# The Vapor and Vapor-M Operational Semantics

Jens Palsberg

October 10, 2016

# 1 What are Vapor and Vapor-M?

Vapor and Vapor-M are languages that we will use as intermediate languages when we compile MiniJava to MIPS. The translation steps are: MiniJava  $\rightarrow$  Vapor  $\rightarrow$  Vapor-M  $\rightarrow$  MIPS.

A Vapor program consists of functions that each operates on a heap, global constants, parameters, local variables, and a stack. We will specify Vapor's abstract syntax and operational semantics.

A Vapor-M program consists of functions that each operates on a heap, global constants, global registers, and a stack. We will specify Vapor-M's abstract syntax and operational semantics. Vapor and Vapor-M are closely related. One difference is that in Vapor-M, each function has no parameters, no local variables, and no return value.

## 2 Notation

#### 2.1 Grammars

The grammars for Vapor and Vapor-M use the following metanotation:

- Nonterminal symbols are words written in this font.
- Terminal symbols are written in this font, except  $\langle STRING \rangle$ ,  $\langle LABEL \rangle$ ,  $\langle IDENTIFIER \rangle$ , and  $\langle INTEGER\_LITERAL \rangle$ .
- A production is of the form lhs ::= rhs, where lhs is a nonterminal symbol and rhs is a sequence of nonterminal and terminal symbols, with choices separated by |, and some times using "..." to denote a possibly empty list.
- We will use superscripts and subscripts to distinguish metavariables.

#### 2.2 Rules

We will use the following notation:

$$\frac{hypothesis_1}{conclusion} \quad \frac{hypothesis_2}{conclusion} \dots \quad \frac{hypothesis_n}{conclusion}$$

This is a *rule* that says that if we can derive all of  $hypothesis_1, hypothesis_2, ..., hypothesis_n$ , then we can also derive *conclusion*.

A special case arises when n = 0: we can write this case as:

conclusion

or we can even omit the horizontal bar and write:

conclusion

We can say that this case is a rule with no hypotheses, or we can call it an axiom.

A derivation happens when we begin with one or more axioms, then perhaps apply some rules, and finally arrive at a conclusion. Notice that we can organize a derivation as a tree that has the axioms as leaves and the conclusion as the root. We can refer to such as tree as a derivation tree.

## 2.3 Maps

A map is a function with finite domain. If M is a map, then dom(M) denotes the domain of M. If  $x_1, \ldots, x_r$  are pairwise distinct, then  $[x_1 \mapsto y_1, \ldots, x_n \mapsto y_n]$  denotes a map with domain  $\{x_1, \ldots, x_n\}$ , which maps  $x_i$  to  $y_i$ , for  $i \in 1..n$ . If  $M_1, M_2$  are maps, then  $M_1 \cdot M_2$  is a map:

$$(M_1 \cdot M_2)(id) = \begin{cases} M_2(id) & \text{if } id \in dom(M_2) \\ M_1(id) & \text{otherwise} \end{cases}$$

Notice that  $M_2$  takes precedence over  $M_1$ .

If M is a map and X is a set, then  $M \setminus X$  denotes M restricted to  $dom(M) \cap X$ . We define a helper function initmap that maps a set to a map.

$$(initmap(X))(x) = \begin{cases} 0 & x \in X \\ undefined & \text{otherwise} \end{cases}$$

If M is a map, then we define the notation  $M^*$  as follows.

$$M^*(x) = \begin{cases} M(x) & \text{if } x \in dom(M) \\ x & \text{otherwise} \end{cases}$$

#### 2.4 Tuples

$$(Tuple)$$
  $t$  ::=  $\langle y_1, \dots, y_n \rangle$ 

We define a helper function *inittuple* that maps a positive integer to a tuple.

$$inittuple(c) = \langle 0, \dots, 0 \rangle$$
 where the number of 0's in the tuple is  $c$ , where  $c > 0$ 

