Abstract Interpretation with Specialized Definitions

Germán Puebla[1] Elvira Albert[2] Manuel V. Hermenegildo[1,3]

[1]School of Computer Science
T. U. of Madrid (UPM)

[2] School of Computer Science
Complutense U. of Madrid

[3]Depts. of CS and ECE
U. of New Mexico

ABSTRACT

The relationship between abstract interpretation and partial evaluation has received considerable attention and (partial) integrations have been proposed starting from both the partial evaluation and abstract interpretation perspectives. In this work we present what we argue is the first generic algorithm for efficient and precise integration of abstract interpretation and partial evaluation from an abstract interpretation perspective. Taking as starting point state-of-the-art algorithms for context-sensitive, polyvariant abstract interpretation and (abstract) partial evaluation of logic programs, we present an algorithm which combines the best of both worlds. Key ingredients include the accurate success propagation inherent to abstract interpretation and the powerful program transformations achievable by partial deduction. In our algorithm, the calls which appear in the analysis graph are not analyzed w.r.t. the original definition of the procedure but w.r.t. *specialized definitions* of these procedures. Such specialized definitions are obtained by applying both unfolding and abstract executability. Also, our framework is parametric w.r.t. different control strategies and abstract domains. Different combinations of these parameters correspond to existing algorithms for program analysis and specialization. Finally, our approach efficiently computes strictly more precise results than those achievable by each of the individual techniques. The algorithm is now one of the key components of CiaoPP, the analysis and specialization system of the Ciao compiler. For concreteness, we have developed the algorithms for logic programming, although our approach is general and can be applied to other programming paradigms.

1. INTRODUCTION AND MOTIVATION

The relationship between abstract interpretation [4] and partial evaluation [13] has received considerable attention (see for example [7, 9, 3, 17, 12, 14, 23, 25, 8, 18, 5, 24, 15] and their references). In order to motivate and illustrate our proposal for an integration of abstract interpretation and partial evaluation, we use the running "challenge" example of Fig. 1. It is a simple Ciao [2] program which uses Peano's arithmetic.[1] We use the Ciao assertion language in order to provide precise descriptions of the initial call patterns. In our case, the entry declaration is used to inform that all calls to the only exported predicate (i.e., main/2) will always be of the form $\leftarrow$ main(s(s(s(L))),R) with L ground and R a variable. The predicate main/2 performs two calls to predicate formula/2. A call formula(X,W) performs mode tests ground(X) and var(W) on its input arguments and returns $W = (X - 2) \times 2$. Predicate two/1 returns s(s(0)), i.e., the natural number 2. A call minus(A,B,C) returns $C = A - B$. However, if the result becomes a negative number, C is left as a free variable. This indicates that the result is not valid. In turn, a call twice(A,B) returns $B = A \times 2$. Prior to computing the result, this predicate checks whether A is valid, i.e., not a variable, and simply returns a variable otherwise.

```
:- module(_,[main/2],[assertions]).
:- entry main(s(s(s(L))),R) : (ground(L),var(R)).

main_1(X,X2) :- formula_{1,1}(X,X1), formula_{1,2}(X1,X2),
                ground_{1,3}(X2).

formula_2(X,W) :- ground_{2,1}(X),var_{2,2}(W),two_{2,3}(T),
                  minus_{2,4}(T,X,X2),twice_{2,5}(X2,W).

two_3(s(s(0))).

minus_4(X,0,X).
minus_5(s(X),s(Y),R) :- minus_{5,1}(X,Y,R).
minus_6(0,s(_Y),_R).

twice_7(X,_Y) :- var_{7,1}(X).
twice_8(X,Y) :- ground_{8,1}(X), tw_{8,2}(X,Y).

tw_9(0,0).
tw_10(s(X),s(s(NX))) :- tw_{10,1}(X,NX).
```

Figure 1: Running Example

By observing the behaviour of the program it can be seen that for initial queries satisfying the entry declaration, all calls to the tests ground$_{1,3}$(X), ground$_{2,1}$(X), and var$_{2,2}$(W) will definitely succeed, and can be replaced by *true*, even if we do not know the concrete values of variable L at compile time. Also, the calls to ground$_{8,1}$(X) will suc-

PEPM '06 Charleston, South Carolina

[1] Rules are written with a unique subscript attached to the head atom (the rule number), and a dual subscript (rule number, body position) attached to each body literal for later reference. We sometimes use this notation for denoting calls to atoms as well.

ceed, while the calls to $\texttt{var}_{7,1}$`(X)` will fail, and can thus be replaced by *fail*. These kinds of optimizations require abstract information from analysis (e.g., groundness and freeness). Thus, the example illustrates the benefits of (1) *exploiting abstract information in order to abstractly execute certain atoms. Furthermore, this may allow unfolding of other atoms.* However, the use of an abstract domain which captures groundness and freeness information will in general not be sufficient to determine that in the second execution of `formula/2` the tests $\texttt{ground}_{2,1}$`(X)` and $\texttt{var}_{2,2}$`(W)` will also succeed. The reason is that, on success of $\texttt{minus}_{2,4}$`(T,X,X2)`, `X2` cannot be guaranteed to be ground since $\texttt{minus}_6$`/3` succeeds with a free variable in its third argument position. It can be observed, however, that for all calls to `minus/3` in executions described by the `entry` declaration, the third clause for `minus/3` is useless. It will never contribute to a success of `minus/3` since such predicate is always called with a value greater than zero in its second argument. Unfolding can make this explicit by fully unfolding calls to `minus/3` since they are sufficiently instantiated (and as a result the "dangerous" third clause is disregarded). This unfolding allows concluding that in our particular context all calls to `minus/3` succeed with a ground third argument. This illustrates the importance of (2) *performing unfolding steps in order to prune away useless branches, and that this will result in improved success information.* By the time execution reaches $\texttt{twice}_{2,5}$`(X2,W)`, we hopefully know that `X2` is ground. In order to determine that upon success of $\texttt{twice}_{2,5}$`(X2,W)` (and thus on success of $\texttt{formula}_{1,1}$`(X,W)`) `W` is ground, we need to perform a fixpoint computation. Since, for example, the success substitution for $\texttt{formula}_{1,1}$`(X,X1)` is indeed the call substitution for $\texttt{formula}_{1,2}$`(X1,X2)`, the success of the second test $\texttt{ground}_{2,1}$`(X)` (i.e., the one reachable from $\texttt{formula}_{1,2}$`(X1,X2)`) cannot be established unless we propagate success substitutions. This illustrates the importance of (3) *propagating (abstract) success information, and performing fixpoint computations when needed, and that this simultaneously will result in an improved unfolding.* Finally, whenever we call `formula(X,W)` `W` is a variable, a property which cannot be captured if we restrict ourselves to downwards-closed domains. This indicates (4) *the usefulness of having information on non* downwards-closed *properties.*

Throughout the paper we show that the framework we propose addresses the issues mentioned, and in the particular case of our challenge example can indeed eliminate all calls to mode tests `ground/1` and `var/1`, and fully unfold predicates `two/1` and `minus/3` so that they no longer appear in the residual code. We have used *sharing–freeness* as abstract domain instead of one based on, say, regular types for two reasons. First, to be able to later illustrate how non-downwards closed information, including freeness and definite independence, can be correctly exploited by our algorithm in order to optimize the program, and second, to show how unfolding can be of great use in order to improve the accuracy of analyses apparently unrelated to partial deduction, such as the classical *sharing–freeness.*

Example 1.1. `CiaoPP`*, which implements our proposed abstract interpretation with specialized definitions, produces the following specialized code for the example of Fig. 1 (rules are renamed using the prefix* `sp`*):*

```
sp_main1(s(s(s(0))),0).
sp_main2(s(s(s(s(B)))),A) :- sp_tw2,1(B,C),
                              sp_formula2,2(C,A).
sp_tw2(0,0).
sp_tw3(s(A),s(s(B))) :- sp_tw3,1(A,B).
sp_formula4(0,s(s(s(s(0))))).
sp_formula5(s(A),s(s(s(s(s(s(B))))))) :- sp_tw5,1(A,B).
```

In addition, the algorithm also produces an accurate analysis for such program. In particular, the success information for `sp_main(X,X2)` *guarantees that* `X2` *is definitely ground on success. Note that this is equivalent to proving* $\forall X \geq 3,\ main(X, X2) \rightarrow X2 \geq 0$*. Furthermore, our system is able to get to that conclusion even if the* `entry` *only informs about* `X` *being any possible ground term and* `X2` *a free variable. This is because, during the computation of the specialized definitions, the branches corresponding to values of* `X` *smaller than 3 are detected to be failing and the residual code is indeed equivalent to the one achieved with the more precise* `entry` *declaration. This illustrates how our proposal is useful for improving the results of analysis even in cases where there are no initial constants in the query which can be propagated through the program.*

The above results cannot be achieved unless all four points mentioned before are addressed by a program analysis/specialization system. For example, if we use traditional partial deduction [21, 10] (PD) with the corresponding *Generalize* and *Unfold* rules followed by abstract interpretation and *abstract specialization* as described in [23, 24] we only obtain a comparable program after four iterations of the: "PD + abstract interpretation + abstract specialization" cycle. This shows the importance of achieving an algorithm which is able to *interleave* PD with abstract interpretation, extended with abstract specialization, in order to communicate the accuracy gains achieved from one to the other as soon as possible. In any case, iterating over "PD + analysis" is not a good idea from the efficiency point of view.

