Efficient Algebraic Effect Handlers for Prolog

AMR HANY SALEH

TOM SCHRIJVERS

KU Leuven, Belgium

Department of Computer Science

(e-mail: {ah.saleh, tom.schrijvers}@cs.kuleuven.be)

submitted 1 January 2003; revised 1 January 2003; accepted 1 January 2003

Abstract

Recent work has provided delimited control for Prolog to dynamically manipulate the program control-flow, and to implement a wide range of control-flow and dataflow effects on top of. Unfortunately, delimited control is a rather primitive language feature that is not easy to use.

As a remedy, this work introduces algebraic effect handlers for Prolog, as a high-level and structured way of defining new side-effects in a modular fashion. We illustrate the expressive power of the feature and provide an implementation by means of elaboration into the delimited control primitives.

The latter add a non-negligible performance overhead when used extensively. To address this issue, we present an optimised compilation approach that combines partial evaluation with dedicated rewrite rules. The rewrite rules are driven by a lightweight effect inference that analyses what effect operations may be called by a goal. We illustrate the effectiveness of this approach on a range of benchmarks.

KEYWORDS: delimited control, algebraic effect handlers, program transformation

1 Introduction

The work of Schrijvers et al. (2013) has introduced delimited control constructs for Prolog. Delimited control is a very powerful means to dynamically manipulate the control-flow of programs that was first explored in the setting of functional programming (Felleisen 1988; Danvy and Filinski 1990). Schrijvers et al. (2013) show its usefulness in Prolog to concisely define implicit state, DCGs and coroutines. More recently, Desouter et al. (2015) have shown that delimited control also concisely captures the control-flow manipulation of tabling.

Unfortunately, there are two prominent downsides to delimited control. Firstly, it is a rather primitive feature that has been likened to the imperative goto, which was labeled harmful for high-level programming by Dijkstra (1968). Secondly, the overhead of delimited control for encoding state-passing features is non-negligible. For example, the delimited control implementation of DCGs is 10 times slower for a tight loop than the traditional implementation.
This paper addresses both issues. In order to provide a high-level structured interface to delimited control, we adapt the algebraic effects and handlers approach to Prolog. Algebraic effects and handlers have been said to relate to delimited control the way structured loops relate to \texttt{goto}. While the structured approach restricts the expressive power, we still show a range of useful applications. Moreover, in exchange for the restricted expressiveness, we provide two benefits. Firstly, multiple handlers can be combined effortlessly to deal with distinct effects, to deal with one effect in terms of another or to customize the behavior of an effect. Secondly, we provide an automated program transformation that eliminates much of the overhead of delimited control. Indeed, compared to the free form of delimited control, the structured approach of effect handlers simplifies the identification of program patterns that can be optimised.

Our specific contributions are as follows:

- We define the syntax of algebraic effects and handlers for Prolog, and provide semantics in terms of an elaboration to the delimited control primitives.
- We illustrate the feature with a number of examples, including DCGs, implicit state and a writer effect.
- We provide a program transformation to eliminate much of the overhead of the elaboration-based implementation. This transformation is formulated in terms of partial evaluation augmented with rewrite rules. These rewrite rules are driven by an effect analysis that characterises which effects may be generated by a goal.
- We have implemented our program analysis and illustrated its effectiveness on a range of benchmarks.

All code of this paper is available at \url{http://github.com/ah-saleh/prologhandlers}.

\section{Algebraic Effect Handlers}

This section introduces our algebraic effect handlers for Prolog.

\subsection{Syntax and Informal Semantics}

We introduce two new syntactic constructs. The \textit{effect operations} are Prolog predicate symbols $op/n$ that are declared as such with the following syntax.

\begin{verbatim}
:- effect op/n.
\end{verbatim}

For instance, we declare operation $c/1$ to consume a token, $\texttt{get}/1$ and $\texttt{put}/1$ to respectively retrieve and overwrite an implicit state, and $\texttt{out}/1$ to output a term.
The handler is a new Prolog goal form that specifies how to interpret effect operations. Its syntax is as follows:

```
handle G₀ with
    op₁(X) → G₁;
    ...
    opₙ(X) → Gₙ;
[finally(G_f)]
[for(P₁ = T₁, ..., Pₙ = Tₙ)]
```

Effect handlers can be thought of as a generalisation of exception handlers, where calling an effect operation corresponds to throwing an exception. The handler "catches" the operations that arise in G₀. Its operation clauses \( op_i(X) \rightarrow G_i \) stipulate that an occurrence of operation \( op_i(X) \) is to be handled by the goal \( G_i \).

Before we explain the optional finally and for clauses, consider a few ways in which the out/1 operation can be handled in hw/0.

```
hw :- out(hello), out(world).
```

In terms of the exception analogy, hw/0 throws two out/1 exceptions. Our first handler intercepts the first out/1 and does nothing.

```
?- handle hw with (out(X) -> true).
true.
```

A more interesting handler prints the argument of out/1.

```
?- handle hw with (out(X) -> writeln(X)).
hello true.
```

Note that only the first out/1 is handled; this aborts the remainder of hw and the second out/1 is never reached. To handle all operations, effect handlers support a feature akin to resumable exceptions: in the lexical scope of \( G_i \), we can call continue to resume the part of the computation after the effect operation (i.e., its continuation). For instance, the next handler resumes the computation after handling the first out/1 operation and intercepts later out/1 operations in the same way.

```
?- handle hw with (out(X) -> writeln(X), continue).
hello world true.
```

Interestingly, we can invoke the same continuation multiple times, for instance both before and after printing the term.

```
?- handle hw with (out(X) -> continue, writeln(X), continue).
world hello world true.
```
The **finally** Clause  
The optional **finally** clause is performed when $G_0$ finishes; if omitted, $G_f$ defaults to **true**.

```prolog
?- handle hw with (out(X) -> writeln(X), continue)
    finally (writeln(done)).
hello
world
done
true.
```

Note that if the goal does not run to completion, the **finally** clause is not invoked.

```prolog
?- handle hw with (out(X) -> writeln(X))
    finally (writeln(done)).
hello
true.
```

The **for** Clause  
All variables in the operation and **finally** clauses are local to that clause, except if they are declared in the **for** clause.\(^1\) Every **Var** = **Term** pair in the **for** clause relates a variable, which we call a parameter, that is in scope of all the operations and **finally** clauses with a term whose variables are in scope in the handler context. For instance, the following handler collects all outputs in a list.

```prolog
?- handle hw with (out(X) -> Lin = [X|Lmid], continue(Lmid,Lout))
    finally (Lin=Lout) for (Lin = List, Lout=[]).
List = [hello,world].
```

Note that **continue** has one argument for each parameter to indicate which values the parameters take in the continuation.