# 3 Vapor

#### 3.1 Syntax

```
(Program) p ::= C_1 \dots C_n F_1 \dots F_m
  (ConstDecl) C ::= const l l_1 \dots l_n
    (FunDecl) F ::= func l (id_1 \ldots id_f) l_1 b_1 \ldots l_q b_q
       (Block) b ::= i_1 \dots i_n j
        (Instr) i ::= id = o \mid id = op (o_1 o_2) \mid id = m \mid m = id
                     | if0 o goto l | id = call o ( o_1 ... o_f )
                        id = HeapAllocZ ( o ) | PrintIntS ( o ) | Error ( s )
       (Jump) j ::= goto l | ret o | ret
   (MemRef) m ::= [ id + c ]
   (Operator) op ::= Add | Sub | MulS | Eq | LtS
     (Operand) o ::= l \mid c \mid id
 (StringLiteral) s ::= \langle STRING \rangle
        (Label) l ::= \langle LABEL \rangle
(IntegerLiteral) c ::= \langle INTEGER\_LITERAL \rangle
   (Identifier) id ::=
                        (IDENTIFIER)
```

## 3.2 Helper Functions

**Defined Variables.** We define a helper function *defined* that maps a block to the set of local variables that the blocks assigns. We will overload *defined* and define it also for instructions.

```
defined(\ i_1 \ldots i_n\ j\ ) \ = \ (defined(i_1) \cup \ldots \ defined(i_n)) defined(\ id = o\ ) \ = \ \{id\} defined(\ id = m\ ) \ = \ \{id\} defined(\ id = m\ ) \ = \ \{id\} defined(\ m = id\ ) \ = \ \emptyset defined(\ if 0\ o\ goto\ l\ ) \ = \ \emptyset defined(\ id = call\ o\ (\ o_1\ \ldots\ o_f\ )\ ) \ = \ \{id\} defined(\ HeapAllocZ\ (\ o\ )\ ) \ = \ \emptyset defined(\ PrintIntS\ (\ o\ )\ ) \ = \ \emptyset defined(\ Error\ (\ s\ )\ ) \ = \ \emptyset
```

**Initialization of Constants.** We define a helper function *initconst* that maps a constant declaration to a map.

$$initconst(\texttt{const}\ l\ l_1\ \dots\ l_n\ )\ =\ [l\mapsto \langle l_1\dots\ l_n\rangle]$$

**Initialization of Functions.** We define a helper function initfun that maps a function declaration to a map.

$$\begin{array}{lll} initfun(F) & = & [l \mapsto F, \ l_1 \mapsto b_1, \ \dots, \ l_q \mapsto b_q] \\ & & \text{where} \\ & F & = & \text{func} \ l \ (id_1 \ \dots \ id_f) \ l_1 \ b_1 \ \dots \ l_q \ b_q \end{array}$$

#### 3.3 Program States

Vapor has three kinds of values: labels l, heap addresses (l,c), and integers c. We use v to range over values. A program state (G,H,E,b) has four components. Intuitively, G is a global table that represents the constants, functions, and blocks; H is the heap; E is an environment that represents the parameters and local variables; and b is the block that is executing right now.

The Global Table. The global table is a map from labels to either tuples, functions, or blocks.

$$(GlobalData)$$
  $d$  ::=  $t \mid F \mid b$   
 $(GlobalTable)$   $G$  ::=  $[l_1 \mapsto d_1, \dots, l_n \mapsto d_n]$ 

**The Heap.** A heap is a map from labels to tuples. We use H to range over heaps. A heap address is a pair of the form (l, c), where l is a label and c is an integer such that  $c \ge 0$  and c is divisible by 4.

**The Environment.** An environment represents the parameters, and local variables. An environment is a map from identifiers to values. We use E to range over environments.

The Initial Program State. Consider a program

$$C_1 \ldots C_n F_1 \ldots F_m$$

where

$$F_1$$
 = func  $id$  ( )  $l_1$   $b_1$  ...  $l_a$   $b_a$ 

Notice that  $F_1$  has no parameters. The initial program state is (G, H, E, b), where:

$$G = initconst(C_1) \cdot \ldots \cdot initconst(C_n) \cdot initfun(F_1) \cdot \ldots \cdot initfun(F_m)$$

$$H = []$$

$$E = initmap( (defined(b_1) \cup \ldots \cup defined(b_q)) )$$

$$b = b_1$$

#### 3.4 Semantics

where  $c' + c \le n$ 

#### Single Steps.

$$(G, H, E, id = o \quad b') \longmapsto (G, H, E \cdot [id \mapsto E^*(o)], b')$$