Figure 2 shows an overview of our *abstract interpretation with specialized definitions* proposal. The main idea is that a generic abstract interpreter, depicted within the outermost box, is equipped with a generator of specialized definitions, depicted within the innermost box. Such generator provides, upon request, the specialized definitions to be analyzed by the interpreter. Certain data structures, which take the form of tables in the figure, are used to communicate between the two processes and achieve a smooth interleaving. The input of the whole process is a program together with a set of calling patterns for it. The output is a specialized program together with the analysis results inferred for it. The scheme can be parameterized with different (abstract) unfolding rules, generalization operators, abstract domains and widenings —which appear within oval boxes in the figure. The different instances give rise to interesting analysis and specialization methods, some of which are well known and others are novel.

The rest of the paper is organized as follows. Section 2 recalls some preliminary concepts. In Sect. 3 we present abstract unfolding which already integrates abstract executability. Section 4 introduces our notion of specialized definition and embeds it within an abstract partial deducer. In Sect. 5 we propose our scheme for abstract interpretation with specialized definitions. Section 6 discusses how to interpret the results of our algorithm. Finally, Sect. 7 compares to related work and Sect. 8 concludes.

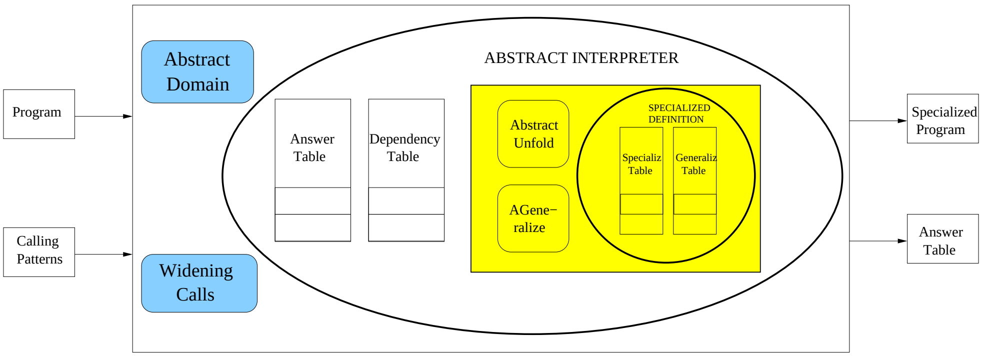

Figure 2: Overview of Abstract Interpreter with Specialized Definition

2. PRELIMINARIES

This section introduces some preliminary concepts on abstract interpretation [4] and partial deduction [21].

We assume some basic knowledge on the terminology of logic programming (see for example [20] for details). Very briefly, an *atom* A is a syntactic construction of the form $p(t_1, \ldots, t_n)$, where p/n, with $n \geq 0$, is a predicate symbol and $t_1, \ldots, t_n$ are terms. A *clause* is of the form $H \leftarrow B$ where its head H is an atom and its body B is a conjunction of atoms. A *definite program* is a finite set of clauses. A *goal* (or query) is a conjunction of atoms.

2.1 The Notions of Unfolding and Resultant

Let G be a goal of the form $\leftarrow A_1, \ldots, A_R, \ldots, A_k$, $k \geq 1$. The concept of *computation rule*, denoted by $\mathcal{R}$, is used to select an atom within a goal for its evaluation. The operational semantics of definite programs is based on derivations [20]. Let $C = H \leftarrow B_1, \ldots, B_m$ be a renamed apart clause in P such that $\exists \theta = mgu(A_R, H)$. Then, the goal $\leftarrow \theta(A_1, \ldots, A_{R-1}, B_1, \ldots, B_m, A_{R+1}, \ldots, A_k)$ is *derived* from G and C via $\mathcal{R}$. As customary, given a program P and a goal G, an *SLD derivation* for $P \cup \{G\}$ consists of a possibly infinite sequence $G = G_0, G_1, G_2, \ldots$ of goals, a sequence $C_1, C_2, \ldots$ of properly renamed apart clauses of P, and a sequence $\theta_1, \theta_2, \ldots$ of mgus such that each G_{i+1} is derived from G_i and C_{i+1} using θ_{i+1}. A derivation step can be non-deterministic when A_R unifies with several clauses in P, giving rise to several possible SLD derivations for a given goal. Such SLD derivations can be organized in *SLD trees*. A finite derivation $G = G_0, G_1, G_2, \ldots, G_n$ is called *successful* if G_n is empty. In that case $\theta = \theta_1\theta_2 \ldots \theta_n$ is called the computed answer for goal G. Such a derivation is called *failed* if G_n is not empty and it is not possible to perform a derivation step from it.

Given an atom A, an *unfolding rule* [21, 10] computes a set of finite SLD derivations $D_1, \ldots, D_n$ (i.e., a possibly incomplete SLD tree) of the form $D_i = A, \ldots, G_i$ with computed answer substitution θ_i for $i = 1, \ldots, n$ whose associated *resultants* (or residual rules) are $\theta_i(A) \leftarrow G_i$.

2.2 Abstract Interpretation

Abstract interpretation [4] provides a general formal framework for computing safe approximations of program behaviour. Programs are interpreted using *abstract values* instead of *concrete values*. An abstract value is a finite representation of a, possibly infinite, set of actual values in the concrete domain D. The set of all possible abstract values constitutes the *abstract domain*, denoted D_α, which is usually a complete lattice or cpo which is ascending chain finite. The subset relation $\subseteq$ induces a partial order on sets of concrete values. The $\subseteq$ relation induces the $\sqsubseteq$ relation on abstract values. Values in the abstract domain $\langle D_\alpha, \sqsubseteq \rangle$ and sets of values in the concrete domain $\langle 2^D, \subseteq \rangle$ are related via a pair of monotonic mappings $\langle \alpha, \gamma \rangle$: the *abstraction* function $\alpha : 2^D \rightarrow D_\alpha$ which assigns to each (possibly infinite) set of concrete values an abstract value, and the *concretization* function $\gamma : D_\alpha \rightarrow 2^D$ which assigns to each abstract value the (possibly infinite) set of concrete values it represents. The following operations on abstract substitutions are domain-dependent and will be used in our algorithms:

- $\mathsf{Arestrict}(\lambda, E)$ performs the abstract restriction (or projection) of a substitution λ to the set of variables in the expression E, denoted $vars(E)$;
- $\mathsf{Aextend}(\lambda, E)$ extends the substitution λ to the variables in the set $vars(E)$;
- $\mathsf{Aunif}(t_1, t_2, \lambda)$ obtains the description which results from adding the abstraction of the unification $t_1 = t_2$ to the substitution λ;
- $\mathsf{Aconj}(\lambda_1, \lambda_2)$ performs the abstract conjunction of two substitutions;
- $\mathsf{Alub}(\lambda_1, \lambda_2)$ performs the abstract disjunction ($\sqcup$) of two substitutions.

