Incremental Analysis of Constraint Logic Programs

MANUEL HERMENEGILDO and GERMAN PUEBLA
Universidad Politécnica de Madrid
KIM MARRIOTT
Monash University
and
PETER J. STUCKEY
University of Melbourne

Global analyzers traditionally read and analyze the entire program at once, in a nonincremental way. However, there are many situations which are not well suited to this simple model and which instead require reanalysis of certain parts of a program which has already been analyzed. In these cases, it appears inefficient to perform the analysis of the program again from scratch, as needs to be done with current systems. We describe how the fixed-point algorithms used in current generic analysis engines for (constraint) logic programming languages can be extended to support incremental analysis. The possible changes to a program are classified into three types: addition, deletion, and arbitrary change. For each one of these, we provide one or more algorithms for identifying the parts of the analysis that must be recomputed and for performing the actual recomputation. The potential benefits and drawbacks of these algorithms are discussed. Finally, we present some experimental results obtained with an implementation of the algorithms in the PLAI generic abstract interpretation framework. The results show significant benefits when using the proposed incremental analysis algorithms.

Categories and Subject Descriptors: D.1.2 [Programming Techniques]: Automatic Programming—Automatic analysis of algorithms, Program transformation; D.1.6 [Programming Techniques]: Logic programming; D.3.4 [Programming Languages]: Compilers; F.3.1 [Logics and Meanings of Programs]: Specifying and Verifying and Reasoning about programs—Logics of programs

General Terms: Languages
Additional Key Words and Phrases: Abstract interpretation, constraint logic programming, incremental computation, static analysis

Authors’ addresses: M. Hermenegildo and G. Puebla, Facultad de Informática, Universidad Politécnica de Madrid, 28660-Boadilla del Monte, Madrid, Spain; email: {herme;german}@fi.upm.es; K. Marriott, School of Computer Science and Software Engineering, Monash University Clayton 3168, Australia; email: marriott@csse.monash.edu.au; P.J. Stuckey, Dept. of Computer Science and Software Engineering, The University of Melbourne, Parkville 3052, Australia; email: pjs@cs.mu.oz.au.
1. INTRODUCTION

Global program analysis is becoming a practical tool in constraint logic program compilation in which information about calls, answers, and the effect of the constraint store on variables at different program points is computed statically [Hermenegildo et al. 1992; Van Roy and Despain 1992; Muthukumar and Hermenegildo 1992; Santos-Costa et al. 1991; Bueno et al. 1994]. The underlying theory, formalized in terms of abstract interpretation [Cousot and Cousot 1977], and the related implementation techniques are well understood for several general types of analysis and, in particular, for top-down analysis of Prolog [Debray 1989; 1992; Bruynooghe 1991; Muthukumar and Hermenegildo 1992; Marriott et al. 1994; Charlier and Van Hentenryck 1994]. Several generic analysis engines, such as PLAI [Muthukumar and Hermenegildo 1992; 1990], GAIA [Charlier and Van Hentenryck 1994], and the CLP(R) analyzer [Kelly et al. 1998b], facilitate construction of such top-down analyzers. These generic engines have the description domain and functions on this domain as parameters. Different domains give analyzers which provide different types of information and degrees of accuracy. The core of each generic engine is an algorithm for efficient fixed-point computation [Muthukumar and Hermenegildo 1990; 1992; Charlier et al. 1993]. Efficiency is obtained by keeping track of which parts of a program must be reexamined when a success pattern is updated. Current generic analysis engines are nonincremental—the entire program is read, analyzed, and the analysis results written out.

Despite the obvious progress made in global program analysis, most logic program and CLP compilers still perform only local analysis (although the &-Prolog [Hermenegildo and Greene 1991], Aquarius [Van Roy and Despain 1992], Andorra-I [Santos-Costa et al. 1991], and CLP(R) [Kelly et al. 1998a] systems are notable exceptions). We believe that an important contributing factor to this is the simple, nonincremental model supported by global analysis systems, which is unsatisfactory for at least three reasons:

—The first reason is that optimizations are often source-to-source transformations; optimization consists of an analyze, perform transformation, then reanalyze cycle. This is inefficient if the analysis starts from scratch each time. Such analyze-transform cycles may occur for example when program optimization and multivariant specialization are combined [Winsborough 1992; Puebla and Hermenegildo 1995; 1999]. This is used, for instance, in program parallelization, where an initial analysis is used to introduce specialized predicate definitions with run-time parallelization tests, and then these new definitions are analyzed and those tests which become redundant in the multiply specialized program removed. It is also the case in optimization of CLP(R) in which specialized predicate definitions are reordered and then reanalyzed.

—The second reason is that incremental analysis supports incremental runtime compilation during the test-debug cycle. Again, for efficiency only those parts of the program which are affected by the changes should be reanalyzed. Incremental

---

1By source-to-source transformation we include transformations on the (high-level) internal compiler representation of the program source, which for (constraint) logic program compilers tend to be very close to the source.
compilation is important in the context of logic programs as traditional environ­ments have been interpretive, allowing the rapid generation of prototypes. Incre­mental analysis is especially important when the system uses analysis information in order to perform compile-time correctness checking of the program [Puebla et al. 2000; Hermenegildo et al. 1999b].

—The third reason is to better handle the optimization of programs in which rules are asserted (added) to or retracted (removed) from the program at runtime.

Clearly, if we modify a program the existing analysis information for it may no longer be correct and/or accurate. However, analysis is often a costly task, and starting analysis again from scratch does not appear to be the best solution. In this article we describe how the fixed-point algorithm in the top-down generic analysis engines for (constraint) logic programs can be extended to support incremental analysis. Guided by the applications mentioned above, we consider algorithms for different types of incrementality. The first, and simplest, type of incrementality is when program rules are added to the original program. The second type of incrementality is rule deletion. We give several algorithms to handle deletion. These capture different trade-offs between efficiency and accuracy. The algorithms for deletion can be easily extended to handle the third and most general type of incrementality, arbitrary change, in which program rules can be deleted or modified in any way. Finally, we consider a restricted type of arbitrary change: local change in which rules are modified, but the answers to the rules are unchanged for the calling patterns they are used with. This case occurs in program optimization, as correctness of the optimization usually amounts to requiring this property. Local change means that changes to the analysis are essentially restricted to recomputing the new call patterns which these rules generate. We give a modification to the fixed-point algorithm which handles this type of incrementality. Finally we give a preliminary empirical evaluation. We argue that the experimental results show that our algorithms are practically important.

In the next section we present the formalization of a fixed-point algorithm which generalizes those used in generic analysis engines. In Section 3 we give an algorithm to handle incremental addition of rules. In Section 4 we give two algorithms to handle incremental deletion of rules. In Section 5 we modify these algorithms to handle arbitrary change of rules. We also give an algorithm to handle the special case of local change. In Section 6 we describe the implementation of the algorithms and our empirical evaluation. Section 7 discusses related work while Section 8 concludes.

2. A GENERIC ANALYSIS ALGORITHM

We start by providing a formalization of a fixed-point algorithm for analysis of (con­straint) logic programs. We assume the reader is familiar with constraint logic pro­gramming (e.g., see Marriott and Stuckey [1998]) and abstract interpretation (see Cousot and Cousot [1977]). The aim of goal-directed top-down program analysis is, for a particular description domain, to take a program and a set of initial calling patterns and to annotate the program with information about the current environment at each program point whenever that point is reached when executing calls described by the calling patterns.
2.1 Program Analysis by Abstract Interpretation

Abstract interpretation [Cousot and Cousot 1977] is a technique for static program analysis in which execution of the program is simulated on a description (or abstract) domain \((D)\) which is simpler than the actual (or concrete) domain \((C)\). Values in the description domain and sets of values in the actual domain are related via a pair of monotonic mappings \((\alpha, \gamma)\): abstraction \(\alpha : C \rightarrow D\) and concretization \(\gamma : D \rightarrow C\) which form a Galois connection. A description \(d \in D\) approximates an actual value \(c \in C\) if \(\alpha(c) \leq d\) where \(\leq\) is the partial ordering on \(D\). Correctness of abstract interpretation guarantees that the descriptions computed approximate all of the actual values which occur during execution of the program.

Different description domains may be used which capture different properties with different accuracy and cost. The description domain that we use in our examples is the definite Boolean functions [Armstrong et al. 1994], denoted \(\text{Def}\). The key idea in this description is to use implication to capture groundness dependencies. The reading of the function \(x \rightarrow y\) is “if the program variable \(x\) is (becomes) ground, so is (does) program variable \(y\).” For example, the best description of the constraint \(f(X, Y) = f(a, g(U, V))\) is \(X \wedge (Y \leftrightarrow (U \wedge V))\). Groundness information is directly useful for many program optimizations such as constraint simplification, parallelization, and simplification of built-ins. It is also indirectly useful for almost all other optimizations of (constraint) logic programs, since it can be combined with many other analysis domains to give more precise analysis information.

We now recall some standard definitions in constraint logic programming. A constraint logic program or program is a set of rules of the form \(A : L_1, \ldots, L_n\), where \(L_1, \ldots, L_n\) are literals and \(A\) is an atom said to be the head of the rule. A literal is an atom or a primitive constraint. We assume that each atom is normalized; that is to say, it is of the form \(p(x_1, \ldots, x_m)\) where \(p\) is an \(m\)-ary predicate symbol and \(x_1, \ldots, x_m\) are distinct variables. A primitive constraint is defined by the underlying constraint domain and is of the form \(c(e_1, \ldots, e_m)\) where \(c\) is an \(m\)-ary predicate symbol and the \(e_1, \ldots, e_m\) are expressions. For simplicity, in the examples we shall restrict ourselves to the Herbrand domain (Prolog) where primitive constraints are of the form \(e_1 = e_2\) where \(e_1\) and \(e_2\) are terms.

As an example of goal-directed top-down program analysis, consider the following program for appending lists:

\[
\text{app}(X,Y,Z) :- X=[] , Y=Z.
\]
\[
\]

Assume that we are interested in analyzing the program for the call \(\text{app}(X,Y,Z)\) with initial description \(Y\) indicating that we wish to analyze it for any call to \(\text{app}\) with the second argument definitely ground. We will denote this as the \textit{calling pattern} \(\text{app}(X,Y,Z) : Y\). In essence the analyzer must produce the \textit{program analysis graph} given in Figure 1, which can be viewed as a finite representation of the (possibly infinite) set of (possibly infinite) AND-OR trees explored by the concrete execution [Bruynooghe 1991]. Finiteness of the program analysis graph (and thus termination of analysis) is achieved by considering description domains with certain characteristics (such as being finite, or of finite height, or without infinite ascending chains) or by the use of a \textit{widening} operator [Cousot and Cousot 1977]. The graph has two sorts of nodes: those belonging to rules (also called “AND-nodes”) and
those belonging to atoms (also called “OR-nodes”). For example, the atom node
\( \text{app}(X, Y, Z) : Y \mapsto Y \land (X \leftrightarrow Z) \) indicates that when the atom \( \text{app}(X, Y, Z) \) is
called with description \( Y \) the resulting description is \( Y \land (X \leftrightarrow Z) \). This answer
description depends on the two rules defining \( \text{app} \) which are attached by arcs to
the node. These rules are annotated by descriptions at each program point of the
constraint store when the rule is executed from the calling pattern of the node
connected to the rules. The program points are at the entry to the rule, the point
between each two literals, and at the return from the call. Atoms in the rule body
have arcs to OR-nodes with the corresponding calling pattern. If such a node is
already in the tree it becomes a recursive call. Thus, the analysis graph in Figure 1
has a recursive call to the calling pattern \( \text{app}(X, Y, Z) : Y \). How this program
analysis graph is constructed is detailed in Example 1.

It is implicit in this approach that the description at each program point rep­
resents sets of constraints rather than sequences of constraints. Although it is
possible to base an analysis on sequences rather than sets, (e.g., see Charlier et al.
[1994]) almost all generic (constraint) logic program analysis engines are set-based
rather than sequence-based, so we shall focus on these.

As we have seen, a program analysis graph is constructed from an initial set of
calling patterns and a program. It is defined in terms of five abstract operations
on the description domain. As is standard these are required to be monotonic
and to approximate the corresponding concrete operations; for more details see for
example García de la Banda et al. [1998]. The abstract operations are

- \( A\text{restrict}(CP, V) \) which performs the abstract restriction of a description \( CP \) to
  the variables in the set \( V \);
- \( A\text{extend}(CP, V) \) which extends the description \( CP \) to the variables in the set \( V \);
- \( A\text{add}(C, CP) \) which performs the abstract operation of conjoining the actual con­
  straint \( C \) with the description \( CP \);
- \( A\text{conj}(CP_1, CP_2) \) which performs the abstract conjunction of two descriptions;
- \( A\text{alub}(CP_1, CP_2) \) which performs the abstract disjunction of two descriptions.
As an example, the abstract operations for the description domain $Def$ are defined as follows. We start by defining the abstraction operation $\alpha_{Def}$ which gives the best description of a constraint. It will be used in the $A\text{add}$ operation and is defined as

$$\alpha_{Def}(x = t) = (x \leftrightarrow \bigwedge \{y \in vars(t)\})$$

where $x$ is a variable, $t$ is a term, and the function $vars$ returns the set of variables appearing in some object. For instance, $\alpha_{Def}(X = [U\vee V])$ is $X \leftrightarrow (U \land V)$. We note that term constraints can always be simplified to conjunctions of this form. Extending to conjunctions, we have

$$\alpha_{Def}(e_1 \land \cdots \land e_k) = \alpha_{Def}(e_1) \land \cdots \land \alpha_{Def}(e_k)$$

where $e_1, \ldots, e_k$ are term equations.