### 2.2 Nested Handlers and Forwarding

Nesting algebraic effect handlers is similar to nesting exception handlers. If an operation is not “caught” by the inner handler, it is **forwarded** to the outer handler. Moreover, if the inner handler catches an operation and, in the process of handling it, raises another operation, then this operation is handled by the outer handler. Let us illustrate both scenarios.

We can easily define a non-deterministic choice operator **or/2** in the style of **TOR** (Schrijvers et al. 2014; Schrijvers et al. 2014) in terms of the primitive **choice/1** effect which returns either of the two boolean values **t** and **f**.

```prolog
:- effect choice/1.

or(G1,G2) :- choice(B), (B = t -> G1 ; B = f -> G2).

chooseAny(G) :- handle G with (choice(B) -> (B = t ; B = f), continue).
```

---

\(^1\) The **for** clause plays a similar role as that in Schimpf’s **logical loops** (Schimpf 2002).
The chooseAny handler interprets choice/1 in terms of Prolog’s built-in disjunction (;)/2.

?- chooseAny(or(X = 1, X = 2)).
    X = 1;
    X = 2.

To obtain more interesting behavior, we can nest this handler with:

flip(G) :- handle G with (choice(B) -> choice(B1), not(B1,B), continue).
not(t,f). not(f,t).

to flip the branches in a goal without touching the goal’s code.

?- chooseAny(flip(or(X = 1, X = 2)));
    X = 2;
    X = 1.

What happens is that the inner flip handler intercepts the choice(B) call of or/2. It produces a new choice(B1) call that reaches the outer chooseAny handler, and unifies B with the negation of B1, which affects the choice in the continue-ation of or/2.

Thanks to forwarding, we can also easily mix different effects. For instance, with:

writeOut(G) :- handle G with (out(T) -> writeln(T), continue).

we can combine output and non-determinism.

?- chooseAny(writeOut(or(out(hello), out(world)))); fail.
   hello
   world
   false.

Note that the inner writeOut handler does not know how to interpret the choice/1 effect. As a consequence, it (implicitly) forwards this operation to the next surrounding handler, chooseAny, who does know what to do.

2.3 Elaboration Semantics

There is a straightforward elaboration of handlers into the shift/1 and reset/3 delimited control primitives for Prolog (Schrijvers et al. 2013). For instance, the last example query of Section 2.1 is elaborated into:

?- handler42(hw, List, []).
   List = [hello,world].

where the declaration of the out/1 operation is elaborated into:

out(X) :- shift(out(X)).

which shifts the term representation of the operation. The actual handler code is elaborated into a predicate (with a fresh name).
This predicate executes the goal in the delimited scope of a \texttt{reset/3}, which captures any \texttt{shift/1} call. If the goal terminates normally (i.e., \texttt{Signal}=0), then the \texttt{finally} code is run. If the goal suspends with a \texttt{shift/1}, the predicate checks whether the operation matches the handler’s operation clause. If so, the clause’s body is run. Note that \texttt{continue(Lmid,Lout)} has been expanded into a recursive invocation of the handler with the actual continuation goal \texttt{Cont}. If the operation does not match, the handler forwards it to the nearest surrounding handler with \texttt{shift/1} and continues with the continuation.

The example above generalizes straightforwardly. Any declaration of an effect operation is elaborated into a predicate definition.

\[
\texttt{:- effect op/n. } \mapsto \texttt{op(X1,...,Xn) :- shift(op(X1,...,Xn)).}
\]

Also every handler goal is substituted with a predicate call.

\[
\texttt{handle } G_0 \texttt{ with }\]
\[
\begin{align*}
\texttt{op}_1 \rightarrow G_1: \\
\vdots \\
\texttt{op}_n \rightarrow G_n \\
\texttt{finally}(G_f) \\
\texttt{for}(P_1 = T_1, \ldots, P_n = T_n)
\end{align*}
\]

where \texttt{h/n + 1} is an auxiliary predicate defined as:

\[
\texttt{h(Goal,P_1,...,P_n) :-}
\begin{align*}
\texttt{reset(Goal,Cont,Signal),} \\
( \texttt{Signal} = 0 \rightarrow G_f) \\
; \texttt{Signal} = \texttt{op}_1 \rightarrow G_1' \\
; \ldots \\
; \texttt{Signal} = \texttt{op}_n \rightarrow G_n' \\
; \texttt{shift(Signal), h(Cont,P_1,...,P_n)}
\end{align*}
\]

Here, each \texttt{G_i'} is derived from \texttt{G_i} by replacing all occurrences of \texttt{continue(S_1,...,S_n)} with recursive calls \texttt{h(Cont,S_1,...,S_n)}. 

\[
\text{handler42(Goal,Lin,Lout) :-}
\begin{align*}
\texttt{reset(Goal,Cont,Signal),} \\
( \texttt{Signal} = 0 \rightarrow \\
\texttt{Lin} = \texttt{Lout}) \\
; \texttt{Signal} = \texttt{out}(X) \rightarrow \\
\texttt{Lin} = [X|\texttt{Lmid}], \\
\texttt{handler42(Cont,\texttt{Lmid},Lout)} \\
; \texttt{shift(Signal),} \\
\texttt{handler42(Cont,Lin,Lout)}
\end{align*}
\]
3 Optimisation

Section 2.3’s elaboration of algebraic effects into the delimited control constructs is conveniently straightforward. Unfortunately, capturing the delimited continuation incurs a non-trivial runtime cost. In many simple cases this cost is quite steep compared to more conventional program transformation approaches. For instance, the implementation of DCGs with delimited control is 10 times slower in a tight loop than the traditional term expansion approach.

