We supply sound over-approximations of "lasso" relations, based on regular alias expressions.

Definition 11 (Lassos). *Consider* r *and* r' *two alias relations, and* x_i, y_i *and* z_i *a set of possibly empty expressions, for* $i \in \{1, 2\}$*. A* lasso *property over* r *and* r' *is defined as:*

$$\text{lasso}(r, r') = ([x_1.y_1.z_1, x_2.y_2.z_2] \in r \;\text{ iff }\; [x_1.y_1.y_1.z_1, x_2.y_2.y_2.z_2] \in r'). \tag{18}$$

Intuitively, there is a lasso between r *and* r' *whenever the elements in* r' *correspond to elements in* r *for which tails of prefixes are repeated.*

Definition 12 (Over-approximating regular aliases). *Consider* r *and* r' *such that* $\text{lasso}(r, r') = true$*. Let* x_i, y_i *and* z_i *be possibly empty expressions, for* $i \in \{1, 2\}$ *The set of regular alias expressions over-approximating* r *and* r' *is defined as:*

$$\begin{aligned} \text{reg}(r, r') \;=\; & \{[x_1.y_1^*.z_1, x_2.y_2^*.z_2] \mid \\ & [x_1.y_1.z_1, x_2.y_2.z_2] \in r \wedge \\ & [x_1.y_1.y_1.z_1, x_2.y_2.y_2.z_2] \in r'\} \end{aligned} \tag{19}$$

As previously indicated, the over-approximation is sound w.r.t. the repeated application of a basic block as in (15), in the way that it does not introduce any false negatives:

Lemma 13. *Consider* r *and* r' *two alias relations, and* p *a basic block as in (15). In a sequential setting, if* $r \gg p = r'$ *and* $\text{lasso}(r, r') = true$*, then the following holds for all* $n \geq 1$*:*

$$r \gg p^n \subseteq \text{reg}(r, r').$$

Proof. We proceed by induction on n.

- *Base case*: $n = 1$. By hypothesis it holds that $\text{lasso}(r, r') = true$. Hence, according to the definition of $\text{lasso}(-,-)$ in (18), the following holds for r and r':

  $$[x_1.y_1.z_1, x_2.y_2.z_2] \in r \;\text{ iff }\; [x_1.y_1.y_1.z_1, x_2.y_2.y_2.z_2] \in r'.$$

  Consequently, by the definition of $\text{reg}(-,-)$ in (19), it is easy to see that

  $$r' \subseteq \text{reg}(r, r').$$

- *Induction step.* Fix a natural number n and suppose that

$$r_k = r \gg p^k \subseteq \mathrm{reg}(r, r') \tag{20}$$

for all $k \in \{1, \dots, n\}$. We want to prove that (20) holds also for $k = n + 1$.

We continue by "reductio ad absurdum". Consider

$$r_n = r \gg p^n \subseteq \mathrm{reg}(r, r'),$$

and assume that

$$r_{n+1} = r_n \gg p \not\subseteq \mathrm{reg}(r, r') \tag{21}$$

Clearly, the execution of p when starting with r_n introduces an alias pair which is not in $\mathrm{reg}(r, r')$. According to (12), only instructions **create** x and **forget** x that appear within p, can remove alias pairs from r_n. Hence, it must be the case that the regular structure of the alias information is broken via a newly added pair $[t', s']$ associated to an assignment $x := s$ within p.

Let $p = C[x := s]$, where C is a context built according to (15), and $x := s$ is the upper-most assignment instruction in the syntactic tree associated to p that introduces a pair $[t', s'] \notin \mathrm{reg}(r, r')$. Assume that $\widetilde{r}_{n-1}$ and $\widetilde{r}_n$, respectively, are the intermediate alias relations obtained by reducing $r_{n-1} \gg C[x := s]$ and $r_n \gg C[x := s]$, respectively, according to (9)–(14), before the application of the assignment axiom corresponding to $x := s$. As only **create** x and **forget** x can remove alias pairs from r_{n-1} and r_n, respectively, and as $x := s$ is the first assignment within p, it follows that $\widetilde{r}_{n-1} \subseteq r_{n-1} \subseteq \mathrm{reg}(r, r')$ and $\widetilde{r}_n \subseteq r_n \subseteq \mathrm{reg}(r, r')$.

Let $\widetilde{r} \subseteq \mathrm{reg}(r, r')$. Based on axiom (14) it follows that $\widetilde{r} \gg x := s \not\subseteq \mathrm{reg}(r, r')$ if and only if:

(i) $[x, s] \notin \mathrm{reg}(r, r')$, or, according to dot-completeness,
(ii.1) $[x.a, s.a] \notin \mathrm{reg}(r, r')$, where $a \in \mathit{dom}(x)$, or,
(ii.2) $[s.a, x_2.y_2^*.z_2] \notin \mathrm{reg}(r, r')$, where $[x.a, x_2.y_2^*.z_2] \in \widetilde{r}$.

Let $[t', s'] \notin \mathrm{reg}(r, r')$ be the pair introduced by $x := s$ when starting from $\widetilde{r}_n$. Assume, based on (i) above, that $[t', s'] = [x, s] \notin \mathrm{reg}(r, r')$. However, $[x, s]$ must have been added when handling $\widetilde{r}_{n-1} \gg x := s$ as well, without affecting the regularity of r_n. Hence, $[x, s]$ must have been removed from r_n via **forget** x or **create** x within p. Consequently, $[x, s]$ will not occur in $\widetilde{r}_n \gg C[x := s]$ either. Therefore, we conclude that all pairs $[x, s] \notin \mathrm{reg}(r, r')$ added to $\widetilde{r}_n$ via assignments $x := s$ within p will be eventually removed via **forget** x or **create** x within p. The reasoning is similar for $[t', s'] = [x.a, s.a]$ as in (ii.1) and $[t', s'] = [s.a, x_2.y_2^*.z_2]$ as in (ii.2) above.

So, it follows that $r_{n+1} = r_n \gg p \subseteq \mathrm{reg}(r, r')$. This contradicts our assumption in (21). We conclude that $r \gg p^n \subseteq \mathrm{reg}(r, r')$.

□

4. A RL-based procedure for collecting aliases

In this section we provide the specification of a RL-based mechanism collecting the alias information. The RL rules are specified using the $\mathbb{K}$ notation for configurations and are implemented in Maude on top of the formalization of SCOOP [1] (we refer to Section 5 for more details on our approach).

In short, our strategy is to start with a program built on top of the control structures in (8), then to apply their corresponding $\mathbb{K}$-rules in order to get the "may aliasing" information in a designated $\mathbb{K}$-cell ($\langle - \rangle_{\mathrm{al}}$). Independently of the setting (sequential or concurrent) one can exploit this approach in order to evaluate the aliases of a given finite length L. We also show that for sequential contexts, the application of the $\mathbb{K}$-like rules is finite and the aliases in the final configuration soundly over-approximate the (infinite) "may alias" relations of the calculus.

$\mathbb{K}$*-like notations for configurations and (RL) rules.* $\mathbb{K}$ [35, 34] is an executable semantic framework based on Rewriting Logic [5]. It is suitable for defining (concurrent) languages and corresponding formal analysis tools, with implementation in $\mathbb{K}$-Maude [36]. $\mathbb{K}$-definitions make use of the so-called *cells*, which are labelled and can be nested, and (rewriting) *rules* describing the intended (operational) semantics. In this section we adopt a $\mathbb{K}$-like notation in order to provide a RL-based specification of the alias calculus.

A *cell* is denoted by $\langle\ -\ \rangle_{\text{[name]}}$, where [name] stands for the *name of the cell.* Intuitively, a construction $\langle\ .\ \rangle_{\text{n}}$ stands for an *empty cell* named n. We use "pattern matching" and write $\langle\ c\ \ \ldots\rangle_{\text{n}}$ for a cell with content c at the top, followed by an arbitrary content $(\ldots)$. We can utilize cells of shape $\langle\ldots\ c\ \rangle_{\text{n}}$ and $\langle\ \ldots c \ldots\ \rangle_{\text{n}}$, defined in the obvious way.

Of particular interest is $\langle\ -\ \rangle_{\text{k}}$ – the *continuation cell* holding the stack of program instructions, in the context of a programming language formalization. We write

$$\langle\ i_1 \curvearrowright i_2\ \ \ldots\rangle_{\text{k}}$$

for a set of instructions to be "executed", starting with instruction i_1, followed by i_2. The associative operation $\curvearrowright$ is the instruction sequencing.

A rule

$$\langle\ \frac{c}{c'}\ \ldots\rangle_{n_1}\ \ \langle\ c'\ \rangle_{n_2}\ \ \langle\ldots\ \frac{.}{c'}\ \rangle_{n_3}$$

reads as: if cell n_1 has c at the top and cell n_2 contains value c', then c is replaced by c' in n_1 and c' is added at the end of the cell n_3. The content of n_2 remains unchanged.

