Lightweight Lemmas in λProlog [1]

Andrew W. Appel
Bell Labs and Princeton University
appel@princeton.edu

Amy P. Felty
Bell Labs, 600 Mountain Avenue, Murray Hill, NJ 07974, USA
felty@research.bell-labs.com

Abstract

λProlog is known to be well-suited for expressing and implementing logics and inference systems. We show that lemmas and definitions in such logics can be implemented with a great economy of expression. We encode a polymorphic higher-order logic using the ML-style polymorphism of λProlog. The terms of the metalanguage (λProlog) can be used to express the statement of a lemma, and metalanguage type-checking can directly type-check the lemma. But to allow polymorphic lemmas requires either more general polymorphism at the meta-level or a less concise encoding of the object logic. We discuss both the Terzo and Teyjus implementations of λProlog as well as related systems such as Elf.

1 Introduction

It has long been the goal of mathematicians to minimize the set of assumptions and axioms in their systems. Implementers of theorem provers use this principle: they use a logic with as few inference rules as possible, and prove lemmas outside the core logic in preference to adding new inference rules. In applications of logic to computer security – such as *proof-carrying code* [12] and distributed authentication frameworks [1] – the implementation of the core logic is inside the trusted code base (TCB), while proofs need not be in the TCB because they can be checked.

Two aspects of the core logic are in the TCB: a set of logical connectives and inference rules, and a program in some underlying programming language that implements proof checking – that is, interpreting the inference rules and matching them against a theorem and its proof.

Definitions and lemmas are essential in constructing proofs of reasonable size and clarity. A proof system should have machinery for checking lemmas, and applying lemmas and definitions, in the checking of proofs. This machinery also is within the TCB. Many theorem-provers support definitions and lemmas and provide a variety of advanced features designed to help with tasks such as organizing definitions and lemmas into libraries, keeping track of dependencies, and providing modularization; in our work we are partic-

[1] In *Proceedings of the 1999 International Conference on Logic Programming, November 1999.*

ularly concerned with separating that part of the machinery necessary for proof checking (i.e., in the TCB) from the programming-environment support that is used in proof development. In this paper we will demonstrate a definition/lemma implementation that is about two dozen lines of code.

The λProlog language [8] has several features that allow concise and clean implementation of logics, proof checkers, and theorem provers [4]. We use λProlog, but many of our ideas should also be applicable in logical frameworks such as Elf/Twelf [14, 17]. An important purpose of this paper is to show which language features allow a small TCB and efficient representation of proofs. We will discuss *higher-order abstract syntax*, *dynamically constructed clauses*, *dynamically constructed goals*, *meta-level formulas as terms*, and *prenex* and *non-prenex polymorphism*.

2 A core logic

The clauses we present use the syntax of the Terzo implementation of λProlog [20]. λProlog is a higher-order logic programming language which extends Prolog in essentially two ways. First, it replaces first-order terms with the more expressive simply-typed λ-terms; λProlog implementations generally extend simple types to include ML-style prenex polymorphism [3, 9], which we use in our implementation. Second, it permits implication and universal quantification (over objects of any type) in goal formulas.

We introduce types and constants using `kind` and `type` declarations, respectively. Capital letters in type declarations denote type variables and are used in polymorphic types. In program goals and clauses, λ-abstraction is written using backslash `\` as an infix operator. Capitalized tokens not bound by λ-abstraction denote free variables. All other unbound tokens denote constants. Universal quantification is written using the constant `pi` in conjunction with a λ-abstraction (e.g., `pi X\` represents universal quantification over variable `X`). The symbols *comma* and `=>` represent conjunction and implication. The symbol `:-` denotes the converse of `=>` and is used to write the top-level implication in clauses. The type `o` is the type of clauses and goals of λProlog. We usually omit universal quantifiers at the top level in definite clauses, and assume implicit quantification over all free variables.

We will use a running example based on a sequent calculus for a higher-order logic. We call this the *object logic* to distinguish it from the *metalogic* implemented by λProlog. We implement a proof checker for this logic that is similar to the one desribed by Felty [4]. We introduce two primitive types: `form` for object-level formulas and `pf` for proofs in the object logic. We introduce constants for the object-level connectives, such as `and` and `imp` of type `form → form → form`, and `forall` of type `(A → form) → form`. We also have `eq` of type `A → A → form` to represent equality at any type. We use infix notation for the binary connectives. The constant `forall` takes a functional argument, and thus object-level binding of variables by quantifiers is defined in terms of meta-level λ-abstraction. This use of higher-order data structures

```
initial proves A :- assume A.
(imp_r Q) proves (A imp B) :- (assume A) => (Q proves B).
(and_l A B Q) proves C :-
   assume (A and B), (assume A) => (assume B) => (Q proves C).
(forall_r Q) proves (forall A) :- pi y\ ((Q y) proves (A y)).
(cut Q1 Q2 A) proves C :-
   Q1 proves A, (assume A) => (Q2 proves C).
(congr X Z H Q P) proves (H X) :-
   Q proves (eq X Z), P proves (H Z).
refl proves (eq X X).
```