The abstract operations for the description domain $Def$ are defined as follows:

- $A\text{restrict}(CP, V) = \exists_{-V} CP$
- $A\text{extend}(CP, V) = CP$
- $A\text{add}(C, CP) = \alpha_{Def}(C) \land CP$
- $A\text{conj}(CP_1, CP_2) = CP_1 \land CP_2$
- $A\text{ub}(CP_1, CP_2) = CP_1 \cup CP_2$

where $\exists_{-V} F$ represents $\exists v_1 \cdots \exists v_k F$ where $\{v_1, \ldots, v_k\} = vars(F) - V$, and $\cup$ is the least upper bound (lub) operation over the $Def$ lattice (e.g., Armstrong et al. [1994]). The top ($T$) of the $Def$ lattice is the formula $true$ while the bottom ($\bot$) is the formula $false$.

For a given program and calling pattern there may be many different analysis graphs. However, for a given set of initial calling patterns, a program and abstract operations on the descriptions, there is a unique least analysis graph which gives the most precise information possible.

For the reader with a formal bent, an alternative way of understanding the analysis graph is in terms of the recursive equations for the general goal-dependent semantics given in García de la Banda et al. [1998]. The least analysis graph corresponds to their least fixed point.

### 2.2 The Generic Algorithm

We will now describe our generic top-down analysis algorithm which computes the least analysis graph. This algorithm captures the essence of the particular analysis algorithms used in systems such as PLAI [Muthukumar and Hermenegildo 1990; 1992], GAIA [Charlier and Van Hentenryck 1994], and the CLP($R$) analyzer [Kelly et al. 1998b]. It will form the basis for our algorithms for incremental analysis. However there are several minor differences between the generic algorithm we present and these systems:

- First, the order in which rules for the same predicate are processed to compute the graph is not fixed, since the algorithm is parametric in the analysis strategy used to determine this order. The reasons for this are two-fold: the first reason is generality. The second reason is that the analysis strategy used for static
analysis is not necessarily good for incremental analysis, and so we need to be able to explicitly refer to and reason about different strategies.

—Second, the algorithm keeps detailed information about dependencies for each literal in the graph. This is finer grained dependency information than that usually maintained in top-down analysis algorithms. We require this extra precision for efficiency in most of the incremental analysis algorithms.

—Third, the algorithm is deliberately simplified. It does not include many minor optimizations, so as not to obscure the core behavior of the algorithm. Also, it is only defined for pure CLP programs. However, standard analysis techniques for handling constructs such as cuts, not, and \textbf{->} and other built-ins can be added without difficulty [Bueno et al. 1996]; indeed the implementation actually handles (almost) full ISO-Prolog.

We first introduce some notation. \textit{CP}, possibly subscripted, stands for a description (in the abstract domain). \textit{AP}, possibly subscripted, stands for a description occurring as an answer description. Each literal in the program is subscripted with an identifier or pair of identifiers. The expression \textit{A : CP} denotes a \textit{calling pattern}. This consists of an atom (unsubscripted or subscripted) together with a calling description for that atom.

As indicated earlier, rules are assumed to be normalized: only distinct variables are allowed to occur as arguments to atoms. Furthermore, we require that each rule defining a predicate \textit{p} has identical sequence of variables \(x_{p_1}, \ldots, x_{p_n}\) in the head atom, i.e., \(p(x_{p_1}, \ldots, x_{p_n})\). We call this the \textit{base form} of \textit{p}. Rules in the program are written with a unique subscript attached to the head atom (the rule number), and dual subscript (rule number, body position) attached to each body literal, e.g.,

\[
H_k : B_{k,1}, \ldots, B_{k,n_k},
\]

where \(B_{k,i}\) is a subscripted atom or constraint. The rule may also be referred to as rule \(k\), the subscript of the head atom. For example, the append program of Section 2.1 is written

\[
\begin{align*}
\text{app}_1(X,Y,Z) & : - X=[\_1,1], Y=Z_{1,2}. \\
\text{app}_2(X,Y,Z) & : - X=[U|V]_{2,1}, Z=[U|W]_{2,2}, \text{app}_{2,3}(V,Y,W).
\end{align*}
\]

The base form of \textit{app} is \textit{app}(\(X, Y, Z\)), and each \textit{app} atom only involves distinct variables as arguments.

The program analysis graph is implicitly represented in the algorithm by means of two data structures, the \textit{answer table} and the \textit{dependency arc table}. Given the information in these it is straightforward to construct the graph and the associated program point annotations. The answer table contains entries of the form \textit{A : CP} \(\mapsto\) \textit{AP}. \textit{A} is always a base form. This corresponds to an OR-node in the analysis graph of the form \(\langle A : CP \mapsto AP \rangle\). It is interpreted as the answer pattern for calls of the form \textit{CP} to \textit{A} is \textit{AP}. A dependency arc is of the form \(H_k : CP_0 \mapsto [CP_1] B_{k,i} : CP_2\). This is interpreted as follows: if the rule with \(H_k\) as head is called with description \textit{CP}_0 then this causes literal \(B_{k,i}\) to be called with description \textit{CP}_2.

\[\text{In fact, as we shall see, the overhead of keeping more detailed information is compensated for by avoiding redundant recomputation when an answer pattern is changed.}\]
The remaining part $CP_1$ is the program annotation just before $B_{k,i}$ is reached and contains information about all variables in rule $k$. $CP_1$ is not really necessary, but is included for efficiency. Dependency arcs represent the arcs in the program analysis graph from atoms in a rule body to an atom node. For example, the program analysis graph in Figure 1 is represented by

**answer table:**

\[
app(X, Y, Z) : Y \rightarrow Y \land (X \leftarrow Z)
\]

**dependency arc table:**

\[
app_2(X, Y, Z) : Y \Rightarrow [Y \land (X \leftarrow (U \land V)) \land (Z \leftarrow (U \land W))] \quad app_2,3(V, Y, W) : Y
\]

Intuitively, the analysis algorithm is just a graph traversal algorithm which places entries in the answer table and dependency arc table as new nodes and arcs in the program analysis graph are encountered. To capture the different graph traversal strategies used in different fixed-point algorithms, we use a priority queue. Thus, the third, and final, structure used in our algorithms is a prioritized event queue. Events are of three forms:

- **newcall($A : CP$)** which indicates that a new calling pattern for atom $A$ with description $CP$ has been encountered.
- **arc($R$)** which indicates that the rule referred to in $R$ needs to be (re)computed from the position indicated.
- **updated($A : CP$)** which indicates that the answer description to calling pattern $A$ with description $CP$ has been changed.

The generic analysis algorithm is given in Figure 2. Apart from the parametric description domain-dependent functions, the algorithm has several other undefined functions. The functions **add_event** and **next_event** respectively add an event to the priority queue and return (and delete) the event of highest priority.

When an event being added to the priority queue is already in the priority queue, a single event with the maximum of the priorities is kept in the queue. When an arc $H_k : CP \Rightarrow [CP']B_{k,i} : CP'$ is added to the dependency arc table, it replaces any other arc of the form $H_k : CP \Rightarrow [CP'B_{k,i} : \_]$ in the table and the priority queue. Similarly when an entry $H_k : CP \leftarrow AP$ is added to the answer table, it replaces any entry of the form $H_k : CP \leftarrow AP$. Note that the underscore (_) matches any description, and that there is at most one matching entry in the dependency arc table or answer table at any time.

The function **initial_guess** returns an initial guess for the answer to a new calling pattern. The default value is $\bot$ but if the calling pattern is more general than an already computed call then its current value may be returned.

The algorithm centers around the processing of events on the priority queue in **main_loop**, which repeatedly removes the highest priority event and calls the appropriate event-handling function. When all events are processed it calls **remove_useless_calls**. This procedure traverses the dependency graph given by the dependency arcs from the initial calling patterns $S$ and marks those entries in the dependency arc and answer table which are reachable. The remainder are removed.

The function **new_calling_pattern** initiates processing of the rules in the definition of atom $A$, by adding arc events for each of the first literals of these rules, and determines an initial answer for the calling pattern and places this in the table. The function **add_dependent_rules** adds arc events for each dependency arc which
analyze(S)
  foreach A : CP ∈ S
    add_event(newcall(A : CP))
main_loop()

main_loop()
  while E := next_event()
    if (E = newcall(A : CP))
      new_calling_pattern(A : CP)
    elseif (E = updated(A : CP))
      add_dependent_rules(A : CP)
    elseif (E = arc(R))
      process_arc(R)
  endwhile
  remove_useless_calls(S)
new_calling_pattern(A : CP)
  foreach rule Ak :- Bk,1,...,Bk,nk
    CP0 :=
      Aextend(CP, vars(Bk,1,...,Bk,nk))
    CP1 := Arestrict(CP0, vars(Bk,1))
    add_event(arc(Ak : CP ⇒ [CP1] Bk,i : CP2))
    AP := initial_guess(A : CP)
    if (AP ≠ ⊥)
      add_event(updated(A : CP))
    add A : CP → AP to answer table
add_dependent_rules(A : CP)
  foreach arc of the form Hk : CP0 ⇒ [CP1] Bk,i : CP2
    in graph
    where there exists renaming σ
    s.t. A : CP = (Bk,i : CP2)σ
    add_event(arc(Hk : CP0 ⇒ [CP1] Bk,i : CP2))
process_arc(Hk : CP0 ⇒ [CP1] Bk,i : CP2)
  if (Bk,i is not a constraint)
    add Hk : CP0 ⇒ [CP1] Bk,i : CP2
to dependency arc table
  W := vars(Ak := Bk,1,...,Bk,nk)
  CP3 := get_answer(Bk,i : CP2, CP1, W)
  if (CP3 / ⊥ and i ≠ nk)
    CP4 := Arestrict(CP3, vars(Bk,i+1))
    add_event(arc(Hk : CP0 ⇒ [CP1] Bk,i+1 : CP4))
  elseif (CP3 / ⊥ and i = nk)
    AP1 := Arestrict(CP3, vars(Hk))
  insert_answer_info(H : CP0 ⇒ AP1)
get_answer(L : CP2, CP1, W)
  if (L is a constraint)
    return Aadd(L, CP1)
  else
    AP0 := lookup_answer(L : CP2)
    AP1 := Aextend(AP0, W)
    return Aconj(CP1, AP1)
lookup_answer(A : CP)
  if (there exists a renaming σ s.t.
    σ(A : CP) = AP in answer table)
    return σ−1(AP)
  else
    add_event(newcall(σ(A : CP)))
    where σ is a renaming s.t.
    σ(A) is in base form
    return ⊥
insert_answer_info(H : CP ⇒ AP)
  AP0 := lookup_answer(H : CP)
  AP1 := Aconj(CP1, AP0)
  if (AP0 ≠ AP1)
    add (H : CP ⇒ AP1) to answer table
  add_event(updated(H : CP))

Fig. 2. Generic analysis algorithm.

depends on the calling pattern (A : CP) for which the answer has been updated.
The function process_arc performs the core of the analysis. It performs a single step
of the left-to-right traversal of a rule body. If the literal Bk,i is an atom, the arc
is added to the dependency arc table. The current answer for the call Bk,i : CP2
is conjoined with the description CP1 from the program point immediately before
Bk,i to obtain the description for the program point after Bk,i. This is either used
to generate a new arc event to process the next literal in the rule if Bk,i is not
the last literal; otherwise the new answer for the rule is combined with the current
answer in insert_answer_info. The function get_answer processes a literal. If it is a
constraint, it is simply abstractly added to the current description. If it is an atom,
the current answer to that atom for the current description is looked up; then this
answer is extended to the variables in the rule the literal occurs in and conjoined
with the current description. The functions lookup_answer and insert_answer_info
lookup an answer for a calling pattern in the answer table, and update the answer table entry when a new answer is found, respectively. The function lookup_answer also generates newcall events in the case that there is no entry for the calling pattern in the answer table.

2.3 Example of the Generic Algorithm

The following example briefly illustrates the operation of the generic fixed-point algorithm. It shows how the app program would be analyzed, to obtain the program analysis graph shown in Figure 1.

**Example 1.** Analysis begins from an initial set $S$ of calling patterns. In our example $S$ contains the single calling pattern $\text{app}(X,Y,Z):Y$. The first step in the algorithm is to add the initial calling patterns as new calling patterns to the priority queue. After this the priority queue contains

$$\text{newcall}(\text{app}(X,Y,Z):Y)$$

and the answer and dependency arc tables are empty. The newcall event is taken from the event queue and processed as follows. For each rule defining app, an arc is added to the priority queue which indicates the rule body must be processed from the initial literal. An entry for the new calling pattern is added to the answer table with an initial guess of false (⊥ for Def) as the answer. The data structures are now

**priority queue:**

<table>
<thead>
<tr>
<th>Arc</th>
</tr>
</thead>
<tbody>
<tr>
<td>arc( app₁(X, Y, Z) : Y ⇒ [Y] X=[_]₁,₁ : true)</td>
</tr>
<tr>
<td>arc( app₂(X, Y, Z) : Y ⇒ [Y] X=[U</td>
</tr>
</tbody>
</table>

**answer table:**

$\text{app}(X, Y, Z) : Y \mapsto \text{false}$

**dependency arc table:** no entries

An arc on the event queue is now selected for processing, say the first. The routine get_answer is called to find the answer pattern to the literal $X=[\_]$ with description true. As the literal is a constraint, the parametric routine $Aadd$ is used. It returns the answer pattern $X$. A new arc is added to the priority queue which indicates that the second literal in the rule body must be processed. The priority queue is now

$$\text{arc( app₁(X, Y, Z) : Y ⇒ [X ∧ Y] Y=Z₁,₂ : X) arc( app₂(X, Y, Z) : Y ⇒ [Y] X=[U|V]₂,₁ : true).}$$

The answer and dependency arc table remain the same.