We further provide the details behind the $\mathbb{K}$-like specification of the alias calculus. As expected, the k-cell retains the instruction stack of the object-oriented program. We utilize cells $\langle - \rangle_{\text{al}}$ to enclose the current alias information, and the so-called *back-tracking cells* $\langle - \rangle_{\text{bkt-}\ldots}$ enabling the sound computation of aliases for the case of **then** – **else** – **end** and, in non-concurrent contexts, for loops and (possibly recursive) function calls. As a convention, we mark with (♣) the rules that are sound only for non-concurrent applications, based on Lemma 13.

The following $\mathbb{K}$-like rules are straightforward, based on the axioms (9)–(14) in Section 3.1. Namely, the rule implementing an instruction $p\,;q$ simply forces the sequential execution of p and q by positioning $p \curvearrowright q$ at the top of the continuation cell:

$$\langle\ \frac{p\,;q}{p \curvearrowright q}\ \ \ldots\rangle_{\text{k}} \tag{22}$$

Handling **create** x and **forget** x complies to the associated definitions. Namely, it updates the current alias relation by removing all the pairs having (at least) one element with x as prefix. In addition, it also pops the corresponding instruction from the continuation stack:

$$\langle\ \frac{r}{r-x}\ \rangle_{\text{al}}\ \ \langle\ \frac{\textbf{create}\ x}{.}\ \ldots\rangle_{\text{k}} \qquad\qquad \langle\ \frac{r}{r-x}\ \rangle_{\text{al}}\ \ \langle\ \frac{\textbf{forget}\ x}{.}\ \ldots\rangle_{\text{k}} \tag{23}$$

The assignment rule restores the current alias relation according to its axiom in (14), and removes the assignment instruction from the top of the k-cell:

$$\langle\ \frac{r}{(r_1-t)[t\ =\ (r_1/s\ -\ t)]-ot}\ \rangle_{\text{al}}\ \ \langle\ \frac{t := s}{.}\ \ldots\rangle_{\text{k}} \quad \text{with } r_1 = r[ot = t] \tag{24}$$

Handling **then** p **else** q **end** statements is more sophisticated, as it requires instrumenting a stack-based mechanism enabling the computation of the union of alias relations $r \gg p\ \cup\ r \gg q$ in three steps. First, we define the $\mathbb{K}$-like rule:

$$\langle\ r\ \rangle_{\text{al}}\ \ \langle\ \frac{\textbf{then}\ p\ \textbf{else}\ q\ \textbf{end}}{p\ \boxed{\text{et}}\ q\ \boxed{\text{ee}}}\ \ldots\rangle_{\text{k}}\ \ \ \langle\ \frac{.}{\langle\ r,p\ \rangle_{\text{t}}\ \langle\ r,q\ \rangle_{\text{e}}}\ \ldots\rangle_{\text{bkt-te}} \tag{25}$$

saving at the top of the back-tracking stack $\langle - \rangle_{\text{bkt-te}}$ the initial alias relation r to be modified by both p and q, via two cells $\langle r,p\rangle_{\text{t}}$ and $\langle r,q\rangle_{\text{e}}$, respectively. Note that the original instruction in the k-cell is replaced by a construction marking the end of the executions corresponding to the **then** and **else** branches with $\boxed{\mathrm{et}}$ and $\boxed{\mathrm{ee}}$, respectively.

Second, whenever the successful execution of p (signaled by $\boxed{\mathrm{et}}$) at the top of the k-cell) builds an alias relation r', the execution of q starting with the original relation r is forced by replacing r' with r in $\langle - \rangle_{\mathrm{al}}$, and by positioning $q\,\boxed{\mathrm{ee}}$ at the top of the k-cell. The new alias information after p, denoted by $\langle r', p \rangle_{\mathrm{t}}$, is updated in the back-tracking cell:

$$\langle\, \frac{r'}{r} \,\rangle_{\mathrm{al}} \quad \langle\, \frac{\boxed{\mathrm{et}}\; q\, \boxed{\mathrm{ee}}}{q\, \boxed{\mathrm{ee}}} \ldots \rangle_{\mathrm{k}} \quad \langle\, \frac{\langle\, r, p \,\rangle_{\mathrm{t}}}{\langle\, r', p \,\rangle_{\mathrm{t}}} \langle\, r, q \,\rangle_{\mathrm{e}} \ldots \rangle_{\mathrm{bkt\text{-}te}} \tag{26}$$

Eventually, if the successful execution of q (marked by $\boxed{\mathrm{ee}}$ at the top of $\langle - \rangle_{\mathrm{k}}$) produces an alias relation r'', then the final alias information becomes $r' \cup r''$, where r' is the aliasing after p, stored as showed in (26). The corresponding back-tracking information is removed from $\langle - \rangle_{\mathrm{bkt\text{-}te}}$, and the next program instruction is enabled in the k-cell:

$$\langle\, \frac{r''}{r' \cup r''} \,\rangle_{\mathrm{al}} \quad \langle\, \frac{\boxed{\mathrm{ee}}}{\cdot} \ldots \rangle_{\mathrm{k}} \quad \langle\, \frac{\langle\, r', p \,\rangle_{\mathrm{t}}\; \langle\, r, q \,\rangle_{\mathrm{e}}}{\cdot} \ldots \rangle_{\mathrm{bkt\text{-}te}} \tag{27}$$

For **loop** p **end**, we utilize a meta-construction $p\,\boxed{\mathrm{l}}$ **loop** p **end** simulating the set union in (11), and a back-tracking stack $\langle - \rangle_{\mathrm{bkt\text{-}l}}$ collecting the alias information obtained after each execution of p. Moreover, the $\mathbb{K}$-like specification exploits the result in Lemma 13. Whenever a "lasso" is reached, the infinite rewriting is prevented by resuming the infinite application of p in terms of a sound over-approximating alias relation. The $\mathbb{K}$-rules are as follows.

First, the aforementioned unfolding is performed, and the alias relation before p is stored in the back-tracking cell as $\langle r \rangle_{\mathrm{al\text{-}o}} \langle p \rangle_{\mathrm{l}}$:

$$\langle\, r \,\rangle_{\mathrm{al}} \quad \langle\, \frac{\textbf{loop}\ p\ \textbf{end}}{p\, \boxed{\mathrm{l}}\ \textbf{loop}\ p\ \textbf{end}} \ldots \rangle_{\mathrm{k}} \quad \langle\, \frac{\cdot}{\langle\, r \,\rangle_{\mathrm{al\text{-}o}} \langle\, p \,\rangle_{\mathrm{l}}} \ldots \rangle_{\mathrm{bkt\text{-}l}} \tag{28}$$

If the alias relation r' obtained after the successful execution of p (marked by $\boxed{\mathrm{l}}$ at the top of the continuation) is not a lasso of the aliasing r before p (previously stored in $\langle - \rangle_{\mathrm{bkt\text{-}l}}$) then p is constrained to a new execution by becoming the top of the k-cell, and r' is memorized for back-tracking:

$$\langle\, r' \,\rangle_{\mathrm{al}} \quad \langle\, \frac{\boxed{\mathrm{l}}\ \textbf{loop}\ p\ \textbf{end}}{p\, \boxed{\mathrm{l}}\ \textbf{loop}\ p\ \textbf{end}} \ldots \rangle_{\mathrm{k}} \quad \langle\, \frac{\langle\, r \,\rangle_{\mathrm{al\text{-}o}} \langle\, p \,\rangle_{\mathrm{l}}}{\langle\, r' \,\rangle_{\mathrm{al\text{-}o}} \langle\, p \,\rangle_{\mathrm{l}}} \ldots \rangle_{\mathrm{bkt\text{-}l}} \quad \text{if not lasso}(r, r') \ \ (\clubsuit) \tag{29}$$

Last, if a lasso is reached after the execution of p, then the current aliasing is soundly replaced by a "regular" over-approximation $\mathrm{reg}(r, r')$, the corresponding back-tracking information is removed from $\langle - \rangle_{\mathrm{bkt\text{-}l}}$ and the **loop** instruction is eliminated from the k-cell:

$$\langle\, \frac{r'}{\mathrm{reg}(r, r')} \,\rangle_{\mathrm{al}} \quad \langle\, \frac{\boxed{\mathrm{l}}\ \textbf{loop}\ p\ \textbf{end}}{\cdot} \ldots \rangle_{\mathrm{k}} \quad \langle\, \frac{\langle\, r \,\rangle_{\mathrm{al\text{-}o}} \langle\, p \,\rangle_{\mathrm{l}}}{\cdot} \ldots \rangle_{\mathrm{bkt\text{-}l}} \quad \text{if lasso}(r, r') \ \ (\clubsuit) \tag{30}$$

For handling function calls such as **call** $f(l)$ we use a meta-construction $\mid f \mid \boxed{\mathrm{f}}$. Here $\mid f \mid$ stands for the body of f and $\boxed{\mathrm{f}}$ marks the end of the corresponding execution. Moreover, a stack $\langle - \rangle_{\mathrm{bkt\text{-}cf}}$ is utilized in order to store the alias information before each (possibly recursive) call of f, with the purpose of identifying the lassos generated by the (possibly repeated) execution of f. In order to guarantee a sound implementation of (mutually) recursive calls, both $\boxed{\mathrm{f}}$ and $\langle - \rangle_{\mathrm{bkt\text{-}cf}}$ are parameterized by f – the name of the function. An example illustrating this reasoning mechanism is provided in Section 4.1.

