Software Verification – Exam

ETH Zürich

20 December 2010

Surname, first name: ..............................................................................................

Student number: ...................................................................................................

I confirm with my signature that I was able to take this exam under regular circumstances and that I have read and understood the directions below.

Signature: ..............................................................................................................

Directions:

• Exam duration: 1 hour 45 minutes.

• Except for a dictionary you are not allowed to use any supplementary material.

• All solutions can be written directly on the exam sheets. If you need more space for your solution ask the supervisors for a sheet of official paper. You are not allowed to use other paper. Please write your student number on each additional sheet.

• Only one solution can be handed in per question. Invalid solutions need to be crossed out clearly.

• Please write legibly! We will only correct solutions that we can read.

• Manage your time carefully (take into account the number of points for each question).

• Please immediately tell the exam supervisors if you feel disturbed during the exam.

    Good luck!
<table>
<thead>
<tr>
<th>Question</th>
<th>Available points</th>
<th>Your points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Axiomatic semantics</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>2) Separation logic</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>3) Data flow analysis</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4) Model checking</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5) Software model checking</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>6) Termination proofs</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>70</strong></td>
<td></td>
</tr>
</tbody>
</table>
1 Axiomatic semantics (9 points)

Consider the following Hoare triple (all variables of type \textit{NATURAL}, assumed to
describe mathematical natural numbers):

\[
\{ y = n \} \\
1 \text{ from} \\
2 \quad z := 1 \\
3 \quad \text{until } y = 0 \text{ loop} \\
4 \quad y := y - 1 \\
5 \quad z := z \times x \\
6 \quad \text{end} \\
7 \{ z = x^n \}
\]

Prove that this triple is a theorem of Hoare’s axiomatic system for partial
correctness.

Solution:

1 \{ y = n \} \\
2 \text{ from} \\
3 \{ 1 = x^{n-y} = x^0 \} \\
4 \quad z := 1 \\
5 \{ z = x^{n-y} \} \\
6 \text{ until } y = 0 \text{ loop} \\
7 \{ (z = x^{n-y}) \land \neg(y = 0) \} \\
8 \{ z \cdot x = x^{n-(y-1)} = x^{n-y} \cdot x \} \\
9 \quad y := y - 1 \\
10 \{ z \cdot x = x^{n-y} \} \\
11 \quad z := z \times x \\
12 \{ z = x^{n-y} \} \\
13 \text{ end} \\
14 \{ (z = x^{n-y}) \land (y = 0) \} \\
15 \{ z = x^n \}
2 Separation Logic (13 points)

Consider the definition of the list binary predicate:

\[
\begin{align*}
\text{list } i & | \equiv \text{empty} \land i = \text{nil} \\
\text{list } i (a : \sigma) & \equiv \exists j \cdot (i \mapsto a, j) * (\text{list } j \sigma)
\end{align*}
\]

where \( \sigma \overset{\text{def}}{=} [] | a : \sigma \) defines a sequence of integers.

2.1 States and semantics (7 points)

Consider the separation logic predicate \( P \), where

\[
\begin{align*}
P & \overset{\text{def}}{=} 3 \mapsto 5, 8 \mapsto 7, 11 \mapsto 6, 1 \mapsto 1 \mapsto 3, \text{nil}
\end{align*}
\]

and answer the following questions:

(1) For every state \((s, h)\) that satisfies \( P \), the heap component \( h \) will be the same. Write such a function \( h \) explicitly as a set of pairs.

Solution:
\[
h = \{(1, 3), (2, \text{nil}), (3, 5), (4, 8), (5, 7), (6, 1), (7, 11), (8, 6), (9, 11), (10, 12), (11, 12)\}.
\]

(2) If \((s, h) \models P\), then \((s, h) \models \text{list } i \sigma * \text{true} \) for several values of \( i \) and \( \sigma \).

Provide all such pairs \((i, \sigma)\).