Again, an arc on the event queue is selected for processing, say the first. As before, get_answer and Aadd are called to obtain the next annotation $X \land Y \land Z$. This time, as there are no more literals in the body, the answer table entry for $\text{app}(X,Y,Z):Y$ is updated. Alub is used to find the least upper bound of the new answer $X \land Y \land Z$ with the old answer false. This gives $X \land Y \land Z$. The entry in the answer table is updated, and an updated event is placed on the priority queue. The data structures are now

**priority queue:**

<table>
<thead>
<tr>
<th>Arc</th>
</tr>
</thead>
<tbody>
<tr>
<td>updated( app(X, Y, Z) : Y) arc( app₂(X, Y, Z) : Y ⇒ [Y] X=[U</td>
</tr>
</tbody>
</table>

**answer table:**

$\text{app}(X, Y, Z) : Y \mapsto X \land Y \land Z$

**dependency arc table:** no entries
The updated event can now be processed. As there are no entries in the dependency arc table, nothing in the current program analysis graph depends on the answer to this call, so nothing needs to be recomputed. The priority queue now contains

\[
arc( \text{app}_2(X, Y, Z) : Y \Rightarrow [Y] \ X=[U|V]_{2,1} : true) .
\]

The answer and dependency arc table remain the same.

Similarly to before we process the arc, giving rise to the new priority queue

\[
arc( \text{app}_2(X, Y, Z) : Y \Rightarrow [Y \wedge (X \leftarrow (U \wedge V))] \ Z=[U|W]_{2,2} : true) .
\]

The arc is processed to give the priority queue

\[
arc( \text{app}_2(X, Y, Z) : Y \Rightarrow [Y \wedge (X \leftarrow (U \wedge V)) \wedge (Z \leftarrow (U \wedge W))]\text{app}(V,Y,W)_{2,3} : Y) .
\]

This time, because \text{app}_{2,3}(V,Y,W) is an atom, the arc is added to the arc dependency table. The call get_answer(\text{app}(V,Y,W)_{2,3} : Y, Y \wedge (X \leftarrow (U \wedge V)) \wedge (Z \leftarrow (U \wedge W)), \{X,Y,Z,U,V,W\}) is made. The answer table is looked up to find the answer to \text{app}(V,Y,W)_{2,3} : Y and, appropriately renamed, gives \(AP_0 = V \wedge Y \wedge W\).

This description is extended to all variables (no change) and then conjoined with the second argument to give the next annotation \(Y \wedge V \wedge W \wedge (X \leftarrow U) \wedge (Z \leftarrow U)\). As this is the last literal in the body, the new answer \(Y \wedge (X \leftarrow Z)\) is obtained.

We take the least upper bound of this answer with the old answer in the table, giving \(Y \wedge (X \leftarrow Z)\). As the answer has changed, an updated event is added to the priority queue. The data structures are now

- **priority queue:** updated( \text{app}(X, Y, Z) : Y )
- **answer table:** \text{app}(X, Y, Z) : Y \leftarrow Y \wedge (X \leftarrow Z)
- **dependency arc table:** \text{app}_2(X, Y, Z) : Y \Rightarrow [Y \wedge (X \leftarrow (U \wedge V)) \wedge (Z \leftarrow (U \wedge W))] \text{app}_{2,3}(V, Y, W) : Y

The updated event is processed by looking in the dependency arc table for all arcs which have a body literal which is a variant of \text{app}(X,Y,Z) : Y and adding these arcs to the priority queue to be reprocessed. We obtain the new priority queue

\[
arc(\text{app}_2(X, Y, Z) : Y \Rightarrow [Y \wedge (X \leftarrow (U \wedge V)) \wedge (Z \leftarrow (U \wedge W))] \text{app}_{2,3}(V, Y, W) : Y).
\]

This arc is reprocessed, and gives rise to the answer \(Y \wedge (X \leftarrow Z)\). Taking the least upper bound of this with the old answer, the result is identical to the old answer, hence no updated event is added to the priority queue. As there are no events on the priority queue, the analysis terminates with the desired answer and dependency arc table.

### 2.4 Correctness

The generic algorithm provides a simple generic description of how top-down goal-directed analysis is performed. It is somewhat less abstract than the semantic equations, since we need to capture the use of dependency information during analysis. The algorithm captures the behavior of several implemented algorithms.
while at the same time is suitable for incremental analysis. Different top-down goal-directed analysis algorithms correspond to different event-processing strategies. In practice these algorithms also incorporate other optimizations. An example event-processing strategy would be to always perform newcall events first, to process nonrecursive rules before recursive rules, and to finish processing a rule before starting another. This strategy would produce an algorithm which is quite close to the one used in PLAI or GAIA (the differences between the proposed algorithm and that used in PLAI are presented in more detail in Section 6).

In essence, the algorithm defines a set of recursive equations whose least fixed point is computed using chaotic iteration [Cousot and Cousot 1977]. We note that even though the order in which events are processed is not fixed, the events themselves encode a left-to-right traversal of the rules, ensuring a unique result. For the least fixed point to be well-defined we require that the abstract operations are monotonic and that initial_guess returns a value below the least fixed point. Under these standard assumptions we have

**Theorem 1.** For a program $P$ and calling patterns $S$, the generic analysis algorithm returns an answer table and dependency arc table which represents the least program analysis graph of $P$ and $S$.

The dependency arc table does not quite capture the annotations on rules in the analysis graph, since program points before constraint literals and the last program point do not correspond to stored arcs. This information can easily be recomputed from the dependency arc table, or indeed the algorithm can be simply modified to save it as it executes.

The corollary of the above theorem is that the priority strategy does not involve correctness of the analysis. This corollary will be vital when arguing correctness of the incremental algorithms in the following sections.

**Corollary 1.** The result of the generic analysis algorithm does not depend on the strategy used to prioritize events.

### 3. incremental addition

If new rules are added to a program which has already been analyzed, we have to compute the success patterns for each new rule, use this to update the answer table information for the atoms defined by the newly added rules, and then propagate the effect of these changes. Note that this propagation is not limited to the new rules, but rather a global fixed point has to be reached in order to ensure correctness of the analysis results. Existing analysis engines for (constraint) logic programming are unable to incrementally compute this new fixed point, and the only safe possibility is to start analysis from scratch. However, the generic algorithm we propose can do this rather simply. Computation of the success patterns for each rule is simply done by adding a set of arcs to the event queue before calling again main_loop. Propagation of the effects corresponds to processing, in the usual way, the updated events for entries in the answer table which are modified due to the newly added rules. When execution of main_loop ends, a new global fixed point has been reached.

The new routine for analysis of programs in which rules are added incrementally is given in Figure 3. The routine takes as input the set of new rules $R$. If these
incremental_addition(R)
  foreach rule $A_k := B_{k,1}, \ldots, B_{k,n_k} \in R$
  foreach entry $A : CP \mapsto AP$ in the answer table
    $CP_0 := Aextend(CP, vars(A_k := B_{k,1}, \ldots, B_{k,n_k}))$
    $CP_1 := Arestrict(CP_0, vars(B_{k,1}))$
    add_event(\text{arc}(A_k : CP \mapsto \{CP_0 | B_{k,1} : CP_1\})$
  main_loop()

Fig. 3. Incremental addition algorithm.

match an atom with a calling pattern of interest, then requests to process the
rule are placed on the priority queue. Subsequent processing is exactly as for the
nonincremental case.

Example 2. As an example, we begin with the program for naive reversal of a
list, rev, already analyzed for the calling pattern rev(X, Y) : true but without
a definition of the append, app, predicate. The initial program is

\[
\begin{align*}
\text{rev}_1(X, Y) & : X = [], Y = []. \\
\text{rev}_2(X, Y) & : X = [U | V], \text{rev}_2(V, W), T = [U | 3], \text{app}_2(W, T, Y). \\
\end{align*}
\]

The answer table and dependency arc tables are (State 1)

answer table:  \(\text{rev}(X, Y) : true \mapsto X \land Y\)
\(\text{app}(X, Y, Z) : X \mapsto false\)

dependency arc table:
\(\text{rev}_2(X, Y) : true \Rightarrow \{X \leftrightarrow (U \land V)\} \ \text{rev}_2(V, W) : true\)
\(\text{rev}_2(X, Y) : true \Rightarrow \{(X \leftrightarrow (U \land V)) \land V \land W \land (T \leftrightarrow U)\} \ \text{app}_2(W, T, Y) : W\)

We now add the rules for app one at a time. The first rule to be added is

\[
\begin{align*}
\text{app}_3(X, Y, Z) & : X = [], Y = Z_3.2. \\
\end{align*}
\]

The incremental analysis begins by looking for entries referring to app in the answer
table. It finds the entry \text{app}(X, Y, Z) : X so the arc

\[
\text{app}_3(X, Y, Z) : X \Rightarrow [X] X = [] : X
\]

is put in the priority queue. After processing this rule, the new answer \(X \land (Y \leftrightarrow Z)\)
for \text{app}(X, Y, Z) : X is obtained. Taking the least upper bound of this with the
current answer we obtain \(X \land (Y \leftrightarrow Z)\), and the answer table entry is updated
(causing an \text{updated(app}(X, Y, Z) : X) event). Examining the dependency arc table,
the algorithm recognizes that the answer from rule 2 must now be recomputed. This
gives rise to the new answer \(X \leftrightarrow (U \land V) \land V \land W \land (T \leftrightarrow U)\) which restricted
to \(\{X, Y\}\) gives \(X \leftrightarrow Y\). Taking the least upper bound of this with the current
answer \(X \land Y\) gives \(X \leftrightarrow Y\). The memo table entry for \(\text{rev}(X, Y) : true\) is updated
appropriately, and an \text{updated} event is placed on the queue. Again the answer to
rule 2 must be recomputed. First we obtain a new calling pattern \text{app}_2(X, Y, Z) :
true. This means that the dependency arc

\[
\text{rev}_2(X, Y) : true \Rightarrow \{(X \leftrightarrow (U \land V)) \land V \land W \land (T \leftrightarrow U)\} \ \text{app}_2(W, T, Y) : W
\]
in the dependency arc table is replaced by

\[
\text{rev}_2(X, Y) : true \Rightarrow \{(X \leftrightarrow (U \land V) \land (V \leftrightarrow W) \land (T \leftrightarrow U))\} \ \text{app}_2(W, T, Y) : true
\]
This sets up a new call \( \text{app}(X, Y, Z) : \text{true} \). The current answer for the old call \( \text{app}(X, Y, Z) : X \) can be used as an initial guess to the new, more general, call. The algorithm examines rule 3 for the new calling pattern. It obtains the same answer \( X \land (Y \leftrightarrow Z) \).

This leads to a new answer for \( \text{rev}_2(X, Y), (X \leftrightarrow (U \lor V)) \land (V \leftrightarrow W) \land (T \leftrightarrow U) \land (T \leftrightarrow Y) \) which restricted to \( \{X, Y\} \) gives \( X \leftrightarrow Y \). This does not change the current answer, so the main loop of the analysis is finished. The reachability analysis removes the entry \( \text{app}(X, Y, Z) : X \leftrightarrow X \land (Y \leftrightarrow Z) \) from the answer table. The resulting answer and dependency arc table entries are (State 2)

\[
\text{answer table: } \begin{align*}
\text{rev}(X, Y) : & \quad \text{true} \quad \mapsto \quad X \leftrightarrow Y \\
\text{app}(X, Y, Z) : & \quad \text{true} \quad \mapsto \quad X \land (Y \leftrightarrow Z)
\end{align*}
\]

\[
\text{dep. arc table: } \begin{align*}
\text{rev}_2(X,Y) : & \quad \text{true} \quad \Rightarrow \quad [X \leftrightarrow (U \lor V)] \quad \text{rev}_2(V,W) : \quad \text{true} \\
\text{rev}_2(X,Y) : & \quad \text{true} \quad \Rightarrow \quad [X \leftrightarrow (U \lor V) \land (V \leftrightarrow W) \land (T \leftrightarrow U)] \quad \text{app}_{2,4}(W,T,Y) : \quad \text{true}
\end{align*}
\]

If the second rule for \( \text{app} \)

\[
\text{app}_4(X,Y,Z) :- X = [U|V]_{4,1}, \quad Z = [U|W]_{4,2}, \quad \text{app}_{4,3}(V,Y,W).
\]
is added, the analysis proceeds similarly. The final memo and dependency arc table entries are (State 3)

\[
\text{answer table: } \begin{align*}
\text{rev}(X, Y) : & \quad \text{true} \quad \mapsto \quad X \leftrightarrow Y \\
\text{app}(X, Y, Z) : & \quad \text{true} \quad \mapsto \quad (X \land Y) \leftrightarrow Z
\end{align*}
\]

\[
\text{dep. arc table: } \begin{align*}
(A) \quad \text{rev}_2(X, Y) : & \quad \text{true} \quad \Rightarrow \quad [X \leftrightarrow (U \lor V)] \quad \text{rev}_2(V,W) : \quad \text{true} \\
(B) \quad \text{rev}_2(X,Y) : & \quad \text{true} \quad \Rightarrow \quad [X \leftrightarrow (U \lor V) \land (V \leftrightarrow W) \land (T \leftrightarrow U)] \quad \text{app}_{2,4}(W,T,Y) : \quad \text{true} \\
(C) \quad \text{app}_4(X,Y,Z) : & \quad \text{true} \quad \Rightarrow \quad [X \leftrightarrow (U \lor V) \land Z \leftrightarrow (U \land W)] \quad \text{app}_{4,3}(V,Y,W) : \quad \text{true}
\end{align*}
\]

\[\square\]

Correctness of the incremental addition algorithm follows from correctness of the original generic algorithm. Essentially, execution of the incremental addition algorithm corresponds to executing the generic algorithm with all rules but with the new rules having the lowest priority for processing. It therefore follows from Corollary 1 that:

**Theorem 2.** If the rules in a program are analyzed incrementally with the incremental addition algorithm, the same answer and dependency arc tables will be obtained as when all rules are analyzed at once by the generic algorithm.