The first $\mathbb{K}$-like rule for handling function calls matches the associated axiom in (13): the alias information is set to $r[f^{\bullet}{:}l]$, whereas the next instructions to be executed are given by $\mid f \mid$. Note that the original aliasing is retained in the (initially empty) back-tracking cell via $\langle r \rangle_{\mathrm{al\text{-}o}}$.

$$\langle\, \frac{r}{r[f^{\bullet}{:}l]} \,\rangle_{\mathrm{al}} \quad \langle\, \frac{\textbf{call}\, f(l)}{\mid f \mid\ \boxed{\mathrm{f}}} \ldots \rangle_{\mathrm{k}} \quad \langle\, \frac{\cdot}{\langle\, r \,\rangle_{\mathrm{al\text{-}o}}} \,\rangle_{\mathrm{bkt\text{-}cf}} \tag{31}$$

Remark 14. *Observe that the back-tracking cell does not need to be parameterized by the actual argument list l of f. Each such argument is anyways replaced in the current alias relation r by its counterpart in the formal argument list of f. In short: r becomes $r[f^{\bullet}{:}l]$.*

A successful execution of **call** $f(l)$ is distinguished by the occurrence of $\boxed{\mathrm{f}}$ at the top of the continuation stack. If this is the case, then the corresponding back-tracking alias information is removed from $\langle - \rangle_{\text{bkt-cf}}$ and the next program instruction (if any) is enabled at the top of the k-cell:

$$\langle\ r'\ \rangle_{\mathrm{al}}\ \ \langle \frac{\boxed{\mathrm{f}}}{\cdot} \dots\rangle_{\mathrm{k}}\ \ \langle \frac{\langle\ r\ \rangle_{\text{al-o}}}{\cdot} \dots\rangle_{\text{bkt-cf}} \tag{32}$$

Recursive calls are treated by means of two rules. Note that a recursive context is identified whenever the current program instruction is of shape **call** $f(l)$ and the associated back-tracking structure is not empty, *i.e.*, rule (31) was previously applied. Then, if the recursive call of f when starting with r produces a lasso r', the execution of $f(l)$ is stopped by soundly over-approximating the alias information with $\mathrm{reg}(r, r')$, according to Lemma 13, and by removing **call** $f(l)$ from the k-cell:

$$\langle \frac{r'}{\mathrm{reg}(r,r')} \rangle_{\mathrm{al}}\ \ \langle \frac{\mathbf{call}\, f(l)}{\cdot} \dots\rangle_{\mathrm{k}}\ \ \langle\ \langle\ r\ \rangle_{\text{al-o}}\ \dots\rangle_{\text{bkt-cf}} \text{ if } \mathrm{lasso}(r,r') \ \ (\clubsuit) \tag{33}$$

If a lasso is not reached, then the body of f is executed once more, and the current aliasing is pushed to the back-tracking cell:

$$\langle\ r'\ \rangle_{\mathrm{al}}\ \ \langle \frac{\mathbf{call}\, f(l)}{\mid f \mid\ \boxed{\mathrm{f}}} \dots\rangle_{\mathrm{k}}\ \ \langle \frac{\cdot}{\langle\ r'\ \rangle_{\text{al-o}}} \langle\ r\ \rangle_{\text{al-o}} \dots\rangle_{\text{bkt-cf}} \quad \text{if not } \mathrm{lasso}(r,r') \ \ (\clubsuit) \tag{34}$$

Qualified calls $x.$**call** $f(l)$ are handled by two rules as follows. First, based on the definition in (13), the "negative variable" x' transposing the context of the call to to the context of the caller is distributed to the elements of the initial alias relation r, and to l – the argument list of f. Moreover, a meta-construction $\boxed{\mathrm{qf}}$ is utilized in order to mark the end of the qualified call in the continuation cell, similarly to the rule (31). The caller is stored in a back-tracking stack $\langle\ .\ \rangle_{\text{bkt-qf}}$ also parameterized by f – the name of the function. The current instruction in the k-cell becomes **call** $f(x'.l)$, as expected:

$$\langle \frac{r}{x'.r} \rangle_{\mathrm{al}}\ \ \langle \frac{x.\mathbf{call}\, f(l)}{\mathbf{call}\, f(x'.l)\ \boxed{\mathrm{qf}}} \dots\rangle_{\mathrm{k}}\ \ \langle \frac{\cdot}{\langle\ x\ \rangle_{\mathrm{f}}} \rangle_{\text{bkt-qf}} \tag{35}$$

Second, when the successful termination of the qualified call is signaled by $\boxed{\mathrm{qf}}$ at the top of the k-cell, the corresponding stored caller is distributed to the current alias relation and removed from the back-tracking cell. The next instruction in the continuation cell is released by eliminating the top $\boxed{\mathrm{qf}}$:

$$\langle \frac{r}{x.r} \rangle_{\mathrm{al}}\ \ \langle \frac{\boxed{\mathrm{qf}}}{\cdot} \dots\rangle_{\mathrm{k}}\ \ \langle \frac{\langle\ x\ \rangle_{\mathrm{f}}}{\cdot} \dots\rangle_{\text{bkt-qf}} \tag{36}$$

In a non-concurrent setting, the machinery orchestrating the $\mathbb{K}$-like rules introduced in this section implements an algorithm that always terminates and provides a sound over-approximation of "may aliasing".

Theorem 15. *Consider p a program built on top of the control structures in (8), that executes in a sequential setting. Then, the application of the corresponding $\mathbb{K}$-like rules when starting with p and an empty alias relation, is a finite rewriting of shape*

$$\langle\ \emptyset\ \rangle_{\mathrm{al}} \langle\ p\ \rangle_{\mathrm{k}} \overset{(*)}{\Longrightarrow} \langle\ r\ \rangle_{\mathrm{al}} \langle\ .\ \rangle_{\mathrm{k}},$$

with r a sound over-approximation of the aliasing information corresponding to the execution of p, and $\overset{(*)}{\Longrightarrow}$ the rewrite relation defined by the rules (22)–(36).

Proof Sketch. First, note that the rewrite system in (22)–(36) is confluent, as there is only one way of performing the rewritings starting with program p, by matching at top (and at the top of the cells $\langle\ -\ \rangle_{[\text{name}]}$). Then, observe that due to the execution of loops and/or recursive calls, expressions can infinitely grow in a *regular* fashion. Hence, a lasso is always reached. Consequently, the control structure generating the infinite behaviour is removed from the k-cell, according to the associated $\mathbb{K}$-like specification for loops and/or recursive calls. This guarantees termination. Moreover, recall that the regular expressions replacing the current alias information are a sound over-approximation, according to Lemma 13. □

Observe that the RL-based machinery can simulate precisely the "cutting at length L" technique in [19]. It suffices to disable the rules (♣) and stop the rewriting after L steps.

The naturalness of applying the resulted aliasing framework is illustrated in the example in Section 4.1, for the case of two mutually recursive functions.

4.1. The $\mathbb{K}$*-machinery by example*

For an example, in this section we show how the $\mathbb{K}$-machinery developed in Section 4 can be used in order to extract the alias information for the case of two mutually recursive functions defined as:

$$f(x)\ \{\ x := x.a\,;\ \ \textbf{call}\, g(x)\ \} \qquad\qquad g(x)\ \{\ x := x.b\,;\ \ \textbf{call}\, f(x)\ \}$$

We assume that x is an object of a class with two fields a and b, respectively. We consider a sequential setting.

The execution of **call** $f(x)$, when starting with an empty alias relation r, produces the alias expressions:

$$[x,\, x.(a.b)^*] \quad [x.a,\, x.(a.b)^*.a] \quad [x.b,\, x.(a.b)^*.b] \tag{37}$$

The associated rewriting is depicted in Figure 4.1. The whole procedure starts with an empty alias relation $r = \emptyset$, and **call** $f(x)$ in the continuation stack. Then, the corresponding $\mathbb{K}$ rules (for handling assignments and function calls) are applied in the natural way.

A lasso is reached after two calls of $f(x)$ that, consequently, determine two calls of $g(x)$ – identified by $\boxed{\text{g}}\ \boxed{\text{f}}\ \boxed{\text{g}}\ \boxed{\text{f}}$ in the k-cell. This triggers the application of rule (33) enabling the "regular" over-approximation as in Lemma 13.