An *abstract atom* of the form $A : CP$ is a concrete atom A which comes equipped with an *abstract substitution* CP which is defined over $vars(G)$ and provides additional information on the context in which the atom will be executed at run-time. In our algorithms, we also use $\mathsf{Atranslate}(A : CP, H \leftarrow B)$ which adapts and projects the information in an abstract atom $A : CP$ to the variables in the clause $C = H \leftarrow B$. This operation can be defined in terms of the operations above as: $\mathsf{Atranslate}(A : CP, H \leftarrow B) = \mathsf{Arestrict}(\mathsf{Aunif}(A, H, \mathsf{Aextend}(CP, C)), C)$.

As customary, the most general abstract substitution is represented as $\top$, and the least general (empty) abstract substitution as $\bot$.

Finally, the following standard operations are used in the algorithms to handle keyed-tables: $\mathsf{Create\_Table}(T)$ initializes a table T. $\mathsf{Insert}(T, Key, Info)$ adds $Info$ associated to Key to T and deletes previous information associated to Key, if any. $\mathsf{IsIn}(T, Key)$ returns true iff Key is currently stored in the table. Finally, $\mathsf{Look\_up}(T, Key)$ returns the information associated to Key in T. For simplicity, we sometimes consider tables as sets and we use the notation $(Key \rightsquigarrow Info) \in T$ to denote that there is an entry in the table T with the corresponding *Key* and associated *Info*.

3. UNFOLDING WITH ABSTRACT SUBSTITUTIONS

We now present our notion of *abstract unfolding* —based on an extension of the SLD semantics which exploits abstract information— which is used later to generate specialized definitions. This will pave the way to overcome difficulties (1) and (2) presented in Section 1.

3.1 SLD with Abstract Substitutions

Our extended semantics handles *abstract goals* of the form $G : CP$, i.e., a concrete goal G equipped with an *abstract substitution* CP. The first rule captures derivation steps.

Definition 3.1 (derivation step). *Let $G : CP$ be an abstract goal where $G = \leftarrow A_1, \ldots, A_R, \ldots, A_k$ and CP is an abstract substitution defined over $vars(G)$. Let $\mathcal{R}$ be a computation rule and let $\mathcal{R}(G) = A_R$. Let $C = H \leftarrow B_1, \ldots, B_m$ be a renamed apart clause in P. Then the abstract goal $G' : CP'$ is* derived *from $G : CP$ and C via $\mathcal{R}$ if $\exists \theta = mgu(A_R, H) \wedge CP_u \neq \bot$, where:*

$$CP_u = \mathsf{Aunif}(A_R, H\theta, \mathsf{Aextend}(CP, C\theta))$$
$$G' = \theta(A_1, \ldots, A_{R-1}, B_1, \ldots, B_m, A_{R+1}, \ldots, A_k)$$
$$CP' = \mathsf{Arestrict}(CP_u, vars(G'))$$

An important difference between the above definition and the standard derivation step is that the use of abstract (call) substitutions allows imposing further conditions for performing derivation steps, in particular, CP_u cannot be $\bot$. This is because if $CP \neq \bot$ and $CP_u = \bot$ then the head of the clause C is incompatible with CP and the unification $A_R = H$ will definitely fail at run-time. Thus, abstract information allows us to remove useless clauses from the residual program. This produces more efficient resultants and increases the accuracy of analysis for the residual code.

Example 3.2. *Consider the goal:* $\mathtt{formula(s^4(X), X2)} : \{\mathtt{X/G, X2/V}\}$ *which appears during the analysis of our running example (c.f. Fig. 3). We abbreviate as* $\mathtt{s^n(X)}$ *the successive application of* `n` *functors* `s` *to variable* `X`*. We have used sharing-freeness as abstract domain in the analysis though, for simplicity, we will represent the results using traditional "modes": the notation* `X/G` *(resp.* `X/V`*) indicates that variable* `X` *is ground (resp. free). After applying a derivation step using the only rule for* `formula`*, we derive:*

$\mathtt{ground(s^4(X)), var(X2), two(T), minus(T, s^4(X), X2'), twice(X2', X2)} :$
$\{\mathtt{X/G, X2/V, T/V, X2'/V}\}$

where the abstract description has been extended with updated information about the freeness of the newly introduced variables. In particular, both `T` *and* `X2'` *are* `V`*.*

The second rule we present makes use of the availability of abstract substitutions to perform *abstract executability* [23] during resolution. This allows replacing some atoms with simpler ones, and, in particular, with the predefined atoms *true* and *false*, provided certain conditions hold. We assume the existence of a predefined *abstract executability table* which contains entries of the form $T : CP \rightsquigarrow T'$ which specify the behaviour of external procedures: builtins, libraries, and other user modules. For instance, for predicate `ground` the abstract execution table contains the information $\mathtt{ground(X)} : \{\mathtt{X/G}\} \rightsquigarrow \mathtt{true}$. For `var`, it contains $\mathtt{var(X)} : \{\mathtt{X/V}\} \rightsquigarrow \mathtt{true}$.[2]

Definition 3.3 (abstract execution). *Let $G : CP$ be an abstract goal where $G = \leftarrow A_1, \ldots, A_R, \ldots, A_k$. Let $\mathcal{R}$ be a computation rule and let $\mathcal{R}(G) = A_R$. Let $(T : CP_T \rightsquigarrow T')$ be a renamed apart entry in the abstract executability table. Then, the goal $G' : CP'$ is* abstractly executed *from $G : CP$ and $(T : CP_T \rightsquigarrow T')$ via $\mathcal{R}$ if $A_R = \theta(T)$ and $CP_A \sqsubseteq CP_T$, where*

$$G' = A_1, \ldots, A_{R-1}, \theta(T'), A_{R+1}, \ldots, A_k$$
$$CP' = \mathsf{Arestrict}(CP, G')$$
$$CP_A = \mathsf{Atranslate}(A_R : CP, T \leftarrow true)$$

Example 3.4. *From the derived goal in Ex. 3.2, we can apply twice the above rule to abstractly execute the calls to* `ground` *and* `var` *and obtain:*

$$\mathtt{two(T), minus(T, s^4(X), X2'), twice(X2', X2)} :$$
$$\{\mathtt{X/G, X2/V, T/V, X2'/V}\}$$

since both calls succeed by using the abstract executability table described above given the information in the abstract substitution.

3.2 Abstract Unfolding

In our framework, resultants for abstract atoms will be obtained using abstract unfolding in a similar way as it is done in the concrete setting using unfolding (see Sect. 2.1).

Definition 3.5 (*AUnfold*). *Let $A : CP$ be an abstract atom and P a program. We define $AUnfold(P, A : CP)$ as the set of* resultants *associated to a finite (possibly incomplete) SLD tree computed by applying the rules of Definitions 3.1 and 3.3 to $A : CP$.*

The so-called *local control* of PD ensures the termination of the above process. For this purpose, the unfolding rule must incorporate some mechanism to stop the construction of SLD derivations (we refer to [16] for details).

Example 3.6. *Consider an unfolding rule AUnfold based on homeomorphic embedding [16] to ensure termination and the initial goal in Ex. 3.2. The derivation continuing from Ex. 3.4 performs several additional derivation steps and abstract executions and branches (we do not include them due to space limitations and also because it is well understood). The following resultants are obtained from the resulting tree:*

```
formula(s(s(s(s(0),s(s(s(0))))).
formula(s(s(s(s(A))))),s(s(s(s(s(B)))))) :-
        tw(A,B)
```

which will later be filtered and renamed as they appear in rules 5 and 6 of Ex. 1.1.

It is important to note that SLD resolution with abstract substitutions is not restricted to the left-to-right computation rule. However, it is well-known that non-leftmost derivation steps can produce incorrect results if the goal contains *impure* atoms to the left of A_R. More details can be found, e.g., in [19]. Also, abstract execution of non-leftmost atoms can be incorrect if the abstract domain used captures properties which are not downwards closed. A simple solution is to only allow leftmost abstract execution stop for non-downwards closed domains (and non-leftmost for derivation steps).

[2] In `CiaoPP` we use assertions to express such information in a domain-independent manner.