$$(G, H, E, id = Add \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto (c_1 + c_2)], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2$$

$$(G, H, E, id = Add \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto (l, c_1 + c_2)], b')$$

$$\text{if } E^*(o_1) = (l, c_1) \text{ and } E^*(o_2) = c_2$$

$$(G, H, E, id = Add \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto (l, c_1 + c_2)], b')$$

$$\text{if } E^*(o_1) = (l, c_1) \text{ and } E^*(o_2) = c_2$$

$$(G, H, E, id = Sub \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto (c_1 - c_2)], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2$$

$$(G, H, E, id = MulS \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto (c_1 \times c_2)], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2$$

$$(G, H, E, id = Eq \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto 1], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2 \text{ and } c_1 = c_2$$

$$(G, H, E, id = Eq \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto 0], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2 \text{ and } c_1 \neq c_2$$

$$(G, H, E, id = LtS \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto 0], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2 \text{ and } c_1 \leq c_2$$

$$(G, H, E, id = LtS \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto 0], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2 \text{ and } c_1 \leq c_2$$

$$(G, H, E, id = LtS \quad (o_1 o_2) \quad b') \longmapsto (G, H, E \cdot [id \mapsto 0], b')$$

$$\text{if } E^*(o_1) = c_1 \text{ and } E^*(o_2) = c_2 \text{ and } c_1 \leq c_2$$

$$(G, H, E, id = [id' + c \mid b') \mapsto (G, H, E \cdot [id \mapsto v_{c'+c}], b')$$

$$\text{where } E^*(id') = (l, c')$$

$$\text{where } E^*(id') = (l, c')$$

$$\text{where } E^*(id) =$$

$$(G, H, E, \text{if0 } o \text{ goto } l \quad b') \longmapsto (G, H, E, b'')$$
  
if  $E^*(o) = 0$  and  $G(l) = b''$ ,  
and  $l \ b''$  and if0  $o \text{ goto } l$  are in the body of the same function (12)

$$(G, H, E, \text{if0 } o \text{ goto } l \quad b') \longmapsto (G, H, E, b')$$
  
if  $E^*(o) = c \text{ and } c \neq 0$  (13)

$$(G, H, E^{init}, b_1) \longmapsto (G, H', E', ret o')$$

$$(G,H,E,id = \texttt{HeapAllocZ} \ (o \ ) \ b') \longmapsto (G,H \cup [l \mapsto t], E \cdot [id \mapsto (l,0)], E,b')$$
 where  $l \not\in dom(G,H)$  and  $E^*(o)$  is a positive integer that is divisible by 4 where  $t = inittuple(\frac{E^*(o)}{4})$  (15)

$$(G, H, E, \text{PrintIntS} (o) b') \longmapsto (G, H, E, b')$$
  
where  $E^*(o) = c$   
and display  $c$  on the screen (16)

$$(G, H, E, \text{Error (} s \text{ )} b')$$
display  $s$  on the screen and stop execution (17)

$$(G, H, E, \mathsf{goto}\ l) \longmapsto (G, H, E, b')$$
  
if  $G(l) = b'$ ,  
and  $l\ b'$  and  $\mathsf{goto}\ l$  are in the body of the same function (18)

#### Multiple Steps.

$$\frac{(G, H, E, b) \longmapsto (G', H', E', b') \qquad (G', H', E', b') \longmapsto (G'', H'', E'', b'')}{(G, H, E, b) \longmapsto (G'', H'', E'', b'')}$$