Our example also illustrates the importance of isolating the back-traced alias information in cells of shape $\langle\ .\ \rangle_{\text{bkt-cf}}$ parameterized by the (possibly recursive) function f. More explicitly, rule (33) is soundly applied by identifying the aforementioned lasso based on: the current alias relation r_4, the recursive call $f(l)$ at the top of the continuation, and the back-traced aliasing $\langle\ \langle\ r_2\ \rangle_{\text{al-o}}\ \ldots\rangle_{\text{bkt-cf}}$ associated to the previous executions of $f(l)$.

As introduced in (18), an alias relation r' is a lasso of a relation r whenever there is a one-to-one correspondence between their elements as follows:

$$[x_1y_1z_1, x_2y_2z_2] \in r \ \text{ iff }\ [x_1y_1y_1z_1, x_2y_2y_2z_2] \in r'.$$

The current alias relation

$$r_4 = \{[x, x.a.b.a.b],\ [x.a, x.a.b.a.b.a],\ [x.b, x.a.b.a.b.b]\},$$

before applying rule (33), is a lasso of

$$r_2 = \{[x, x.a.b],\ [x.a, x.a.b.a],\ [x.b, x.a.b.b]\}.$$

The aforementioned one-to-one correspondence is summarized in the following table:

$[x_1y_1z_1, x_2y_2z_2] \in r_2$ iff $[x_1y_1y_1z_1, x_2y_2y_2z_2] \in r_4$	x_1	y_1	z_1	x_2	y_2	z_2
$[x, x.a.b] \in r_2$ iff $[x, x.a.b.a.b] \in r_4$	x	ε	ε	x	$a.b$	ε
$[x.a, x.a.b.a] \in r_2$ iff $[x.a, x.a.b.a.b.a] \in r_4$	x	ε	a	x	$a.b$	a
$[x.b, x.a.b.b] \in r_2$ iff $[x.b, x.a.b.a.b.b] \in r_4$	x	ε	b	x	$a.b$	b

Here ε stands for the *empty alias expression.*

Moreover, according to rule (33), the lasso shaped by r_2 and r_4 also causes the (otherwise infinite) recursive calls to stop, as $\mathbf{call}\, f(l)$ is eliminated from the top of the k-cell. Hence, the rewriting process finishes with a sound over-approximation $\text{reg}(r_2, r_4)$ replacing the current alias relation (cf. Lemma 13), defined precisely as in (37).

$$\begin{gathered}
\langle\ r\ \rangle_{\text{al}}\ \langle\ \mathbf{call}\, f(x)\ \rangle_{\text{k}} \\
\langle\ \cdot\ \rangle_{\text{bkt-cf}}\ \langle\ \cdot\ \rangle_{\text{bkt-cg}} \\
\Downarrow (31) \\
\langle\ r\ \rangle_{\text{al}}\ \langle\ x := x.a; \mathbf{call}\, g(x)\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \cdot\ \rangle_{\text{bkt-cg}} \\
\Downarrow (24) \\
\langle\ r_1\ \rangle_{\text{al}}\ \langle\ \mathbf{call}\, g(x)\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \cdot\ \rangle_{\text{bkt-cg}} \\
\text{where } r_1 = \{[x, x.a],\ [x.a, x.a.a],\ [x.b, x.a.b]\} \\
\Downarrow (34) \\
\langle\ r_1\ \rangle_{\text{al}}\ \langle\ x := x.b; \mathbf{call}\, f(x)\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\Downarrow (24) \\
\langle\ r_2\ \rangle_{\text{al}}\ \langle\ \mathbf{call}\, f(x)\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\text{where } r_2 = \{[x, x.a.b],\ [x.a, x.a.b.a],\ [x.b, x.a.b.b]\} \\
\Downarrow (34) \\
\langle\ r_2\ \rangle_{\text{al}}\ \langle\ x := x.a; \mathbf{call}\, g(x)\ \boxed{\text{f}}\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r_2\ \rangle_{\text{al-o}}\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\Downarrow (24) \\
\langle\ r_3\ \rangle_{\text{al}}\ \langle\ \mathbf{call}\, g(x)\ \boxed{\text{f}}\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r_2\ \rangle_{\text{al-o}}\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\text{where } r_3 = \{[x, x.a.b.a],\ [x.a, x.a.b.a.a],\ [x.b, x.a.b.a.b]\} \\
\Downarrow (34) \\
\langle\ r_3\ \rangle_{\text{al}}\ \langle\ x := x.b; \mathbf{call}\, f(x)\ \boxed{\text{g}}\ \boxed{\text{f}}\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r_2\ \rangle_{\text{al-o}}\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_3\ \rangle_{\text{al-o}}\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\Downarrow (24) \\
\langle\ r_4\ \rangle_{\text{al}}\ \langle\ \mathbf{call}\, f(x)\ \boxed{\text{g}}\ \boxed{\text{f}}\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r_2\ \rangle_{\text{al-o}}\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_3\ \rangle_{\text{al-o}}\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\text{where } r_4 = \{[x, x.a.b.a.b],\ [x.a, x.a.b.a.b.a],\ [x.b, x.a.b.a.b.b]\} \\
\Downarrow (33) \\
\langle\ reg(r_2, r_4)\ \rangle_{\text{al}}\ \langle\ \boxed{\text{g}}\ \boxed{\text{f}}\ \boxed{\text{g}}\ \boxed{\text{f}}\ \rangle_{\text{k}} \\
\langle\ \langle\ r_2\ \rangle_{\text{al-o}}\ \langle\ r\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cf}}\ \langle\ \langle\ r_3\ \rangle_{\text{al-o}}\ \langle\ r_1\ \rangle_{\text{al-o}}\ \rangle_{\text{bkt-cg}} \\
\Downarrow (^*)(32) \\
\langle\ \{[x, x.(a.b)^*],\ [x.a, x.(a.b)^*.a],\ [x.b, x.(a.b)^*.b]\}\ \rangle_{\text{al}} \langle\ .\ \rangle_{\text{k}} \langle\ .\ \rangle_{\text{bkt-cf}} \langle\ .\ \rangle_{\text{bkt-cg}}
\end{gathered}$$

Figure 1: Over-approximating aliasing.

5. Aliasing in SCOOP

In this section we provide a brief overview on the integration and applicability of the alias calculus in SCOOP. First, recall from Section 2 that the Maude semantics of SCOOP in [1] is defined over tuples of shape

$$\langle p_1 :: St_1 \mid \ldots \mid p_n :: St_n, \sigma \rangle$$

where, p_i and St_i stand for processors and their call stacks, respectively. σ is the state of the system and it holds information about the heap and the store.

The assignment instruction, for instance, is formally specified as the transition rule:

$$\frac{\text{a is fresh}}{\Gamma \vdash \langle p :: t := s; St,\ \sigma \rangle \rightarrow \langle p :: \mathrm{eval}(a, s); \mathrm{wait}(a); \mathrm{write}(t, a.data); St,\ \sigma \rangle} \tag{38}$$

where, intuitively, "eval(a, s)" evaluates s and puts the result on channel a, "wait(a)" enables processor p to use the evaluation result, "write$(t, a.data)$" sets the value of t to $a.data$, St is a call stack, and Γ is a typing environment [37] containing the class hierarchy of a program and all the type definitions.

The $\mathbb{K}$-like rule for assignments

$$\langle \frac{r}{(r_1 - t)[t \ = \ (r_1/s \ - \ t)] - ot} \rangle_{\mathrm{al}} \ \langle \frac{t := \ s}{\cdot} \cdots \rangle_{\mathrm{k}} \quad \text{with } r_1 = r[ot = t] \tag{24}$$

can be straightforwardly integrated in (38) by enriching the SCOOP configuration with a new component *alias_* encapsulating the alias information, and considering instead the transition:

$$\Gamma \vdash \langle p ::\ t := s; St,\ \sigma, alias_{old} \rangle \rightarrow$$
$$\langle p :: \mathrm{eval}(a, s); \mathrm{wait}(a); \mathrm{write}(t, a.data); St,\ \sigma, alias_{new} \rangle$$

where

$$alias_{old} = r \qquad alias_{new} = (r_1 - t)[t \ = \ (r_1/s \ - \ t)] - ot$$

with r and r_1 as in (24). The integration of all the $\mathbb{K}$-rules of the alias calculus on top of the Maude formalization of SCOOP is achieved by following a similar approach.