Program 1: Some type declarations and inference rules of the object logic.

is called *higher-order abstract syntax* [16]; with it, we don't need to describe the mechanics of substitution explicitly in the object logic [4]. Program 1 shows λProlog clauses for some of the inference rules. The following two declarations illustrate the types of proof constructors.

```
type  forall_r  (A → pf) → pf.
type  congr     A → A → (A → form) → pf → pf → pf.
```

To implement assumptions (that is, formulas to the left of the sequent arrow) we use implication. The goal A `=>` B adds clause A to the λProlog clause database, evaluates B, and then (upon either the success or failure of B) removes A from the clause database. It is a dynamically scoped version of Prolog's `assert` and `retract`. For example, suppose we use `(imp_r initial)` to prove `((eq x y) imp (eq x y))`; then λProlog will execute the (instantiated) body of the `imp_r` clause:

```
(assume (eq x y))  =>  (initial proves (eq x y))
```

This adds `(assume (eq x y))` to the database; then the subgoal

```
initial proves (eq x y)
```

generates a subgoal `(assume (eq x y))` which matches our dynamically added clause.

We have used λProlog's ML-style prenex polymorphism to reduce the number of inference rules in the TCB. Instead of a different `forall` constructor at each type – and a corresponding pair of inference rules – we have a single polymorphic `forall` constructor. Our full core logic (not shown in this paper) uses a base type `exp` of machine integers, and a type `exp` → `exp` of functions, so if we desire quantification both at expressions and at predicates (let alone functions at several types) we have already saved one constructor and two inference rules.

We have also used polymorphism to define a general congruence rule on the `eq` operator, from which many other desirable facts (transitivity and symmetry of equality, congruence at specific functions) may be proved as lemmas.

Theorem 1 shows the use of our core logic to check a simple proof.

It is important to show that our encoding of higher-order logic in λProlog is *adequate*. To do so, we must show that a formula has a sequent proof if and only if its representation as a term of type `form` has a proof term that can

```
 (forall_r I\ forall_r J\ forall_r K\
  (imp_r (and_l (eq J I) (eq J K)
          (congr I J (X\ (eq X K))
           (congr J I (eq I) initial refl) initial))))
proves
 (forall I\ forall J\ forall K\ (eq J I and eq J K) imp eq I K).
```

Theorem 1. $\forall I\ \forall J\ \forall K\ (J = I \wedge J = K) \rightarrow I = K$.

```
type  lemma  (A → o) → A → (A → pf) → pf.

(lemma Inference Proof Rest) proves C :-
 pi Name\ (valid_clause (Inference Name),
           Inference Proof,
           (Inference Name) => ((Rest Name) proves C)).
```

Program 2: The `lemma` proof constructor.

be checked using the inference rules of Program 1. Proving such a theorem should be straightforward. In particular, since we have encoded our logic using prenex polymorphism, we can expand out instantiated copies of all of the polymorphic expressions in terms of type `pf`; the expanded proof terms will then map directly to sequent proof trees. Although we do not discuss it, it should be easy to extend such an adequacy proof to account for the extensions to this core logic that we discuss in the rest of the paper.

3 Lemmas

In mathematics the use of lemmas can make a proof more readable by structuring the proof, especially when the lemma corresponds to some intuitive property. For automated proof checking (in contrast to automated or traditional theorem proving) this use of lemmas is not essential, because the computer doesn't need to understand the proof in order to check it. But lemmas can also reduce the *size* of a proof (and therefore the time required for proof checking): when a lemma is used multiple times it acts as a kind of "subroutine." This is particularly important in applications like proof-carrying code where proofs are transmitted over networks to clients who check them.

The heart of our lemma mechanism is the clause shown in Program 2. The proof constructor `lemma` takes three arguments: (1) a derived inference rule `Inference` (of type `A → o`) parameterized by a proof constructor (of type `A`), (2) a term of type `A` representing a proof of the lemma built from core-logic proof constructors (or using other lemmas), and (3) a proof of the main theorem `C` that is parameterized by a proof constructor (of type `A`).

