Program Verification: Lecture 25

José Meseguer

Computer Science Department
University of Illinois at Urbana-Champaign
In the case of deterministic programs, we first studied the verification of declarative deterministic programs such as Maude functional modules. Then, in a sense, we reduced to this case the verification of imperative programs.

Indeed, we can specify the semantics of a deterministic imperative language $\mathcal{L}$ as an equational theory $\mathcal{E}(\mathcal{L})$ (in fact, a Maude functional module).

Then, reasoning about the correctness of imperative programs in $\mathcal{L}$ reduces (perhaps through decomposition by means of a Hoare logic) to proving inductive properties satisfied by the initial model $T_{\mathcal{E}(\mathcal{L})}$. 
What should the analogous situation be in the case of concurrent imperative programs? We should of course specify the semantics of a concurrent imperative language $\mathcal{L}$ as a rewrite theory $\mathcal{R}(\mathcal{L})$ (in fact, a Maude system module).

Then, the correctness of imperative programs in $\mathcal{L}$ can be reduced to proving inductive properties satisfied by the initial model $(T_{\Sigma_\mathcal{L}/E_\mathcal{L}}, \rightarrow_{\mathcal{R}_\mathcal{L}})$. If such properties are specified in temporal logic (resp. reachability logic) then we can use methods such as model checking (resp. deductive proof).

We can illustrate this general method by defining the rewriting logic semantics of a simple threaded language called THREADED-IMP extending IMP.
THREADED-IMP extends the semantics of IMP with threads that can communicate with each other through shared variables.

mod THREADED-IMP-SEMANTICS is pr IMP-EVAL + IMP-SYNTAX + IMP-NUMBERS .
sort ThreadedImpState Thread ThreadSet .
subsort Thread < ThreadSet .
op none : -> ThreadSet .

Each thread has the form \{S | J\}, where S is a sequential IMP program and J a number (identifier). The global state is a tripe consisting of an AC set of threads, a shared memory, and the identifier of the last executed thread (used for model checking purposes). The semantics is defined by:
sort TermStmt .

subsort TermStmt < Stmt .

var NE : NatExp . var S S' : Stmt . var I : Id . var TS : ThreadSet .
var M : Memory . var N J K : Nat . var BR : BoolRedex .
var B : Bool . var BE : BoolExp . var PM : PreMemory .

rl {T:TermStmt ; S' | J} TS | M | K => {S' | J} TS | M | J .
rl {I := NE ; S' | J} TS | {[I,N] PM} | K
   => {S' | J} TS | {[I,eval({[I,N] PM},NE)] PM} | J .
rl {if BR then S fi ; S' | J} TS | M | K
   => {if eval(M,BR) then S fi ; S' | J} TS | M | J .
rl {if true then S fi ; S' | J} TS | M | K => {S ; S' | J} TS | M | J .
rl {if false then S fi ; S' | J} TS | M | K => {S' | J} TS | M | J .
rl {while BE do S od ; S' | J} TS | M | K
   => {if BE then S ; while BE do S od fi ; S' | J} TS | M | J .
endm

Call two processes, p_0 and p_1. To indicate that it wants to enter the critical section p_0 (resp. p_1) sets flag a (resp. b) to 1. If one process, after setting its variable to 1 finds that the variable of its competitor is 0, then it enters its critical section crit right away. In case of a tie (both a and b set to 1) the tie is broken using a variable t that takes values in \{0, 1\}. After executing its critical code p_0 (resp. p_1) sets its flag a (resp. b) to 0, and t to 1 (resp. 0). Here is the code:
mod PETERSON is pr THREADED-IMP-SEMANTICS .
  ops crit rest : -> TermStmt [ctor] .
  ops p0 p1 : -> Thread .
  op init-mem : -> Memory .
  eq init-mem = {{a,0} [b,0] [t,0]} .

  eq p0 = {while true do
    a := 1 ;
    while b = 1 do
      if t = 1 then a := 0 ;
      while t = 1 do skip od ;
      a := 1
      fi
    od ;
    crit ;
    t := 1 ; a := 0 ;
    rest
    od | 0} .
eq p1 = {while true do
    b := 1 ;
    while a = 1 do
        if t = 0 then b := 0 ;
            while t = 0 do skip od ;
        b := 1
    fi
    od ;
    crit ;
    t := 0 ; b := 0 ;
    rest
        od | 1} .
endm
The most important property guaranteed by Peterson’s algorithm is **mutual exclusion**. This property is violated by any state in which **both** processes are in their critical section.