Algorithm 1 Abstract Partial Deduction with Specialized Definitions

```
 1: procedure PARTIAL_EVALUATION_WITH_SPEC_DEFS($P, \{A_1 : CP_1, \ldots, A_n : CP_n\}$)
 2:     Create_Table($\mathcal{GT}$); Create_Table($\mathcal{ST}$)
 3:     for $j = 1..n$ do
 4:         PROCESS_CALL_PATTERN($A_j : CP_j$)
 5: procedure PROCESS_CALL_PATTERN($A : CP$)
 6:     if not IsIn($\mathcal{GT}, A : CP$) then
 7:         $(A_1, A'_1) \leftarrow$ SPECIALIZED_DEFINITION($P, A : CP$)
 8:         $A_1 : CP_1 \leftarrow$ Look_up($\mathcal{GT}, A : CP$)
 9:         for all ren. apart clause $C_k = H_k \leftarrow B_k \in P$ s.t. $H_k$
                unifies with $A'_1$ do
10:             $CP_k \leftarrow$ Atranslate($A'_1 : CP_1, C_k$)
11:             PROCESS_CLAUSE($CP_k,\ B_k$)
12: procedure PROCESS_CLAUSE($CP,\ B$)
13:     if $B = [L|R]$ then
14:         $CP_L \leftarrow$ Arestrict($CP, L$)
15:         PROCESS_CALL_PATTERN($L : CP_L$)
16:         PROCESS_CLAUSE($CP,\ R$)
17: function SPECIALIZED_DEFINITION($P, A : CP$)
18:     $A' : CP' \leftarrow AGeneralize(\mathcal{ST}, A : CP)$
19:     Insert($\mathcal{GT}, A : CP, A' : CP'$)
20:     if IsIn($\mathcal{ST}, A' : CP'$) then
21:         $A'' \leftarrow$Look_up($\mathcal{ST}, A' : CP'$)
22:     else
23:         $Def \leftarrow AUnfold(P, A' : CP')$
24:         $A'' \leftarrow$ new_filter($A'$)
25:         Insert($\mathcal{ST}, A' : CP', A''$)
26:         $Def' \leftarrow \{(H' \leftarrow B) \mid (H \leftarrow B) \in Def \wedge H' =$
                ren$(H, \{A'/A''\})\}$
27:         $P \leftarrow P \bigcup Def'$
28:     return $(A', A'')$
```

4. SPECIALIZED DEFINITIONS

Typically, PD is presented as an iterative process in which partial evaluations are computed for the new generated atoms until they *cover* all calls which can appear in the execution of the residual program. This is formally known as the *closedness* condition of PD [21]. In order to ensure termination of this global process, the so-called *global* control defines a *Generalize* operator (see [16]) which guarantees that the number of SLD trees computed is kept finite, i.e., it ensures the finiteness of the set of atoms for which partial evaluation is produced. However, the residual program is not generated until such iterative process terminates.

We now define an Abstract Partial Deduction (APD) algorithm whose execution can later be *interleaved* in a seamless way with a state-of-the-art abstract interpreter. For this it is essential that the APD process be able to generate residual code for each call pattern as soon as we finish processing it. This will make it possible for the analysis algorithm to have access to the improved definition. As a consequence, the accuracy of the analyzer may be increased and objective (2) described in Sect. 1 achieved.

4.1 Abstract Partial Deduction

Algorithm 1 presents an APD algorithm. The main difference with standard algorithms for APD is that the resultants computed by *AUnfold* (L23) are added to the program during execution of the algorithm (L27) rather than in a later code generation phase. In order to avoid conflicts among the new clauses and the original ones, clauses for specialized definitions are renamed with a fresh predicate name (L26) prior to adding them to the program (L27). The algorithm uses two global data structures. The *specialization table* contains entries of the form $A : CP \rightsquigarrow A'$. The atom A' provides the link with the clauses of the specialized definition for $A : CP$. The *generalization table* stores the results of the $AGeneralize$ function and contains entries $A : CP \rightsquigarrow A' : CP'$ where $A' : CP'$ is a generalization of $A : CP$, in the sense that $A = A'\theta$ and $(A : CP) \sqsubseteq (A' : CP')$.

Procedure PARTIAL_EVALUATION_WITH_SPEC_DEFS (L1-4) initiates the computation. It first initializes the tables and then calls PROCESS_CALL_PATTERN for each abstract atom $A_j : CP_j$ in the initial set to be partially evaluated. The task of PROCESS_CALL_PATTERN is, if the atom has not been processed yet (L6), to compute a specialized definition for it (L7) and then process all clauses in its specialized definition by means of calls to PROCESS_CLAUSE (L9-11). For simplicity of the presentation, we assume that clause bodies returned by SPECIALIZED_DEFINITION are represented as lists rather than conjunctions. Procedure PROCESS_CLAUSE traverses clause bodies, processing their corresponding atoms by means of calls to PROCESS_CALL_PATTERN, in a depth-first, left-to-right fashion. In contrast, the order in which pending call patterns (atoms) are handled is usually not fixed in APD algorithms. They are often all put together in a set. The purpose of the two procedures PROCESS_CLAUSE and PROCESS_CALL_PATTERN is to traverse the clauses in the left-to-right order and add the corresponding call patterns. In principle, this does not have additional advantages w.r.t. existing APD algorithms because success propagation has not been integrated yet. However, the reason for our presentation is to be as close as possible to our analysis algorithm with success propagation which enforces a depth-first, left-to-right traversal of program clauses.

The correctness of Algorithm 1 can be established using the framework for APD in [15].

4.2 Limitations of APD

It is important to note that Algorithm 1 does not perform success propagation yet (difficulty 3). In L16, it becomes apparent that all atom(s) in R will be analyzed with the same call pattern CP as L, which is to their left in the clause. This may clearly lead to substantial precision loss. For instance, the abstract pattern `formula(C,A)` : `{C/G,A/V}` which is necessary to obtain the last two resultants of Ex. 1.1 cannot be obtained with this algorithm. In particular, we cannot infer the groundness of `C` which, in turn, prevents us from abstractly executing the next call to `ground` and, thus, from obtaining this optimal specialization.

In addition, this lack of success propagation makes it difficult or even impossible to work with non downwards closed domains (difficulty 4), since CP may contain information which holds before execution of the leftmost atom L but which can be uncertain or even false after that. In fact, in our example CP contains the info `C/V`, which becomes false after execution of `tw(B,C)`, since now `C` is ground. This problem is solved in the algorithm we present in the next section, where analysis information flows from left to right, adding more precise information and eliminating information which is no longer safe or even definitely wrong.

4.3 Integration with Abstract Interpreter

For the integration we propose, the most relevant part of the algorithm comprises L17-28, as it is the code fragment which is *directly* executed from our abstract interpreter. The

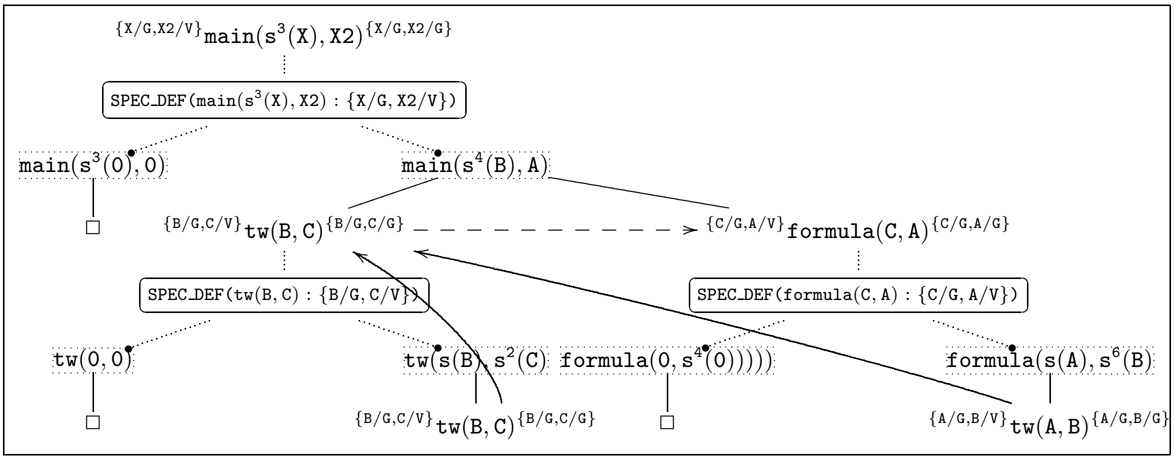
{X/G,X2/V}main(s3(X), X2){X/G,X2/G}
SPEC_DEF(main(s3(X), X2) : {X/G, X2/V})
main(s3(0), 0)
main(s4(B), A)
□
{B/G,C/V}tw(B, C){B/G,C/G}
{C/G,A/V}formula(C, A){C/G,A/G}
SPEC_DEF(tw(B, C) : {B/G, C/V})
SPEC_DEF(formula(C, A) : {C/G, A/V})
tw(0, 0)
tw(s(B), s2(C))
formula(0, s4(0)))))
formula(s(A), s6(B))
□
{B/G,C/V}tw(B, C){B/G,C/G}
□
{A/G,B/V}tw(A, B){A/G,B/G}