For example, we can prove a lemma about the symmetry of equality; the proof uses congruence and reflexivity of equality:

```
pi A\ pi B\ pi P\ (P proves (eq B A) =>
                   ((congr B A (eq A) P refl) proves (eq A B))).
```

This theorem can be checked as a successful λProlog query to our proof

```
(lemma
  (Symmx\ pi A\ pi B\ pi P\
     (Symmx A B P) proves (eq A B) :- P proves (eq B A))
  (A\B\P\(congr B A (eq A) P refl))
  (symmx\ (forall_r I\ forall_r J\ imp_r (symmx J I initial))))
proves (forall I\ forall J\ eq I J imp eq J I).
```

Theorem 2. $\forall I\, \forall J\, (I = J \rightarrow J = I)$.

checker: for an arbitrary `P`, add `(P proves (eq B A))` to the logic, then check the proof of congruence using this fact. The syntax F `=>` G means exactly the same as G `:-` F, so we could just as well write this query as:

```
pi A\ pi B\ pi P\ ((congr B A (eq A) P refl) proves (eq A B) :-
                   P proves (eq B A)).
```

Now, suppose we abstract the proof (roughly, `congr B A (eq A) P refl`) from this query:

```
(Inference = (PCon\ pi A\ pi B\ pi P\
              (PCon A B P) proves (eq A B) :- P proves (eq B A)),
 Proof = (A\B\P\ congr B A (eq A) P refl),
 Query = (Inference Proof),
 Query)
```

The solution of this query proceeds in four steps: the variable `Inference` is unified with a λ-term; `Proof` is unified with a λ-term; `Query` is unified with the application of `Inference` to `Proof` (which is a term β-equivalent to the query of the previous paragraph), and finally `Query` is solved as a goal (checking the proof of the lemma).

Once we know that the lemma is valid, we make a new λProlog atom `symmx` to stand for its proof, and we prove some other theorem in a context where the clause `(Inference symmx)` is in the clause database; remember that `(Inference symmx)` is β-equivalent to

```
pi A\ pi B\ pi P\ (symmx A B P proves eq A B :- P proves eq B A).
```

This looks remarkably like an inference rule! With this clause in the database, we can use the new proof constructor `symmx` just as if it were primitive.

To "make a new atom" we simply `pi`-bind it. This leads to the recipe for lemmas shown in Program 2 above: first execute `(Inference Proof)` as a query, to check the proof of the lemma itself; then `pi`-bind `Name`, and run `Rest` (which is parameterized on the lemma proof constructor) applied to `Name`. Theorem 2 illustrates the use of the `symmx` lemma. The `symmx` proof constructor is a bit unwieldy, since it requires `A` and `B` as arguments. We can imagine writing a primitive inference rule

```
(symm P) proves (eq A B) :- P proves (eq B A).
```

using the principle that the proof checker doesn't need to be told `A` and `B`, since they can be found in the formula to be proved.

Therefore we add three new proof constructors – `elam`, `extract`, and `extractGoal` – as shown in Program 3. These can be used in the following stereotyped way to extract components of the formula to be proved. First bind variables with `elam`, then match the target formula with `extract`. The-

```
type elam         (A → pf) → pf.
type extract      form → pf → pf.
type extractGoal  o → pf → pf.

(elam Q) proves B :- (Q A) proves B.
(extract B P) proves B :-  P proves B.
(extractGoal G P) proves B :- valid_clause G, G, P proves B.
```

Program 3: Proof constructors for implicit arguments of lemmas.

```
(lemma
  (Symm\ pi A\ pi B\ pi P\
    (Symm P) proves (eq A B) :- P proves (eq B A))
  (P\ elam A\ elam B\ extract (eq A B) (congr B A (eq A) P refl))
  (symm\ (forall_r I\ forall_r J\ imp_r (symm initial))))
proves (forall I\ forall J\ eq I J imp eq J I).
```

Theorem 3. $\forall I\, \forall J\, (I = J \rightarrow J = I)$.

orem 3 is a modification of Theorem 2 that makes use of these constructors. The `extractGoal` asks the checker to run λProlog code to help construct the proof. Of course, if we want proof checking to be finite we must restrict what kinds of λProlog code can be run, and this is accomplished by `valid_clause` (see below). The proof of lemma `def_l` in Section 4 is an example of `extractGoal`.

Of course, we can use one lemma in the proof of another.

Since the type of `(Inference Proof)` is `o`, the lemma `Inference` might conceivably contain any λProlog clause at all, including those that do input/output. Such λProlog code cannot lead to unsoundness – if the resulting proof checks, it is still valid. But there are some contexts where we wish to restrict the kind of program that can be run inside a proof. For example, in a proof-carrying-code system, the code consumer might not want the proof to execute λProlog code that accesses private local resources.

