ÉCOLE POLYTECHNIQUE
PROMOTION X2011
Férey Gaspard

RAPPORT DE STAGE D'OPTION SCIENTIFIQUE

Titre

NON CONFIDENTIEL

Option :	INFORMATIQUE
Champ de l'option :	Math-Informatique
Directeur de l'option :	Olivier Bournez
Directeur de stage :	Olivier Bournez
Dates du stage :	7 avril - 22 aout 2014
Nom et adresse de l'organisme :	SRI International Computer Science Laboratory (CSL) 333 Ravenswood Avenue Menlo Park, CA 94025-3493 United States

July 9, 2014

Contents

1 Introduction

2 SRI

2.1 The HACMS Project

2.2 PVS

2.3 Translating PVS

2.3.1 Parsing and typechecking PVS

These two task we leave to PVS native parser and typechecker.

The parser generates objects representing the expressions of the theory.

We only convert a subset of PVS. This subset is defined by a subset of expression objects we can translate. The objective is, of course, to be able to translate the maximum of (if not all) PVS expression objects.

2.3.2 Other translator

- Common Lisp (native) - Clean - Yices

3 Translating PVS Syntax

3.1 PVS Syntax

We describe here the syntax of PVS and the objects system used to represent them in Lisp. Some slots of the classes are voluntarily omitted. For a full description of PVS parser representation, refer to [5].

Expr ::= *Number*
| *Name*
| *Expr* *Arguments*
| *Expr* *Binop* *Expr*
| *Unaryop* *Expr*
| *Expr* ` { *Id* | *Number* }
| (*Expr*$^{+}_{,}$)
| (# *Assignment*$^{+}_{,}$ #)
| *IfExpr*
| `LET` *LetBinding*$^{+}_{,}$ `IN` *Expr*
| *Expr* `WHERE` *LetBinding*$^{+}_{,}$
| *Expr* `WITH` [*Assignment*$^{+}_{,}$]
|

Number ::= *Digit*$^{+}$

Id ::= *Letter* *IdChar*$^{+}$

IdChar ::= *Letter* | *Digit*

Letter ::= `A` | ... | `Z`

Digit ::= `0` | ... | `9`

Arguments ::= (*Expr*$^{+}_{,}$)

IfExpr ::= `IF` *Expr* `THEN` *Expr*
{ `ELSIF` *Expr* `THEN` *Expr* } * `ELSE` *Expr* `ENDIF`

Name ::= `true` | `false` | `number_field_pred` | `real_pred`
| `integer_pred` | `integer?` | `rational_pred`
| `floor` | `ceiling` | `rem` | `ndiv` | `even?` | `odd?`
| `cons` | `car` | `cdr` | `cons?` | `null` | `null?`
| `restrict` | `length` | `member` | `nth` | `append` | `reverse`

Binop ::= `=` | `\=` | `OR` | `\/` | `AND` | `&` | `/\`
| `IMPLIES` | `=>` | `WHEN` | `IFF` | `<=>`
| `+` | `-` | `*` | `/` | `<` | `<=` | `>` | `>=`

Unaryop ::= `NOT` | `-`

Assignment ::= *AssignArg*$^{+}$ { `:=` | `|->` } *Expr*

AssignArg ::= (*Expr*$^{+}_{,}$)
| ` *Id*
| ` *Number*

LetBinding ::= { *LetBind* | (*LetBind*$^{+}_{,}$) } `=` *Expr*

LetBind ::= *Id* [`:` *TypeExpr*]

`expr` ⊂ `syntax` [*abstract class*]
type the type of the expression

`name` ⊂ `syntax` [*mixin class*]
id .... the identifier
actuals .... a list of actual parameters
resolutions singleton

This is a mixin for names, i.e., `name-exprs`, `type-names`, etc.

`name-expr` ⊂ `name` `expr` *[class]*

`number-expr` ⊂ `expr` *[class]*

number a nonnegative integer

`tuple-expr` ⊂ `expr` *[class]*

exprs a list of expressions

`application` ⊂ `expr` *[class]*

operator an expr

argument an expr (maybe a tuple-expr)

`field-application` ⊂ `expr` *[class]*

id identifier

actuals a list of actuals

argument the argument

A field application is the internal representation for record extraction, e.g., `r‘a`

`record-expr` ⊂ `expr` *[class]*

assignments a list of assignments

`lambda-expr` ⊂ `binding-expr` *[class]*

This is the subclass of `binding-expr` used for LAMBDA expressions.

`if-expr` ⊂ `application` *[class]*

When an application has an operator that resolves to the `if_def` it is changed to this class.