Fortunately, the runtime overhead is not inherent in the algebraic effects and handlers approach, and we can obtain competitive performance through optimised compilation. This section presents our optimisation approach, which aims to eliminate most uses of delimited control. The optimisation consists of two collaborating transformation approaches: rewrite rules (Section 3.2) and partial evaluation (Section 3.3). We use term rewrite rules to simplify handler goals and possibly eliminate the handler construct altogether. These rules depend on an effect system (Section 3.1) that infers which effects can or cannot be generated by a goal. Partial evaluation complements the rewrite rules by specialising handled predicate calls. This enables in particular the specialisation of (mutually) recursive predicates.

3.1 Effect System

Driving our optimisation is an effect system that associates with each goal $G$ an effect set $E$ that denotes which effects the goal may call.

Effect Sets In order to cater for modular programs, effect sets $E$ are not elements of the powerset lattice over the closed set $OP$ of locally known effect operation symbols $op/n$. Instead we use the powerset lattice over an open-ended set of effect operations augmented with the additional top element $All$. This allows us to express the effects of unknown goals and unknown effect operations in an abstract manner.

Hence, we denote effect sets in one of two forms: $\bigcup_i op_i/n_i$ or $All - \bigcup_i op_i/n_i$. The former is an explicit enumeration of effect operations, while the latter expresses the dual: all but the given effect operations.

The $\in$ relation as well as the functions $\cup$ and $-$ are extended in the obvious way from the powerset lattice to our augmented version.

Effect System We use these functions over effect sets in the definition of our effect system judgment $E_c \vdash G : E$. This judgment expresses that goal $G$ calls only effect operations from the effect set $E$, provided that $continue$ calls only effect operations from the effect set $E_c$. Since $continue$ is not defined for a top-level goal, we may assume any value for its $E_c$. Hence, for convenience, we always take $E_c = \emptyset$ for top-level goals $G$ and just write $\vdash G : E$.

Figure 1 defines this judgment by means of inference rules. Rule (E-Var) expresses that a variable (i.e., unknown) goal may call $All$ effect operations. Rule (E-Op) states that known effect operation calls itself. Rules (E-Conj) and (E-Disj) combine the effects of their subgoals. Rule (E-True) expresses that the goal $true$,
as an example for other built-ins, is op-free. Rule (E-Cont) captures the invariant that \texttt{continue} has the \( E_c \) effect. In Rule (E-Pred) the effect of a user-defined predicate is the effect of its body. Finally, the most of the complexity of the inference system is concentrated in Rule (E-Handle) that deals with a handler goal. The rule expresses that the handler goal forwards all the effect operations \( E_0 \) of the goal \( G_0 \) it handles, except for the ones that the handler takes care of, \( \bigcup_i \text{op}_i \). In addition, the handler may introduce additional calls to effect operations in its operation and \texttt{finally} clauses. Also note that calls to \texttt{continue} in the operation clauses have exactly the same effect as the handler goal itself; they are essentially recursive calls after all.

Here are a few examples:

\[
\begin{align*}
\vdash \text{hw} : \text{out/1} \\
\vdash \text{handle hw with (out(X) \rightarrow writeln(X))} : \emptyset \\
\vdash \text{handle Y with (out(X) \rightarrow writeln(X))} : \text{All} - \text{out/1}
\end{align*}
\]

### 3.2 Rewrite Rules

We use the information of the effect system and the syntactic structure of goals to perform a number of handler-specific optimisations. We denote these optimisations in terms of semantics-preserving equivalences \( G_1 \equiv G_2 \) that we use as left-to-right rewrite rules. Figure 2 lists our rewrite rules in the form of inference rules where conditions on the inferred effects are written above the bar.

Rule (O-Disj) captures the fact that effect handling is orthogonal to disjunction to specialise the branches of a disjunction separately. There are two rules for con-
Fig. 2: Optimisation Rules for effect handlers

junction. Rule (O-Conj) pulls the first goal \( G_1 \) of a conjunct out of the handler provided that it does not call any of the handler’s operations. This covers both the case where \( G_1 \) is an op-free goal and the case where all the operations in \( G_1 \) are forwarded by the handler. The second rule for conjunction, Rule (O-Op), statically evaluates the special case where the first goal is an operation dealt with by the handler. This consists of three parts: 1) the unification of the formal and actual parameters, 2) the unification of the formal and actual operation arguments, and 3) calling the operation clause’s goal. Note that we substitute all calls to \( \text{continue}(U) \) (for any \( U \)) in this last goal with the second conjunct wrapped in the handler; note
that the arguments $\bar{U}$ become the new actual parameters. In the process we are careful to freshen all the local logical variables that are used.

Rule (O-DROP) removes spurious operation clauses from the handler; it only retains those that correspond to operations that the goal may actually call. In the case that no operation clauses remain, Rule (O-Triv) dispenses with the handler altogether. This amounts to unifying the formal and actual parameters and calling the finally goal.

Finally, the most complex rule of all, Rule (O-Merge), merges two nested handlers into one single handler and thereby eliminates expensive forwarding of operations. At first it might seem trivial to merge two handlers: We simply merge all the components of the two handlers pairwise. There is an obvious simplification to perform in the process: we can drop all outer handler’s operation clauses that overlap with any of the inner handler’s clauses, as the inner handler takes precedence over the outer one.

Yet, there is a further subtle issue that have to be taken account in order to preserve the original semantics. The finally goal $G_{1,f}$ and the operation clause goals $\bar{V}_1$ may call operations that are originally intercepted by the outer handler. We have to make sure that this remains the case. For that reason we adjust those goals to $G'_{1,f}$ and $G'_{1}$ in the merged handler. Let us explain these adjustments for the different forms of operation clause goals $G_{1,i}$ that we consider.