In a sense, therefore, the cost of performing the analysis incrementally can be no worse than performing the analysis all at once, as the generic analysis could have used a priority strategy which has the same cost as the incremental strategy. We will now formalize this intuition. Our cost measure will be the number of calls to the underlying parametric functions. This is a fairly simplistic measure, but our results continue to hold for reasonable measures.

Let \( C_{\text{noninc}}(F, R, S) \) be the worst-case number of calls to the parametric functions \( F \) when analyzing the rules \( R \) and call patterns \( S \) for all possible priority strategies with the generic analysis algorithm.
Let $C_{\text{add}}(F, R, R', S)$ be the worst-case number of calls to the parametric functions $F$ when analyzing the new rules $R'$ for all possible priority strategies with the incremental addition algorithm after already analyzing the program $R$ for call patterns $S$.

**Theorem 3.** Let the set of rules $R$ be partitioned into $R_1, \ldots, R_n$ rule sets. For any call patterns $S$ and parametric functions $F$,

$$C_{\text{noninc}}(F, R, S) \geq \sum_{i=1}^{n} C_{\text{add}}(F, (\bigcup_{j=1}^{j<i} R_j), R_i, S).$$

The theorem holds because the priority strategies which give the worst-case behavior for each of the $R_1, \ldots, R_n$ can be combined to give a priority strategy for analyzing the program nonincrementally.

We note that our theorems comparing relative complexity of incremental and nonincremental analysis (Theorems 3, 5, and 8) are rather weak, since they relate only the worst-case complexity. Unfortunately, it is difficult to provide more insightful analytic comparisons; instead we will provide an empirical comparison in Section 6.

4. **INCREMENTAL DELETION**

In this section we consider deletion of rules from an already analyzed program and how to incrementally update the analysis information. The first thing to note is that, unlike incremental addition, we need not change the analysis results at all. The current approximation is trivially guaranteed to be correct, because an answer table entry $A : CP \rightarrow AP$ is computed by taking the least upper bound of the contribution of all rules in the definition of $A$. If some rules with $A$ as head are deleted, the previously computed answer pattern $AP$ is still clearly correct for the remaining rules. This approach is obviously inaccurate but simple.

4.1 **Refinement**

More accuracy can be obtained by applying a strategy similar to narrowing [Cousot and Cousot 1979]. Narrowing is a generic fixed-point approximation technique in which analysis proceeds from above the least fixed point and iterates downward until a fixed point (not necessarily the least fixed point) is reached. We can use this approach because the current approximations in the answer table are greater than or equal to those in the answer table of the program analysis graph for the modified program. Applying the analysis engine as usual except taking the greatest lower bound (glb), written $\sqcap$, of new answers with the old rather than the least upper bound is guaranteed to produce a correct, albeit perhaps imprecise, result. We can let this process be guided from the initial changes using the dependency graph information. Care must be taken to treat new calling patterns that arise in this process correctly. Note this narrowing-like strategy is correct in part because of the existence of a Galois connection between the concrete and abstract domains, as this means that glb on the abstract domain approximates glb on the underlying concrete domain.
Example 3. Consider the program in Example 2 after both additions. The current answer table and dependency graph entries are given by State 3. Deleting rule 4 results in the following process.

First we delete all dependency arcs which correspond to deleted rules. In this case we remove the arc \( \text{app4}(X, Y, Z) \rightarrow \text{true} \rightarrow \text{app4}(V, Y, W) \rightarrow \text{true}. \) In general we may subsequently delete other dependency arcs which are no longer required.

We recompute the answer information for all (remaining) rules for \( \text{app}(X, Y, Z) \) for all calling patterns of interest using the current answer information. We obtain \( \text{app}(X, Y, Z) \rightarrow \text{true} \rightarrow X \land (Y \leftarrow Z). \)

Because this information has changed we now need to consider recomputing answer information for any calling patterns that depend on \( \text{app}(X, Y, Z) \rightarrow \text{true}, \) in this case \( \text{rev}(X, Y) \rightarrow \text{true}. \) Recomputing using rules 1 and 2 obtains the same answer information \( X \leftarrow Y. \) The result is State 2 (with the useless entry for \( \text{app}(X, Y, Z) \rightarrow X \) removed).

Deleting rule 3 subsequently leads back to State 1 as expected. In contrast, removing rule 3 from the program consisting of rules 1 to 4 does not result in recovering State 1 as might be expected. This highlights the possible inaccuracy of the narrowing method. In this case rule 4 prevents more accurate answer information from being acquired. □

The disadvantage of this method is its inaccuracy. Starting the analysis from scratch will often give a more accurate result. We now give two algorithms which are incremental yet are as accurate as the nonincremental analysis.

4.2 “Top-Down” Deletion Algorithm

The first accurate method we explore for incremental analysis of programs after deletion is to remove all information in the answer and dependency arc tables which depends on the rules which have been deleted and then to restart the analysis. Not only will removal of rules change the answers in the answer table, it will also mean that subsequent calling patterns may change. Thus we will also remove entries in the dependency arc table for those rules which are going to be reanalyzed.

Information in the answer table and dependency arc table allows us to find these no longer valid entries. Let \( D \) be the set of deleted rules and \( H \) be the set of atoms which occur as the head of a deleted rule, i.e., \( H = \{A|(A :- B) \in D\}. \) We let \( \text{depend}(H) \) denote the set of calling patterns whose answers depend on some atom...
in \( H \). More precisely, \( \text{depend}(H) \) is the smallest superset of

\[
\{(A : CP) | (A : CP \rightarrow AP) \in \text{answer table and } A \in H\}
\]

such that if \( A : CP \) is in \( \text{depend}(H) \) and there is a dependency arc of the form \( B : CP_0 \Rightarrow [A'] : CP \) such that \( A' : CP' \) is a renaming of \( A : CP \) then \( B : CP_0 \) is also in \( \text{depend}(H) \). After entries for these dependent calling patterns which are no longer valid are deleted, the usual generic analysis is performed. The routine for top-down rule deletion is given in Figure 4. It is called with the set of deleted rules \( D \) and a set of initial calling patterns \( S \).

**Example 4.** Consider the program

\[
\begin{align*}
q_1 & : p_{1,1}(X, Y), r_{1,2}(X, Y, Z), s_{1,3}(Y, Z). \\
p_2(X, Y) & : X = a_{2,1}, Y = b_{2,2}. \\
p_3(X, Y) & : X = Y_{3,1}. \\
r_4(X, Y, Z) & : X = Z_{4,1}. \\
r_5(X, Y, Z) & : Y = Z_{5,1}. \\
s_6(Y, Z) & : Y = c_{6,1}. 
\end{align*}
\]

After program analysis we obtain *(State 5)*

**answer table:**

\[
\begin{align*}
q & : \text{true} \Rightarrow \text{true} \\
p(X, Y) & : \text{true} \Rightarrow X \leftrightarrow Y \\
r(X, Y, Z) & : X \leftrightarrow Y \Rightarrow (X \leftrightarrow Y) \land (Y \leftrightarrow Z) \\
s(Y, Z) & : Y \leftrightarrow Z \Rightarrow Y \land Z
\end{align*}
\]

**dependency arc table:**

\[
\begin{align*}
(D) & : q_1 : \text{true} \Rightarrow [\text{true}] \\
& : p_{1,1}(X, Y) : \text{true} \\
(E) & : q_1 : \text{true} \Rightarrow [X \leftrightarrow Y] \\
& : r_{1,2}(X, Y, Z) : X \leftrightarrow Y \\
(F) & : q_1 : \text{true} \Rightarrow [(X \leftrightarrow Y) \land (Y \leftrightarrow Z)] \\
& : s_{1,3}(Y, Z) : Y \leftrightarrow Z
\end{align*}
\]

Now consider the deletion of rule \( r_5 \). \( H = \{r(X, Y, Z)\} \) and \( \text{depend}(H) = \{r(X, Y, Z) : X \leftrightarrow Y, q : \text{true}\} \). Hence, the answer table entries for \( r \) and \( q \) are deleted and also all the dependency arcs are deleted since they are all for predicate \( q \) which is going to be reanalyzed. The initial state *(State 6)* when we start the main loop is

**answer table:**

\[
\begin{align*}
p(X, Y) & : \text{true} \Rightarrow X \leftrightarrow Y \\
s(Y, Z) & : Y \leftrightarrow Z \Rightarrow Y \land Z
\end{align*}
\]

**dependency arc table:** no entries

The only entry in the priority queue is \text{newcall}(q : \text{true}). Then, the execution of main loop processes this event by adding the entry \( q : \text{true} \Rightarrow \text{false} \) to the answer table and an arc event for \((D)\). This is selected; arc \((D)\) is added again to the dependency arc table; and arc \((E)\) is placed on the priority queue. This is selected; arc \((E)\) is placed back in the dependency arc table, and the event \text{newcall}(r(X, Y, Z) : X \leftrightarrow Y)\) is placed on the queue. This generates an answer entry \( r(X, Y, Z) : X \leftrightarrow Y \Rightarrow \text{false} \) and arc \( r_4(X, Y, Z) : X \leftrightarrow Y \Rightarrow [X \leftrightarrow Y] \times = Z : \text{true} \). This in turn generates new answer information \( (X \leftrightarrow Y) \land (Y \leftrightarrow Z) \) and the event \text{updated}(r(X, Y, Z) : X \leftrightarrow Y). This is replaced with arc \((E)\), which is replaced with arc \((F)\), which results in arc \((F)\) being added again to the dependency graph and new answer info \( q : \text{true} \Rightarrow \text{true} \) and an event \text{updated}(q : true \Rightarrow true).
true) which is removed with no effect. The resulting state is identical to the starting state (State 5).

Example 5. Consider again the rev and app program in Example 2. After analysis of the entire program we are in State 3. Now consider the deletion of rule 3 from the program consisting of rules 1 to 4. \(T = \text{depend}(\text{app}(X, Y, Z))\) is all the calling patterns, so the answer table and dependency arc table are emptied. Reanalysis is complete starting from the initial calling pattern \(\text{rev}(X, Y) : true\) and results in State 1 as expected. Note that this is the case in Example 3 for which the refinement method yielded an inaccurate answer.

Correctness of the incremental top-down deletion algorithm follows from correctness of the original generic algorithm. Execution of the top-down deletion algorithm is identical to that of the generic algorithm except that information about the answers to some call patterns which do not depend on the deleted rules is already in the data structures.

**Theorem 4.** If a program \(P\) is first analyzed and then rules \(R\) are deleted from the program and the remaining rules are reanalyzed with the top-down deletion algorithm, the same answer and dependency arc tables will be obtained as when the rules \(P \setminus R\) are analyzed by the generic algorithm.

The cost of performing the actual analysis incrementally can be no worse than performing the analysis all at once. Let \(C_{\text{del--td}}(\bar{F}, R, R', S)\) be the worst-case number of calls to the parametric functions \(\bar{F}\) when analyzing the program \(R\) with rules \(R'\) deleted for all possible priority strategies with the top-down deletion algorithm after already analyzing the program \(R\) for call patterns \(S\).

**Theorem 5.** Let \(R\) and \(R'\) be sets of rules such that \(R' \subseteq R\). For any call patterns \(S\) and parametric functions \(\bar{F}\),

\[
C_{\text{noninc}}(\bar{F}, R \setminus R', S) \geq C_{\text{del--td}}(\bar{F}, R, R', S).
\]

### 4.3 "Bottom-Up" Deletion Algorithm

The last theorem shows that the top-down deletion algorithm is never worse than starting the analysis from scratch. However, in practice it is unlikely to be that much better, as on average deleting a single rule will mean that half of the dependency arcs and answers are deleted in the first phase of the algorithm. The reason is that the top-down algorithm is very pessimistic—deleting everything unless it is sure that it will be both correct and useful. For this reason we now consider a more optimistic algorithm. The algorithm assumes that calling patterns to changed predicate definitions are still likely to be useful. In the worst case it may spend a large amount of time reanalyzing calling patterns that end up being useless. But in the best case we do not need to reexamine large parts of the program above changes when no actual effect is felt.