Figure 3: Analysis Graph computed by ABS_INT_WITH_SPEC_DEF

remaining procedures (L1-L16) will be overridden by more accurate ones later. The procedure of interest is SPECIALIZED_DEFINITION. As it is customary, it performs (L18) a generalization of the call $A : CP$ using the abstract counterpart of the *Generalize* operator, denoted by *AGeneralize*, and which is in charge of ensuring termination at the global level. The result of the generalization, $A' : CP'$, is inserted (L19) in the generalization table $\mathcal{GT}$. Correctness of the algorithm requires that $(A : CP) \sqsubseteq (A' : CP')$. If $A' : CP'$ has been previously treated (L20), then its specialized definition A'' is looked up in $\mathcal{ST}$ (L21) and returned. Otherwise, a specialized definition *Def* is computed for it by using the *AUnfold* operator of Def. 3.5 (L23).

As already mentioned, the specialized definition *Def* for the abstract atom $A : CP$ is used to extend the original program P. First, the atom A' is renamed by using new_filter which returns an atom with a fresh predicate name, A'', and optionally filters constants out (L24). Then, function ren is applied to rename the clause heads using atom A' (L26). The function $\mathsf{ren}(A, \{B/B'\})$ returns $\theta(B')$ where $\theta = mgu(A, B)$. Finally, the program P is extended with the new, *renamed* specialized definition, Def'.

EXAMPLE 4.1. *Three calls to* SPECIALIZED_DEFINITION *appear (within an oval box) during the analysis of our running example in Fig. 3 from the following abstract atoms, first* $\mathtt{main(s^3(X), X2) : \{X/G, X2/V\}}$*, then* $\mathtt{tw(B, C) : \{B/G, C/V\}}$ *and finally* $\mathtt{formula(C, A) : \{C/G, A/V\}}$*. The output of such executions is used later (with the proper renaming) to produce the resultants in Ex. 1.1. For instance, the second clause obtained from the first call to* SPECIALIZED_DEFINITION *is*

sp_main$_2$(s(s(s(s(B))))),A) :- tw$_{2,1}$(B,C),
formula$_{2,2}$(C,A).

where only the head is renamed. The renaming of the body literals is done in a later code-generation phase (see Section 6.1). As already mentioned, Alg. 1 is not able to obtain the three abstract atoms above due to the absence of success propagation.

5. ABSTRACT INTERPRETATION WITH SPECIALIZED DEFINITIONS

We now present our final algorithm for abstract interpretation with specialized definitions. This algorithm extends both the APD Algorithm 1 and the abstract interpretation algorithms in [22, 11]. The main improvement w.r.t. Algorithm 1 is the addition of success propagation. Unfortunately, this requires computing a global fixpoint. It is an important objective for us to be able to compute an accurate fixpoint in an efficient way. The main improvements w.r.t the algorithms in [22, 11] are the following. (1) It deals directly with non-normalized programs. This point, which does not seem very relevant in a pure analysis system, becomes crucial when combined with a specialization system in order to profit from constants propagated by unfolding. (2) It incorporates a hardwired efficient graph traversal strategy which eliminates the need for maintaining priority queues explicitly [11]. (3) The algorithm includes a widening operation for calls, *Widen_Call*, which limits the amount of multivariance in order to keep the number of call patterns analyzed finite. This is required in order to be able to use abstract domains which are infinite, such as regular types. (4) It also includes a number of simplifications to facilitate understanding, such as the use of the keyed-table ADT, which we assume encapsulates proper renaming apart of variables and the application of renaming transformations when needed. (5) It interleaves program analysis and specialization in a way that is efficient, accurate, and practical.

5.1 The Program Analysis Graph

In order to compute and propagate success substitutions, Algorithm 2 computes a *program analysis graph* in a similar fashion as state of the art analyzers such as the CiaoPP analyzer [22, 11]. For instance, the analysis graph computed by Algorithm 2 for our running example is depicted in Fig. 3.

The graph has two sorts of nodes. Those which correspond to atoms are called “OR-nodes”. An OR-node of the form ${}^{CP}A^{AP}$ is interpreted as the answer (success) pattern for the abstract atom $A : CP$ is AP. For instance, the OR-node ${}^{\{X/G,X2/V\}}\mathtt{main(s^3(X), X2)}^{\{X/G,X2/G\}}$ indicates that when the atom $\mathtt{main(s^3(X), X2)}$ is called with description $\{\mathtt{X/G, X2/V}\}$ the answer (or success) substitution computed is $\{\mathtt{X/G, X2/G}\}$.

Those nodes which correspond to rules are called “AND-nodes”. In Fig. 3, they appear within a dashed box and contain the head of the corresponding clause. Each AND-node has as children as many OR-nodes as literals there are in the body. If a child OR-node is already in the tree, it is not expanded any further and the currently available answer is used. For instance, the analysis graph in Figure 3 contains three occurrences of the abstract atom $\texttt{tw(B,C)} : \{\texttt{B/G,C/V}\}$ (modulo renaming), but only one of them has been expanded. This is depicted by arrows from the two non-expanded occurrences of $\texttt{tw(B,C)} : \{\texttt{B/G,C/V}\}$ to the expanded one. More information on the efficient construction of the analysis graph can be found in [22, 11, 1].

5.2 Answer and Dependency Tables

The program analysis graph is implicitly represented in the algorithm by means of two data structures, the *answer table* ($\mathcal{AT}$) and the *dependency table* ($\mathcal{DT}$).

The answer table contains entries of the form $A : CP \rightsquigarrow AP$ which are interpreted as the answer (success) pattern for $A : CP$ is AP. For instance, there exists an entry of the form $\texttt{main(s}^3\texttt{(X),X2)} : \{\texttt{X/G,X2/V}\} \rightsquigarrow \{\texttt{X/G,X2/G}\}$ associated to the OR-node discussed above.

Dependencies indicate direct relations among OR-nodes. An OR-node $A_F : CP_F$ *depends on* another OR-node $A_T : CP_T$ iff in the body of some clause for $A_F : CP_F$ there appears the OR-node $A_T : CP_T$. The intuition is that in computing the answer for $A_F : CP_F$ we have used the answer pattern for $A_T : CP_T$. In our algorithm we store *backwards* dependencies,[3] i.e., for each OR-node $A_T : CP_T$ we keep track of the set of OR-nodes which depend on it. I. e., the keys in the dependency table are OR-nodes and the information associated to each node is the set of other nodes which depend on it, together with some additional information required to iterate when an answer is modified (updated). Each element of a *dependency set* for an atom $B : CP_2$ is of the form $\langle H : CP \Rightarrow [H_k : CP_1]\ k, i\rangle$. It should be interpreted as follows: the OR-node $H : CP$ through the literal at position k, i depends on the OR-node $B : CP_2$. Also, the remaining information $[H_k : CP_1]$ encodes the fact that the head of this clause is H_k and the substitution (in terms of all variables of clause k) just before the call to $B : CP_2$ is CP_1. Such information avoids having to reprocess atoms in the clause k to the left of position i.