To limit the kind and amount of execution possible in the executable part of a lemma, we introduce the `valid_clause` predicate of type `o → o` (Program 4). A clause is valid if contains `pi`, comma, `:-`, `=>`, `proves`, `assume`, and nothing else. Of course, a `proves` clause contains subexpressions of type `pf` and `form`, and an `assume` clause has a subexpression of type `form`, so all the connectives in proofs and formulas are also permitted. Absent

```
valid_clause (pi C) :-   pi X\ valid_clause (C X).
valid_clause (A,B) :-    valid_clause A, valid_clause B.
valid_clause (A :- B) :- valid_clause A, valid_clause B.
valid_clause (A => B) :- valid_clause A, valid_clause B.
valid_clause (P proves A).
valid_clause (assume A).
```

Program 4: Valid clauses.

from this list are λProlog input/output (such as `print`) and the semicolon (backtracking search).

In principle, we do not need lemmas at all. Instead of the symmetry lemma, we can prove `(forall A\ forall B\ (eq B A imp eq A B))` and then cut it into the proof of a theorem using the ordinary `cut` of sequent calculus. To make use of the fact requires two `forall_l`'s and an `imp_l`. This approach adds undesirable complexity to proofs.

4 Definitions

Definitions are another important mechanism for structuring proofs to increase clarity and reduce size. If some property (of a base-type object, or of a higher-order object such as a predicate) can be expressed as a logical formula, then we can make an abbreviation to stand for that formula.

For example, we can express the fact that f is an associative function by the formula $\forall X\ \forall Y\ \forall Z\ f\,X\,(f\,Y\,Z) = f\,(f\,X\,Y)\,Z$. Putting this formula in λProlog notation and abstracting over f, we get the predicate:

```
F\ forall X\ forall Y\ forall Z\ eq (F X (F Y Z)) (F (F X Y) Z)
```

A definition is just an association of some name with this predicate:

```
eq associative
  (F\ forall X\ forall Y\ forall Z\ eq (F X (F Y Z)) (F (F X Y) Z))
```

To use definitions in proofs we introduce three new proof rules: (1) `define` to bind a λ-term to a name, (2) `def_r` to replace a formula on the right of a sequent arrow with the definition that stands for it (or viewed in terms of backward sequent proof, to replace a defined name with the term it stands for), and (3) `def_l` to expand a definition on the left of a sequent arrow during backward proof. All three of these proof constructors are just lemmas provable in our system using congruence of equality, as Program 5 shows.

To check a proof `(define Formula (Name\ (RestProof Name)))` the system interprets the `pi D` within the `define` lemma to create a new atom `D` to stand for the `Name`. It then adds `(assume(eq D Formula))` to the clause database. Finally it substitutes `D` for `Name` within `RestProof` and checks the resulting proof. If there are occurrences of `(def_r D)` or `(def_l D)` within `(RestProof D)` then they will match the newly added clause.

To check that `(def_r associative (A\ A f) P)` is a proof of the formula `(associative f)` the prover checks that `(A\ A f)(associative)` matches `(associative f)` and that `(assume (eq associative Body))` is in the assumptions for some formula, predicate, or function `Body`. Then it applies `(A\ A f)` to `Body`, obtaining the subgoal `(Body f)`, of which `P` is required to be a proof.

To check that `(def_l associative (A\ A f) P)` proves some formula `D`, the checker first reduces `(A\ A f)(associative)` to `associative f`, and checks that `(assume (associative f))` is among the assumptions in the λProlog database. Then it verifies that `(assume (eq associative Body))` is in the assumption database for some `Body`. Finally the checker introduces

```
(lemma (Define\ pi F\ pi P\ pi B\
          ((Define F P) proves B :-
           pi D\ (assume (eq d F) => (P D) proves B)))
       (F\P\ (cut refl (P F) (eq F F)))
  define\

(lemma (Def_r\ pi Name\ pi B\ pi F\ pi P\
          ((Def_r Name B P) proves (B Name) :-
           assume (eq Name F), P proves (B F)))
       (Name\B\P\ elam F\ (extract (B Name)
        (extractGoal (assume (eq Name F))
         (congr Name F B initial P))))
  def_r\

(lemma (Def_l\ pi Name\ pi B\ pi D\ pi F\ pi Q\
          ((Def_l Name B Q) proves D :- assume (B Name),
           assume (eq Name F), (assume (B F) => Q proves D)))
        (Name\B\Q\ elam F\ (extractGoal (assume (eq Name F))
         (cut (congr F Name B (symm initial) initial) Q (B F))))
  def_l\ ...
```

Program 5: Machinery for definitions.

`(assume (Body f))` into the assumptions and verifies that, under that assumption, `Q` proves `D`.