The algorithm proceeds by computing new answers for calling patterns in the lowest strongly connected component\(^3\) (SCC) of the set \(\text{depend}(H)\) of calling patterns...
bottom_up_delete(D, S)
    H := ∅
    foreach rule A_k := B_k,1,...,B_k,nk ∈ D
        foreach A : CP ⇒ AP in answer table
            H := H ∪ {A : CP}
            delete each arc A_ : CP ⇒ B_ : CP from dependency arc table
        while H is not empty
            let B : CP be such that B is of minimum predicate SCC level
            T := calling patterns in program analysis graph for predicates in
            the same predicate SCC level as B
            foreach A : CP ∈ T
                delete each arc A_ : CP ⇒ B_ : CP from dependency arc table
            foreach A : CP ∈ external_calls(T, S)
                move entry A : CP ⇒ AP from answer table to old answer table
                foreach arc B_k : CP0 ⇒ [CP1] B_k,j : CP2 in dependency arc table
                    where there exists renaming σ s.t. (A : CP) = (B_k,j : CP2)σ
                    move B_k : CP0 ⇒ [CP1] B_k,j : CP2 to old dependency table
                    add_event(newcall(A : CP))
            main_loop()
        foreach A : CP ∈ external_calls(T, S)
            foreach arc B_k : CP0 ⇒ [CP1] B_k,j : CP2 in old dependency arc table
                where there exists renaming σ s.t. (A : CP) = (B_k,j : CP2)σ
                if answer pattern for A : CP in old answer table and answer table agree
                    move B_k : CP0 ⇒ [CP1] B_k,j : CP2 to dependency arc table
                else
                    H := H ∪ {B : CP0}
            H := H − T
            empty old answer table and old dependency arc table
    external_calls(T, S)
        U := ∅
        foreach A : CP ∈ T
            where exists arc B_k : CP0 ⇒ [CP1] B_k,j : CP2
            and B : CP0 ∉ T
            and there exists renaming σ s.t. (A : CP) = (B_k,j : CP2)σ
            % this means there is an external call
            U := U ∪ {A : CP}
        return U ∪ (T ∩ S)

Fig. 5. Bottom-up incremental deletion algorithm.

terns which could be affected by the rule deletion. After evaluating the lowest SCC, the algorithm moves upward to higher SCCs. At each stage the algorithm recomputes or verifies the current answers to the calls to the SCC without considering dependency arcs from SCCs in higher levels. This is possible because if the answer changes, the arc events they would generate are computed anyway. If the answers are unchanged then the algorithm stops; otherwise it examines the SCCs which depend on the changed answers (using the dependency arcs). For obvious reasons we call the algorithm Bottom-Up Deletion. It is shown in Figure 5.

Rather than using the program analysis graph to determine SCCs, an object which changes as the analysis proceeds, the algorithm uses the predicate call graph of the program P before the rule deletions to determine the calling patterns that can affect one another. This is static, and gives overestimates on the SCCs of
the program analysis graph (at any stage), and hence this approach is safe. The *predicate call graph* has predicate names as nodes, and an arc from node p to q if there exists a rule with predicate p in the head and an atom for predicate q appearing in the body. We can use this graph to define a *predicate SCC level* to each predicate (and atom) where predicates in lower levels cannot reach any predicate in a higher level.

The algorithm begins by collecting all the calling patterns for atoms with deleted rules, and deleting all dependency arcs for deleted rules, since they can play no part in the reanalysis. It calculates as H the initial calling patterns which must be reevaluated since they have had a rule removed.

It then chooses the minimum predicate SCC level for predicates in H and collects all the affected calling patterns in this level in T. All the dependency arcs for pairs in T are deleted; the current answers for pairs in T are moved to the old table for later comparing with the new answers.

Next the external calls to calling patterns in T are calculated by external_calls(T,S). These are the calling patterns for which there is dependency arc from a higher SCC levels, or an external call from S. All the dependency arcs which call T from higher SCC are moved temporarily into the old dependency arc table. During analysis of this SCC, *updated* events may be generated which do not need to be propagated outside the SCC, as the answer computed after deleting some rules may (finally) be equal to the one computed for the original program. In effect this isolates the calls T from the rest of the higher SCC levels of the program analysis graph. Each external call is then treated as a *newcall* event.

After the SCC has been fully analyzed, a comparison is made for each external calling pattern A : CP. If the new and old answers in the table agree then no recomputation of calling patterns which depend on this A : CP is required. Otherwise the dependent calling patterns are added to H, the calling patterns which need recomputation. Finally the calling patterns T are deleted from H, and the process proceeds.

**Example 6.** Consider the same deletion as in Example 4. Initially the set H is \{r(X,Y,Z) : X <-> Y\}, and T is the same set. There are no dependency arcs for pairs in T, and the single pattern is an external call because of arc (E). The answer table entry r(X,Y,Z) : X <-> Y \rightarrow (X <-> Y) \land (Y <-> Z) is moved to the old answer table. The dependency arc (E) is moved to the old dependency table. The event *newcall*(r(X,Y,Z) : X <-> Y) is placed on the queue. This (re)generates new answer information r(X,Y,Z) : X <-> Y \rightarrow (X <-> Y) \land (Y <-> Z) and an event *updated*(r(X,Y,Z) : X <-> Y). As the dependency arc table has no arc that needs to be recomputed we stop. Because the answer for r(X,Y,Z) : X <-> Y is unchanged the arc (E) is moved to the dependency arc table and the algorithm terminates, without recomputing q : true. □

**Example 7.** Consider the **rev** and **app** program. After analysis of the entire program we are in **State 3**. Now consider the deletion of rule 3 from the program. H is initially \{app(X,Y,Z) : true\}. So is T. We remove the arc (C). We move the answer pattern for app(X,Y,Z) : true and the arc (B) to the old tables. The event *newcall*(app(X,Y,Z) : true) is placed in the queue. The analysis proceeds to compute answer app(X,Y,Z) : true \rightarrow false. Since this has changed, rev(X,Y) :
true is added to \( H \). \( \text{app}(X,Y,Z) : \text{true} \) is removed from \( H \). In the next iteration \( H = \{ \text{rev}(X,Y) : \text{true} \} \). The answer pattern for \( \text{rev}(X,Y) : \text{true} \) is moved to the old table, and the arc (\( A \)) is removed. Reanalysis proceeds as before including building a new call to \( \text{app}(X,Y,Z) : X \). This gives the answer \( \text{rev}(X,Y) : \text{true} \leftarrow X \land Y \). The resulting state is \textbf{State 1} as expected. Note that the reanalysis of \( \text{app}(X,Y,Z) : \text{true} \) was unnecessary for computing the answers to the call to \( \text{rev} \) (this was avoided by the top-down deletion). □

Proving correctness of the incremental bottom-up deletion algorithm requires an inductive proof on the SCCs. Correctness of the algorithm for each SCC follows from correctness of the generic algorithm.

\textbf{Theorem 6.} If a program \( P \) is first analyzed for calls \( S \), and then rules \( R \) are deleted from the program, while the remaining rules are reanalyzed with the bottom-up deletion algorithm, then the same answer and dependency arc tables will be obtained as when the rules \( P \setminus R \) are analyzed by the generic algorithm for \( S \).

Unfortunately, in the worst case, reanalysis with the bottom-up deletion algorithm may take longer than reanalyzing the program from scratch using the generic algorithm. This is because the bottom-up algorithm may do a lot of work recomputing the answer patterns to calls in the lower SCCs which are no longer made. In practice, however, if the changes are few and have local extent, the bottom-up algorithm will be faster than the top-down.

5. \textbf{ARBITRARY CHANGE}

In this section we consider the most general case of program modification, in which rules can both be deleted from and added to an already analyzed program and how to incrementally update the analysis information. Given the above algorithms for addition and deletion of rules we can handle any possible change of a set of rules by first deleting the original and then adding the revised version. This is inefficient, since the revision may not involve very far reaching changes, while the deletion and addition together do. Moreover we compute two fixed points rather than one.

Instead, we can use the top-down and bottom-up deletion algorithms in order to perform incremental analysis while only computing a single fixed point. For the case of top-down deletion it suffices to execute main_loop with the updated set of rules. For using the bottom-up deletion algorithm, and unlike in the case of incremental deletion, we must recompute the SCCs in order to ensure correctness, as new cycles may have been introduced in the call dependency graph due to the newly added rules. Then, we can use the bottom-up deletion algorithm as usual with the updated set of rules.

\textit{Example 8.} Consider the following program:

\begin{align*}
q_1(X, Y) &\leftarrow p_{1,1}(X, Y) .
p_2(X, Y) &\leftarrow X=a_{2,1}, Y=b_{2,2} .
\end{align*}

The complete analysis information for the initial call \( q(X, Y) : X \) is

\begin{align*}
\text{answer table:} & \quad q(X, Y) : X \leftarrow X \land Y \\
p(X, Y) : X \leftarrow X \land Y
\end{align*}

\begin{align*}
\text{dependency arc table:} & \quad q_1(X, Y) : X \Rightarrow p_{1,1}(X, Y) : X
\end{align*}
local_change(S, R)
let R be of the form $A_k := D_{k,1}, \ldots, D_{k,n_k}$
T := ∅
foreach A : CP ⇒ AP in answer table
T := T ∪ {A : CP}
T := T plus all B : CP₀ in same SCCs of program analysis graph
delete each arc of the form $A_k := \_ \Rightarrow \_ \_ : \_$ from the dependency arc table
foreach A : CP ∈ external_calls(T, S)
    CP₀ := Aextend(CP, vars(D_{k,1}, \ldots, D_{k,n_k}))
    CP₁ := Arestrict(CP₀, vars(D_{k,1}))
    add_event(arc(A_k : CP ⇒ [CP₀ \ D_{k,1} : CP₁]))
main_loop()

Fig. 6. Local change algorithm.

Consider replacing the rule for $p$ by $p₃(X, Y) :- U = a₃₁, q₃₂(U, Y)$. If we do not recompute the SCCs, the bottom-up algorithm would analyze the rule $p₃(X, Y) :- U = a₃₁, q₃₂(U, Y)$ with entry description $X$. Using the (no longer correct) entry $q(X, Y) : X \rightarrow X \wedge Y$ in the answer table, analysis would compute the incorrect entry $p(X, Y) : X \rightarrow X \wedge Y$ which is consistent with the old table and thus terminate. However, if we had recomputed the SCCs, since $p$ and $q$ are in the same SCC, the entry $q(X, Y) : X \rightarrow X \wedge Y$ would not be in the answer table but rather in the old table. □

5.1 Local Change

One common reason for incremental modification to a program is optimizing compilation. Changes from optimization are special in the sense that usually the answers to the modified rule do not change. This means that the changes caused by the modification are local in that they cannot affect SCCs above the change. Thus, changes to the analysis are essentially restricted to computing the new call patterns that these rules generate. This allows us to obtain an algorithm for local change (related to bottom-up deletion) which is more efficient than arbitrary change.

The algorithm for local change is given in Figure 6. It takes as arguments the original calling patterns $S$ and a modified rule $R$, which we assume has the same number as the rule it replaces.

The local change algorithm resembles bottom_up_delete in that only (part of) an SCC is reanalyzed. An important difference is that in local change it is guaranteed that we do not need to reanalyze the SCCs above the modified one. First all possibly affected calling patterns are collected in $T$. Because it is a local change we do not need to consider calling patterns outside the SCC of the program analysis graph. The arcs corresponding to the deleted rule are then deleted, and new arc events are added to process the new version of the changed rule.

Correctness of the local change algorithm essentially follows from correctness of the bottom-up deletion algorithm. Let $A :- B$ and $A :- B'$ be two rules. They are local variants with respect to the calls $S$ and program $P$ if for each call pattern in $S$ the program $P \cup \{A :- B\}$ has the same answer patterns as $P \cup \{A :- B'\}$.

**Theorem 7.** Let $P$ be a program analyzed for the initial call patterns $S$. Let $R$ be a rule in $P$ which in the analysis is called with call patterns $S'$, and let $R'$ be a
local variant of $R$ with respect to $S'$ and $P\setminus\{R\}$. If the program $P$ is reanalyzed with the routine $\text{local\_change}(S', R')$ the same answer and dependency arc tables will be obtained as when the rules $P \cup \{R'\} \setminus \{R\}$ are analyzed by the generic algorithm.

The cost of performing the actual analysis incrementally can be no worse than performing the analysis all at once. Let $C_{\text{local}}(\hat{F}, P, R, R', S)$ be the worst-case number of calls to the parametric functions $F$ when analyzing the program $P$ with rule $R$ changed to $R'$ for all possible priority strategies with the local change algorithm after already analyzing the program $P$ for call patterns $S$.