1. The operation goal $G_{1,i}$ is of the form $G_{1,i,a},\text{continue}(\bar{V})$ where $G_{1,i,a}$ does not contain any call to continue. We wrap the initial part of the goal in the outer handler and finally proceed with continue.

$$G'_{1,i} = \left( \begin{array}{l} \text{handle } G_{1,i,a} \text{ with } \bar{V}_2 \rightarrow G_2 \\ \text{finally continue}(\bar{V},P_{2,F}) \text{ for } (P_{2,F},P_{2,F}') \end{array} \right)$$

2. The operation goal $G_{1,i}$ does not contain a call to continue. In this case we wrap the entire goal in the outer handler and make sure to call the outer handler’s final goal.

$$G'_{1,i} = \left( \begin{array}{l} \text{handle } G_{1,i} \text{ with } \bar{V}_2 \rightarrow G_2 \\ \text{finally } G_{2,f} \text{ for } (P_{2,F},P_{2,F}') \end{array} \right)$$

Similarly, we adapt the final goal $G_{1,f}$ to

$$G'_{1,f} = \left( \begin{array}{l} \text{handle } G_{1,f} \text{ with } \bar{V}_2 \rightarrow G_2 \\ \text{finally } G_{2,f} \text{ for } (P_{2,F},P_{2,F}') \end{array} \right)$$

### 3.3 Partial Evaluation

We use a custom partial evaluation approach to expose more optimisation opportunities for the rewrite rules and to deal with recursive predicates. Our partial evaluation is targeted at predicate calls that are handled. Consider the following simple DCG example that checks if a phrase is a succession of the terminals $\text{ab}$:

```prolog
:- effect c/1.
ab.               ab : - c(a), c(b), ab.
```
query(Lin) :-
  handle ab with
  (c(X) -> Lin1=[X|Lmid], continue(Lmid,Lout1))
finally (Lin1 = Lout1) for (Lin1=Lin,Lout1=[]).. Here we abstract the goal handle ab with ... into a fresh predicate (say ab0/2), which makes abstraction of the actual handler parameters. This yields the new definition of query/1:

query(Lin) :- ab0(Lin,[]).

At the same time we unfold the definition of ab/0 in the newly created predicate ab0/2. Because ab/0 has two clauses, this means that ab0/2 bifurcates similarly.

ab0(Lin,Lout) :- ab0(Lin,Lout) :-
  handle true with
  handle (c(a), c(b), ab) with
  (c(X) -> Lin1=[X|Lmid],
   continue(Lmid,Lout1))
finally (Lin1 = Lout1) finally (Lin1 = Lout1)
for (Lin1=Lin,Lout1=Lout). for (Lin1=Lin,Lout1=Lout).

This unfolding exposes new rewriting opportunities. Using the Rules (O-Drop) and (O-Triv), the first clause specialises to Lin1=Lin, Lout1=Lout, Lin1=Lout1. In the second clause, a double use of Rule (O-Op) deals with the c/1 operations. This leaves a recursive invocation of ab/0, wrapped in the handler. Now the partial evaluation kicks in again, realises that this is a variant of the earlier specialisation and ties the knot with a recursive call to ab0/2. After further clean-up of the unifications, we get:

ab0(L,L).
ab0([a,b|Lmid],Lout) :- ab0(Lmid,Lout).

There is no trace of delimited control left. Moreover, this is precisely the tight code that the traditional DCG yields.

4 Evaluation

We evaluate the usefulness of our optimisation approach experimentally on a set of benchmarks. All results were obtained on an Intel Core i7 with 8 GB RAM running hProlog 3.2.38 on Ubuntu 14.04.

The first experiment concerns the ab program of Section 3.3. Table 1 lists the timings (in ms) for different input sizes obtained with three different versions of the program: the traditional DCG implementation (based on SICStus), the elaborated handler implementation and the optimised handler implementation. Clearly, the naive use of delimited control slows the program down by more than an order of magnitude. Fortunately, our optimisation eliminates all uses of delimited control and matches the traditional implementation’s performance.\(^2\)

\(^2\) Thanks to more aggressive inlining it is even slightly faster.
The second experiment considers three scenarios with nested handlers. Table 2 lists the runtime results (in ms) for different input sizes of different versions: the plain elaborated program, the program optimised with only the rewrite rules and the program optimised with both rewrite rules and partial evaluation.

The first benchmark, `state_dcg`, extends the `ab` example with an implicit state that is incremented with every occurrence of `ab` in the input. Because the rewrite rules merge the two handlers in this benchmark, they generate an almost two-fold speed-up. With partial evaluation, the speed-up of around two orders of magnitude is much more dramatic. The main reason is that delimited control is again eliminated.

The second benchmark adds an inner-most dummy handler for an unused `foo/0`. The aim of this benchmark is to assess the cost of forwarding. In the plain elaborated version, we can see there is a significant overhead. Thanks to the rewriting, the three handlers are again merged and most of the overhead of the spurious handler disappears – the only remaining cost is the spurious `foo/0` operation clause. Finally, with partial evaluation, all trace of the `foo/0` is eliminated.

The third benchmark re-implements the calculator example of Dragan et al. (2009) with two handlers, one to manage an implicit stack and one to one for an implicit register. The behavior is similar to the other two benchmarks: merging the handlers roughly halves the runtime and partially evaluating them speeds up the code by two orders of magnitude.

## 5 Related Work

### Language Extensions

Various Prolog language extensions have been proposed in terms of program transformations. Van Roy has proposed Extended DCGs (Roy 1989) to thread multiple named accumulators. Similarly, Ciao Prolog’s structured state threading (Ivanovic et al. 2009) enables different implicit states. Algebraic effects and handlers can easily provide similar functionality.

Schimpf’s logical loops (Schimpf 2002) approach has been very influential on our handler design, in particular regarding the elaboration into recursive predicates and the notions of locally fresh variables and parameters. Of course, both features originate in distinct paradigms: logical loops are inspired by imperative loops, while handlers originate in the functional programming paradigm.