$$(19)$$

# 4 Vapor-M

## 4.1 Syntax

```
(Program) p ::= C_1 \dots C_n F_1 \dots F_m
  (ConstDecl) C ::= const l l_1 \dots l_n
    (\mathit{FunDecl}) F ::= func l [in c_1, out c_2, local c_3] l_1 b_1 ... l_q b_q
        (Block) b ::= i_1 \dots i_n j
        (Instr) i ::= id = o \mid id = op (o_1 o_2) \mid id = m \mid m = id
                          if 0 o goto l | call o
                         id = \text{HeapAllocZ } (o) \mid \text{PrintIntS } (o) \mid \text{Error } (s)
        (Jump) j ::= goto l | ret
    (MemRef) m ::= [ id + c ] | in [ c ] | out [ c ] | local [ c ]
   (Operator) op ::= Add | Sub | MulS | Eq | LtS
     (Operand) o ::= l \mid c \mid id
 (StringLiteral) s ::= \langle STRING \rangle
        (Label) l ::= \langle LABEL \rangle
(IntegerLiteral) c ::= \langle INTEGER\_LITERAL \rangle
   (Identifier) id ::=
                         (IDENTIFIER)
```

## 4.2 Helper Functions

**Defined Variables.** We define a helper function *defined* that maps a block to the set of local variables that the blocks assigns. We will overload *defined* and define it also for instructions.

```
defined(\ i_1 \ldots i_n\ j\ ) \ = \ (defined(i_1) \cup \ldots \ defined(i_n))
defined(\ id = o\ ) \ = \ \{id\}
defined(\ id = m\ ) \ = \ \{id\}
defined(\ id = m\ ) \ = \ \{id\}
defined(\ m = id\ ) \ = \ \emptyset
defined(\ if0\ o\ goto\ l\ ) \ = \ \emptyset
defined(\ call\ o\ ) \ = \ \emptyset
defined(\ HeapAllocZ\ (\ o\ )\ ) \ = \ \emptyset
defined(\ PrintIntS\ (\ o\ )\ ) \ = \ \emptyset
defined(\ Error\ (\ s\ )\ ) \ = \ \emptyset
```

**Initialization of Constants.** We define a helper function *initconst* that maps a constant declaration to a map.

$$initconst(\texttt{const}\ l\ l_1\ \dots\ l_n\ )\ =\ [l\mapsto \langle l_1\dots\ l_n\rangle]$$

**Initialization of Functions.** We define a helper function initfun that maps a function declaration to a map.

$$\begin{array}{rcl} initfun(F) & = & [l\mapsto F,\; l_1\mapsto b_1,\; \dots,\; l_q\mapsto b_q] \\ & & \text{where} \\ & F & = & \text{func}\; l\; [\text{in}\; c_1\text{, out}\; c_2\text{, local}\; c_3] \;\; l_1\;\; b_1\; \dots \; l_q\;\; b_q \end{array}$$

#### 4.3 Program States

Vapor-M has three kinds of values: labels l, heap addresses (l, c), and integers c. We use v to range over values.

A program state (G, H, R, S, b) has five components. Intuitively, G is a global table that represents the constants, functions, and blocks; H is the heap; R is the register file; S is the stack; and b is the block that is executing right now.

The Global Table. The global table is a map from labels to either tuples, functions, or blocks.

$$(GlobalData)$$
  $d$  ::=  $t \mid F \mid b$   
 $(GlobalTable)$   $G$  ::=  $[l_1 \mapsto d_1, \dots, l_n \mapsto d_n]$ 

**The Heap.** A heap is a map from labels to tuples. We use H to range over heaps. A heap address is a pair of the form (l, c), where l is a label and c is an integer such that  $c \ge 0$  and c is divisible by 4.

The Register File. The set registers consists of 23 identifiers that are akin to the names of 23 MIPS registers.

$$registers = \{s0, \dots, s7, t0, \dots, t8, a0, \dots, a3, v0, v1\}$$

Intuitively, the elements of registers are the global registers. A register file is a map from registers to values. We use R to range over register files.

The Stack.

$$(Stack)$$
  $S$   $::=$   $empty \mid S \odot t$ 

The Initial Program State. Consider a program

$$C_1 \ldots C_n F_1 \ldots F_m$$

where

$$F_1$$
 = func  $id$  [in 0, out  $c_2$ , local  $c_3$ ]  $l_1$   $b_1$  ...  $l_q$   $b_q$ 

Notice that  $F_1$  has no parameters and has "[in 0]". The initial program state is (G, H, R, S, b), where:

$$G = initconst(C_1) \cdot \ldots \cdot initconst(C_n) \cdot initfun(F_1) \cdot \ldots \cdot initfun(F_m)$$

H = []

R = initmap(registers)