Nevertheless, it is important to point out that in order of the sound over-approximating aliases to be constructed as in Section 4, Theorem 15, one needs to simulate a SCOOP *sequential* setting. This is possible in SCOOP by creating all the objects in the program to be analysed as objects handled by the same processor. An example computing the aliasing information is provided at:
`https://dl.dropboxusercontent.com/u/1356725/SCOOP-SCP.zip`
Simply run the command

```
> maude SCOOP.maude ..\examples\linked_list-test.maude
```

corresponding to the execution of the code in (6), Section 3

```
create $x_0$
loop
  $i := i + 1$
  create $x_i$
  $x_i.set\_next(x_{i-1})$
end
```

for five iterations of the loop. As can be observed based on the code in `linked_list-test.maude`, in order to implement our applications in Maude, we use intermediate (still intuitive) representations. For instance, the class structure defining a node in a simple linked list, with filed *next* and setter *set_next* is declared as:

```
class 'NODE
    create {'make} (
    attribute { 'ANY } 'next : [?, . , 'NODE] ;
    procedure { 'ANY } 'set_next ( 'a_next : [?, ., 'NODE] ;) [...]
   )
end ;
```

where `’next : [?, . , ’NODE]` stands for an object of type `NODE`, that is handled by the current processor (`.`) and that can be Void (`?`), and `’make` plays the role of a constructor. The “entry point” of the program corresponds to the function `’make` in the (main) class `’LINKED_LIST_TEST` and is set via:

```
settings(’LINKED_LIST_TEST, ’make, aliasing-on) .
```

Observe that the flag for performing the alias analysis is switched to “on”.

The relevant parts of the corresponding Maude output after executing the aforementioned command are as follows:

```
rewrite in SYSTEM :
[...] settings(’LINKED_LIST_TEST, ’make, aliasing-on))

|-
  {0}proc(0) :: nil | {0}proc(1) :: nil,
  {[’x0 ; ’x0]} U {[’x0 ; ’x1.’next]} U
  {[’x0 ; ’x2.’next.’next]} U {[’x0 ; ’x3.’next.’next.’next]} U
  {[’x0 ; ’x4.’next.’next.’next.’next]} U
  [...]
  {[’x3 ; ’x3]} U {[’x3 ; ’x4.’next]}

state
  heap [...]
  store [...]
end
```

In short, one can see that two processors that were created finished executing the instructions on their corresponding call stacks: `{0}proc(0) :: nil` and `{0}proc(1) :: nil`. The aliased expressions include, as expected based on (7), pairs of shape $[x_i\,;x_{i+k}.next^k]$. Moreover, the output displays the contents of the current system state, by providing information on the *heap* and *store*, as formalized in [1].

6. Deadlocking in SCOOP

In what follows we provide the details behind the formalization and the implementation of a deadlock detection mechanism for SCOOP. We discuss how Maude can be exploited in order to test and, respectively, model-check deadlocks in the context of SCOOP programs, we analyze the associated challenges and propose a series of corresponding solutions.

6.1. Formalizing deadlocks in SCOOP

Recall that the key idea of SCOOP is to associate to each object declared as *separate* a processor, or handler in charge of executing the routines of that object. Assume a processor p that performs a call $o.f(a_1, a_2, \ldots)$ on a separate object o, with separate arguments a_i $(i \geq 1)$. As previously stated in Section 2, according to the SCOOP semantics, the application of the call $f(\ldots)$ will *wait* until it has been able to *lock* all the separate objects associated to $a_1, a_2, \ldots$. This reservation mechanism enables deadlocks to occur whenever a set of processors reserve each other circularly.

Definition 16 (Deadlock)**.** *For a processor p, let $W(p)$ be the set of handlers p* waits *for its asynchronous execution, and $H(p)$ represent the set of resources p already acquired. A* deadlock *exists if for some set D of processors the following holds:*

$$(\forall p \in D).(\exists p' \in D).(p \neq p') \wedge (W(p) \cap H(p') \neq \emptyset). \tag{39}$$

A deadlock can happen, for instance, in a Dining Philosophers scenario, where both philosophers and forks are objects residing on their own processors. Consider, for an example, two separate objects p_1 and p_2 denoting two philosophers, and two separate objects f_1 and f_2 denoting two forks. Assume that p_1 picks f_1, whereas p_2 picks f_2. Then, the following hold whenever p_1 and p_2 want to acquire both forks: $H(p_1) = \{f_1\} = W(p_2)$ and $H(p_2) = \{f_2\} = W(p_1)$. Hence, (16) is satisfied for $D = \{p_1, p_2\}$.

The integration of a deadlock detection mechanism based on Definition 16 on top of the SCOOP formalization in [1] is immediate. As already presented in Section 2, the operational semantics of SCOOP is defined over tuples of shape:

$$\langle p_1 :: St_1 \mid \ldots \mid p_n :: St_n, \sigma\rangle$$

where, p_i and St_i stand for processors and their call stacks, respectively, and σ is the state of the system. Given a processor p' as in (39), the set $H(p')$ corresponds, based on [1], to $\sigma.rq\_locks(p')$. Whenever the top of the instruction stack of a processor p is of shape $lock(\{q_i, \ldots, q_n\})$, we say that the wait set $W(p)$ is the set of processors $\{q_1, \ldots, q_n\}$. Hence, assuming a predefined system configuration $\langle deadlock\rangle$, the SCOOP transition rule corresponding to (39) can be written as:

$$\frac{\begin{array}{c}(\exists D \subseteq \sigma.procs).(\forall p \in D).(\exists p' \in D).(p \neq p')\wedge \\ (aqs := \ldots \mid p :: lock(\{q_i, \ldots\}); St \mid \ldots) \ \wedge \ (\sigma.rq\_locks(p').has(q_i))\end{array}}{\langle aqs, \sigma\rangle \rightarrow \langle deadlock\rangle} \tag{40}$$

Note that $\sigma.procs$ in (40) returns the set of processors in the system, whereas aqs stands for the list of these processors and their associated instruction stacks (separated by the associative and commutative operator "|"). We use "..." to represent an arbitrary sequence of processors and processor stacks.

6.2. Testing deadlocks

We implemented (40) and tested the deadlock detection mechanism on top of the formalization in [1] for the Dining Philosophers problem. A case study considering two philosophers sharing two forks can be run by downloading the SCOOP formalization at:
`https://dl.dropboxusercontent.com/u/1356725/SCOOP-SCP-deadlock.zip`
and executing the command

```
> maude SCOOP.maude ..\examples\dining-philosophers-rewrite.maude
```

The class implementing the *philosopher* concept is briefly given below:

```
class 'PHILOSOPHER
    create { 'make } (
        attribute {'ANY} 'left : [!,T,'FORK] ;
        attribute {'ANY} 'right : [!,T,'FORK] ;

        procedure { 'ANY } 'make ( 'fl : [!,T,'FORK] ;
                                   'fr : [!,T,'FORK] ; )
            do  ( assign ('left, 'fl) ; assign ('right, 'fr) ; )
            [...]
        end ;

        procedure { 'ANY } 'eat_wrong (nil)
            do ( command ('Current . 'pick_in_turn('left ;)) ; )
            [...]
        end ;

        procedure { 'ANY } 'pick_in_turn ('f : [!,T,'FORK] ; )
            do ( command ('Current . 'pick_two('f ; 'right ;)) ; )
            [...]
        end ;
        procedure { 'ANY } 'pick_two ('fa : [!,T,'FORK] ;
                                      'fb : [!,T,'FORK] ; )
            do ( command ('fa . 'use(nil)) ;  command ('fb . 'use(nil)) ; )
            [...]
        end ;
[...] end
```

It declares two forks – `'left` and `'right` of type `'FORK`, that can be handled by any processor (`T`) and that cannot be `Void` (`!`). Assume two philosophers `p1` and `p2` (of *separate* type `PHILOSOPHER`) and two forks `f1` and `f2` (of *separate* type `FORK`). Moreover, assume that `'left` and `'right` for `p1` correspond to `'f1` and `'f2`. For the case of `p2` they correspond to `'f2` and `'f1`, respectively. Asynchronously, `p1` and `p2` can execute `eat_wrong`, which calls `pick_in_turn(left)`. In the context of `p1`, the actual value of `left` is `f1`, whereas for `p2` is `f2`. Consequently, both resources `f1` and `f2`, respectively, may be locked "at the same time" by `p1` and `p2`, respectively. Note that `pick_in_turn` subsequently calls `pick_two` that, intuitively, should enable the philosophers to use both forks. Thus, if `f1` and `f2`, respectively, are locked by `p1` and `p2`, respectively, the calls `pick_two(f2, f1)` and `pick_two(f1, f2)` corresponding to `p1` and `p2` will (circularly) wait for each other to finish. According to the SCOOP semantics, `pick_two(f1, f2)` is waiting for `p2` to release `f2`, whereas `pick_two(f2, f1)` is waiting for `p1` to release `f1`, as the forks are passed to `pick_two(...)` as *separate* types. In the context of SCOOP, this corresponds to a Coffman deadlock [38].

The entry point of the program implementing the Dining Philosophers example is the function `'make` in the class `APPLICATION`, which asks the two philosophers `p1` and `p2` to adopt a wrong eating strategy as above, possibly leading to a deadlock situation. The flag enabling the deadlock analysis as in (40) is set to "on". This information is specified using a Dining Philosophers configuration (DP-config) specified by `settings('APPLICATION, 'make, deadlock-on)`.