### Control Primitives

Various works have considered extensions of Prolog that enable control-flow manipulation. Before the work of Schrijvers et al. (Schrijvers et al.
Efficient Algebraic Effect Handlers for Prolog

<table>
<thead>
<tr>
<th>Program Name</th>
<th>Input Size</th>
<th>Elaborated</th>
<th>Rewriting</th>
<th>Rewriting + PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>$1 \times 10^9$</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>state</td>
<td>$1 \times 10^4$</td>
<td>20</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>state</td>
<td>$1 \times 10^5$</td>
<td>151</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>state</td>
<td>$1 \times 10^6$</td>
<td>1879</td>
<td>604</td>
<td>37</td>
</tr>
<tr>
<td>state</td>
<td>$2 \times 10^6$</td>
<td>2814</td>
<td>1208</td>
<td>75</td>
</tr>
<tr>
<td>state</td>
<td>$5 \times 10^6$</td>
<td>7919</td>
<td>4348</td>
<td>186</td>
</tr>
<tr>
<td>state</td>
<td>$1 \times 10^7$</td>
<td>29695</td>
<td>18094</td>
<td>375</td>
</tr>
<tr>
<td>state_foo</td>
<td>$1 \times 10^3$</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>state_foo</td>
<td>$1 \times 10^4$</td>
<td>23</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>state_foo</td>
<td>$1 \times 10^5$</td>
<td>358</td>
<td>61</td>
<td>3</td>
</tr>
<tr>
<td>state_foo</td>
<td>$1 \times 10^6$</td>
<td>4666</td>
<td>670</td>
<td>37</td>
</tr>
<tr>
<td>state_foo</td>
<td>$2 \times 10^6$</td>
<td>8777</td>
<td>1350</td>
<td>75</td>
</tr>
<tr>
<td>state_foo</td>
<td>$5 \times 10^6$</td>
<td>30026</td>
<td>4551</td>
<td>186</td>
</tr>
<tr>
<td>calculator</td>
<td>$1 \times 10^3$</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>calculator</td>
<td>$1 \times 10^4$</td>
<td>30</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>calculator</td>
<td>$1 \times 10^5$</td>
<td>307</td>
<td>78</td>
<td>10</td>
</tr>
<tr>
<td>calculator</td>
<td>$1 \times 10^6$</td>
<td>1195</td>
<td>761</td>
<td>57</td>
</tr>
<tr>
<td>calculator</td>
<td>$2 \times 10^6$</td>
<td>3015</td>
<td>1525</td>
<td>110</td>
</tr>
<tr>
<td>calculator</td>
<td>$5 \times 10^6$</td>
<td>12326</td>
<td>6114</td>
<td>247</td>
</tr>
</tbody>
</table>

Table 2. Runtimes of nested-handler benchmarks in ms

In (2013), Tarau and Dahl already allowed the users of BinProlog to access and manipulate the program’s continuation.

Various coroutine-like features have been proposed in the context of Prolog for implementing alternative execution mechanisms, such as constraint logic programming and delay. Nowadays most of these are based on a single primitive concept: attributed variables (Holzbaur 1992). Like delimited control, attributed variables are a very low-level feature that is meant to be used directly, but is often used by library writers as the target for much higher-level declarative features.

Algebraic Effects and Handlers The work in this paper adapts the existing work in the functional programming community on algebraic effects and handlers to Prolog. Both algebraic effects (Plotkin and Power 2002) and handlers (Plotkin and Matija 2013) were first explored at a theoretical level, before giving rise to a whole range of implementations in functional programming languages, such as Eff (Bauer and Pretnar 2015), Multicore OCaml (Dolan et al. 2015) and Haskell (Kammar et al. 2013; Kiselyov et al. 2013; Wu and Schrijvers 2015) to name a few.

Schrijvers et al. (2014) have previously appealed to a functional model of algebraic effects and handlers to derive a Prolog implementation of search heuristics (2014). This paper enables a direct Prolog implementation that avoids this detour.

6 Conclusion

This paper has defined algebraic effects and handlers for Prolog as a high-level alternative to delimited control for implementing custom control-flow and dataflow effects. In order to avoid undue runtime overhead of capturing delimited continuations, we provide an optimised compilation approach based on partial evaluation and rewrite rules. Our experimental evaluation shows that this approach greatly reduces the runtime overhead.
A. H. Saleh and T. Schrijvers

Acknowledgments

This work is partly funded by the Flemish Fund for Scientific Research (FWO). We are grateful to Bart Demoen for his support of hProlog.

References


Appendix A Detailed Optimisation Example Walk-Through

This appendix elaborates the optimisation example of Section 3.3 in more depth. We start from the following program:

```prolog
:- effect c/1.
ab.
ab :- c(a), c(b), ab.
query(Lin) :-
    handle ab with
    (c(X) -> Lin1=[X|Lmid], continue(Lmid,Lout1))
    finally (Lin1 = Lout1)
    for (Lin1=Lin,Lout1=[]).
```

**Step 1** We abstract the goal `handle ab with ...` into a new predicate `ab0/2`. This new predicate takes two arguments: one for every parameter in the handler’s `for` clause. The original call is replaced by a call to the new predicate, supplying the actual parameters of the handler as actual arguments.

```prolog
query(Lin) :- ab0(Lin,[]).
```

The predicate `ab0/2` is a copy of `ab/0`’s definition, with the handler wrapped around each clause’s body.

```prolog
ab0(Lin,Lout) :-
    handle true with
    (c(X) -> Lin1=X[Lmid], continue(Lmid,Lout1))
    finally (Lin1 = Lout1)
    for (Lin1=Lin,Lout1=Lout).
```

```prolog
ab0(Lin,Lout) :-
    handle (c(a), c(b), ab) with
    (c(X) -> Lin1=X[Lmid], continue(Lmid,Lout1))
    finally (Lin1 = Lout1)
    for (Lin1=Lin,Lout1=Lout).
```

**Step 2** The optimiser now applies rewrite rules to the two clauses. In the first clause, Rule (O-DROP) can be applied because the effect system provides the information that the goal `true` has no effects. Hence, we drop the operation clause:

```prolog
ab0(Lin,Lout) :-
    handle true with
    finally (Lin1 = Lout1)
    for (Lin1=Lin,Lout1=Lout).
```

**Step 3** The handler currently handles no operations\(^3\). The optimizer proceeds with applying (O-TRIV):

---

\(^3\) This syntax is only allowed during the compilation process.
ab0(Lin,Lout) :-
    true,
    Lin1 = Lout1,
    Lin1 = Lin,
    Lout1 = Lout.