5 Dynamically constructed clauses and goals

Our technique allows lemmas and definitions to be contained *within* the proof. We do not need to install new "global" lemmas and definitions into the proof checker. The dynamic scoping also means that the lemmas of one proof cannot interfere with the lemmas of another, even if they have the same names. This machinery uses several interesting features of λProlog:

Metalevel formulas as terms. As we have seen, the `symm` lemma

`(Symm\ pi A\ pi B\ pi P\ (Symm P) proves eq A B :- P proves eq B A)`

occurs inside the proofs as an argument to the `lemma` constructor and so is just a data structure (parameterized by `Symm`); it does not "execute" anything, in spite of the fact that it contains the λProlog connectives `:-` and `pi`. This gives us the freedom to write lemmas using the same syntax as we use for writing primitive inference rules.

Dynamically constructed goals. When the clause from Program 2 for the `lemma` proof constructor checks the validity of a lemma by executing the goal `(Inference Proof)`, we are executing a goal that is built from a run-time-constructed data structure. `Inference` will be instantiated with terms such as the one above representing the `symm` lemma. It is only when such a term is applied to its proof and thus appears in "goal position" that it becomes the current subgoal on the execution stack.

```
(lemma Inference Proof Rest) proves C :-
 pi Name\ (valid_clause (Inference Name),
           Inference Proof,
           cl (Inference Name) => ((Rest Name) proves C)).

P proves A :- cl Cl, backchain (P proves A) Cl.

backchain G G.
backchain G (pi D) :- backchain G (D X).
backchain G (A,B) :-  backchain G A; backchain G B.
backchain G (H <<== G1) :-  backchain G H, G1.
backchain G (G1 ==>> H) :-  backchain G H, G1.

(D ==>> G) :- (cl D) => G.
(G <<== D) :- (cl D) => G.
```

Program 6: An interpreter for dynamic clauses.

Dynamically constructed clauses. When, having successfully checked the proof of a lemma, the `lemma` clause executes

`(Inference Name) => ((Rest Name) proves C))`

it is adding a dynamically constructed clause to the λProlog database.

The Teyjus system does not allow `=>` or `:-` to appear in arguments of predicates. It also does not allow variables to appear at the head of the left of an implication. These restrictions come from the theory underlying λProlog [7]; without this restriction, a runtime check is needed to insure that every dynamically created goal is an acceptable one. We now show that it is possible to relax the requirements on dynamically constructed clauses and goals to accommodate Teyjus's restrictions.

We can avoid putting `:-` inside arguments of predicates by writing the lemma as

```
(Symm\ pi A\ pi B\ pi P\
  (Symm P) proves (eq A B) <<== P proves (eq B A))
```

where `<<==` is a new infix operator of type $\mathtt{o} \to \mathtt{o}$ But this, in turn, means that the clause for checking lemmas cannot add `(Inference Name)` as a new clause, since `<<==` has no operational meaning. Instead, Program 6 contains a modifed `lemma` clause that adds the clause `(cl (Inference Name))` where `cl` is a new atomic predicate of type $\mathtt{o} \to \mathtt{o}$. The rest of Program 6 implements an interpreter to handle clauses of the form `(cl A)` and goals of the form `(A <== B)` and `(A ==>> B)`. The use of `cl` is the only modification to the `lemma` clause. The new clause for the `proves` predicate is used for checking nodes in a proof representing lemma applications and illustrates the use of the new atomic clauses. The `(cl Cl)` subgoal looks up the lemmas that have been added one at a time and tries them out via the `backchain` predicate. This predicate processes the clauses in a manner similar to the λProlog language itself. The remaining two clauses are needed in both checking lemmas and in checking the rest of the proof for interpreting the new implication operators when they occur at the top level of a goal.

```
(lemma (Symm\ pi A\ pi B\ pi P\
         (Symm P) proves (eq A B) :- P proves (eq B A))
       (P\ elam A\ elam B\
         (extract (eq A B) (congr B A (eq A) P refl)))
       (symm\ (forall_r f\ forall_r g\ forall_r x\
         (imp_r (imp_r (and_r (symm initial) (symm initial)))))))
proves (forall f\ forall g\ forall x\
        (eq f g) imp (eq (f x) x) imp ((eq g f) and (eq x (f x)))).
```

Theorem 6. $\forall f, g, x. f = g \to f(x) = x \to (g = f \wedge x = f(x))$.

Handling new constants for `:-` and `=>` is easy enough operationally. However, it is an inconvenience for the user, who must use different syntax in lemmas than in inference rules.