EXAMPLE 5.1. *For instance, the dependency set for the abstract atom* $\texttt{formula(C,A)} : \{\texttt{A/V,C/G}\}$ *is* $\{\langle\texttt{main(s}^3\texttt{(X),X2)} : \{\texttt{X/G,X2/V}\} \Rightarrow [\,\texttt{main(s}^4\texttt{(B),A)} : \{\texttt{B/G,A/V,C/G}\}\,]\ 2,2\rangle\}$ *It indicates that the OR-node* $\texttt{formula(C,A)} : \{\texttt{A/V,C/G}\}$ *is only used in the OR-node* $\texttt{main(s}^3\texttt{(X),X2)} : \{\texttt{X/G,X2/V}\}$ *via literal 2,2 (see Example 1.1). Thus, if the answer pattern for* $\texttt{formula(C,A)} : \{\texttt{A/V,C/G}\}$ *is ever updated, then we must reprocess the OR-node* $\texttt{main(s}^3\texttt{(X),X2)} : \{\texttt{X/G,X2/V}\}$ *from position 2,2.*

5.3 The Algorithm

Algorithm 2 presents our proposed algorithm. Procedure ABS_INT_WITH_SPEC_DEFS initializes the four tables used by the algorithm and calls PROCESS_CALL_PATTERN for each abstract atom in the initial set. PROCESS_CALL_PATTERN applies, first of all (L7), the *Widen_Call* function to $A : CP$ taking into account the set of entries already in $\mathcal{AT}$. This returns a substitution CP_1 s.t. $CP \sqsubseteq CP_1$. The most precise *Widen_Call* function possible is the identity function, but it can only be used with abstract domains with a finite number of abstract values. This is the case with *sharing–freeness* and thus we will use the identity function in our example. If the call pattern $A : CP_1$ has not been processed before, it places (L9) $\bot$ as initial answer in $\mathcal{AT}$ for $A : CP$ and sets to empty (L10) the set of OR-nodes in the graph which depend on $A : CP_1$. It then computes (L11) a specialized definition for $A : CP_1$. We do not show in Algorithm 2 the definition of SPECIALIZED_DEFINITION, since it is identical to that in Algorithm 1. In the graph, we show within an oval box the calls to SPECIALIZED_DEFINITION which appear during the execution of the running example (see the details in Sect. 4). The heads of the clauses in the specialized definition are linked to the box with a dotted arc. Then (L13-15) calls to PROCESS_CLAUSE are launched for the clauses in the specialized definition w.r.t. which $A : CP_1$ is to be analyzed. Only after this, the *Parent* OR-node is added (L16-17) to the dependency set for $A : CP_1$.

Algorithm 2 Abstract Interpretation with Specialized Definitions

```
 1: procedure ABS_INT_WITH_SPEC_DEFS($P, \{A_1 : CP_1, \ldots, A_n : CP_n\}$)
 2:     Create_Table($\mathcal{AT}$); Create_Table($\mathcal{DT}$)
 3:     Create_Table($\mathcal{GT}$); Create_Table($\mathcal{ST}$)
 4:     for $j = 1..n$ do
 5:         PROCESS_CALL_PATTERN($A_j : CP_j, \langle A_j : CP_j \Rightarrow [A_j : CP_j], j, entry\rangle$)
 6: function PROCESS_CALL_PATTERN($A : CP, Parent$)
 7:     $CP_1 \leftarrow Widen\_Call(\mathcal{AT}, A : CP)$
 8:     if not IsIn($\mathcal{AT}, A : CP_1$) then
 9:         Insert($\mathcal{AT}, A : CP_1, \bot$)
10:         Insert($\mathcal{DT}, A : CP_1, \emptyset$)
11:         $(A', A'_1) \leftarrow$ SPECIALIZED_DEFINITION($P, A : CP_1$)
12:         $A'' \leftarrow$ ren($A, \{A'/A'_1\}$)
13:         for all renamed apart clause $C_k = H_k \leftarrow B_k \in P$
                s.t. $H_k$ unifies with $A''$ do
14:             $CP_k \leftarrow$ Atranslate($A'' : CP_1, C_k$)
15:             PROCESS_CLAUSE($A : CP_1 \Rightarrow [H_k : CP_k]\ B_k, k, 1$)
16:     $Deps \leftarrow$ Look_up($\mathcal{DT}, A : CP_1$) $\bigcup\{Parent\}$
17:     Insert($\mathcal{DT}, A : CP_1, Deps$)
18:     return Look_up($\mathcal{AT}, A : CP_1$)
19: procedure PROCESS_CLAUSE($H : CP \Rightarrow [H_k : CP_1]\ B, k, i$)
20:     if $CP_1 \neq \bot$ then
21:         if $B = [L|R]$ then
22:             $CP_2 \leftarrow$ Arestrict($CP_1, L$)
23:             $AP_0 \leftarrow$ PROCESS_CALL_PATTERN($L : CP_2,$
                    $\langle H : CP \Rightarrow [H_k : CP_1], k, i\rangle$)
24:             $CP_3 \leftarrow$ Aconj($CP_1$, Aextend($AP_0, CP_1$))
25:             PROCESS_CLAUSE($H : CP \Rightarrow [H_k : CP_3]R, k, i+1$)
26:         else
27:             $AP_1 \leftarrow$ Atranslate($H_k : CP_3, H \leftarrow true$)
28:             $AP_2 \leftarrow$ Look_up($\mathcal{AT}, H : CP$)
29:             $AP_3 \leftarrow$ Alub($AP_1, AP_2$)
30:             if $AP_2 \neq AP_3$ then
31:                 Insert($\mathcal{AT}, H : CP, AP_3$)
32:                 $Deps \leftarrow$ Look_up($\mathcal{DT}, H : CP$)
33:                 PROCESS_UPDATE($Deps$)
34: procedure PROCESS_UPDATE($Updates$)
35:     if $Updates = \{A_1, \ldots, A_n\}$ with $n \geq 0$ then
36:         $A_1 = \langle H : CP \Rightarrow [H_k : CP_1], k, i\rangle$
37:         if $i \neq entry$ then
38:             $B \leftarrow$ get_body($P, k, i$)
39:             REMOVE_PREVIOUS_DEPS($H : CP \Rightarrow [H_k : CP_1]$
                    $B, k, i$)
40:             PROCESS_CLAUSE($H : CP \Rightarrow [H_k : CP_1]\ B, k, i$)
41:             PROCESS_UPDATE($Updates - \{A_1\}$)
```

The function PROCESS_CLAUSE performs the success prop-

[3] In the implementation, for efficiency, both forward and backward dependencies are stored. We do not include them in the algorithm for simplicity of the presentation.

agation and constitutes the core of the analysis. First, the current answer (AP_0) for the call to the literal at position k, i of the form $B : CP_2$ is (L24) conjoined (Aconj), after being extended (Aextend) to all variables in the clause, with the description CP_1 from the program point immediately before B in order to obtain the description CP_3 for the program point after B. If B is not the last literal, CP_3 is taken as the (improved) calling pattern to process the next literal in the clause in the recursive call (L25). This corresponds to left-to-right success propagation and is marked in Fig. 3 with a dashed horizontal arrow. If we are actually processing the last literal, CP_3 is (L27) adapted (Atranslate) to the initial call pattern $H : CP$ which started PROCESS_CLAUSE, obtaining AP_1. This value is (L29) disjoined (Alub) with the current answer, AP_2, for $H : CP$ as given by Look_up. If the answer changes, then its dependencies, which are readily available in $\mathcal{DT}$, need to be recomputed (L33) using PROCESS_UPDATE. This procedure restarts the processing of all body postfixes which depend on the calling pattern for which the answer has been updated by launching new calls to PROCESS_CLAUSE. There is no need of recomputing answers in our example. The procedure REMOVE_PREVIOUS_DEPS eliminates (L39) entries in $\mathcal{DT}$ for the clause postfix which is about to be re-computed. We do not present its definition here due to lack of space. Note that the new calls to PROCESS_CLAUSE may in turn launch calls to PROCESS_UPDATE. On termination of the algorithm a global fixpoint is guaranteed to have been reached. Note that our algorithm also stores in the dependency sets calls from the initial entry points (marked with the value *entry* in L5). These do not need to be reprocessed (L37) but are useful for determining the specialized version to use for the initial queries after code generation.

The CiaoPP analysis and specialization system implements abstract interpretation with specialized definitions as introduced in Algorithm 2. For our running example, the system is able to obtain the specialized code and the accurate analysis results of Example 1.1. Due to space limitations, we have not traced all the steps performed during the execution of the algorithm, though the analysis graph in Fig. 3 shows the clauses obtained by SPECIALIZED_DEFINITION and the call/success patterns inferred by the analysis of such clauses.

5.4 Termination

If we compose a terminating analysis strategy (abstract domain plus widening operator) with a terminating PD strategy (local control plus global control), then Algorithm 2 also terminates for such strategies. Intuitively, if we have a terminating *AUnfold* rule and the abstract domain is ascending chain finite, non-termination can only occur if the set of call patterns handled by the algorithm is infinite. Since the *Widen_Call* function guarantees that a given concrete atom A can only be analyzed w.r.t. a finite number of abstract substitutions CP, non-termination can only occur if the set of atoms has an infinite number of elements with different concrete parts. If the *AGeneralize* function guarantees that an infinite number of different concrete atoms cannot occur, then termination is guaranteed.