**Theorem 8.** Let $P$ be a program analyzed for the initial call patterns $S$. Let $R$ be a rule in $P$ which in the analysis is called with call patterns $S_1$, and let $R'$ be a local variant of $R$ with respect to $S'$ and $P \setminus \{R\}$. For any parametric functions $\hat{F}$,

$$C_{\text{noninc}}(\hat{F}, P \cup \{R'\} \setminus \{R\}, S) \geq C_{\text{local}}(\hat{F}, P, R, R', S).$$

6. EXPERIMENTAL RESULTS

We have conducted a number of experiments using the PLAI generic abstract interpretation system in order to assess the practicality of the techniques proposed in the previous sections. PLAI can be seen as an efficient restricted instance of the generic algorithm of Section 2 specialized for the case of analyzing the whole program at once. As mentioned in Section 2, PLAI uses the concrete strategy of always performing new call events first, processing nonrecursive rules before recursive rules, and finishing processing a rule before starting another. Prior to the invocation of the fixed-point algorithm a step is performed in which the set of predicates in the program is split into the SCCs based on the call graph of the program found using Tarjan’s algorithm [Tarjan 1972]. This information, among other things, allows determining which predicates and which rules of a predicate are recursive. PLAI (and its incremental extensions) also incorporates some additional optimizations such as dealing directly with nonnormalized programs and filtering out noneligible rules using concrete unification (or constraint solving) when possible.

In one way, however, the original PLAI differed from the algorithm given in Section 2: in order to simplify the implementation, the original fixed-point algorithm did not keep track of dependencies at the level of literals, but rather, in a coarser way, at the level of rules. As a result, when an updated event takes place, the dependent arcs have to be recomputed from the head of the rule. However, all subsequent iterations do not affect those rules which are detected not to be recursive. Since keeping track of dependencies at the literal level allows the analysis to only recompute from the point where the changed answer can first affect the analysis information for the rule, it can avoid a significant amount of recomputation. This idea has been applied as an optimization technique (in a nonincremental analysis setting) in the prefix version [Englebert et al. 1993] of GAIA. The technique proved quite relevant in practice for GAIA, as it allowed avoiding recomputation of nonrecursive rules and starting reanalysis of a rule from the first literal possibly affected. This prefix version of GAIA can also be taken as a starting point for implementing an instance of the generic algorithm. We modified PLAI in order to keep track of dependencies at the literal level and to store such dependencies explicitly, i.e., in such a way that it computes the dependency arc table of the generic algorithm.
Table I. Summary of Benchmark Statistics

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Mean</th>
<th>Max</th>
<th># of</th>
<th># of</th>
<th>% direct</th>
<th>% mutual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vars/rule</td>
<td>vars/rule</td>
<td>preds</td>
<td>rules</td>
<td>recursion</td>
<td>recursion</td>
</tr>
<tr>
<td>aiakl</td>
<td>4.58</td>
<td>9</td>
<td>7</td>
<td>12</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>ann</td>
<td>3.17</td>
<td>14</td>
<td>65</td>
<td>170</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>bid</td>
<td>2.20</td>
<td>7</td>
<td>19</td>
<td>50</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>boyer</td>
<td>2.36</td>
<td>7</td>
<td>26</td>
<td>133</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>browse</td>
<td>2.63</td>
<td>5</td>
<td>8</td>
<td>29</td>
<td>62</td>
<td>25</td>
</tr>
<tr>
<td>deriv</td>
<td>3.70</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>fib</td>
<td>2.00</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>grammar</td>
<td>2.13</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>hanoiapp</td>
<td>4.25</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>mmatrix</td>
<td>3.17</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>occur</td>
<td>3.12</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>peephole</td>
<td>3.15</td>
<td>7</td>
<td>26</td>
<td>134</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td>progeom</td>
<td>3.59</td>
<td>9</td>
<td>9</td>
<td>18</td>
<td>66</td>
<td>0</td>
</tr>
<tr>
<td>qplan</td>
<td>3.18</td>
<td>16</td>
<td>46</td>
<td>148</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>qsortapp</td>
<td>3.29</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>query</td>
<td>0.19</td>
<td>6</td>
<td>4</td>
<td>52</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>rdtok</td>
<td>4.20</td>
<td>13</td>
<td>24</td>
<td>54</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>read</td>
<td>3.07</td>
<td>7</td>
<td>22</td>
<td>88</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>serialize</td>
<td>4.18</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>tak</td>
<td>7.00</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>warplan</td>
<td>2.47</td>
<td>7</td>
<td>29</td>
<td>101</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>witt</td>
<td>4.57</td>
<td>18</td>
<td>77</td>
<td>160</td>
<td>35</td>
<td>22</td>
</tr>
<tr>
<td>zebra</td>
<td>2.06</td>
<td>25</td>
<td>6</td>
<td>18</td>
<td>33</td>
<td>0</td>
</tr>
</tbody>
</table>

PLAI, as most analysis engines not designed for incremental analysis, does compute the answer table, but loses the dependency information when analysis terminates. As a result, the modified version of PLAI, which uses the dependency arc table for guiding recomputations, constitutes precisely an instance of the generic algorithm. As we will see below, and unlike for the prefix version of GAIA, such modification does not speed up PLAI much, since the original PLAI algorithm already avoids reanalysis of nonrecursive rules.

A relatively wide range of programs have been used as benchmarks. These benchmarks are the de-facto standard for logic program analysis. Some statistics on their size and complexity are given in Table I. Dead code, i.e., code which is unreachable from the original calling patterns, has been removed from the benchmarks. We give the average and maximum number of variables in each rule analyzed; total number of predicates and rules in the program; the percentage of directly and mutually recursive predicates.

All the analysis algorithms we experiment with have been implemented as extensions to the PLAI generic abstract interpretation system, which is an efficient, highly optimized, state-of-the-art analysis system and which is part of a working compiler.\(^4\) We argue that this makes the comparisons meaningful, since the algorithms have been implemented using the same technology, with many data struc-

---

\(^4\)PLAI is currently integrated in CiaoPP [Hermenegildo et al. 1999a], the preprocessor of the Ciao Prolog System. The analysis information produced by PLAI is used for static debugging [Puebla et al. 2000], program specialization, and program parallelization.
tures in common. They also share the domain-dependent functions, which are those of the sharing + freeness domain [Muthukumar and Hermenegildo 1991] in all the experiments. The whole system is implemented in Prolog and has been compiled using Ciao Prolog 0.9 [Bueno et al. 1997] with compilation to bytecode. All of our experiments have been performed on a Pentium II at 400mH and 512MB RAM running RedHat linux 5.2. Execution times are given in milliseconds and memory usage in kilobytes. In order to have an accurate measure of memory usage, garbage collection is turned off during analysis. Memory usage has been computed as follows. We measure the size of the heap (where dynamic terms are built) plus the size of dynamic code (where asserted information is stored) both before and after analyzing each benchmark. The difference between those two values is taken as the memory used in the process. Other memory structures such as the environment stack are not considered, since their size has been measured to be irrelevant when compared to the heap space. Though the resulting memory usage is high, it is important to note that run-time garbage collection actually reduces memory consumption by a very large amount.

6.1 Efficiency of the Generic Algorithm

Our first experiment compares the original PLAI fixed-point approach, where dependencies are kept at the level of rules, versus the approach with dependency tracking at the level of literals. This also serves to show that our baseline, the PLAI instance of the generic algorithm, is competitive with other top-down analysis engines. The results are shown in Table II.

In this table, as in the other tables in this section, there are basically two kinds of columns. Those which present the cost of analysis both in terms of time and memory usage for a given algorithm, and those which compare the costs of two analysis algorithms. Columns of the first kind are labeled with the name of the algorithm, say \( A_{alg1} \), while columns of the second kind are labeled by \( A_{alg1} / A_{alg2} \). The values in these columns are computed as the cost of analysis using \( A_{alg1} \) divided by cost using \( A_{alg2} \). In addition, below columns which compare algorithms we summarize the results for the different benchmarks using two different measures: the arithmetic mean of the values in the column above and a weighted mean which place more importance on those benchmarks with relatively larger analysis times / memory usage figures. We use as the weight for each program the actual analysis time or memory usage for it using \( A_{alg2} \). We believe that the weighted mean is more informative than the arithmetic mean, as, for example, doubling the speed in which a large and complex program is analyzed is more important than doing this for small, simple programs.

As shown in Table II, it turns out that the additional cost of keeping track of more detailed dependencies is offset by savings on recomputation. The value of 0.97 for the weighted mean for both time and memory consumption indicates that the original algorithm is slightly faster and uses slightly less memory than the modified algorithm. In any case, the differences in analysis times are due to the

---

6Obtained using statistics(global_stack,[G,\_]).
7Obtained using statistics(program,[P,\_]).
Table II. Cost of Literal Based Dependency Tracking

| Benchmark   | Modified PLAI Time | Original PLAI Time | Original / Modified
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Memory</td>
<td>Memory</td>
<td></td>
</tr>
<tr>
<td>aiakl</td>
<td>867 3304</td>
<td>1003 3712</td>
<td>1.16 1.12</td>
</tr>
<tr>
<td>ann</td>
<td>1655 6324</td>
<td>1743 6347</td>
<td>1.05 1.00</td>
</tr>
<tr>
<td>bid</td>
<td>178 668</td>
<td>171 602</td>
<td>0.96 0.90</td>
</tr>
<tr>
<td>boyer</td>
<td>573 2190</td>
<td>761 2659</td>
<td>1.33 1.21</td>
</tr>
<tr>
<td>browse</td>
<td>87 294</td>
<td>85 270</td>
<td>0.98 0.92</td>
</tr>
<tr>
<td>deriv</td>
<td>105 408</td>
<td>177 683</td>
<td>1.69 1.67</td>
</tr>
<tr>
<td>fib</td>
<td>2 22</td>
<td>4 21</td>
<td>2.00 0.95</td>
</tr>
<tr>
<td>grammar</td>
<td>25 98</td>
<td>26 96</td>
<td>1.04 0.98</td>
</tr>
<tr>
<td>hanoiapp</td>
<td>113 475</td>
<td>115 432</td>
<td>1.02 0.91</td>
</tr>
<tr>
<td>mmatrix</td>
<td>57 244</td>
<td>51 222</td>
<td>0.89 0.91</td>
</tr>
<tr>
<td>occur</td>
<td>68 258</td>
<td>69 262</td>
<td>1.01 1.02</td>
</tr>
<tr>
<td>peephole</td>
<td>1453 5639</td>
<td>1611 5843</td>
<td>1.11 1.04</td>
</tr>
<tr>
<td>progeom</td>
<td>39 144</td>
<td>37 137</td>
<td>0.95 0.95</td>
</tr>
<tr>
<td>qplan</td>
<td>433 1285</td>
<td>413 1218</td>
<td>0.95 0.95</td>
</tr>
<tr>
<td>qsortapp</td>
<td>64 273</td>
<td>61 256</td>
<td>0.95 0.94</td>
</tr>
<tr>
<td>query</td>
<td>19 46</td>
<td>21 45</td>
<td>1.11 0.98</td>
</tr>
<tr>
<td>rdtok</td>
<td>137 399</td>
<td>143 379</td>
<td>1.04 0.95</td>
</tr>
<tr>
<td>read</td>
<td>10695 38396</td>
<td>9475 35086</td>
<td>0.89 0.91</td>
</tr>
<tr>
<td>serialize</td>
<td>147 621</td>
<td>115 468</td>
<td>0.78 0.75</td>
</tr>
<tr>
<td>tak</td>
<td>11 79</td>
<td>14 73</td>
<td>1.27 0.92</td>
</tr>
<tr>
<td>warplan</td>
<td>823 2970</td>
<td>979 3452</td>
<td>1.19 1.16</td>
</tr>
<tr>
<td>witt</td>
<td>491 1651</td>
<td>509 1662</td>
<td>1.04 1.01</td>
</tr>
<tr>
<td>zebra</td>
<td>591 2172</td>
<td>579 2144</td>
<td>0.98 0.99</td>
</tr>
<tr>
<td></td>
<td>Arithmetic mean</td>
<td>1.10</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>Weighted mean</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

difference in dependency tracking rather than the cost of supporting incrementality. Hence modifying the analysis algorithm in order to support incremental analysis has not significantly slowed nonincremental analysis nor posed additional memory requirements. All subsequent experiments use the modified PLAI as the basic analysis algorithm, both for the incremental and nonincremental case. In Tables III and VI we denote by $\text{Orig}$ the cost of analyzing the whole program at once using this modified PLAI algorithm.