`update-expr` ⊂ `expr` *[class]*

expression an expr

assignments a list of assignments

An update expression of the form `e WITH [x := 1, y := 2]`, maps to an `update-expr` instance, where the `expression` is `e`, and the `assignments` slot is set to the list of generated `assignment` instances. Note that these are very succinct representations, but correspondingly difficult to typecheck or to translate to other systems (e.g., decision procedures). See the description of the function `translate-update-to-if!` for more details.

`assignment` ⊂ `syntax` *[class]*

arguments the list of arguments

expression the value expression

Assignments occur in both record-exprs and update-exprs. The `arguments` form is a list of lists. For example, given the assignment `‘a(x, y)‘1 := 0`, the `arguments` are `((a) (x y) (1))` and the `expression` is `0`.

3.2 Translator architecture

Describe here the Lisp functions and data structures

Skeleton

Expected input

Output objects

Assertions that we (try to) maintain
Main steps :

- Typechecking: The PVS typechecker perform a type analysis on the PVS code to associate a PVS type to each expression. This might generates some proof obligations (TCC).
- Lexical and syntactic analysis: The PVS parser transforms PVS code into a Lisp internal representation.
- Translation: The translator generates a different representation from PVS expressions and functions declarations. Typically, an expression e is translated into a tuple of four elements (t, n, i, d), where t represents a C type used to describe the expression, n is a string representing the expression, i is a list of instructions supposed to be executed prior to using n (initialisation of n) and d is a list of instructions to be executed when n isn't needed anymore (destruction of n).
- Optimizations
- Code generation: C code is generated.

We first define a function T to translate an expression e.

$$T(e) = (\ T^t(e)\ ,\ T^n(e)\ ,\ T^i(e)\ ,\ T^d(e)\ )$$

$$\begin{aligned}
T(\texttt{2}) &= (\ \texttt{int}, \text{"2"}, [], []) \\
T(\texttt{4294967296}) &= (\ \texttt{mpz_t},\ ?\ , \\
&\quad [\texttt{mpz_init(?);}\ | \\
&\quad\ \texttt{mpz_set_str(?, "4294967296");}], \\
&\quad [\texttt{mpz_clear(?);}]) \\
T(\texttt{lambda(x:below(10)):x}) &= (\ \texttt{int*},\ ?\ , \\
&\quad [\texttt{? = malloc(10 * sizeof(int));}\ | \\
&\quad\ \texttt{int i;}| \\
&\quad\ \texttt{for(i = 0; i < 10; i++)} \\
&\quad\quad \texttt{?[i] = i;}] \\
&\quad [\texttt{free(?);}])
\end{aligned}$$

Figure 1: Translation examples: number expressions

It may occur that $T^n(e) = ?$. In that case, the symbol ? appearing in $T^i(e)$ and $T^d(e)$ needs to be replaced by a proper variable name.

We then define two other operators:

- R wich take an expression and a type and may add an extra conversion in the instructions to make sure its result has the expected type. Also the result of this function has a proper name.
- S which take an expression, a type and a name. It makes sure that the given variable (type + name) is set to a value representing the expression.

3.3 A few translation rules

Translation rules :

```
number-expr "2"
(C-int, "2", [], [])

number-expr "12315468453213"
(C-mpz, nil,
        [mpz_t ~a; | mpz_t_init("12315468453213"); ],
        [mpz_clear ~a;])

application "f(e1, e2)"
(C-mpz, nil,
       [ instr(e1) | instr(e2)
                   | mpz(~a); | f(~a, e1, e2) ]
       [mpz_clear(~a);]
```

4 Types

4.1 PVS Types

A PVS theory can be typechecked using the emacs interface `M-x typecheck` or with Lisp function `(tc name-theory)`. This first runs the PVS parser on the code and generates CLOS objects to represent it. Then, the PVS typechecker is run on this internal representation of the theory and tries to give a type to all expressions generating TCC when needed.

Here we describe how PVS types are represented in Lisp. The syntax of PVS we allow

TypeExpr	::=	*Name*
	|	*EnumerationType*
	|	*Subtype*
	|	*TypeApplication*
	|	*FunctionType*
	|	*TupleType*
	|	*CotupleType*
	|	*RecordType*
EnumerationType	::=	`{` *IdOps* `}`
Subtype	::=	`{` *SetBindings* `|` *Expr* `}`
	|	`(` *Expr* `)`
TypeApplication	::=	*Name Arguments*
FunctionType	::=	[`FUNCTION` | `ARRAY`]
		`[` −[*IdOp* `:`] *TypeExpr* $''^{+}_{,}$ `->` *TypeExpr* `]`
TupleType	::=	`[` −[*IdOp* `:`] *TypeExpr* $''^{+}_{,}$ `]`
CotupleType	::=	`[` −[*IdOp* `:`] *TypeExpr* $''^{+}_{+}$ `]`
RecordType	::=	`[#` *FieldDecls* $^{+}_{,}$ `#]`
FieldDecls	::=	*Ids* `:` *TypeExpr*

`type-expr` ⊂ `syntax` [*abstract class*]

. . . .