Solution:
\[
(\text{nil}, []), (1, 3:[]), (11, 6:3:[]), (8, 7:6:3:[]), (3, 5:7:6:3:[]).
\]

It is also fine to write \([5, 7, 6, 3]\) instead of \(5:7:6:3:[]\), etc.
2.2 Separation logic and verification (6 points)

Consider the signature and separation logic specification for a routine that adds a value to the front of a linked list. It returns a pointer to the new head node by storing it in the Result variable:

\[
\text{add_front } (\text{list}_\text{pointer} : \text{INTEGER} ; \text{value} : \text{INTEGER}) : \text{INTEGER} \\
\text{require list list}_\text{pointer} \sigma \\
\text{ensure list Result } (\text{value} : \sigma)
\]

(1) Write a body for the routine. Use the \textit{cons} command, whose semantics is given by the axiom:

\[
\text{ConsAxiom} \\
\{\text{empty}\} x := \text{cons}(e_1, \ldots, e_n) \{x \mapsto e_1, \ldots, e_n\}
\]

provided that \(1 \leq n\) and \(x\) is not free in any of \(e_1, \ldots, e_n\).

\textbf{Solution:}

Result := cons(value, list_pointer)

(2) Prove your routine body correct.

\textbf{Solution:}

The proof looks as follows in outline form (other forms are also acceptable):

\[
\{\text{list list}_\text{pointer} \sigma\}
\{\text{empty}\}
\text{Result := cons(value, list}_\text{pointer}\}
\{\text{Result} \mapsto \text{value, list}_\text{pointer}\} \quad // \text{By the axiom for cons.}
\{\text{Result} \mapsto \text{value, list}_\text{pointer} \ast \text{list list}_\text{pointer} \sigma\} \quad // \text{By the frame rule.}
\{\text{list Result} (\text{value} : \sigma)\} \quad // \text{By the rule of consequence.}
\]

(3) Write down the schemas of all the inference rules that you used in the proof above.

\textbf{Solution:}

The rule names may differ.

\[
\text{Frame} \\
\{P\} c \{Q\} \\
\{P \ast R\} c \{Q \ast R\}
\]

provided that no free variable of \(R\) is assigned by \(c\).

\[
\text{Consequence} \\
\{P\} c \{Q\} \\
\{P'\} c \{Q'\}
\]

provided that \(P' \Rightarrow P\) and \(Q \Rightarrow Q'\).
### 3 Data flow analysis (12 points)

An arithmetic expression is called *trivial* if it consists only of a single variable or constant; it is called *non-trivial* otherwise. Let $\mathbf{AExp}_a$ denote the set of all non-trivial arithmetic expressions that occur in a given program fragment, and let $\mathbf{AExp}(a)$ denote the set of all non-trivial arithmetic subexpressions of an expression $a$. Furthermore, let $\mathbf{Vars}(a)$ denote the set of variables occurring in $a$.

With this terminology, recall the definition of the *available expressions analysis* from the lecture:

$$
\begin{align*}
\text{AE}_{\text{entry}}(\ell') &= \begin{cases}
\emptyset & \text{if } \ell' \text{ is the initial label} \\
\bigcap_{(\ell, \ell') \in \mathbf{CFG}} \text{AE}_{\text{exit}}(\ell) & \text{otherwise}
\end{cases} \\
\text{AE}_{\text{exit}}(\ell) &= (\text{AE}_{\text{entry}}(\ell) \setminus \text{kill}_{\text{AE}}(B')) \cup \text{gen}_{\text{AE}}(B')
\end{align*}
$$

where $B$ is an elementary block of the form $[x := a]$ or $[b]$, and the $\text{kill}$ and $\text{gen}$ functions are given by

$$
\begin{align*}
\text{kill}_{\text{AE}}([x := a]) &= \{ a' \in \mathbf{AExp}_a \mid x \in \mathbf{Vars}(a') \} \\
\text{kill}_{\text{AE}}([b]) &= \emptyset \\
\text{gen}_{\text{AE}}([x := a]) &= \{ a' \in \mathbf{AExp}(a) \mid x \not\in \mathbf{Vars}(a') \} \\
\text{gen}_{\text{AE}}([b]) &= \mathbf{AExp}(b)
\end{align*}
$$

