CS243 Homework 7

Winter 2019

Due: March 13, 2019 at 4:30 pm

Directions:

- Submit via Gradescope.
- You may use up to two of your remaining late days for this assignment, for a late deadline of March 15, 2019 at 4:30 pm.
- Remember to complete the corresponding Gradiance quizzes by the start of class on the due date. **There are no late days for Gradiance.**
- This is an individual assignment. You are allowed to discuss the homework with others, but you must write the solution individually. If you look up any material in the textbook or online, you should cite it appropriately.
Problem 1. Pointer Analysis

Perform pointer analysis on the following Java snippet, and answer the following questions.

```java
public class A {
    public A foo() {
        return new A(); // h1
    }
}

public class B extends A {
    public A foo() {
        return new C(); // h2
    }
}

public class C extends A {
    public A foo() {
        return new A(); // h3
    }
}

public class Main {
    public static void main(String[] args) {
        A a = new B(); // h4
        a = a.foo();
    }
}
```

1. What is the true set of allocations that `a` can refer to at any point during the program’s execution.

2. What are the `pts` tuples inferred from this code in a context-insensitive, flow-insensitive pointer analysis? To get you started here’s one: `pts(a, h4).`
Problem 2. Binary Decision Diagrams

Draw the optimal (minimum number of nodes) BDD for the following expression:

$$\exists (x_0, (x_0 \ \text{XOR} \ x_2) \land (x_1 \ \text{XOR} \ x_3))$$

1. What is the variable order you chose to create this BDD?

2. What is the minimum number of nodes?

Problem 3. Path Sensitive Analysis With Satisfiability Modulo Theories

In this problem you will use an SMT solver to find test cases exhibiting a bug in the following C function:

```c
int func(int x, int[] data, int N) {
    int v, z, mid;
    if (0 <= x && x < N) {
        if (x <= N/2) {
            x = 2 * x;
        }
        v = data[x]; // line 7
        if (v >= 0 && v < N) {
            z = data[v]; // line 9
        } else {
            z = 0;
        }
        if (z >= 0 && z < x) {
            mid = (z + x) / 2;
            data[z] = data[mid]; // line 15
        }
        return data[data[z]]; // line 17
    } else {
        return data[0]; // line 19
    }
}
```

Assume we care about ‘crashes’ due to array out-of-bounds accesses (even if they don’t cause the program to segfault), and that data has length \(N\). First you will translate this function into a formula. Then you will run a solver and interpret its output. We will use the Z3 tool available at [https://rise4fun.com/z3/?menu=0](https://rise4fun.com/z3/?menu=0). There is a introductory guide at [rise4fun.com/Z3/tutorial/guide](http://rise4fun.com/Z3/tutorial/guide) which also serves as a language reference sufficient for our purposes. The full language specification is available at [http://smtlib.cs.uiowa.edu/papers/smt-lib-reference-v2.6-r2017-07-18.pdf](http://smtlib.cs.uiowa.edu/papers/smt-lib-reference-v2.6-r2017-07-18.pdf).
Your submission will consist of the input to Z3 (copy and pasted from the web editor), the output from running it (sat or unsat, and, if sat, the produced model), along with comments explaining your interpretation of the result. Follow these steps to complete this problem:

1. **Rewrite the program in SSA form.** The first step is to write the function in Static Single Assignment form by assigning all definitions a unique suffix. You need to generate phi-nodes for each branch. At each join point in the CFG you will need to introduce new definitions for the variables that are defined on either path. At this point, the program is still imperative code, but each variable is defined exactly once.

2. **Translation to SMT.** The second step is to translate into an SMT formula. An assignment $x_3 = e$ becomes an assertion \(\text{assert (= x_3 E)}\) in SMT, where \(E\) is the translation of \(e\).

You need to translate \texttt{int} operations at the C level into bit vector operations at the SMT level. Do not use \texttt{Int} in Z3, as these model mathematical integers. Assume 32-bit 2's complement representation for \texttt{int}.

For example, the translation of $x = y + 1$ is:

```z3
(declare-const x (_ BitVec 32))
(declare-const y (_ BitVec 32))
(assert (= x (bvadd y #x00000001)))
(check-sat)
(get-model)
```

Z3 responds with:

```
sat
(model
 (define-fun y () (_ BitVec 32) #x00000000)
 (define-fun x () (_ BitVec 32) #x00000001)
 )
)```

Here the program is satisfiable with model $x = 1$, $y = 0$. Note that the variables $x$ and $y$ are given as functions of no arguments (which must be constants because there are no side effects), and the constants themselves are hexadecimal.

To translate arrays, you will use variables of the sort \((\text{Array (_ BitVec 32) (_ BitVec 32)})\). Array dereferences like \texttt{data[i]} become \(\text{select data i}\) when translating a read, and \(\text{store data i x}\) when translating a write. Note that \(\text{store data i x}\) returns a new array, whose \(i\)'s element is now equal to \(x\), and does not modify the original array.

To translate phi-nodes, you must use a logical expression that captures the condition under which the phi node is evaluated. For example, given the following code in SSA form:
if (c) {
  b1:
    x1 = ...;
} else {
  b2:
    x2 = ...;
}
x3 = phi(x1 from b1, x2 from b2);

the translation of x3 is (ite c x1 x2). ite is short for if-then-else, and evaluates to the second or third argument based on the first.

3. **Bounds checks.** The final step is to add an assertion to check each of the three particular array accesses. You need to check that the signed value of the index is in bounds. Further, not all accesses are accessible on all paths, so you need to guard the assertion for a particular access. The assertion should express the execution reaches this access, and it is out of bounds. Note that this will possibly constrain some path variables if the access is nested inside an if statement, for example. You should use the sequence (push)(assert C)(check-sat)(pop) for each access, where C is the check for that access. Push/pop allows us to add C to our set of assertions, check satisfiability, and then remove it to add a different C. If you add all the assertions together you will find a path which crash all points simultaneously rather than just at least one.

4. **Interpretation.** Once you find one or more bug, add (get-model) after (check-sat) within the push/pop sequence for each satisfiable assertion, to print the model found for that bug.

In writing, interpret the results. Does the result indicate a crash can occur on some concrete path? If not, does this mean there can be no crash for this access? If there is a crash, translate the model into a concrete input represented by a call data = ...; func(...) which causes a crash at the corresponding access. Hint: what happens if the program has more than one bug, and how should the model be interpreted in that case?

5. **Submission.** Provide the Z3 input and output in your homework submission.