6 Meta-level types

In the encoding we have presented, ML-style prenex polymorphism is used in the `forall_r` and `congr` rules of Program 1 and in implementing lemmas as shown in Program 2. We now discuss the limitations of prenex polymorphism for implementing lemmas which are themselves polymorphic; and we discuss ways to overcome these limitations both at the meta-level and at the object level. The `symm` lemma is naturally polymorphic: it should express the idea that $a = 3 \to 3 = a$ (at type int) just as well as $f = \lambda x.3 \to (\lambda x.3) = f$ (at type int $\to$ int). But Theorem 6, which uses `symm` at two different types, fails to type-check in our implementation. When the λProlog type-checker first encounters `symm` as a λ-bound variable, it creates an uninstantiated type metavariable to hold its type. The first use of `symm` unifies this metavariable type variable with the type `T` of `x`, and then the use of `symm` at type `T` $\to$ `T` fails to match. Prohibiting λ-bound variables from being polymorphic is the essence of prenex polymorphism. On the other hand, the proof of Theorem 3 type-checks because `symm` is used at only one type.

We can generalize the prenex polymorphism of the metalanguage by removing the restriction that all type variables are bound at the outermost level and allow such binding to occur anywhere in a type, to obtain the second-order λ-calculus. We start by making the bindings clear in our current version by annotating terms with fully explicit bindings and quantification. The result will not be λProlog code, as type quantification and type binding are not supported in that language. So we will use the standard λProlog `pi` and `\` to quantify and abstract term variables; but we'll use Π and Λ to quantify and abstract type variables, and use *italics* for type arguments and other nonstandard constructs.

```
type congr     ΠT.  T → T → (T → form) → pf → pf → pf.
type forall_r  ΠT.  (T → pf) → pf.

ΠT. pi X: T\ pi Z: T\ pi H: T → form\ pi Q: pf\ pi P: pf\
  (congr T X Z H Q P) proves (H X) :-
    Q proves (eq T X Z), P proves (H Z).
```

```
type  lemma Π T.  (T → o) → T → (T → pf) → pf.

(lemma T Inference Proof Rest) proves C :-
 pi Name:T\ (valid_clause (Inference Name),
             Inference Proof,
             (Inference Name) => ((Rest Name) proves C)).
(lemma  T
        (Symm: Π T.  pf → pf \        ← here!
          Π T.  pi A:T\ pi B:T\ pi P:pf\
            (Symm T P) proves (eq T A B) :- P proves (eq T B A))
        (Λ T.  P:pf\ elam A:T\ elam B:T\
          (extract (eq T A B) (congr T B A (eq T A) P refl)))
        (symm\ (forall_r I:int\ forall_r J:int\
          (imp_r (symm int initial)))))
proves (forall I\ forall J\ (eq int I J) imp (eq int J I)).
```

Figure 7: Explicitly typed version of Theorem 3.

```
Π T.  pi A:  T → form\ pi Q:  T → pf\
  (forall_r T Q) proves (forall T A) :- pi Y:T\ (Q Y proves A Y).
```

Every type quantifier is at the outermost level of its clause; the ML-style prenex polymorphism of λProlog can typecheck this program. However, we run into trouble when we try to write a polymorphic lemma. The lemma itself is prenex polymorphic, but the lemma definer is not.

Figure 7 is pseudo-λProlog in which type quantifiers and type bindings are shown explicitly. The line marked *here* contains a λ-term, λ`Symm`.body, in which the type of `Symm` is Π*T*.`pf` → `pf`. Requiring a function argument to be polymorphic is an example of non-prenex polymorphism, which is permitted in second-order λ-calculus but not in an ML-style type system.

Polymorphic definitions (using `define`) run into the same problems and also require non-prenex polymorphism. Thus prenex polymorphism is sufficient for polymorphic inference rules; non-prenex polymorphism is necessary to directly extend the encoding of our logic to allow polymorphic lemmas, although one can scrape by with monomorphic lemmas by always duplicating each lemma at several different types within the same proof.

There are also several ways to encode our polymorphic logic and allow for polymorphic lemmas without changing the metalanguage. One possibility is to encode object-level types as meta-level terms. The following encoding of the `congr` rule illustrates this approach.

```
kind tp     type.
kind tm     type.
type arrow  tp → tp → tp.
type form   tp.
type eq     tp → tm → tm → tm.
type congr  tp → pf → pf → (A → tm) → A → A → pf.

congr T Q P H X Z proves H X :-
  typecheck X T, typecheck Z T, Q proves (eq T X Z), P proves H Z.
```

This encoding also requires the addition of explicit `app` and `abs` constructors, primitive rules for β- and η-reduction, and typechecking clauses for terms of types `exp` and `form`, but not `pf`. To illustrate, the new constructors and corresponding type checking clauses are given below.

```
type app  tp → tm → tm → tm.
type lam  (tm → tm) → tm.
typecheck (app T1 F X) T2 :-
  typecheck F (arrow T1 T2),  typecheck X T1.
typecheck (lam F) (arrow T1 T2) :-
  pi X\ (typecheck X T1 => typecheck (F X) T2).
```

This encoding loses some economy of expression because of the extra constructors needed for the encoding, and requires a limited amount of typechecking, though not as much as would be required in an untyped framework. For instance, in addition to typechecking subgoals such as the ones in the `congr` rule, it must also be verified that all the terms in a particular sequent to be proved have type `form`. In this encoding, polymorphism at the meta-level is no longer used to encode formulas, although it is still used for the `lemma` constructor. Lemma polymorphism can also be removed by using an application constructor at the level of proofs, though this would require adding typechecking for proofs also.