We can verify this property by giving the search command:

```
search in PETERSON :
p0 p1 | init-mem | 0 =>*  
   {crit ; S0:Stmt | 0} {crit ; S1:Stmt | 1} | M:Memory | J:Natural .
```

No solution.

states: 258  rewrites: 1241 in 7ms cpu (7ms real) (174714 rewrites/second)
We should also check non-starvation of p0 and p1. We need to define suitable state predicates:

mod THREADED-IMP-CHECK is inc PETERSON . inc SATISFACTION .
  subsort ThreadedImpState < State .
  ops exec crit : Nat -> Prop .
  var M : Memory . vars S S' : Stmt . vars J K L : Nat .
  eq ({S | J} {S' | L} | M | J) |= exec(J) = true .
  eq ({S | J} {S' | L} | M | K) |= exec(J) = false [owise] .
  eq ({crit ; S | J} {S' | L} | M | K) |= crit(J) = true .
  eq ({S | J} {S' | L} | M | K) |= crit(J) = false [owise] .
endm
The following property of non-starvation of p0 and p1 holds, among others:

reduce in CHECK : modelCheck(p0 p1 | init-mem | 0, []<> exec(0) /\ []<> exec(1) -> []<> crit(0)) .
rewrites: 1991 in 6ms cpu (7ms real) (326019 rewrites/second)
result Bool: (true).Bool

reduce in CHECK : modelCheck(p0 p1 | init-mem | 0, []<> exec(0) /\ []<> exec(1) -> []<> crit(1)) .
rewrites: 1995 in 13ms cpu (13ms real) (151262 rewrites/second)
result Bool: (true).Bool
THREADED–IMP is a toy language. Can the rewriting logic approach scale up to real concurrent languages? The answer is “yes.” We can define the semantics of a concurrent programming language $L$ by a rewrite theory $\mathcal{R}_L = (\Sigma_L, E_L, R_L)$, where:

1. $\Sigma_L$ specifies $L$’s syntax and the auxiliary operators needed in semantic definitions (memory, environment, etc.)

2. the equations $E_L$ specify the semantics of all the deterministic features of $L$ and of the auxiliary semantic operations.

3. the rewrite rules $R_L$ specify the semantics of all the concurrent features of $L$. 
Once a definition of a language is given in Maude, we get an interpreter for free and we also get:

1. a semi-decision procedure to find failures of safety properties in a (possibly infinite-state) concurrent program using Maude’s search command;

2. an LTL model checker for finite-state programs or program abstractions;

3. a theorem prover (Maude’s RL Tool) that can be used to semi-automatically prove programs correct.
Java was defined at UIUC by Feng Chen, using a CPS semantics as above, with 600 equations and 15 rewrite rules. Azadeh Farzan developed a more direct specification for the JVM, not based on continuations, with around 300 equations and 40 rewrite rules.

Both the Java and the JVM specifications include multithreading, inheritance, polymorphism, object references, and dynamic object allocation. Native methods and most Java libraries are not supported at present.
Based on Maude rewriting logic specifications of Java and JVM, we are developing JavaFAN (Java Formal ANalyzer), a tool in which Java and JVM code can be executed and analyzed.
## Performance of JavaFAN

<table>
<thead>
<tr>
<th>Tests</th>
<th>JVM</th>
<th>Java</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote Agent (s)</td>
<td>0.3</td>
<td>0.1</td>
<td>2 (Stanford)</td>
</tr>
<tr>
<td>2-stage Pipeline</td>
<td>17m</td>
<td>—</td>
<td>100m+ (Stanford)</td>
</tr>
<tr>
<td>DinPhil (4)</td>
<td>0.64</td>
<td>1.2</td>
<td>—</td>
</tr>
<tr>
<td>DinPhil (6)</td>
<td>33.3</td>
<td>81.7</td>
<td>—</td>
</tr>
<tr>
<td>DinPhil (8)</td>
<td>13.7m</td>
<td>98m</td>
<td>—</td>
</tr>
<tr>
<td>DinPhil (9)</td>
<td>803.2m</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Deadlock-free DinPhil (5)</td>
<td>3.2m</td>
<td>19.2</td>
<td>∞ (JPF)</td>
</tr>
<tr>
<td>Deadlock-free DinPhil (7)</td>
<td>686.4m</td>
<td>27m</td>
<td>∞ (JPF)</td>
</tr>
<tr>
<td>Thread Game (100) (s)</td>
<td>17.1</td>
<td>6.6</td>
<td>—</td>
</tr>
<tr>
<td>Thread Game (1000) (s)</td>
<td>10.1m</td>
<td>5.1m</td>
<td>—</td>
</tr>
</tbody>
</table>
Performance of JavaFAN: Some discussion

There are essentially two reasons for JavaFAN to compare favorably with more conventional Java analysis tools: (1) the high performance of Maude for execution, search, and model checking; and (2) optimized equational and rule definitions.

The second reason is the use of performance-enhancing specification techniques at the Maude level, including:

- expressing as equations $E$ the semantics of all deterministic computations, and as rules $R$ only concurrent computations.
- favoring unconditional equations and rules over less efficient conditional versions.
- using a continuation passing style in semantic equations.
Other Language Case Studies

Similar positive experience in using rewriting logic and Maude to give semantics definitions of concurrent programming languages and getting interpreters and program analysis tools for free for those languages is reported in several papers, including the surveys by Meseguer and Roșu in: (i) Proc. IJCAR’04, Springer LNCS 3097; and (ii) Proc. SOS’05, Elsevier ENTCS.

In particular, semantic definitions have already been given in Maude for substantial subsets of the following languages: ABEL, bc, Beta, CCS, CIAO, CML, Creol, ELOTOS, Haskell, Lisp, LLVM, MSR, Pi-Calculus, Pict, PLAN, Python, Ruby, SIMPLE, Verilog, and Smalltalk. And full definitions have been given in K-Maude to C and Scheme.