6.2 Incremental Addition

The next experiment compares the relative performance of incremental and nonincremental analysis in the context of addition. To do so for each benchmark we measured the cost of analyzing the program adding one rule at a time, rather than analyzing the whole program at once. That is, the analysis was first run for the first rule only. Then the next rule was added and the resulting program (re-)analyzed. This process was repeated until all rules had been added. Table III shows the cost of this process using the incremental approach of Section 3, and using a nonincremental approach, where analysis starts from scratch whenever a new rule is added. For the nonincremental case the same implementation was used but the tables were erased between analyses. This factors out any differences in fixed-point algorithms. Since for the nonincremental approach the analysis tables are erased after each analysis, we can reuse memory space. Thus, we take as memory usage
Table III. Incremental vs. Non-incremental Addition

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>Incremental</th>
<th>Non-incremental</th>
<th>Non / Inc</th>
<th>Inc / Orig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time Mem</td>
<td>Time Mem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aiakl</td>
<td>973 1399</td>
<td>1750 3361</td>
<td>1.80 2.40</td>
<td>1.12 0.42</td>
</tr>
<tr>
<td>ann</td>
<td>5136 6470</td>
<td>23688 7333</td>
<td>4.61 1.13</td>
<td>3.10 1.02</td>
</tr>
<tr>
<td>bid</td>
<td>861 260</td>
<td>3617 831</td>
<td>4.29 3.20</td>
<td>4.84 0.39</td>
</tr>
<tr>
<td>boyer</td>
<td>3632 1054</td>
<td>59867 2554</td>
<td>16.48 2.42</td>
<td>6.34 0.48</td>
</tr>
<tr>
<td>browse</td>
<td>395 145</td>
<td>2881 433</td>
<td>7.29 2.99</td>
<td>4.54 0.49</td>
</tr>
<tr>
<td>deriv</td>
<td>681 1482</td>
<td>369 442</td>
<td>0.54 0.30</td>
<td>6.49 3.63</td>
</tr>
<tr>
<td>fib</td>
<td>9 29</td>
<td>13 34</td>
<td>1.44 1.17</td>
<td>4.50 1.32</td>
</tr>
<tr>
<td>grammar</td>
<td>84 40</td>
<td>223 136</td>
<td>2.65 3.40</td>
<td>3.36 0.41</td>
</tr>
<tr>
<td>hanoiapp</td>
<td>189 392</td>
<td>358 494</td>
<td>1.89 1.26</td>
<td>1.67 0.83</td>
</tr>
<tr>
<td>mmatrix</td>
<td>153 191</td>
<td>216 266</td>
<td>1.41 1.39</td>
<td>2.68 0.78</td>
</tr>
<tr>
<td>occur</td>
<td>74 263</td>
<td>151 285</td>
<td>2.04 1.08</td>
<td>1.09 1.02</td>
</tr>
<tr>
<td>peephole</td>
<td>12930 1752</td>
<td>136145 6118</td>
<td>10.53 3.49</td>
<td>8.90 0.31</td>
</tr>
<tr>
<td>progeom</td>
<td>69 159</td>
<td>168 206</td>
<td>2.43 1.30</td>
<td>1.77 1.10</td>
</tr>
<tr>
<td>qplan</td>
<td>7970 1147</td>
<td>35492 1898</td>
<td>4.45 1.65</td>
<td>18.41 0.89</td>
</tr>
<tr>
<td>qsortapp</td>
<td>127 175</td>
<td>263 298</td>
<td>2.07 1.70</td>
<td>1.98 0.64</td>
</tr>
<tr>
<td>query</td>
<td>365 31</td>
<td>839 139</td>
<td>2.30 4.48</td>
<td>19.21 0.67</td>
</tr>
<tr>
<td>rdtok</td>
<td>609 227</td>
<td>4112 725</td>
<td>6.75 3.19</td>
<td>4.45 0.57</td>
</tr>
<tr>
<td>read</td>
<td>42195 25673</td>
<td>246754 51361</td>
<td>5.85 2.00</td>
<td>3.95 0.67</td>
</tr>
<tr>
<td>serialize</td>
<td>322 793</td>
<td>719 667</td>
<td>2.23 0.84</td>
<td>2.19 1.28</td>
</tr>
<tr>
<td>tak</td>
<td>22 84</td>
<td>25 94</td>
<td>1.14 1.12</td>
<td>2.00 1.06</td>
</tr>
<tr>
<td>warplan</td>
<td>9147 2151</td>
<td>26429 3335</td>
<td>2.89 1.55</td>
<td>11.11 0.72</td>
</tr>
<tr>
<td>witt</td>
<td>3808 1543</td>
<td>13536 2523</td>
<td>3.55 1.64</td>
<td>7.76 0.93</td>
</tr>
<tr>
<td>zebra</td>
<td>1685 1426</td>
<td>5836 2240</td>
<td>3.46 1.57</td>
<td>2.85 0.66</td>
</tr>
</tbody>
</table>

Arithmetic mean 4.09 1.97 5.40 0.88
Weighted mean 6.16 1.83 4.91 0.69

The maximum amount of memory used when analyzing rules 1 to i of the benchmark for i \in \{1...n\}, where n is the number of rules in that benchmark. Usually, the maximum is reached when all the rules are analyzed, i.e., i = n. In the next column we compare the relative costs of nonincremental and incremental analysis. Finally we compare the cost of analyzing the entire program one rule at a time using the incremental approach (column 1) with analyzing the entire program once using the nonincremental approach (column 2 of Table II).

The results are quite encouraging: using an incremental analysis was 6.16 times faster than using nonincremental analysis on average, while the cost of incrementally analyzing a program rule by rule as opposed to all at once was only 4.91 times worse on average (5.40 arithmetic mean). Though not explicitly shown in the table, using a nonincremental approach is over 30 times worse than analyzing the program at once.

An important observation is that incremental analysis performed when adding rules one by one (Inc), although slower than when performing the analysis nonincrementally all at once (Orig), on average requires significantly less memory (only 69% as much). This indicates that those programs which are too large to be analyzed at once (because the system runs out of memory) could be tackled by splitting them in several parts and performing incremental addition, albeit at the price of increasing analysis time somewhat.
Table IV. Top-Down Incremental Deletion

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Top-down incr.</th>
<th>Non-incr.</th>
<th>Non / Top-down</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>Mem</td>
<td>Time</td>
</tr>
<tr>
<td>aialkl</td>
<td>650</td>
<td>1407</td>
<td>869</td>
</tr>
<tr>
<td>ann</td>
<td>4955</td>
<td>28</td>
<td>21542</td>
</tr>
<tr>
<td>bid</td>
<td>1124</td>
<td>241</td>
<td>3425</td>
</tr>
<tr>
<td>boyer</td>
<td>17478</td>
<td>1400</td>
<td>59731</td>
</tr>
<tr>
<td>browse</td>
<td>529</td>
<td>238</td>
<td>2843</td>
</tr>
<tr>
<td>deriv</td>
<td>230</td>
<td>336</td>
<td>264</td>
</tr>
<tr>
<td>fib</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>grammar</td>
<td>80</td>
<td>39</td>
<td>194</td>
</tr>
<tr>
<td>hanoiapp</td>
<td>149</td>
<td>255</td>
<td>144</td>
</tr>
<tr>
<td>mmatrix</td>
<td>139</td>
<td>178</td>
<td>142</td>
</tr>
<tr>
<td>occur</td>
<td>9</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>peephole</td>
<td>30761</td>
<td>2454</td>
<td>134902</td>
</tr>
<tr>
<td>progeom</td>
<td>19</td>
<td>4</td>
<td>128</td>
</tr>
<tr>
<td>qplan</td>
<td>27380</td>
<td>629</td>
<td>34954</td>
</tr>
<tr>
<td>qsortapp</td>
<td>83</td>
<td>100</td>
<td>173</td>
</tr>
<tr>
<td>query</td>
<td>399</td>
<td>37</td>
<td>818</td>
</tr>
<tr>
<td>rdtk</td>
<td>1300</td>
<td>340</td>
<td>3927</td>
</tr>
<tr>
<td>read</td>
<td>201059</td>
<td>38748</td>
<td>234405</td>
</tr>
<tr>
<td>serialize</td>
<td>114</td>
<td>112</td>
<td>564</td>
</tr>
<tr>
<td>tak</td>
<td>2</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>warplan</td>
<td>18737</td>
<td>2032</td>
<td>25977</td>
</tr>
<tr>
<td>witt</td>
<td>2432</td>
<td>152</td>
<td>12974</td>
</tr>
<tr>
<td>zebra</td>
<td>2040</td>
<td>2021</td>
<td>5233</td>
</tr>
</tbody>
</table>

| Arithmetic mean | 2.98 | 6.42 |
| Weighted mean   | 1.75 | 1.51 |

6.3 Incremental Deletion

In order to test the relative performance of incremental and nonincremental analysis in the context of deletion, we timed the analysis of the same benchmarks where each rule was deleted one by one. Starting from an already analyzed program, the last rule was deleted and the resulting program (re-)analyzed. This process was repeated until no rules were left. This experiment has been performed using a nonincremental approach and using the top-down deletion algorithm of Section 4.2 and the bottom-up deletion algorithm of Section 4.3. Table IV presents the results for top-down deletion and the nonincremental approach, while Table V contains the figures for the bottom-up algorithm together with a comparison of the two incremental deletion algorithms.

The results are also very encouraging. The improvement of analyzing rule by rule in an incremental fashion gave an average speedup with respect to the nonincremental algorithm of 1.75 for the top-down deletion algorithm and 7.30 for the bottom-up deletion algorithm. The results favor the bottom-up algorithm in this experiment, as shown in the comparison column in Table V, where bottom-up deletion is 4.16 times faster than top-down deletion.

Examining memory usage, we see that both the top-down and bottom-up deletion algorithms require less memory than the nonincremental approach (by a factor of 1.51 and 1.40 respectively). Between the two incremental deletion algorithms, top-
Table V. Bottom-Up Incremental Deletion

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Bottom-up incr.</th>
<th>Non / Bottom-up</th>
<th>Top-down / Bottom-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>Mem</td>
<td>Time</td>
</tr>
<tr>
<td>aiakl</td>
<td>514</td>
<td>1110</td>
<td>1.69</td>
</tr>
<tr>
<td>ann</td>
<td>979</td>
<td>469</td>
<td>22.00</td>
</tr>
<tr>
<td>bid</td>
<td>331</td>
<td>138</td>
<td>10.35</td>
</tr>
<tr>
<td>boyer</td>
<td>4535</td>
<td>1372</td>
<td>13.17</td>
</tr>
<tr>
<td>browse</td>
<td>196</td>
<td>155</td>
<td>14.51</td>
</tr>
<tr>
<td>deriv</td>
<td>230</td>
<td>347</td>
<td>1.15</td>
</tr>
<tr>
<td>fib</td>
<td>1</td>
<td>4</td>
<td>4.00</td>
</tr>
<tr>
<td>grammar</td>
<td>47</td>
<td>35</td>
<td>4.13</td>
</tr>
<tr>
<td>hanoiapp</td>
<td>147</td>
<td>248</td>
<td>0.98</td>
</tr>
<tr>
<td>mmatrix</td>
<td>39</td>
<td>57</td>
<td>3.64</td>
</tr>
<tr>
<td>occur</td>
<td>17</td>
<td>18</td>
<td>1.59</td>
</tr>
<tr>
<td>peephole</td>
<td>8754</td>
<td>2562</td>
<td>15.41</td>
</tr>
<tr>
<td>progeom</td>
<td>14</td>
<td>5</td>
<td>9.14</td>
</tr>
<tr>
<td>qplan</td>
<td>623</td>
<td>392</td>
<td>56.11</td>
</tr>
<tr>
<td>qsortapp</td>
<td>57</td>
<td>64</td>
<td>3.04</td>
</tr>
<tr>
<td>query</td>
<td>110</td>
<td>13</td>
<td>7.44</td>
</tr>
<tr>
<td>rdtok</td>
<td>301</td>
<td>224</td>
<td>13.05</td>
</tr>
<tr>
<td>read</td>
<td>55517</td>
<td>45162</td>
<td>4.22</td>
</tr>
<tr>
<td>serialize</td>
<td>84</td>
<td>122</td>
<td>6.71</td>
</tr>
<tr>
<td>tak</td>
<td>6</td>
<td>18</td>
<td>1.17</td>
</tr>
<tr>
<td>warplan</td>
<td>1135</td>
<td>1437</td>
<td>22.89</td>
</tr>
<tr>
<td>witt</td>
<td>528</td>
<td>446</td>
<td>24.57</td>
</tr>
<tr>
<td>zebra</td>
<td>259</td>
<td>707</td>
<td>20.20</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>11.35</td>
<td>4.82</td>
<td>5.11</td>
</tr>
<tr>
<td>Weighted mean</td>
<td>7.30</td>
<td>1.40</td>
<td>4.16</td>
</tr>
</tbody>
</table>

down deletion, though slower, requires somewhat less memory than the bottom-up algorithm. In our experiments, top-down requires (only) 92% of the amount of memory used by the bottom-up algorithm. As for the incremental addition experiment, we take as memory usage for the nonincremental approach the maximum amount of memory when analyzing rules 1 to \(i\) of the benchmark with \(i \in \{1\ldots n\}\) where \(n\) is the number of rules in that benchmark. Note that we start with the program already analyzed, and thus analysis of the complete program (i.e., with rules 1–\(n\)) is not considered.

6.4 Local Change

Although we have implemented it, we do not report explicitly on the performance of arbitrary change because of the difficulty in modeling in a meaningful way the types of changes that are likely to occur in the circumstances in which this type of change occurs (as, for example, during an interactive program development session). We have studied, however, the case of local change in a context in which it occurs in a way which is more amenable to systematic study: within the &-Prolog compiler. The &-Prolog system is capable of executing in parallel goals which are independent [Conery and Kibler 1981; Hermenegildo and Rossi 1995]. The &-Prolog compiler includes an automatic parallelizer which replaces some conjunctions of literals in

---

5Currently also integrated into the Ciao System preprocessor.
the body of rules which are possibly independent with parallel expressions [Bueno et al. 1999b]. The conditions for independence can often be proved at compile-time by the use of global analysis [Bueno et al. 1999a]. However, there are cases in which run-time tests have to be introduced in the program in order to ensure independence. If the run-time tests fail, the goals are executed sequentially. Though the system does not introduce redundant tests, the parallelized programs can be further optimized if a *multivariant* analysis (of which both PL Al and its incremental versions are capable) is performed on such programs. The additional optimization (e.g., recursive invariant extraction) is possible because more precise information can be propagated beyond the program points in which tests have been introduced [Puebla and Hermenegildo 1995; 1999]. The interesting point in our context is that in this process the program is analyzed once, parallelized (introducing conditionals), and then reanalyzed before performing the multiple specialization step, and that between these two analysis phases changes are made to the program that correspond to our local change scenario.