Another alternative is to use an encoding similar to one by Harper et al. [5] (for a non-polymorphic higher-order logic) in a metalanguage such as Elf/Twelf [14, 17]. The extra expressiveness of dependent types allows object-level types to be expressed more directly as meta-level types, eliminating the need for any typechecking clauses. This encoding still requires explicit constructors for `app` and `abs` as well as primitive rules for $\beta\eta$-reduction. The following Twelf clauses, corresponding to λProlog clauses above, illustrate the use of dependent types for this kind of encoding.

```
tp : type.
tm : tp → type.
form : tp.
pf : tm form → type.
arrow : tp → tp → tp.
eq : {T:tp}tm T → tm T → tm form.
congr : {T:tp}{X:tm T}{Z:tm T}{H:tm T → tm form}
          pf (eq T X Z) → pf (H Z) → pf (H X).
```

Elf [14] and Twelf [17] are both implementations of LF [5], the Edinburgh logical framework. Elf 1.5 has full (nonprenex) statically checked polymorphism with explicit type quantification and explicit type binding, which we have used to implement polymorphic lemmas approximately as shown in Figure 7. But polymorphism in Elf 1.5 is undocumented and discouraged [15], so we recommend the above encoding instead. Twelf is the successor to Elf. Like Elf, it has higher-order data structures with a static type system, but Twelf is monomorphic. Thus, the above encoding is the only possibility.

Both of the above λProlog and Twelf encodings look promising as a basis for a proof system with polymorphic lemmas [2].

7 Other issues

Although we are focusing on the interaction of the meta-level type system with the object-logic lemma system, are other aspects of metalanguage implementation are also relevant to our needs for proof generation and checking.

Type abbreviations In the domain of proof-carrying code, we encode types as predicates which themselves take predicates as arguments. For example, our program has declarations like this one:

```
type hastype (exp → form) → (exp → exp) → exp →
                ((exp → form) → (exp → exp) → exp → form) → form.
```

Neither Terzo nor Teyjus allow such abbreviations and this is rather an inconvenience. ML-style (nongenerative) type abbreviations would be very helpful. In the object-types-as-meta-terms encoding (Section 6), Twelf definitions can act as type abbreviations, which is a great convenience.

Arithmetic. For our application, proof-carrying code, we wish to prove theorems about machine instructions that add, subtract, and multiply; and about load/store instructions that add offsets to registers. Therefore we require some rudimentary integer arithmetic in our logic.

Some logical frameworks have powerful arithmetic primitives, such as the ability to solve linear programs [13] or to handle general arithmetic constraints [6]. For example, Twelf will soon provide a complete theory of the rationals, implemented using linear programming [18]. Some such as Elf 1.5 have no arithmetic at all, forcing us to define integers as sequences of booleans. On the one hand, linear programming is a powerful and general proof technique, but we fear that it might increase the complexity of the trusted computing base. On the other hand, synthesizing arithmetic from scratch is no picnic. The standard Prolog `is` operator seems a good compromise and has been adequate for our needs.

Representing proof terms. Parameterizable data structures with higher-order unification modulo β-equivalence provide an expressive way of representing formulas, predicates, and proofs. We make heavy use of higher-order data structures with both direct sharing and sharing modulo β-reduction. The implementation of the metalanguage must preserve this sharing; otherwise our proof terms will blow up in size.

Any logic programming system is likely to implement sharing of terms obtained by copying multiple pointers to the same subterm. In Terzo, this can be seen as the implementation of a reduction algorithm described by Wadsworth [19]. But we require even more sharing. The similar terms obtained by applying a λ-term to different arguments should retain as much sharing as possible. Therefore some intelligent implementation of higher-order terms within the metalanguage—such as Teyjus's use of explicit substitutions [10, 11]—seems essential.