Now consider the following program fragment:

```plaintext
1  a := b * c
2  d := e + f
3  f := a - d
4  if f > 0 then
5     f := b * c
6  else
7      from
8          g := 1
9      until a * g > 10 loop
10     a := a * f
11    g := g + 1
12  end
13  end
14  b := a + b * c
```

(1) Draw the control flow graph of the program fragment and label each elementary block. (3 points)

(2) Annotate your control flow graph with the analysis result of an available expressions analysis of the program fragment. (7 points)
Solution to (1) and (2):

\[
\begin{align*}
  &\{a := b \ast c\} \\
  &\{d := e + f\} \\
  &\{f := a - d\} \\
  &\{f > 0\} \\
  &\{f := b \ast c\} \\
  &\{g := 1\} \\
  &\{a \ast g > 10\} \\
  &\{a := a \ast f\} \\
  &\{g := g + 1\} \\
  &\{b := a + b \ast c\}
\end{align*}
\]

(3) How can you use your analysis result to optimize the program fragment? (2 points)

Solution:
The analysis result can be used to eliminate common subexpressions, i.e. expressions which are always computed at least twice on a computation path.

As the expression \(b \ast c\) is available at the entries to blocks 5 and 10 where it is also recomputed, it may be worth for optimization purposes to introduce a temporary variable \(tmp\) holding the computed value. The transformed code looks as follows:

\[
\begin{align*}
  &\text{tmp} := b \ast c \\
  &a := \text{tmp} \\
  &d := e + f \\
  &f := a - d \\
  &\text{if } f > 0 \text{ then} \\
  &f := \text{tmp} \\
  &\text{else} \\
  &\text{from} \\
  &\text{until } a \ast g > 10 \text{ do} \\
  &a := a \ast f \\
  &g := g + 1 \\
  &\text{end} \\
  &\text{end} \\
  &b := a + \text{tmp}
\end{align*}
\]
4. Model Checking (10 points)

Recall the semantics of LTL over finite words with alphabet \( \mathcal{P} \). For a word \( w = w(1)w(2) \cdots w(n) \in \mathcal{P}^* \) with \( n \geq 0 \) and a position \( 1 \leq i \leq n \) the satisfaction relation \( \models \) is defined recursively as follows for \( p, q \in \mathcal{P} \).

\[
\begin{align*}
  w, i \models p & \quad \text{iff} \quad p = w(i) \\
  w, i \models \neg \phi & \quad \text{iff} \quad w, i \not\models \phi \\
  w, i \models \phi_1 \land \phi_2 & \quad \text{iff} \quad w, i \models \phi_1 \quad \text{and} \quad w, i \models \phi_2 \\
  w, i \models \phi & \quad \text{iff} \quad i < n \quad \text{and} \quad w, i + 1 \models \phi \\
  w, i \models \phi_1 \mathbin{\text{U}} \phi_2 & \quad \text{iff} \quad \text{there exists } i \leq j \leq n \text{ such that: } w, j \models \phi_2 \\
  & \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \text{and for all } i \leq k < j \text{ it is the case that } w, k \models \phi_1 \\
  w, i \models \Diamond \phi & \quad \text{iff} \quad \text{there exists } i \leq j \leq n \text{ such that: } w, j \models \phi \\
  w, i \models \Box \phi & \quad \text{iff} \quad \text{for all } i \leq j \leq n \text{ it is the case that: } w, j \models \phi \\
  w \models \phi & \quad \text{iff} \quad w, 1 \models \phi
\end{align*}
\]

4.1 Automata and LTL formulas (6 points)

Consider the automata \( T_A \) (with states \( A, B, C \)) and \( T_X \) (with states \( X, Y, Z \)) in Figure 1 over the alphabet \( \{p, q\} \). Notice that \( T_A \) is nondeterministic but \( T_X \) is deterministic.

![Automata](image)

For each of the following LTL formulas say whether every run of \( T_A \) or \( T_X \) satisfies the formula. If it does, argue informally (but precisely) why this is the case; if it does not, provide a counterexample.
(1) $T_A \models \Box(\Diamond p)$
No: the word $w_1 = p q$ is a counterexample because $w_1, 2 \not\models p$ and hence $w_1, 2 \not\models \Diamond p$

(2) $T_X \models \Box(\Diamond p)$
No, with the same counterexample as in question (1).