Unfortunately, none of the executions of the Dining Philosophers scenario by simply invoking the Maude `rewrite` command lead to a deadlock situation. Each of our tests displayed the output:

```
rewrite in SYSTEM :
[...]  settings('APPLICATION, 'make, deadlock-on)

|- {0}proc(0) :: nil | {0}proc(1) :: nil | {0}proc(3) :: nil
   {0}proc(5) :: nil | {0}proc(7) :: nil
   {0}proc(9) :: nil | {0}proc(11) :: nil
```

consisting of a list of processors (including the handlers of both the philosophers and the forks) with empty call stacks (`:: nil`). This indicates that every time, the two philosophers proceeded by lifting their forks simultaneously, hence no deadlock was possible.

One possible workaround is to use predefined strategies [39] in order to guide the rewriting of the Maude rules formalizing SCOOP towards a $\langle deadlock \rangle$ system configuration. An example of applying such a strategy for the Dining Philosophers case can be tested by running the command:

```
> maude SCOOP.maude ..\examples\dining-philosophers-strategy.maude
```

The command `srew [...] using init ; parallelism{lock} ; [...] ; deadlock-on` forces the execution of a `pick_in_turn` approach as above. This determines Maude to first trigger the rule `[init]` in the SCOOP formalization in [1]. This makes all the required initializations of the *bootstrap* processor. Then, one of the processors that managed to *lock* the necessary resources is ("randomly") enabled to proceed to the asynchronous execution of its instruction stack, according to the strategy `parallelism{lock}` . The last step of the strategy calls the rule `[deadlock-on]` implementing the Coffman deadlock detection as in (40).

This time the guided rewriting leads, indeed, to one solution identifying a deadlock. The relevant parts of the corresponding Maude output are as follows:

```
srewrite in SYSTEM :
[...] settings('APPLICATION, 'make, deadlock-on)
```

```
using init ; parallelism{lock} ; [...] ; deadlock-on .

Solution 1
result Configuration: deadlock
```

Nevertheless, such an approach requires lots of ingeniousness (our strategy has more than 300 rules!) and, moreover, is not automated.

6.3. Model-checking deadlocks

In this section we provide an overview on our approach to model-checking deadlocks for SCOOP, using the LTL Maude model-checker [40]. As mentioned in the beginning of the current paper, the idea behind our work is to exploit the unifying flavor of the Maude executable semantics of SCOOP [1]. The latter integrates both the formalization of the language and its concurrency mechanisms, thus enabling using the semantic framework for program analysis purposes, "for free".

One possible way to proceed is by simply running the Maude LTL model-checker on a Dining Philosophers configuration (DP-config) as thoroughly discussed in Section 6.2:

```
red modelCheck(DP-config, [ ] no-deadlock-mck) .
```

Intuitively, `[ ] no-deadlock-mck` is a safety property stating the freedom from deadlocks. The predicate witnessing the absence of deadlocks is defined as:

```
eq < deadlock > |= no-deadlock-mck = false .
eq sys::Configuration |= no-deadlock-mck = true [owise] .
```

Above, $\langle deadlock \rangle$ is the predefined system configuration "signalling" deadlocks in the SCOOP rule (40), whereas *sys::Configuration* is an arbitrary SCOOP system configuration.

Unfortunately, running the LTL model checker led to the state explosion problem. At a first look, the issue was caused by the size of the SCOOP formalization in [1] which includes all the aspects of a real concurrency model.

As a first step, we reduced this formalization by eliminating the parts that are not relevant in the context of deadlocking; examples include the garbage collection and the exception handling mechanisms.

In addition, we provided a simplified, abstract semantics of SCOOP based on aliasing. This idea stems from the fact that SCOOP processors are known from object references, that may be aliased. Therefore, the SCOOP semantics can be simplified by retaining within the corresponding transition rules only the information important for aliasing. Consider, for instance, the rules (2)–(4) specifying "**if**" instructions in Section 2. The abstract transition rule omits the evaluation of the conditional and computes the aliasing information similarly to the semantics of **then . . . else . . . end** in (10), in Section 3. The abstraction collects the aliases resulted after the execution of both "**if**" and "**else**" branches:

$$\frac{\cdot}{\begin{array}{c}\langle p :: \textbf{if } e \textbf{ then } St_1 \textbf{ else } St_2 \textbf{ end} \, ; St,\, \sigma, alias_{old}\rangle \rightarrow \\ \langle p :: St,\, \sigma, alias_{new}\rangle\end{array}} \tag{41}$$

Observe that the SCOOP system configurations in (1) are enriched with a new component $alias\_$ consisting of a set of alias expressions. Above, $alias_{old}$ is the aliasing before the execution of the "**if**" instruction, whereas, intuitively, $alias_{new}$ stands for $alias_{old} \gg St_1 \cup alias_{old} \gg St_2$.

As a second step, we analyzed the implementation in [1] from a more engineering perspective, and identified a series of design decisions that either slowed down considerably the rewriting or made the search space grow unnecessarily large.

After running some experiments, we understood that the parallelism rule

$$\frac{\langle p_1 :: St_1,\, \sigma\rangle \rightarrow \langle p_1' :: St_1',\, \sigma'\rangle}{\langle p_1 :: St_1 \mid p_2 :: St_2,\, \sigma\rangle \rightarrow \langle p_1' :: St_1' \mid p_2 :: St_2,\, \sigma'\rangle} \tag{42}$$

in [1] was increasing the rewriting time. Though elegant from the formalization perspective, the usage of this rule was not efficient. Therefore, we eliminated it from the SCOOP semantics and made the remaining rules apply directly, by matching at top. For instance, the abstract rule (41) formalizing “**if**” instructions in the context of one processor p becomes:

$$\frac{\cdot}{\begin{array}{c}\langle p :: \textbf{if } e \textbf{ then } St_1 \textbf{ else } St_2 \textbf{ end}\,; St \mid aqs,\, \sigma, alias_{old}\rangle \to \\ \langle p :: St \mid aqs,\, \sigma, alias_{new}\rangle\end{array}} \tag{43}$$

for an arbitrary list aqs of processors and their instruction stacks. For Dining Philosophers, for example, this modification reduced the rewriting time from around 10s to less than 1s.

Recall that SCOOP processors communicate via channels. The implementation in [1] creates *fresh* channels (as in (2), for instance) parameterized by natural indexes. This was inefficient for model-checking purposes, as the state space contained many identical states up-to channel naming.

The implementation of the above observations enabled us to successfully identify a deadlock situation in a Dining Philosophers scenario, by using the Maude LTL model-checker. The new (reduced) formalization of SCOOP can be found at:
`https://dl.dropboxusercontent.com/u/1356725/SCOOP-SCP.zip`.

This approach successfully identifies a SCOOP deadlock configuration:

```
  [...]
  proc(7) ::
  lock(({proc(3)} U {proc(5)})) ;
  command('fa . 'use(nil)) ; command('fb . 'use(nil)) ;
  [...]
  proc(9) ::
  lock(({proc(3)} U {proc(5)})) ;
  command('fa . 'use(nil)) ; command('fb . 'use(nil)) ;
  [...]

state
  regions(
      proc(3) --> {ref(4)} #  proc(5) --> {ref(6)} #
          proc(7) --> {ref(8)} # proc(9) --> {ref(10)} #
  [...]
  )
  heap
    objects
        'f1 --> ref(4) # 'f2 --> ref(6) #
            'p1 --> ref(8) # 'p2 --> ref(10) #
    [...]
  end
```

At a closer look, the information within the SCOOP deadlock state above identifies two processors `proc(7)` and `proc(9)` that cannot process the `lock(({proc(3)} U {proc(5)}))` statement at the top of their corresponding instruction stacks. Observe, based on the content of `regions` above, that `proc(7)` is the processor handling the reference `ref(8)` which corresponds to the philosopher object `'p1`. Similarly, `proc(9)` corresponds to the philosopher `'p2`, whereas `proc(3)` and `proc(5)` are associated to the fork objects `'f1` and `'f2`, respectively. Hence, it must be the case that each of the philosophers asynchronously picked (*i.e.*, locked) one of the two forks by executing `pick_in_turn(...)`. This way, locking both fork resources (`lock(({proc(3)} U {proc(5)}))`) by any of the philosophers, in order to execute `pick_two(...)`, is impossible according to the SCOOP semantics of locks.

Unfortunately, model-checking deadlocks in a scenario with five philosophers, for instance, takes unacceptably long time. Further improvements may be obtained by following the same recipe of collapsing semantically equivalent states, from the deadlocking perspective. A major source of redundancy is represented by the so-called *regions* in [1] that, intuitively, manage all the objects handled by the same processor. Their elimination from the SCOOP abstract state could improve the overall time performance by enabling the model-checker to make less identifications.

An additional observation is that most of the rewrite rules formalizing SCOOP are conditional and have rewrites in their conditions. This makes their execution expensive and inefficient for model-checking. A reduction semantics of SCOOP [30, 41] is therefore worth investigating as future work.