Programming the prover. In this paper, we have concentrated on an encoding of the logic used for proof checking. But of course, we will also need to construct proofs. For the proof-carrying code application, we need an automatic theorem prover to prove the safety of programs. For implementing this prover, we have found that the Prolog-style control primitives (such as the cut (`!`) operator and the `is` predicate), which are also available in λProlog, are quite important. λProlog also provides an environment for implementing tactic-style interactive provers [4]. This kind of prover is useful for proving the lemmas that are used by the automatic prover. Neither Elf nor Twelf have any control primitives. However, there are plans to add an operator to Twelf similar to Prolog cut [15], which would allow us to implement the automatic prover in the same way as in λProlog. It is not possible to build interactive provers in Elf or Twelf, so proofs of lemmas used by the automatic prover must be constructed by hand.

8 Conclusion

The logical frameworks discussed in this paper are promising vehicles for proof-carrying code, or in general where it is desired to keep the proof checker as small and simple as possible. We have proposed a representation for lemmas and definitions that should help keep proofs small and well-structured, and it appears that each of these frameworks has features that are useful in implementing, or implementing efficiently, our machinery.

Although the lemma system shown in this paper is particularly lightweight and simple to use, its lack of polymorphic definitions and lemmas has led us to further investigate the encodings (sketched in Section 6) that use object-level polymorphic types [2].

Acknowledgements

We thank Robert Harper, Frank Pfenning, Carsten Schürmann for advice about encoding polymorphic logics in a monomorphic dependent-type metalanguage; Robert Harper and Daniel Wang for discussions about untyped systems; Ed Felten, Neophytos Michael, Kedar Swadi, and Daniel Wang for providing user feedback; Gopalan Nadathur and Dale Miller for discussions about λProlog.

References

[1] Andrew W. Appel and Edward W. Felten. Proof-carrying authentication. In *6th ACM Conf. on Computer and Communications Security*, Nov. 1999.

[2] Andrew W. Appel and Amy P. Felty. Polymorphic lemmas in LF and λProlog. In preparation, 1999.

[3] Luis Damas and Robin Milner. Principal type-schemes for functional programs. In *Ninth ACM Symposium on Principles of Programming Languages*, pages 207–12, New York, 1982. ACM Press.

[4] Amy Felty. Implementing tactics and tacticals in a higher-order logic programming language. *J. Automated Reasoning*, 11(1):43–81, August 1993.

[5] Robert Harper, Furio Honsell, and Gordon Plotkin. A framework for defining logics. *Journal of the ACM*, January 1993.

[6] Joxan Jaffar and Jean-Louis Lassez. Constraint logic programming. In *Proceedings of the SIGACT-SIGPLAN Symposium on Principles of Programming Languages*, pages 111–119. ACM, January 1987.

[7] Dale Miller, Gopalan Nadathur, Frank Pfenning, and Andre Scedrov. Uniform proofs as a foundation for logic programming. *Annals of Pure and Applied Logic*, 51:125–157, 1991.

[8] Gopalan Nadathur and Dale Miller. An overview of λProlog. In K. Bowen and R. Kowalski, editors, *Fifth International Conference and Symposium on Logic Programming*. MIT Press, 1988.

[9] Gopalan Nadathur and Frank Pfenning. *The Type System of a Higher-Order Logic Programming Language*, pages 243–283. MIT Press, 1992.

[10] Gopalan Nadathur and Debra Sue Wilson. A representation of lambda terms suitable for operations on their intensions. In *Proc. 1990 ACM Conf. on Lisp and Functional Programming*, pages 341–348. ACM Press, 1990.

[11] Gopalan Nadathur and Debra Sue Wilson. A notation for lambda terms: A generalization of environments. *Theoretical Computer Science*, 198(1-2):49–98, 1998.

[12] George Necula. Proof-carrying code. In *24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 106–119, New York, January 1997. ACM Press.

[13] George Ciprian Necula. *Compiling with Proofs*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, September 1998.

[14] Frank Pfenning. Logic programming in the LF logical framework. In Gérard Huet and Gordon Plotkin, editors, *Logical Frameworks*, pages 149–181. Cambridge University Press, 1991.

[15] Frank Pfenning. personal communication, June 1999.

[16] Frank Pfenning and Conal Elliot. Higher-order abstract syntax. In *Proceedings of the ACM-SIGPLAN Conference on Programming Language Design and Implementation*, pages 199–208, 1988.

[17] Frank Pfenning and Carsten Schürmann. System description: Twelf — a meta-logical framework for deductive systems. In *The 16th International Conference on Automated Deduction*. Springer-Verlag, July 1999.

[18] Roberto Virga. Twelf(X): Extending Twelf to rationals and beyond. In preparation, 1999.

[19] C. P. Wadsworth. *Semantics and Pragmatics of the Lambda Calculus*. PhD thesis, Oxford University, 1971.

[20] Philip Wickline. The Terzo implementation of λProlog. http://www.cse.psu.edu/~dale/lProlog/terzo/index.html, 1999.