`type-name` ⊂ `type-expr name` [*class*]

adt

. . . .

subtype ⊂ type-expr *supertype* *predicate*	[*class*]

funtype ⊂ type-expr *domain* *range.*	[*class*]

tupletype ⊂ type-expr *types*	[*class*]

recordtype ⊂ type-expr *fields*	[*class*]

4.2 Translating types

PVS types:boolean, number, number_field, real, rational, integer, $A \to B$, restricted types below(10) := $\{x : \text{int}|0 \le x < 10\}$) enum datatype

This requires a type analysis to decide on the type of a PVS expression. For example the PVS `int` type can be represented by the `int`, `unsigned long` or `mpz_t` C types. In that case, we study the range of the expression to decide which types are allowed to represent it. Then we take the context in which the expression appears to decide. For instance in

```
incr(x:below(10)):int = x+1
```

the `x` expression, result of the function `incr` can always be represented by an `int` or `unsigned long` in C but we choose here to represent it using a `mpz_t`.

Auxiliary type system : C-type with a flag : mutable (meaning that the expression it describes only has one pointer pointing to it.

```
int a = 2;      a : int[mutable]
int* a = malloc( 10 * sizeof(int*) );
```

destructive addition:

```
d_add(*mpz_t res, mpz_t[mutable] a, long b) {
  mpz_add(a, a, b);
  (*res) = a;
}
```

Rq : `d_add` is given a mutable `mpz_t`, meaning that it can modify it and is responsible for freeing it. It is also responsible for allocating memory for the result. Here it uses the memory to assign res.

Use an auxiliary language :

```
( expr, C-type[mutable] )
```

Conversions and copies create mutables types (at a cost) : `a[mutable]_from_b` [2]

C types:[3]

```
// integer and floating point types
[unsigned] char, int, long, double
type*  //arrays
char*  // strings
struct types  // structures with fields
enum types
```

```
short int, float, union, size_t // etc...
```

Listing 1: C types

We can only translate a subset of all PVS types. What's missing ?

4.3 Translating PVS syntax

We can only translate a subset of PVS syntax. What's missing ?

5 Difficulties and successes

5.1 `if` expressions

Represented by `if-expr`

5.2 Integer, rationnals

In PVS, the `integer` represent the whole set $\mathbb{Z}$ of all relative numbers (and `rational` also describe $\mathbb{Q}$). In C, we have finite types int, long, ...

We need the GMP library which introduces the types `mpz_t`and `mpq_t`. These types are arrays and should be used just as integer (not as pointers except they still need to be freed).

5.3 Garbage collection

We implement a very simple "Reference Counting Garbage Collector" as described in [4].

We maintain a hashtable of pointer counters. Each pointer in the code is a key in the hastable to which we associate an int counter as value.

Pointers only occurs in arrays or struct.

Arrays are created in the code.

```
T* a = b;
```

becomes

```
T* a = GC( b ); // Should not happen often...
```

```
t[0] = b;  // with b of type T*
```

becomes

```
GC_free( t[0] );
t[0] = (T*) GC( b );
```

Examples

```
int* f() {
  int* res;
  res = (int*) GC_alloc( 10 * sizeof(int) );
  [... init res...]
  return res; // pointer count = 1
}

void main() {
   int* a = f();  // pointer counter of a = 1
   int** b = (int**) GC_alloc( sizeof( int*) );
   GC_free( b[0] );  // useless
   b[0] = (int*) GC( a );  // pointer counter of a = 2
   printf("f(0) = %s", b[0][0]);
```

```
    GC_free(b); // frees b, pointer count of a = 1
    GC_free(a); // frees a
}
```

5.4 Performance

5.5 Update expressions

Update expressions are represented by PVS as `update-expr` objects.

$$E := \texttt{ t with [e1 := e2]}$$

Problem : `t` is an expression typed as a function. Therefore it might be represented in C as an array (if domain type is `below(`n`)`. We want to know if we can update `t` in place to obtain a C object representing `E` or if we have to make a copy of `t`.

We consider three solutions to this problem.

5.5.1 Pointer counting

We keep track of the number of pointer pointing to an array or a struct.