7. Conclusions and pointers to future work

The focus of this paper is on building a toolbox for the analysis of SCOOP programs, with emphasis on an alias analyzer and a deadlock detector. The naturalness of our approach consists in exploiting the recent formalization of SCOOP in [1], that is executable and implemented in Maude [6]. This provides a unifying framework that can be used not only to reason about the SCOOP model and its design as in [1], but also to analyze SCOOP programs via Maude's analysis tools.

Of particular interest for the aliasing tool is the calculus introduced in [19], which abstracts the aliasing information in terms of explicit access paths referred to as "alias expressions". We provide an extension of this calculus from finite alias relations to infinite ones corresponding to loops and recursive calls. Moreover, we devise an associated RL-based executable specification using $\mathbb{K}$ notation [34]. In Theorem 15 we show that the RL-based machinery implements an algorithm that always terminates with a sound over-approximation of "may aliasing", in non-concurrent settings. This is achieved based on the sound (finitely representable) over-approximation of alias expressions in terms of regular expressions, as in Lemma 13. The integration of the alias calculus on top of the Maude formalization of SCOOP [1] is straightforward, based on the aforementioned executable specification of the calculus. Executions of SCOOP programs can be simulated by simply exploiting the Maude rewriting capabilities, hence the computation of the corresponding aliasing information is immediate.

We agree that it could be worth presenting our analysis as an abstract interpretation (AI) [43]. A modeling exploiting the machinery of AI (based on abstract domains, abstraction and concretization functions, Galois connections, fixed-points, *etc.*) is an interesting topic left for future investigation.

An immediate direction for future work is to identify interesting (industrial) case studies to be analyzed using the framework developed in this paper. We are also interested in devising heuristics comparing the efficiency and the precision (*e.g.*, the number of false positives introduced by the alias approximations) between our approach and other aliasing techniques.

Another research direction is to derive alias-based abstractions for analyzing concurrent programs. We foresee possible connections with the work in [44] on concurrent Kleene algebra formalizing choice, iteration, sequential and concurrent composition of programs. The corresponding definitions exploit abstractions of programs in terms of traces of events that can depend on each other. Thus, obvious challenges in this respect include: (i) defining notions of dependence for all the program constructs in this paper, (ii) relating the concurrent Kleene operators to the semantics of the SCOOP concurrency model and (iii) checking whether fixed-points approximating the aliasing information can be identified via fixed-point theorems.

Furthermore, it would be worth investigating whether the graph-based model of alias relations introduced in [19] can be exploited in order to derive finite $\mathbb{K}$ specifications of the extended alias calculus. In case of a positive answer, the general aim is to study whether this type of representation increases the speed of the reasoning mechanism, and why not – its accuracy. With the same purpose, we refer to a possible integration with the technique in [45] that handles point-to graphs via a stack-based algorithm for fixed-point computations.

Related to deadlock detection, the second contribution of this paper, we provided a formalization based on sets of acquired resources and sets of handlers processors wait to lock in their attempt to execute asynchronously. This definition corresponds to Coffman deadlocks [38] in the context of SCOOP, occurring whenever there is a set of processors reserving each other circularly. We introduced the equivalent SCOOP semantic rule and discussed the results of using Maude in order to analyze deadlocks in the context of a Dining

Philosophers scenario. On the one hand, the SCOOP semantics in [1] is very large, as it incorporates all the aspects of a real concurrency model. On the other hand, the formalization in [1] was tailored for reasoning about the SCOOP model, and not necessarily about SCOOP applications, hence it includes design decisions (such as index-based naming of communication channels) that make the state space grow unacceptably large for model-checking purposes. As a workaround for the model-checking issue, we presented the idea behind building an abstract semantics of SCOOP based on aliases, together with a series of implementation improvements that eventually enabled the Maude LTL model-checker to correctly identify deadlocks. A survey on abstracting techniques on top of Maude executable semantics is provided in [5].

We leave investigating the relationship between the original SCOOP semantics and the aliasing-based one as future work. At this point we observe that the abstract alias semantics introduces false positives (w.r.t. deadlocking as well) due to the over-approximating nature of the alias calculus.

As a clear direction for future work we consider designing and analyzing deadlock situations for more SCOOP applications. Based on the experience so far, this would help better understand and observe the SCOOP state space, thus providing hints for further improvements in the context of model-checking, for instance. As we could already see, major advances in this regard are obtained when semantically equivalent states are identified and collapsed within the same equivalence classes. This was the case of the indexed-based communication channels in Section 6.3. Similar redundant states are introduced by the so-called "regions" in SCOOP "administrating" objects handled by the same processor. Nevertheless, as the rewrite rules defining the semantics of SCOOP are conditional and have rewrites in their conditions, model-checking can still be expensive. A reduction semantics [30, 41] of SCOOP is therefore worth investigating.

Acknowledgements. We are grateful for valuable comments to the anonymous reviewers of SCP, FTSCS'14, Măriuca Asăvoae, Alexander Kogtenkov, José Meseguer, Benjamin Morandi and Sergey Velder.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement no. 291389.

References

[1] B. Morandi, M. Schill, S. Nanz, B. Meyer, Prototyping a concurrency model, in: ACSD, 2013, pp. 170–179.

[2] B. Meyer, Eiffel: The Language, Prentice-Hall, 1991.

[3] F. A. Torshizi, J. S. Ostroff, R. F. Paige, M. Chechik, The SCOOP concurrency model in java-like languages, in: P. H. Welch, H. W. Roebbers, J. F. Broenink, F. R. M. Barnes, C. G. Ritson, A. T. Sampson, G. S. Stiles, B. Vinter (Eds.), The thirty-second Communicating Process Architectures Conference, CPA 2009, organised under the auspices of WoTUG, Eindhoven, The Netherlands, 1-6 November 2009, Vol. 67 of Concurrent Systems Engineering Series, IOS Press, 2009, pp. 7–27. doi:10.3233/978-1-60750-065-0-7.
URL `http://dx.doi.org/10.3233/978-1-60750-065-0-7`

[4] A. Rusakov, J. Shin, B. Meyer, Simple concurrency for robotics with the Roboscoop framework, in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14-18, 2014, 2014, pp. 1563–1569. doi:10.1109/IROS.2014.6942763.
URL `http://dx.doi.org/10.1109/IROS.2014.6942763`

[5] J. Meseguer, G. Rosu, The rewriting logic semantics project: A progress report, Inf. Comput. 231 (2013) 38–69. doi:10.1016/j.ic.2013.08.004.
URL `http://dx.doi.org/10.1016/j.ic.2013.08.004`

[6] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer, C. L. Talcott (Eds.), All About Maude - A High-Performance Logical Framework, How to Specify, Program and Verify Systems in Rewriting Logic, Vol. 4350 of Lecture Notes in Computer Science, Springer, 2007.

[7] W. Landi, B. G. Ryder, Pointer-induced aliasing: A problem classification, in: Proceedings of the 18th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '91, ACM, New York, NY, USA, 1991, pp. 93–103. doi:10.1145/99583.99599.
URL `http://doi.acm.org/10.1145/99583.99599`

[8] W. Landi, Undecidability of static analysis, ACM Lett. Program. Lang. Syst. 1 (4) (1992) 323–337. doi:10.1145/161494.161501.
URL `http://doi.acm.org/10.1145/161494.161501`

[9] E. M. Myers, A precise inter-procedural data flow algorithm, in: Proceedings of the 8th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '81, ACM, New York, NY, USA, 1981, pp. 219–230. doi:10.1145/567532.567556.
URL `http://doi.acm.org/10.1145/567532.567556`

[10] M. Hind, M. Burke, P. Carini, J.-D. Choi, Interprocedural pointer alias analysis, ACM Trans. Program. Lang. Syst. 21 (4) (1999) 848–894. doi:10.1145/325478.325519.
URL http://doi.acm.org/10.1145/325478.325519

[11] A. Diwan, K. S. McKinley, J. E. B. Moss, Type-based alias analysis, SIGPLAN Not. 33 (5) (1998) 106–117. doi:10.1145/277652.277670.
URL http://doi.acm.org/10.1145/277652.277670

[12] M. Burke, P. Carini, J.-D. Choi, M. Hind, Flow-insensitive interprocedural alias analysis in the presence of pointers, in: K. Pingali, U. Banerjee, D. Gelernter, A. Nicolau, D. Padua (Eds.), Languages and Compilers for Parallel Computing, Vol. 892 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 1995, pp. 234–250. doi:10.1007/BFb0025882.
URL http://dx.doi.org/10.1007/BFb0025882

[13] J.-D. Choi, M. Burke, P. Carini, Efficient flow-sensitive interprocedural computation of pointer-induced aliases and side effects, in: Proceedings of the 20th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '93, ACM, New York, NY, USA, 1993, pp. 232–245. doi:10.1145/158511.158639.
URL http://doi.acm.org/10.1145/158511.158639