Step 4 By partially evaluating true and the remaining unifications, the first clause is simplified to:

ab0(L,L).

Step 5 In the second clause the handler's goal starts with the c/1 operation. The optimiser applies (O-Op) to the handler, producing the following code:

ab0(Lin,Lout) :-
    Lin1 = [a|Lmid],
    Lin1 = Lin,
    Lout1 = Lout,
    handle (c(b), ab) with
      (c(X1) -> Lin11=X1|Lmid1, continue(Lmid1,Lout11))
    finally (Lin11 = Lout11)
    for (Lin11=Lmid,Lout11=Lout1).

All the variables in the new handler goal are fresh variables. Observe that the actual arguments in the newly generated for clause are taken from the continue call of the previous handler. This is to ensure the correct state threading of the handler, and to keep the correct semantics of the program.

Step 6 The optimiser re-applies (O-Op) for c(b), generating the following code:

ab0(Lin,Lout) :-
    Lin1 = [a|Lmid],
    Lin1 = Lin,
    Lout1 = Lout,
    Lin11 = [b|Lmid1],
    Lin11 = Lmid,
    Lout11 = Lout1,
    handle (ab) with
      (c(X2) -> Lin12=X1|Lmid2, continue(Lmid2,Lout12))
    finally (Lin12 = Lout12)
    for (Lin12=Lmid1,Lout12=Lout1).

Step 7 The remaining handler goal is now a variant of the original one, which was already abstracted into ab0/2. Therefore, we can replace it with ab0/2.

ab0(Lin,Lout) :-
    Lin1 = [a|Lmid],
Efficient Algebraic Effect Handlers for Prolog

Lin1 = Lin,
Lout1 = Lout,
Lin11 = [b|Lmid1],
Lin11 = Lmid,
Lout11 = Lout1,
ab0(Lmid1,Lout11).

Step 8 The clause now consists of several unifications followed by a tail-recursive call. Partially evaluating the unifications leads to the final optimised code:

\[
\begin{align*}
& \text{ab0([a,b|Lmid1],Lout)} : - \\
& \text{ab0(Lmid1,Lout)}.
\end{align*}
\]

Appendix B State-DCG Handler Example in Detail

This appendix shows the result of optimizing a program that consists of two handlers. We first show the elaboration into delimited control. Then, we show how the original program can be optimised by means of the rewrite rules and partial evaluation.

We use the following program, which was used to generate the results of the first benchmarks in Table 2. As described in Section 4, there are two handlers: one handles the implicit state operations and the other handles the DCG operations.

\[
\begin{align*}
& \text{abinc.} \\
& \text{abinc :- c(a), c(b), get_state(St), St1 is St+1, put_state(St1), abinc.} \\
& \text{state_phrase_handler(Sin,Sout,Lin,Lout) :-} \\
& \quad \text{handle} \\
& \quad \quad \text{(handle abinc} \\
& \quad \quad \quad \quad \text{with} \\
& \quad \quad \quad \quad \quad \text{(get_state(Q) -> Q = Sin1, continue(Sin1,Sout1) \\
& \quad \quad \quad \quad \quad \quad ; put_state(NS) -> continue(NS,Sout1))} \\
& \quad \quad \quad \quad \text{finally} \\
& \quad \quad \quad \quad \quad \text{Sout1 = Sin1} \\
& \quad \quad \quad \quad \text{for} \\
& \quad \quad \quad \quad \quad \text{(Sin1 = Sin, Sout1 = Sout)} \\
& \quad \quad \quad \quad \)} \\
& \quad \quad \quad \quad \text{with} \\
& \quad \quad \quad \quad \quad \text{(c(X) -> Lin1 = [X|Lmid], continue(Lmid,Lout1))} \\
& \quad \quad \quad \quad \text{finally} \\
& \quad \quad \quad \quad \quad \text{Lin1 = Lout1} \\
& \quad \quad \quad \quad \text{for} \\
& \quad \quad \quad \quad \quad \text{(Lin1=Lin, Lout1=Lout).}
\end{align*}
\]

The inner handler’s goal is \text{abinc}, which consumes two elements, \text{a} and \text{b}, by using
the operation \( c/1 \) and then increments the state using the operations \texttt{get state/1} and \texttt{put state/1}.

\[
\text{- state_phase_handler}(0, \text{Sout}, [a,b,a,b,a,b], \text{Lout}).
\]

\[
\begin{align*}
\text{Sout} & = 0 \\
\text{Lout} & = [a,b,a,b,a,b]; \\
\text{Sout} & = 1 \\
\text{Lout} & = [a,b,a,b]; \\
\text{Sout} & = 2 \\
\text{Lout} & = [a,b]; \\
\text{Sout} & = 3 \\
\text{Lout} & = [].
\end{align*}
\]

The immediate elaboration into delimited control yields:

\[
\text{state_phrase_handler}(A, B, C, D) :-
\]

\[
\begin{align*}
\text{handler} \_0(\text{handler} \_1(\text{abinc}, A, B), C, D).
\end{align*}
\]

\[
\text{handler} \_1(A, B, C) :-
\]

\[
\begin{align*}
\text{reset}(A, D, E), \\
( D == 0 -> \\
C = B \\
; E = \text{get state}(E) -> \\
F = B, \\
\text{handler} \_1(D, B, C) \\
; E = \text{put state}(E) -> \\
\text{handler} \_1(D, B, C) \\
; \text{shift}(E), \\
\text{handler} \_1(D, B, C) \\
). \\
\text{handler} \_0(A, B, C) :-
\]

\[
\begin{align*}
\text{reset}(A, D, E), \\
( D == 0 -> \\
B = C \\
; E = \text{c}(E) -> \\
B = [F|G], \\
\text{handler} \_0(D, G, C) \\
; \text{shift}(E), \\
\text{handler} \_0(D, B, C) \\
).
\end{align*}
\]

The predicates \texttt{handler} \_0/3 and \texttt{handler} \_1/3 correspond to the elaborated DCG and state handlers respectively. They follow the semantics described in Section 2.3.