 $S = empty \odot inittuple(c_3) \odot inittuple(c_2)$ 

 $b = b_1$ 

#### 4.4 Semantics

Single Steps.

where  $t' = \langle v_0, v_4, \dots v_{c-4} \ R^*(id') \ v_{c+4} \dots \ v_n \rangle$ 

where  $c' + c \le n$ 

$$(G, H, R, S, \text{in } [c] = id \quad b') \longmapsto (G, H, R, S \odot t'_{in} \odot t_{local} \odot t_{out}, b')$$
where  $S = S' \odot t_{in} \odot t_{local} \odot t_{out}$ 
where  $t_{in} = \langle v_m \dots v_c \dots v_1, v_0 \rangle$ 
where  $t'_{in} = \langle v_m \dots R^*(id) \dots v_1, v_0 \rangle$ 

$$(34)$$

$$(G, H, R, S, \text{out } [c] = id \quad b') \longmapsto (G, H, R, S \odot t_{in} \odot t_{local} \odot t'_{out}, b')$$
where  $S = S' \odot t_{in} \odot t_{local} \odot t_{out}$ 
where  $t_{out} = \langle v_m \dots v_c \dots v_1, v_0 \rangle$ 
where  $t'_{out} = \langle v_m \dots R^*(id) \dots v_1, v_0 \rangle$ 

$$(35)$$

$$(G, H, R, S, \text{local } [c] = id \quad b') \longmapsto (G, H, R, S \odot t_{in} \odot t'_{local} \odot t_{out}, b')$$
where  $S = S' \odot t_{in} \odot t_{local} \odot t_{out}$ 
where  $t_{local} = \langle v_m \dots v_c \dots v_1, v_0 \rangle$ 
where  $t'_{local} = \langle v_m \dots R^*(id) \dots v_1, v_0 \rangle$ 

$$(36)$$

$$(G, H, R, S, \text{if0 } o \text{ goto } l \quad b') \longmapsto (G, H, R, S, b'')$$
  
if  $R^*(o) = 0$  and  $G(l) = b''$ ,  
and  $l \ b''$  and if0  $o \text{ goto } l$  are in the body of the same function (37)

$$(G, H, R, S, \text{if0 } o \text{ goto } l \quad b') \longmapsto (G, H, R, S, b')$$
  
if  $R^*(o) = c$  and  $c \neq 0$  (38)

$$\frac{(G, H, R, S \odot inittuple(c_3) \odot inittuple(c_2), b_1) \longmapsto (G, H', R', S' \odot t_3 \odot t_2, \text{ret})}{(G, H, R, S, \text{call } o \quad b') \longmapsto (G, H', R', S', b')} \\
\text{where } R^*(o) = l \\
\text{where } G(l) = \text{func } l \text{ [in } c_1, \text{ out } c_2, \text{ local } c_3] \quad l_1 \quad b_1 \quad \dots \quad l_q \quad b_q$$
(39)

$$(G,H,R,S,id = \texttt{HeapAllocZ} \ (o \ ) \ b') \longmapsto (G,H \cup [l \mapsto t],R \cdot [id \mapsto (l,0)],S,b')$$
 where  $l \not\in dom(G,H)$  and  $R^*(o)$  is a positive integer that is divisible by 4 where  $t = inittuple(\frac{R^*(o)}{4})$ 

$$(G, H, R, S, \text{PrintIntS} \ (o) \ b') \longmapsto (G, H, R, S, b')$$
  
where  $R^*(o) = c$   
and display  $c$  on the screen  $(41)$ 

$$(G, H, R, S, \text{Error } (s) b')$$
  
display  $s$  on the screen and stop execution (42)

$$(G, H, R, S, \text{goto } l) \longmapsto (G, H, R, S, b')$$
  
if  $G(l) = b'$ ,  
and  $l \ b'$  and goto  $l$  and in the body of the same function (43)

#### Multiple Steps.

$$\frac{(G,H,R,S,b)\longmapsto (G',H',R',S',b')}{(G,H,R,S,b)\longmapsto (G'',H'',R'',S'',b'')} (G',H',R',S',b')\longmapsto (G'',H'',R'',S'',b'')$$
(44)