This requires to build our own C struct (heavy)

```
struct array_int {
    int pointer_count = 1;
    int *data;
};
```

When we update the struct, if the pointer is 0, we update in place.

Besides every update require now to read the structure and make a test (small compared to a copy but no so small compared to a single in place update)

Besides, the creation / destruction gets more complicated

Passing argument to function :

```
array_int f(array_int arg) {
    arg.pointer_count++;  \\ Since now f also have a pointer to the struct
    if (arg.pointer_count == 1) {
        arg.data[0] = 0;
        return arg;
    } else {
        array_int res;
        res.data = malloc( 10 * sizeof(int*) );
        copy(res, arg);  // Very long...
        res.pointer_count --; // This function is about to lose its pointer to res
        return res;
    }
}

void main() {
    array_int t;
    init(t);  // somehow...

    t.pointer_counter --;  // We assure we won't use the pointer "t" to the array anymore
    array_int r = f(t);
    t = null; // This way we garantee the variable "t" won't be used later in the code

    [...]
}
```

This add quite some code compared to the simple :

```
array_int f(array_int arg) {
   arg[0] = 0;
   return arg;
}

void main() {
   array_int t;
   init(t);  // somehow...
   array_int r = f(t);

   [...]
}
```

5.5.2 Using a different data structure

PVS uses arrays in a very particular way, we might then represent them with an other structure than just only a C array. For example :

```
struct r_list_int {
   int key;
   int value;
   r_list_int tl;
};
struct array_int {
   int *data;
   r_list_int replacement_list;
};
```

Each structure represent the array `data` with the modifications contained in the linked list `r_list_int`

Problems : Just as the previous solution : - add some extra code - add some extra computation (runtime tests, reading the replacement list) - require to create as many structures and associated functions as there are range types fo the manipulated arrays - Very dependent on the GC

5.5.3 Flow analysis on the PVS code

[1]

5.5.4 Analysis of the C code

Trying to avoid copying arrays by analyzing the C code generated.

2 different functions (destructive and non destructive)

Algorithm :

Always have a "non destructive" version of any function. A "cautious" version that never modify the arguments in place and always make copies when necessary (when a "mutable" version of an array is necessary (for instance updates)).

In destructive versions of all functions : Flag all array arguments to "mutable". Then for each of these arguments : - If it never occurs destructively, then remove flag (function just observe the arg) - If it occurs destructively, it can never occur at all AFTER. -¿ Need to define the order of evaluation of expression (easy rules on simple expressions) -¿ Need to be able to detect occurences of a name-expr -¿ Otherwise, unflag the arg

What is a destructive occurence :

$$E := \texttt{f(t with [e1 := e2] , t(0))}$$

order of eval : e1 and e2 (t can occur non destr) t (expression of an update : destr) t(0) (occurence of t (even non destr)) `f(x:Arr):int = g( h(t), t)` is destructively translated to

```
1 int f_d(int* t) {    // t has type ! since this is destructive f
2   int* arg1 = h(t); // h can't be called destructively because
3                     // even though t is !, it appears later (line 4)
4   int* arg2 = t;    // t is ! and never appears later => arg2 is !
5   return g( arg1, arg2); // arg2 is ! but g can only be called
6 }                        // destructively if arg1 is
```

Listing 2: Example

if g has type `[Array! -> ?]` then t can't be destructive
if g has type `[Array -> ?]` then t can be destructive
First algorithm:
Need multiple passes as the flags disappear

6 Conclusion

6.1 What's left to be done ?

Use a C structure to represent a closure

```
struct r_list_int {
   int (*body)(void* env, void* args);
   void* env;
   void* args;
};
```

6.2 My stay at SRI

Discovering PVS : Translating Coq proofs to PVS PVS library for basic linear algebra

Robin project, HACMS Contest week-end 14-15 June Summer School Parsing Lisp code -¿ generate HTML architecture fileCorrecting translator PVS to SMT-LIB [1]

References

[1] Pavol Černý. Static analyses for guarded optimizations of high level languages.

[2] Jean-Christophe Filliâtre. Langages de programmation et compilation. MPRI, 2013-14. https://www.lri.fr/~filliatr/ens/compil/.

[3] Eric Huss. The c library reference guide. *Webmonkeys: A Special Interest*, 2004.

[4] Richard Jones and Rafael Lins. Garbage collection: Algorithms for automatic dynamic memory management, 1996. *John Wiliey & Sons Ltd., England*.

[5] N. Shankar and S. Owre. *PVS API Reference*. Computer Science Laboratory, SRI International, Menlo Park, CA, September 2003.