Using the rewrite rules first, yields the following elaborated program instead:

\[
\text{state_phrase_handler}(A, B, C, D) :-
\]

\[
\begin{align*}
\text{handler} \_2(\text{abinc}, A, B, C, D).
\end{align*}
\]

\[
\text{handler} \_2(A, B, C, D, E) :-
\]
reset(A, F, G),
( F == 0 ->
   C = B,
   D = E
 ; G = get_state(H) ->
   H = B,
   handler_2(F, B, C, D, E)
 ; G = put_state(I) ->
   handler_2(F, I, C, D, E)
 ; G = c(J) ->
   D = [J|K],
   handler_2(F, B, C, K, E)
 ; shift(G),
   handler_2(F, B, C, D, E)
).

The two handlers have been merged into one, with the corresponding performance improvement.

When partial evaluation is enabled as well, the optimisation goes one step further and yields the following final program:

state_phrase_handler(A, B, C, D) :-
   abinc0(A, B, C, D).
abinc0(A, A, B, B).
abinc0(A, B, [a,b|C], D) :-
   E is A+1,
   abinc0(E, B, C, D).

Partial evaluation has pushed the handlers into the definition of \texttt{abcinc/0} where the rewrite rules have been able to replace the operations by the corresponding handler clauses. As a consequence, the handlers are eliminated and no delimited control primitives are generated.

**Appendix C Soundness of Rule (O-Disj)**

This appendix proves the soundness of the (O-Disj) rewrite rule. Our proof relies on the elaboration of the handler syntax into delimited control and the corresponding semantics for delimited control given by Schrijvers et al. (2013). This semantics is expressed in terms of a Prolog meta-interpreter that we show in Figure C 1.

We start from the left-hand side of the rewrite rule and turn it into the right-hand side by means of a number of equivalence preserving transformations.

\[
\text{handle } (G_1; G_2) \text{ with } \frac{op \rightarrow G;}{\text{finally } G_f} \text{ for } G_s. \quad (\text{C1})
\]
eval(G) :- eval(G,Signal),
    ( Signal = shift(Term,Cont) ->
      fail
    ; true).

eval(shift(Term),Signal) :- !, Signal = shift(Term, true).

eval(reset(G,Cont,Term),Signal) :- !, eval(G,Signal1),
    ( Signal1 = ok -> Cont = 0, Term = 0
    ; Signal1 = shift(Term, Cont)),
    Signal = ok.

eval((G1,G2),Signal) :- !, eval(G1,Signal1),
    ( Signal1 = ok -> eval(G2,Signal)
    ; Signal1 = shift(Term, Cont),
      Signal = shift(Term, (Cont,G2))).

eval((G1;G2),Signal) :- !, ( eval(G1,Signal) ; eval(G2,Signal)).

eval((C->G1;G2),Signal) :- !, ( eval(C,Signal1) ->
    ( Signal1 = ok -> eval(G1,Signal)
    ; fail )
    ; eval(G2,Signal)).

eval(Goal,Signal) :- built_in_predicate(Goal), !, call(Goal), Signal = ok.

eval(Goal,Signal) :- clause(Goal,Body), eval(Body,Signal).

Fig. C 1: Delimited Control Meta-Interpreter

The elaboration of this handler goal into delimited control yields the following auxiliary predicate:

\[
\begin{align*}
\text{h}(\text{Goal}, P_1, \ldots, P_n) & := \\
\text{reset}(\text{Goal}, \text{Cont}, \text{Term}), \\
& \quad (\text{Term} = 0 \rightarrow G_1) \\
& \quad ; (\text{Term} = \text{op} \rightarrow G) \\
& \quad ; \text{shift(Signal), h(Cont, P_1, \ldots, P_n)}
\end{align*}
\]

Here the variables \( P_i \) are the formal parameters of \( G \). The goal itself is then by definition equivalent to

\[
\begin{align*}
\text{h}((G1;G2), A_1, \ldots, A_n)
\end{align*}
\]  

(C2)

where the \( A_i \) are the actual parameters of \( G \).

This is equivalent to evaluation the goal in the meta-interpreter:

\[
\begin{align*}
eval(\text{h}((G1;G2), A_1, \ldots, A_n))
\end{align*}
\]  

(C3)

We can now unfold the \text{eval/1} call and subsequently unfold the resulting call to the auxiliary predicate \text{eval/2} which selects the last clause. After also evaluating
the call to clause/2 to unfold $h/n + 1$ we get:

$$\text{eval}((\text{reset}((G1;G2),\text{Cont},\text{Term}),
\begin{cases}
\text{Term} == 0 & \rightarrow G_f \\
\text{Term} = op & \rightarrow G \\
\text{shift}(\text{Signal}), h(\text{Cont}, P_1, \ldots, P_n)
\end{cases})
\text{, Signal}
\),
\begin{cases}
\text{Signal} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{fail} ; \text{true}
\end{cases}
\text{(C4)}$$

For the sake of space, we refer to the if-then-else block after the reset/3 call as $<\text{Switches}>$. We can thus abbreviate the above as:

$$\text{eval}((\text{reset}((G1;G2),\text{Cont},\text{Term}), <\text{Switches}>)
\text{, Signal}
\),
\begin{cases}
\text{Signal} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{fail} ; \text{true}
\end{cases}
\text{(C5)}$$