We have applied the process to our set of benchmarks, and the results are shown in Table VI. Benchmarks for which no run-time tests were added in the parallelized program during the first step of the process have been omitted, since the specialization step is not performed for them, and no reanalysis is needed in that case. The calling patterns S used for these tests were just program entry points with the most general calling pattern, i.e., no information about constraints in calls. This represents the likely situation where the user provides little information to the analyzer. The case is also interesting in our context because it produces more run-time tests and thus more specializations and reanalyses, which allows us to study more benchmarks (note that if very precise information is given by the user

<table>
<thead>
<tr>
<th>Benchm</th>
<th>CGEs</th>
<th>Incremental</th>
<th>Non-Incr.</th>
<th>Non / Inc</th>
<th>Non/(Orig+Inc)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Time Mem</td>
<td>Time Mem</td>
<td>Time Mem</td>
<td>Time Mem</td>
</tr>
<tr>
<td>aiakl</td>
<td>2</td>
<td>97 347</td>
<td>823 3404</td>
<td>8.48 9.81</td>
<td>0.85 1.02</td>
</tr>
<tr>
<td>ann</td>
<td>12</td>
<td>739 2605</td>
<td>2035 8193</td>
<td>2.75 3.15</td>
<td>0.85 1.24</td>
</tr>
<tr>
<td>bid</td>
<td>6</td>
<td>111 363</td>
<td>257 948</td>
<td>2.32 2.61</td>
<td>0.89 1.27</td>
</tr>
<tr>
<td>boyer</td>
<td>2</td>
<td>181 710</td>
<td>975 3876</td>
<td>5.39 5.46</td>
<td>1.29 1.68</td>
</tr>
<tr>
<td>browse</td>
<td>4</td>
<td>93 331</td>
<td>179 697</td>
<td>1.92 2.11</td>
<td>0.99 1.96</td>
</tr>
<tr>
<td>deriv</td>
<td>4</td>
<td>211 821</td>
<td>243 980</td>
<td>1.15 1.19</td>
<td>0.77 1.16</td>
</tr>
<tr>
<td>hanoiapp</td>
<td>1</td>
<td>42 187</td>
<td>135 567</td>
<td>3.21 3.03</td>
<td>0.87 1.12</td>
</tr>
<tr>
<td>mmatrix</td>
<td>2</td>
<td>81 337</td>
<td>120 501</td>
<td>1.48 1.49</td>
<td>0.87 1.44</td>
</tr>
<tr>
<td>occur</td>
<td>2</td>
<td>81 328</td>
<td>131 558</td>
<td>1.62 1.70</td>
<td>0.88 1.63</td>
</tr>
<tr>
<td>peephole</td>
<td>2</td>
<td>267 938</td>
<td>1579 6286</td>
<td>5.91 6.70</td>
<td>0.92 1.10</td>
</tr>
<tr>
<td>progeom</td>
<td>1</td>
<td>11 30</td>
<td>47 156</td>
<td>4.27 5.20</td>
<td>0.94 1.04</td>
</tr>
<tr>
<td>qplan</td>
<td>2</td>
<td>51 203</td>
<td>415 1430</td>
<td>8.14 7.04</td>
<td>0.86 1.05</td>
</tr>
<tr>
<td>query</td>
<td>2</td>
<td>13 62</td>
<td>31 79</td>
<td>2.38 1.27</td>
<td>0.97 1.20</td>
</tr>
<tr>
<td>read</td>
<td>1</td>
<td>13 35</td>
<td>10979 38425</td>
<td>844.54 1097.86</td>
<td>1.03 1.00</td>
</tr>
<tr>
<td>serialize</td>
<td>1</td>
<td>10 32</td>
<td>161 666</td>
<td>16.10 20.81</td>
<td>1.03 1.06</td>
</tr>
<tr>
<td>warplan</td>
<td>8</td>
<td>721 2496</td>
<td>4693 17359</td>
<td>6.51 6.95</td>
<td>3.04 5.14</td>
</tr>
<tr>
<td>zebra</td>
<td>1</td>
<td>379 1325</td>
<td>627 2382</td>
<td>1.65 1.80</td>
<td>0.65 1.05</td>
</tr>
</tbody>
</table>

| Arithmetic mean | 53.99 | 69.31 | 1.04 | 1.48 |
| Weighted mean   | 7.56  | 7.76  | 1.12 | 1.28 |
then many benchmarks are parallelized without any run-time tests, and then no specialization—and thus no reanalysis—occurs).

Table VI presents the results for this experiment. The first column shows the number of conditional graph expressions (CGEs), which are parallel expressions “guarded” by run-time tests. It is an indicator of the number of changes introduced during parallelization. The following columns present the cost of performing the second analysis using the incremental algorithm for local change and by simply reanalyzing from scratch. The next column compares the incremental and nonincremental approaches, giving the ratio of nonincremental reanalysis costs to incremental reanalysis costs. The final column compares the cost of analyzing the parallelized program either by a nonincremental analysis of the parallelized program, or by analyzing the original program and then after parallelization, incrementally reanalyzing.

The results of incremental analysis of local change are even more encouraging than the previous ones. The improvement in analysis time and memory usage is very high overall (7.56 and 7.76 respectively). For large programs with little modifications, such as read, the incremental approach is three orders of magnitude better than reanalyzing from scratch. In addition, for a number of benchmarks, in particular boyer and warplan, it is actually less costly both in terms of time and memory to analyze first the original program and then update the analysis information after the modifications introduced by the parallelizer than analyzing the specialized program alone. This is because parallelization has introduced parallel expressions, and thus complicated even further, rules for highly recursive and complex predicates for which analysis requires several iterations until a fixed point is reached. Because the incremental approach already starts with an answer for such predicates which is already a fixed point (which has been computed with a simpler but equivalent version of the predicates), the algorithm for local change requires much less work to complete the analysis than analyzing the entire specialized program from scratch. As a result, analysis plus local change is faster by a factor of 1.12. The results for memory usage are even better (1.28). This is because we do not have to add the memory used in both analyses, as we can reuse the heap space. All the information required by the second (incremental) analysis is asserted. Thus, memory usage is the maximum heap space usage for the two analysis phases plus the sum of the program space used by each phase.

7. RELATED WORK

Surprisingly, there has been little research into incremental analysis for (constraint) logic programs apart from previous work by the authors [Hermenegildo et al. 1995; Puebla and Hermenegildo 1996]. Also, Krall and Berger [1995a; 1995b] define a compiled analysis approach for logic programs that use the Vienna abstract machine model. They briefly mention that their approach can be used for incremental analysis, though the only kind of incremental change they consider is incremental addition. Several researchers have looked at compositional analysis of modules in (constraint) logic programs [Codish et al. 1993; Bossi et al. 1994], but this does not consider incremental analysis at the level of rules. Also, preliminary studies of full modular analysis of (constraint) logic programs [Puebla and Hermenegildo 2000]
indicate that the techniques presented in this article may be very valuable in that context.

There has, however, been considerable research into incremental analysis for other programming paradigms (see for example the bibliography of Ramalingam and Reps [1993]). The need for incremental analysis techniques was realized very early in research on data flow analysis of traditional languages. The first algorithm was proposed by Rosen [1981], and since then a bewildering array of algorithms for the incremental data flow analysis of traditional languages has been suggested.

Generally, these algorithms can be separated into two approaches which reflect the underlying mechanism used to solve the data flow equations. Elimination-based methods use variable elimination, much like in Gauss-Jordan elimination, to solve the equations, while iterative methods find the least fixed point essentially by computing the Kleene sequence. Incremental analysis algorithms based on each of these approaches have been suggested. Those based on elimination include Burke [1990], Carroll and Ryder [1988], and Ryder [1988]; those based on iteration methods include Cooper and Kennedy [1984] and Pollock and Soffa [1989], while a hybrid approach is described in Marlowe and Ryder [1990].

Our algorithms are formulated in terms of the standard top-down abstract interpretation framework for (constraint) logic programs. Like iteration-based data flow analysis algorithms this framework also computes the Kleene sequence in order to find the least fixed point of the recursive equations. Thus our algorithms are most closely related to those using iteration. Early incremental approaches such as Cooper and Kennedy [1984] were based on restarting iteration. That is, the fixed point of the new program’s data flow equations is found by starting iteration from the fixed point of the old program’s data flow equations. This is always safe, but may lead to unnecessary imprecision if the old fixed point is not below the least fixed point of the new equations [Ryder et al. 1988]. Reinitialization approaches such as Pollock and Soffa [1989] improve the accuracy of this technique by reinitializing nodes in the data flow graph to bottom if they are potentially affected by the program change. Thus they are as precise as if the new equations had been analyzed from scratch.

Our first algorithm, that for incremental addition, works by restarting iteration from the old fixed point. However, because the contribution of each rule is lubbed together, if rules are added to the program, the least fixed point will always increase. Thus restarting iteration is guaranteed to be as precise as starting from scratch in the case rules are incrementally added. Of course, this “monotonicity” of rule addition does not apply to traditional programming languages. The top-down deletion algorithm can be viewed as a variant of the reinitialization approach in which dependency arcs are used to keep track of which information will be affected by deletion of a rule. The bottom-up deletion and local change algorithms are to the best of our knowledge novel.

Despite the similarities between our work and previous research into incremental analysis there are a number of important differences. First, other research has

---

5For consistency with the rest of this article, we have reversed the usual data flow terminology so as to accord with abstract interpretation terminology in which the description lattice is ordered by generality, with the most general element at the top of the lattice.
concentrated on traditional programming languages. To our knowledge this is the first article to identify the different types of incremental change which are useful in (constraint) logic program analysis and to give practical algorithms which handle these types of incremental change. Second, our research is formalized in terms of the generic data flow analysis technique, abstract interpretation. This means that our algorithms are usable with any abstract domain approximating constraints. This contrasts to earlier work in which the algorithms applied only to a single type of analysis or at best to a quite restricted class of analyses. Another contribution of the paper is a simple formalization of the nonincremental fixed-point algorithms used in generic analysis engines. We formalize the analysis as a graph traversal and couch the algorithm in terms of priority queues. Different priorities correspond to different traversals of the program analysis graph. This simple formalization greatly facilitates the description of our incremental algorithms and their proofs of correctness. Finally, we have given a detailed empirical evaluation of our algorithms.

8. CONCLUSIONS

We have described extensions to the fixed-point algorithms used in current top-down generic analysis engines for constraint logic programming languages in order to support incremental analysis. We have classified the possible changes to a program into addition, deletion, arbitrary change, and local change, and proposed, for each one of these, algorithms for identifying the parts of the analysis that must be recomputed and for performing the actual recomputation. We have also discussed the potential benefits and drawbacks of these algorithms. Finally, we have presented some experimental results obtained with an implementation of the algorithms in the PLAI generic abstract interpretation framework.

Our empirical evaluation shows that the incremental analysis algorithms have little overhead compared with the standard nonincremental analysis algorithms but offer considerable benefits both in terms of analysis times and memory usage over these algorithms if the program must be reanalyzed after modification. Thus optimizing compilers for constraint logic programming which rely on a repeated source-to-source transformation and reanalysis cycle should make use of incremental algorithms, such as those presented in this article. Indeed we have successfully employed them [Puebla and Hermenegildo 1999] in an automatic parallelizer for logic programs [Bueno et al. 1999b] and an optimizing compiler for CLP(\mathcal{R}) [Kelly et al. 1998a]. Modifying PLAI to support our algorithms for incremental analysis was relatively straightforward. The only real difficulty was the addition of dependency arc tracking, which also enabled handling updated events. We believe it would also be straightforward to modify other analysis engines for (constraint) logic programs in a similar way.

We believe that our work contributes to opening the door to practical, everyday use of global analysis in the compilation of (constraint) logic programs, even in the interactive environment which is often preferred by the users of such systems. Furthermore, while current analyzers can deal correctly with dynamic program modification primitives, this implies having to give up on many optimizations not only for the dynamic predicates themselves but also for any code called from such predicates. The ability to update global information incrementally (and thus with
reduced overhead) down to the level of single rule additions and deletions makes it possible to deal with these primitives in a much more accurate way.

Throughout the article we have concentrated on the analysis of pure CLP programs. There is nothing precluding the use of standard analysis techniques for handling logical and nonlogical built-ins within the incremental framework. Indeed the implementation [Bueno et al. 1996] actually handles (almost) full ISO-Prolog, and many such built-ins occur in the benchmarks. For simplicity we have also ignored the abstract operation of widening [Cousot and Cousot 1979] in our analysis framework. However, it is straightforward to modify the algorithms to include widening of call and answer patterns.

Although we have focussed on top-down goal-dependent analysis of constraint logic programs, our results are also applicable to goal-independent analysis. Again the key is to keep track of dependencies between the head of a rule and the literals in the rule and answers to each atom. It is straightforward to modify the generic algorithm of Figure 2 and our incremental algorithms to define an incremental top-down goal-independent analysis (by restricting the possible calling patterns in the answer table). Similar ideas can also be used to give incremental algorithms for bottom-up goal-independent analysis, which are far simpler, since information flow is only in one direction (upward).

ACKNOWLEDGMENTS
The authors would like to thank María García de la Banda, Harald Søndergaard, and the anonymous referees for useful comments.

REFERENCES