[14] M. Emami, R. Ghiya, L. J. Hendren, Context-sensitive interprocedural points-to analysis in the presence of function pointers, in: Proceedings of the ACM SIGPLAN 1994 Conference on Programming Language Design and Implementation, PLDI '94, ACM, New York, NY, USA, 1994, pp. 242–256. doi:10.1145/178243.178264.
URL http://doi.acm.org/10.1145/178243.178264

[15] R. P. Wilson, M. S. Lam, Efficient context-sensitive pointer analysis for C programs, in: Proceedings of the ACM SIGPLAN 1995 Conference on Programming Language Design and Implementation, PLDI '95, ACM, New York, NY, USA, 1995, pp. 1–12. doi:10.1145/207110.207111.
URL http://doi.acm.org/10.1145/207110.207111

[16] A. Miné, Field-sensitive value analysis of embedded C programs with union types and pointer arithmetics, in: Proceedings of the 2006 ACM SIGPLAN/SIGBED Conference on Language, Compilers, and Tool Support for Embedded Systems, LCTES '06, ACM, New York, NY, USA, 2006, pp. 54–63. doi:10.1145/1134650.1134659.
URL http://doi.acm.org/10.1145/1134650.1134659

[17] E. Albert, P. Arenas, S. Genaim, G. Puebla, Field-sensitive value analysis by field-insensitive analysis, in: Proceedings of the 2Nd World Congress on Formal Methods, FM '09, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 370–386.

[18] M. Hind, Pointer analysis: haven't we solved this problem yet?, in: PASTE, 2001, pp. 54–61.

[19] A. Kogtenkov, B. Meyer, S. Velder, Alias calculus, change calculus and frame inference, Sci. Comput. Program. 97 (2015) 163–172. doi:10.1016/j.scico.2013.11.006.
URL http://dx.doi.org/10.1016/j.scico.2013.11.006

[20] J. R. Larus, P. N. Hilfinger, Detecting conflicts between structure accesses, in: PLDI, 1988, pp. 21–34.

[21] I. M. Asavoae, Abstract semantics for alias analysis in K, Electr. Notes Theor. Comput. Sci. 304 (2014) 97–110. doi:10.1016/j.entcs.2014.05.005.
URL http://dx.doi.org/10.1016/j.entcs.2014.05.005

[22] M. Hills, P. Klint, J. J. Vinju, Program analysis scenarios in rascal, in: F. Durán (Ed.), Rewriting Logic and Its Applications - 9th International Workshop, WRLA 2012, Held as a Satellite Event of ETAPS, Tallinn, Estonia, March 24-25, 2012, Revised Selected Papers, Vol. 7571 of Lecture Notes in Computer Science, Springer, 2012, pp. 10–30.

[23] C. Shih, J. A. Stankovic, Survey of deadlock detection in distributed concurrent programming environments and its application to real-time systems, Tech. rep., Amherst, MA, USA (1990).

[24] G. R. Andrews, G. M. Levin, On-the-fly deadlock prevention, in: Proceedings of the First ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing, PODC '82, ACM, New York, NY, USA, 1982, pp. 165–172.

[25] T. Minoura, Deadlock avoidance revisited, J. ACM 29 (4) (1982) 1023–1048.

[26] K. Chandy, J. Misra, L. Haas, T. U. A. A. D. O. C. SCIENCES., A Distributed Deadlock Detection Algorithm and Its Correctness Proof, Defense Technical Information Center, 1982.
URL http://books.google.ch/books?id=AH-3NwAACAAJ

[27] D. Badal, M. Gehl, M. Gehl, N. P. S. M. CA., N. P. S. (U.S.), On Deadlock Detection in Distributed Computing Systems, Defense Technical Information Center, 1983.
URL http://books.google.ch/books?id=JjrqOAAACAAJ

[28] C. Ellison, G. Rosu, An executable formal semantics of C with applications, in: J. Field, M. Hicks (Eds.), Proceedings of the 39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2012, Philadelphia, Pennsylvania, USA, January 22-28, 2012, ACM, 2012, pp. 533–544.

[29] C. Ellison, A formal semantics of C with applications, Ph.D. thesis, University of Illinois (July 2012).

[30] M. AlTurki, J. Meseguer, Reduction semantics and formal analysis of Orc programs, Electr. Notes Theor. Comput. Sci. 200 (3) (2008) 25–41.

[31] J. Misra, W. R. Cook, Computation orchestration, Software and System Modeling 6 (1) (2007) 83–110.

[32] G. Caltais, Expression-based aliasing for OO-languages, in: C. Artho, P. C. Ölveczky (Eds.), Formal Techniques for Safety-Critical Systems - Third International Workshop, FTSCS 2014, Luxembourg, November 6-7, 2014. Revised Selected Papers, Vol. 476 of Communications in Computer and Information Science, Springer, 2014, pp. 47–61. doi:10.1007/978-3-319-17581-2.
URL http://dx.doi.org/10.1007/978-3-319-17581-2

[33] M. O. Rabin, D. Scott, Finite automata and their decision problems, IBM J. Res. Dev. 3 (2) (1959) 114–125. doi:10.1147/rd.32.0114.
URL http://dx.doi.org/10.1147/rd.32.0114

[34] G. Rosu, T. F. Serbanuta, K overview and SIMPLE case study, in: Proceedings of International K Workshop (K'11), ENTCS, Elsevier, 2013, To appear.

[35] V. Rusu, D. Lucanu, T. Serbanuta, A. Arusoaie, A. Stefanescu, G. Rosu, Language definitions as rewrite theories, J. Log. Algebr. Meth. Program. 85 (1) (2016) 98–120. doi:10.1016/j.jlamp.2015.09.001.
URL `http://dx.doi.org/10.1016/j.jlamp.2015.09.001`

[36] T.-F. Serbanuta, G. Rosu, K-Maude: A rewriting based tool for semantics of programming languages, in: P. C. Ölveczky (Ed.), WRLA, Vol. 6381 of Lecture Notes in Computer Science, Springer, 2010, pp. 104–122. doi:10.1007/978-3-642-16310-4.
URL `http://dx.doi.org/10.1007/978-3-642-16310-4`

[37] P. Nienaltowski, Practical Framework for Contract-based Concurrent Object-oriented Programming, ETH, 2007.
URL `http://books.google.ch/books?id=ZDcEkgAACAAJ`

[38] E. G. Coffman, M. Elphick, A. Shoshani, System deadlocks, ACM Comput. Surv. 3 (2) (1971) 67–78.

[39] N. Martí-Oliet, J. Meseguer, A. Verdejo, Towards a Strategy Language for Maude, Electronic Notes in Theoretical Computer Science 117 (2005) 417–441.

[40] S. Eker, J. Meseguer, A. Sridharanarayanan, The maude LTL model checker, Electr. Notes Theor. Comput. Sci. 71 (2002) 162–187.

[41] M. Felleisen, D. P. Friedman, A reduction semantics for imperative higher-order languages, in: J. W. de Bakker, A. J. Nijman, P. C. Treleaven (Eds.), PARLE, Parallel Architectures and Languages Europe, Volume II: Parallel Languages, Eindhoven, The Netherlands, June 15-19, 1987, Proceedings, Vol. 259 of Lecture Notes in Computer Science, Springer, 1987, pp. 206–223.

[42] A. Bouajjani, J. Esparza, O. Maler, Reachability analysis of pushdown automata: Application to model-checking, in: CONCUR, 1997, pp. 135–150.

[43] P. Cousot, R. Cousot, Abstract interpretation and application to logic programs, J. Log. Program. 13 (2&3) (1992) 103–179.

[44] C. A. R. Hoare, B. Möller, G. Struth, I. Wehrman, Concurrent Kleene algebra, in: CONCUR 2009 - Concurrency Theory, 20th International Conference, CONCUR 2009, Bologna, Italy, September 1-4, 2009. Proceedings, 2009, pp. 399–414.

[45] D. R. Chase, M. N. Wegman, F. K. Zadeck, Analysis of pointers and structures, in: PLDI, 1990, pp. 296–310.

[46] E. Giachino, C. A. Grazia, C. Laneve, M. Lienhardt, P. Y. H. Wong, Deadlock analysis of concurrent objects: Theory and practice, in: Integrated Formal Methods, 10th International Conference, IFM 2013, Turku, Finland, June 10-14, 2013. Proceedings, 2013, pp. 394–411.

[47] A. Heußner, C. M. Poskitt, C. Corrodi, B. Morandi, Towards practical graph-based verification for an object-oriented concurrency model, in: Proc. Graphs as Models (GaM 2015), Vol. 181 of Electronic Proceedings in Theoretical Computer Science, 2015, pp. 32–47.

[48] A. H. Ghamarian, M. de Mol, A. Rensink, E. Zambon, M. Zimakova, Modelling and analysis using groove, Int. J. Softw. Tools Technol. Transf. 14 (1) (2012) 15–40.