6. INTERPRETING THE RESULTS OF THE ALGORITHM

We first discuss whether we can interpret the results of Algorithm 2 in terms of analysis. We use $\theta|_{\{X_1,\ldots,X_n\}}$ to denote the projection of substitution θ onto the set of variables $\{X_1, \ldots, X_n\}$. We denote by $success(A : CP, P)$ the set of computed answers for initial queries described by the abstract atom $A : CP$ in a program P, i.e., $success(A : CP, P) = \{\theta'' \mid \exists \theta \in \gamma(CP) \wedge \exists \theta'\}$ s.t. θ' is a computed answer for $A\theta$ and $\theta'' = \theta\theta'|_{vars(A)}\}$.

THEOREM 6.1 (CORRECTNESS OF SUCCESS). *Let P be a program and let $S = \{A_1 : CP_1, \ldots, A_n : CP_n\}$ be a set of abstract atoms. For all $A_i : CP_i \in S$, after termination of* ABS_INT_WITH_SPEC_DEFS(P, S)*, there exists $(A_i : CP_i' \rightsquigarrow AP_i) \in \mathcal{AT}$ s.t. $CP_i \sqsubseteq CP_i' \wedge success(A_i : CP_i, P) \subseteq \gamma(AP_i)$.*

Intuitively, correctness holds since Algorithm 2 computes an abstract and–or graph and, thus, we inherit a generic correctness result for success substitutions. However, now we analyze the call patterns in S w.r.t. specialized definitions rather than their original definition in P. Since the transformation rules in Definitions 3.1 and 3.3 are semantics preserving, then analysis of each specialized definition is guaranteed to produce a safe approximation of its success set, which is also a safe approximation of the success of the original definition.

6.1 The Framework as a Specializer

Before presenting the algorithm for code generation, we introduce some notation. We denote by $spec\_defs(P, \mathcal{ST})$ the subset of clauses in P which correspond to specialized definitions, as stored in $\mathcal{ST}$. It is defined as $spec\_defs(P, \mathcal{ST}) = \{(H \leftarrow B) \in P \mid \exists (\_ : \_ \rightsquigarrow A') \in \mathcal{ST}$ s.t. H unifies with $A'\}$.

Each non-root OR-node in the analysis graph has been generated by a call of the form PROCESS_CALL_PATTERN$(B : CP_2, \langle H : CP \Rightarrow [H_k : CP_1], k, i\rangle)$, see L23 in Algorithm 2. Thus, each non-root OR-node is uniquely identified by a pair of the form $(B : CP_2, \langle H : CP \Rightarrow [\_ : \_], k, i\rangle)$. We can classify the OR-nodes in an analysis graph according to the program point they correspond to, i.e., k, i. We denote by $OR\_nodes(k, i)$ the set of OR-nodes of the form $(\_ : \_, \langle \_ : \_ \Rightarrow [\_ : \_], k, i\rangle)$.

We denote by $SD((B : CP_2, Id), \mathcal{DT}, \mathcal{GT})$ the abstract atom $B' : CP_2'$ which has been used for generating the specialized definition w.r.t. which the atom $(B : CP_2, Id)$ has been analyzed, and it is defined as:

$SD((B : CP_2, \langle H : CP \Rightarrow [\_ : \_], k, i\rangle), \mathcal{DT}, \mathcal{GT}) = B' : CP_2'$ s.t. $\exists (B : CP_1 \rightsquigarrow Deps) \in \mathcal{DT}$ s.t. $(H : CP \Rightarrow [\_ : \_], k, i) \in Deps \wedge \exists (B : CP_1 \rightsquigarrow B' : CP_2') \in \mathcal{GT}$.

Algorithm 3 presents the procedure for code generation. Since the specialized definitions generated already have different predicate names, the heads of the new clauses do not need to be renamed. Function RENAME_BODY simply traverses the body of the clauses in the specialized definitions and replaces atoms for predicates in the original program with atoms for predicates in the specialized definitions. Deciding which predicate to use is done by function RENAME_ATOM. Note that since (optionally) constants are filtered out by function new_filter, this renaming can remove constants from the original program.

We now present two *AGeneralize* functions which can be used in Alg. 2 when using it as a specializer. In both of them, the decision on whether to lose information in a call

Algorithm 3 Code Generation

```
function $\textsc{Codegen}(P, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
    return $\{(H_k \leftarrow B'_k) \mid \exists(H_k \leftarrow B_k) \in \mathit{spec\_defs}(P, \mathcal{ST}) \wedge$
        $B'_k = \textsc{rename\_body}(B_k, k, 1, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
function $\textsc{rename\_body}(B, k, i, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
    if $B = (L, R)$ then
        $L' \leftarrow \textsc{rename\_atom}(L, k, i, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
        $R' \leftarrow \textsc{rename\_body}(R, k, i+1, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
        $B' \leftarrow (L', R')$
    else
        $B' \leftarrow \textsc{rename\_atom}(B, k, i, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
    return $B'$
function $\textsc{rename\_atom}(L, k, i, \mathcal{DT}, \mathcal{GT}, \mathcal{ST})$
    $L' : CP' \leftarrow SD((L : \_, \langle \_ : \_ \Rightarrow [\_ : \_], k, i \rangle), \mathcal{DT}, \mathcal{GT})$
    return $\mathsf{Look\_up}(\mathcal{ST}, L' : CP')$
```

$\mathit{AGeneralize}(\mathcal{ST}, A : CP)$ is based on the concrete part of the atom, A. This allows easily defining $\mathit{AGeneralize}$ operators in terms of existing *Generalize* operators. Let *Generalize* be a (concrete) generalization function. Then we define $\mathit{AGeneralize}_\alpha(\mathcal{ST}, A : CP) = (A', CP')$ where $A' = \mathit{Generalize}(\mathcal{ST}, A)$ and $CP' = \mathsf{Atranslate}(A : CP, A' \leftarrow \mathit{true})$. Function $\mathit{AGeneralize}_\alpha$ only assigns the same specialized definition for different abstract atoms when we know that after adapting the analysis info of both $A_1 : CP_1$ and $A_2 : CP_2$ to the new atom A' the same entry substitution CP' will be obtained in either case. Similarly, we define $\mathit{AGeneralize}_\gamma(\mathcal{ST}, A : CP) = (A', CP')$ where $A' = \mathit{Generalize}(\mathcal{ST}, A)$ and $CP' = \top$. The function $\mathit{AGeneralize}_\gamma$ assigns generalizations taking into account the concrete part of the abstract atom only, which is the same for all OR-nodes which correspond to a literal k, i. These functions are in fact two extremes. In $\mathit{AGeneralize}_\alpha$ we try to keep as much abstract information as possible, whereas in $\mathit{AGeneralize}_\gamma$ we lose all abstract information. The latter is useful when we do not have an unfolding system which can exploit abstract information or when we do not want the specialized program to have different implemented specialized definitions for atoms with the same concrete part (modulo renaming) but different abstract substitution.

7. DISCUSSION AND RELATED WORK

The versatility of our framework (and of our implementation) can be seen by recasting well-known specialization and analysis frameworks as instances in which the different parameters: unfolding rule, widen call rule, abstraction operator, and analysis domain, take the following values.

Polyvariant Abstract Interpretation. Our algorithm can behave as the analysis algorithm described in [11, 22] for polyvariant static analysis by defining an *AGeneralize* operator which returns the base form of an expression (i.e., it loses all constants) and an *AUnfold* operator which performs a single derivation step (i.e., it returns the original definition). Thus, the resulting framework would always produce a residual program which coincides with the original one and can be analyzed with any abstract domain of interest.

Multivariant Abstract Specialization. The specialization power of the framework described in [24, 23] can be obtained by using the same *AGeneralize* described in the above point plus an *AUnfold* operator which always performs a derive step followed by zero or more abstract execution steps. It is interesting to note that in the original framework, abstract executability is performed as an offline optimization phase, i.e., after analysis, while it is performed online, i.e., during analysis, in our framework.