(3) $T_A \models \Diamond (p \wedge X(p \vee q))$
Yes: every accepting run reaches the state $C$; to do so it must end with the events $p p$ or $p q$.

(4) $T_X \models \Diamond (p \wedge Xp)$
No: the word $w_2 = p q$ is clearly accepted but $w, 1 \not\models Xp$ because $w_2(1+1) = q \neq p$.

(5) $T_X \models p U q$
Yes: every accepted word begins with $q$ or with $p^n q$, with $n \geq 1$, which satisfy $p U q$.

4.2 Automata-based model checking (4 points)
Let $\langle T_A \rangle$ and $\langle T_X \rangle$ respectively denote the set of all words accepted by $T_A$ and $T_X$. Show that $\langle T_A \rangle \nsubseteq \langle T_X \rangle$ by constructing the intersection automaton $T_A \times \neg T_X$ of $T_A$ and the complement of $T_X$, and by showing that the intersection automaton accepts some word.
(Remember that the complement automaton of $T_X$ is identical to $T_X$ except for the accepting states which are $X$ and $Y$ in the complement, with $Z$ becoming a rejecting state in the complement).

Solution:

The accepting state $C, Y$ is reachable with the word $p p$ which is therefore in the intersection of $\langle T_A \rangle$ and $\neg \langle T_X \rangle$. 
5 Software model checking (13 points)

Consider the following code snippet $C$, where $x$, $y$ are integer variables.

```plaintext
1 assume $x + y > 0$ end
2 $x := x + y$
```

Remember that the Boolean abstraction of an `assume c end` statement is `assume not Pred (not c) end` followed by a parallel conditional assignment updating the predicates with respect to the original `assume` statement. $Pred(f)$ denotes the weakest under-approximation of the expression $f$ in terms of the given predicates.

5.1 Boolean abstractions (10 points)

Build the Boolean abstraction $A$ of the code snippet $C$ with respect to the following predicates:

- $p = x > 0$
- $q = y > 0$

Solution:

The abstraction is:

```plaintext
1 assume not (not p and not q) end
2 if (not p and not q) or p then p := True
3 elseif (not p and not q) or not p then p := False
4 else p := ? end
5 if (not p and not q) or q then q := True
6 elseif (not p and not q) or not q then q := False
7 else q := ? end
8
9 if p and q then p := True
10 elseif not p and not q then p := False
11 else p := ? end
12 if q then q := True
13 elseif not q then q := False
14 else q := ? end
```

After simplifications, we get:

```plaintext
1 assume p or q end
2 if not q then p := True end
3
4 if p and q then p := True end
```

5.2 Abstract and concrete traces (3 points)

Provide an annotated trace for the Boolean abstraction $A$, and a corresponding annotated trace for the concrete program $C$ which is feasible. Note that in general there are multiple traces of $C$ corresponding to the same trace of $A$: you must select one which is feasible.

The trace of $A$ should be in the form of a valid sequence of statements and branch conditions in $A$ which reaches the bottom of $A$. Each statement in the sequence must be preceded and followed by a complete description of the abstract program state in terms of values of the Boolean predicates $p$, $q$. Similarly, the trace of $C$ should be in the form of a valid sequence of statements.
and branch conditions in $C$ which reaches the bottom of $C$ without violating any assertion. Each statement in the sequence must be preceded and followed by a concrete value for the variables $x, y$ which satisfies the corresponding state in the abstract trace of $A$.

**Solution:**

1. $\{p, \text{not } q\}$
2. assume $p$ or $q$ end
3. $\{p, \text{not } q\}$
4. if not $q$ then $p := \text{True}$ end
5. $\{p, \text{not } q\}$
6. if $p$ and $q$ then $p := \text{True}$ end
7. $\{p, \text{not } q\}$

A matching concrete trace which is feasible is, for example, the following.