Unfolding eval/2 using the appropriate clause for conjunction, yields:

$$\text{eval}(\text{reset}((G1;G2),\text{Cont},\text{Term}), \text{Signal1}),
\begin{cases}
\text{Signal1} = \text{ok} & \rightarrow \text{eval}(<\text{Switches}>, \text{Signal}) \\
\text{Signal1} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{Signal} = \text{shift}(\text{Term},(\text{Cont},<\text{Switches}>))
\end{cases}
\),
\begin{cases}
\text{Signal} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{fail} ; \text{true}
\end{cases}
\text{(C6)}$$

Now we unfold the first call to eval/2 using the clause for reset/3:

$$\text{eval}((G1;G2), \text{Signal2}),
\begin{cases}
\text{Signal2} = \text{ok} & \rightarrow \text{Cont} = 0, \text{Term} = 0 \\
\text{Signal2} = \text{shift}(\text{Term},\text{Cont})
\end{cases}
\),
\text{Signal1} = \text{ok},
\begin{cases}
\text{Signal1} = \text{ok} & \rightarrow \text{eval}(<\text{Switches}>, \text{Signal}) \\
\text{Signal1} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{Signal} = \text{shift}(\text{Term},(\text{Cont},<\text{Switches}>))
\end{cases}
\),
\begin{cases}
\text{Signal} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{fail} ; \text{true}
\end{cases}
\text{(C7)}$$

Again, we unfold the first call to eval/2 using the clause for disjunction:

$$\begin{cases}
\text{eval}(G1, \text{Signal2}) ; \text{eval}(G2, \text{Signal2})
\end{cases}
\),
\begin{cases}
\text{Signal2} = \text{ok} & \rightarrow \text{Cont} = 0, \text{Term} = 0 \\
\text{Signal2} = \text{shift}(\text{Term},\text{Cont})
\end{cases}
\),
\text{Signal1} = \text{ok},
\begin{cases}
\text{Signal1} = \text{ok} & \rightarrow \text{eval}(<\text{Switches}>, \text{Signal}) \\
\text{Signal1} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{Signal} = \text{shift}(\text{Term},(\text{Cont},<\text{Switches}>))
\end{cases}
\),
\begin{cases}
\text{Signal} = \text{shift}(\text{Term},\text{Cont}) & \rightarrow \text{fail} ; \text{true}
\end{cases}
\text{(C8)}$$
We now distribute what comes after the first disjunction into both branches.

\[
\begin{align*}
&\text{eval}(G1, Signal2), \quad \text{eval}(G2, Signal2), \\
&\quad \text{Signal2} = \text{ok} \rightarrow \text{Cont} = 0, \text{Term} = 0 \\
&\quad \text{Signal2} = \text{shift}((\text{Term}, \text{Cont})) \\
&\quad \text{Signal1} = \text{ok}, \quad \text{Signal1} = \text{ok} \rightarrow \text{eval}(<\text{Switches}>, \text{Signal}) \\
&\quad \text{Signal1} = \text{shift}((\text{Term}, \text{Cont})) \rightarrow \text{Signal} = \text{shift}((\text{Term}, (\text{Cont}, <\text{Switches})))) \\
&\quad \text{Signal} = \text{shift}((\text{Term}, \text{Cont})) \rightarrow \text{fail} \text{ ; true} \\
&\quad \text{Signal} = \text{shift}((\text{Term}, \text{Cont})) \rightarrow \text{fail} \text{ ; true} \\
\end{align*}
\]

At this point we change gear and start folding again. First we fold the reset/2 clause of eval/2 twice, once in each branch.

\[
\begin{align*}
&\text{eval}(\text{reset}(G1, \text{Term}, \text{Cont}), \text{Signal1}), \\
&\quad \text{Signal1} = \text{ok} \rightarrow \text{eval}(<\text{Switches}>, \text{Signal}) \\
&\quad \text{Signal1} = \text{shift}((\text{Term}, \text{Cont})) \rightarrow \text{Signal} = \text{shift}((\text{Term}, (\text{Cont}, <\text{Switches})))) \\
&\quad \text{Signal} = \text{shift}((\text{Term}, \text{Cont})) \rightarrow \text{fail} \text{ ; true} \\
&\quad \text{Signal} = \text{shift}((\text{Term}, \text{Cont})) \rightarrow \text{fail} \text{ ; true} \\
\end{align*}
\]
Then we fold the conjunction clause of `eval/2` in each branch.

\[
\begin{align*}
\text{eval}((\text{reset}(G_1, \text{Term}, \text{Cont}), <\text{Switches}>), \text{Signal}), \\
\quad (\text{Signal} = \text{shift}(<\text{Term}, \text{Cont}) \rightarrow \text{false} ; \text{true}) \\
\end{align*}
\]

\[
(C11)
\]

Subsequently, we fold `eval/1` twice.

\[
\begin{align*}
\text{eval}((\text{reset}(G_1, \text{Term}, \text{Cont}), <\text{Switches}>)) \\
\end{align*}
\]

\[
(C12)
\]

Now we can drop the meta-interpretation layer again.

\[
\begin{align*}
(\text{reset}(G_1, \text{Term}, \text{Cont}), <\text{Switches}>) \\
\end{align*}
\]

\[
(C13)
\]

Then we fold \( h/n + 1 \) twice.

\[
( h(G_1, A_1, \ldots, A_n); h(G_2, A_1, \ldots, A_n))
\]

\[
(C14)
\]

Finally, we invert the elaboration to obtain the right-hand side of the rewrite rule.

\[
\begin{align*}
\text{handle } G_1 \text{ with } \text{op} \rightarrow C; \\
\text{finally}(G_f) ; \\
\text{for}(G_s) \\
& \quad \text{handle } G_2 \text{ with } \text{op} \rightarrow C; \\
\text{finally}(G_f) \\
& \quad \text{for}(G_s)
\end{align*}
\]

\[
(C15)
\]