Classical Partial Deduction. Our method can be used to perform classical PD in the style of [21, 10] by using an abstract domain with the single abstract value $\top$ and the identity function as *Widen_Call* rule. This corresponds to the $\mathcal{PD}$ domain of [15] in which an atom with variables represents all its instances. Let us note that, in spite of the fact that the algorithm follows a left-to-right computation flow at the global control level, the computation order is irrelevant since the $\mathcal{PD}$ domain conveys no information on variables. However, the process of generating specialized definitions (as discussed in Section 3) can perform *non-leftmost* unfolding steps at the local control level and achieve the same optimizations as in PD.

Abstract Partial Deduction. Several approaches have been proposed which extend PD with SLDNF-trees by using abstract substitutions [14, 8, 18, 15]. In essence, such approaches are very similar to the abstract partial deduction with call propagation shown in Algorithm 1. Though all those proposals identify the need of propagating success substitutions, they either fail to do so or propose means for propagating success information which are not fully integrated within the APD algorithm and, in our opinion, do not fit in as nicely as the use of and–or trees. Also, these proposals are either strongly coupled to a particular (downward closed) abstract domain, i.e., regular types, as in [8, 18] or do not provide the exact description of operations on the abstract domain which are needed by the framework, other than general correctness criteria [14, 15]. However, the latter allow Conjunctive PD [6], which is not available in our framework yet. It remains as future work to investigate the extension of our framework in order to analyze conjunctions of atoms and in order to achieve optimizations like tupling and deforestation.

Finally, the work in [25] identifies the need for including unfolding in abstract interpretation frameworks in order to increase their power. Then, four different alternatives for doing so (Section 5.3) are discussed. Note that the framework we propose in this work does not correspond to any of those alternatives and is in fact more powerful than any of them. Some of the main differences between our approach and that in [25] are: (1) [25] proposes performing (individual) unfolding steps directly on the and–or graph computed by analysis, whereas here we propose to compute specialized definitions by a separate component. This has both theoretical and practical advantages. It allows separation of concerns, which results in a clearer specification. In addition, it allows directly reusing the important body of work in control of PD. (2) The work in [25] cannot handle abstract domains which are infinite, such as regular types, since there is no notion of widening on calls. (3) There is no separation between global and local control. This separation is essential in order to guarantee termination of the specialization process. In particular, global control allows reusing already specialized OR–nodes. It is unclear how one would reuse an OR–node in an analysis graph for another call not exactly identical. Finally, (4) in contrast to [25] we provide a precise algorithm which implements the framework.

8. CONCLUSIONS

We have proposed a novel scheme for a seamless integration of the techniques of abstract interpretation and partial deduction. Our scheme is parametric w.r.t. the abstract domain and the control issues which guide the partial deduction process. Existing proposals for the integration use abstract interpretation as a *means* for improving partial evaluation rather than as a *goal*, at the same level as producing a specialized program. This implies that, as a result, their objective is to yield a set of atoms which determines a partial evaluation rather than to compute a safe approximation of its success. Unlike them, a main objective of our work is to improve success information by analyzing the specialized code, rather than the original one. We achieve this objective by smoothly *interleaving* both techniques which, on one hand, improves success information—even for abstract domains which are not related directly to partial evaluation. Furthermore, with more accurate success information, we can improve further the quality of partial evaluation. The overall method thus yields not only a specialized program but also a safe approximation of its behaviour.

Acknowledgments

The authors would like to thank John Gallagher and Michael Leuschel for useful discussions on the integration of abstract interpretation and partial deduction.

9. REFERENCES

[1] M. Bruynooghe. A Practical Framework for the Abstract Interpretation of Logic Programs. *Journal of Logic Programming*, 10:91–124, 1991.

[2] F. Bueno, D. Cabeza, M. Carro, M. Hermenegildo, P. López-García, and G. Puebla (Eds.). The Ciao System. Reference Manual (v1.10). Technical report, School of Computer Science (UPM), 2004. Available at `http://clip.dia.fi.upm.es/Software/Ciao/`.

[3] C. Consel and S.C. Koo. Parameterized partial deduction. *ACM Transactions on Programming Languages and Systems*, 15(3):463–493, July 1993.

[4] P. Cousot and R. Cousot. Abstract Interpretation: a Unified Lattice Model for Static Analysis of Programs by Construction or Approximation of Fixpoints. In *Proc. of POPL'77*, pages 238–252, 1977.

[5] P. Cousot and R. Cousot. Systematic Design of Program Transformation Frameworks by Abstract Interpretation. In *Proc. of POPL'02*, pages 178–190. ACM, 2002.

[6] D. De Schreye, R. Glück, J. Jørgensen, M. Leuschel, B. Martens, and M.H. Sørensen. Conjunctive Partial Deduction: Foundations, Control, Algorihtms, and Experiments. *Journal of Logic Programming*, 41(2&3):231–277, 1999.

[7] J. Gallagher, M. Codish, and E. Shapiro. Specialisation of Prolog and FCP Programs Using Abstract Interpretation. *New Generation Computing*, 6(2–3):159–186, 1988.

[8] J. P. Gallagher and J. C. Peralta. Regular tree languages as an abstract domain in program specialisation. *Higher Order and Symbolic Computation*, 14(2,3):143–172, 2001.

[9] J.P. Gallagher. Static Analysis for Logic Program Specialization. In *Workshop on Static Analysis WSA'92*, pages 285–294, 1992.

[10] J.P. Gallagher. Tutorial on specialisation of logic programs. In *Proc. of PEPM'93*, pages 88–98. ACM Press, 1993.

[11] M. Hermenegildo, G. Puebla, K. Marriott, and P. Stuckey. Incremental Analysis of Constraint Logic Programs. *ACM TOPLAS*, 22(2):187–223, March 2000.

[12] N. D. Jones. Combining Abstract Interpretation and Partial Evaluation. In *Static Analysis Symposium*, number 1140 in LNCS, pages 396–405. Springer-Verlag, 1997.

[13] N.D. Jones, C.K. Gomard, and P. Sestoft. *Partial Evaluation and Automatic Program Generation.* Prentice Hall, New York, 1993.

[14] M. Leuschel. Program Specialisation and Abstract Interpretation Reconciled. In *Joint International Conference and Symposium on Logic Programming*, June 1998.

[15] M. Leuschel. A framework for the integration of partial evaluation and abstract interpretation of logic programs. *ACM Transactions on Programming Languages and Systems*, 26(3):413 – 463, May 2004.

[16] M. Leuschel and M. Bruynooghe. Logic program specialisation through partial deduction: Control issues. *Theory and Practice of Logic Programming*, 2(4 & 5):461–515, July & September 2002.

[17] M. Leuschel and D. De Schreye. Logic program specialisation: How to be more specific. In *Proc. of PLILP'96*, LNCS 1140, pages 137–151, 1996.

[18] M. Leuschel and S. Gruner. Abstract conjunctive partial deduction using regular types and its application to model checking. In *Proc. of LOPSTR*, number 2372 in LNCS. Springer, 2001.

[19] M. Leuschel, J. Jørgensen, W. Vanhoof, and M. Bruynooghe. Offline specialisation in Prolog using a hand-written compiler generator. *Theory and Practice of Logic Programming*, 4(1):139–191, 2004.

[20] J.W. Lloyd. *Foundations of Logic Programming.* Springer, second, extended edition, 1987.

[21] J.W. Lloyd and J.C. Shepherdson. Partial Evaluation in Logic Programming. *Journal of Logic Programming*, 11(3–4):217–242, 1991.

[22] G. Puebla and M. Hermenegildo. Optimized Algorithms for the Incremental Analysis of Logic Programs. In *Proc. of SAS'96*, pages 270–284. Springer LNCS 1145, 1996.

[23] G. Puebla and M. Hermenegildo. Abstract Multiple Specialization and its Application to Program Parallelization. *J. of Logic Programming.*, 41(2&3):279–316, November 1999.

[24] G. Puebla and M. Hermenegildo. Abstract Specialization and its Applications. In *Proc. of PEPM'03*, pages 29–43. ACM Press, 2003. Invited talk.

[25] G. Puebla, M. Hermenegildo, and J. Gallagher. An Integration of Partial Evaluation in a Generic Abstract Interpretation Framework. In *Proc. of PEPM'99*, number NS-99-1 in BRISC Series, pages 75–85. University of Aarhus, Denmark, 1999.