1. $\{x = 3, y = -1\}$
2. assume $x + y > 0$ end
3. 4
5. $\{x = 3, y = -1\}$
6. $x := x + y$
7. $\{x = 2, y = -1\}$
6 Termination proofs (13 points)

Consider the following implementation of binary search, where // denotes integer division.

\[ \text{binary_search} \ (v \ : \ G \ ; \ \text{list} \ : \ \text{LIST} \ [G] \ ; \ n \ : \ \text{INTEGER}) \ : \ \text{BOOLEAN} \]

\[ \text{-- Is 'v' contained in 'list' in the range [1..n]?} \]

\[ \text{require } n > 0 \text{ and } \text{list.is_sorted} \]

\[ \text{do} \]

\[ l := 1 \]

\[ u := n \]

\[ \text{Result} := \text{False} \]

\[ \text{until } l > u \]

\[ \text{loop} \]

\[ m := (l + u) \ // 2 \]

\[ \text{if } \text{list} \ [m] = v \text{ then} \]

\[ \text{-- Element found} \]

\[ \text{Result} := \text{True} \]

\[ l := u + 1 \]

\[ \text{elseif } \text{list} \ [m] > v \text{ then} \]

\[ \text{-- Continue search on left side} \]

\[ u := m - 1 \]

\[ \text{else} \]

\[ \text{-- Continue search on right side} \]

\[ l := m + 1 \]

\[ \text{end} \]

\[ \text{end} \]

\[ (1) \] Consider the loop invariant

\[ I \triangleq u - l + 1 \geq 0 \]

Find a suitable \textit{variant} function \( V \) which decreases along all branches of the loop body, and describe how \( V \) and \( I \) can be combined to prove that the loop always terminates. You do not have to provide a formal proof, but only to outline a termination argument for the given program with a suitable variant \( V \). (7 points)

\[ \text{Solution:} \]

A termination proof can be carried out using the variant

\[ V \triangleq u - l + 1 \]

Termination can be established from the observation that \( V \) decreases along each branch, because either \( u \) is decreased and \( l \) stays the same, or \( l \) is increased and \( u \) stays the same. The invariant \( I \) then guarantees that \( V \) has a lower bound, hence the loop must terminate when \( V \) reaches the lower bound.

\[ (2) \] Provide a proof that \( I \) is an invariant of the loop. For full credit, it is enough if you consider only the \textbf{else} branch of the conditional and prove invariance (consecution) along it. (6 points)

\[ \text{Solution:} \]

14
from
  \{ n > 0 \}
  \{ n - 1 + 1 = n \geq 0 \}
  l := 1
  u := n
Result := False
  \{ u - l + 1 \geq 0 \}
until \( l > u \)
loop
  \{ u - l + 1 \geq 0 \text{ and } l \leq u \}
  m := (l + u) // 2
  \{ u - l + 1 \geq 0 \text{ and } l \leq u \text{ and } \text{list}[m] = v \}
  Result := True
  \{ u - l + 1 \geq 0 \}
elseif \( \text{list}[m] > v \) then
  \{ m = (l + u) // 2 \text{ and } u - l + 1 \geq 0 \text{ and } l \leq u \text{ and } \text{list}[m] > v \}
  \{ m - 1 - l + 1 = m - l \geq 0 \}
  u := m - 1
  \{ u - l + 1 \geq 0 \}
else
  \{ m = (l + u) // 2 \text{ and } u - l + 1 \geq 0 \text{ and } l \leq u \text{ and } \text{list}[m] > v \}
  \{ u - m - 1 + 1 = u - m \geq 0 \}
  l := m + 1
  \{ u - l + 1 \geq 0 \}
end

To discharge the verification condition in the first branch of the elseif, notice that \( m = (l + u) // 2 \) implies \( u \geq 2m - l \), which combined with \( u - l + 1 \geq 0 \) implies \( 2m - l - 1 + 1 = 2m - l + 1 \geq 0 \). The latter also implies \( m - l \geq 0 \) because \( m, l \) are of integer type.

A similar reasoning discharges the verification condition in the second branch of the elseif: \( m = (l + u) // 2 \) implies \( -l \leq u - 2m \), which combined with \( u - l + 1 \geq 0 \) implies \( u + u - 2m + 1 = 2(u - m) + 1 \geq 0 \). The latter also implies \( u - m \geq 0 \) because \( m, u \) are of integer type.