Outline

Overview

Program understanding

Program verification and property checking

Bug finding

How static analysis works

Dynamic analysis: taint tracking

Summary
Recap

We’re looking at

- principles and tools

for ensuring software security.

This lecture looks at:

- further example uses of static analysis
- an example of dynamic analysis for information flow
- some hints about how static analysis works
Advanced static analysis jobs

Static analysis is used for a range of tasks that are useful for ensuring secure code.

Basic tasks include **type checking** and **style checking**, described last lecture.

More advanced tasks are:

- **Program understanding**: inferring meaning
- **Property checking**: ensuring no bad behaviour
- **Program verification**: ensuring correct behaviour
- **Bug finding**: detecting likely errors
Program understanding tools

Help developers understand and manipulate large codebases.

- Navigation swiftly inside the code
  - finding definition of a constant
  - finding call graph for a method
- Support *refactoring* operations
  - re-naming functions or constants
  - move functions from one module to another
  - needs internal model of whole code base
- Inferring *design* from *code*
  - Reverse engineer or check informal design

**Outlook:** may become increasingly used for security review, with dedicated tools. Close relation to tools used for malware analysis (reverse engineering).
Commercial example: Structure101
Research example: Fujaba and Reclipse
How Reclipse works

We’ll explain some of these processes later.

See Fujaba project at University of Paderborn
Outline

Overview

Program understanding

Program verification and property checking

Bug finding

How static analysis works

Dynamic analysis: taint tracking

Summary
Program verification

- The gold standard, ultimate guarantee
- Uses **formal methods** techniques, e.g.,
  - theorem proving
  - model checking
- Drawback: needs precise **formal specification** to verify against
- Very expensive to industry
  - time consuming
  - needs experts (logic/maths)
- Currently only used in safety critical domains
  - e.g., railway, nuclear, aeronautics
  - emerging: automobile, security

Example tools: SPARK, Event-B. See also general purpose **interactive theorem provers**. Many other research-quality and/or unmaintained tools.
Property checking

*Lightweight formal methods*

- Make specifications be *standard* and *generic*
- This program cannot raise NullPointerException
- All database connections are closed after use

**Static checking** (not verification)

- Prevent many violations of specification, not all
- May produce *counterexamples* to explain violations
- Chain pre-conditions (*requires*) and post-conditions (*ensures*)
  - Allows *inter-procedural* analysis

Examples: Code Contracts, Splint, JML, Grammatech CodeSonar, PolySpace, ThreadSafe, PRQA, Facebook Infer.
Assertion checking

Many languages have support for assertions. These are dynamic (runtime) checks that can be used to test properties the programmer expects to be true.

```
assert(exp)
```

- fails if exp evaluates to false
- assertion tests **usually disabled**
  - treated as comments
  - may be enabled for testing during development
  - or when running unit tests

**Question.** What is the risk with running tests only with assertions enabled?
private static int addHeights(int ah, int bh) {
    assert ah > 0 && bh > 0 : "parameters should be positive";
    return ah+bh;
}

Notice above method is private.

- API (public) functions should *always* test constraints
  - throw exceptions if not met
  - eliminate clients (or attackers) who break API contract
- Internal functions may rely on local properties
  - if maintained in same class, easier to check/ensure
Assertions for security

We could use assertions as safety checks for functions that are at risk of being used in a buggy way.

```c
assert(alloc_size(dest) > strlen(src));
strcpy(dest, src);
```

[alloc_size() is not a standard C function, but GCC, for example, has support for trying to track the size of allocated functions with function attributes]
From dynamic to static

With static analysis, we *may* be able to automatically determine whether assertions (if enabled) will:

1. always succeed
2. may sometimes fail (unknown)
3. will always fail

Easy cases:

1. `assert(true);`
2. `x=readint(); assert(x>0);`
3. `assert(false);`

The perfect case would be showing that assertions in a program can only succeed: thus they do not need to be checked dynamically.

**Question.** what troubles can you see with case 2?
Reasoning with assertions

How does a static analyser reason?

Computations about assertions can be chained through the program, using a program logic inside the tool.

E.g., build up a set of facts known before each statement:

```
x = 1;       // { }   (nothing known)
y = 1;       // { x = 1 }
assert (x < y);  // { x = 1, y = 1 }
```

```
assert (x < y);  // FAIL
```
Symbolic evaluation

This can work also with variables, whose value is not known statically:

```
x = z;         // { }  (nothing known)
y = z+1;       // { x = z }
assert (x < y); // { x = z, y = z+1 }
```

```
// SUCCEED  (provided z<MAXINT)
```
Conditionals and loops

These make static analysis much harder, of course.

```
x = v; // {x=v}
if (x < y) // {x=v, x<y}
y = v;    // {x=v, y=v}: FAIL
assert (x < y) // Or: {x=v,¬(x<y)}: FAIL
```

For conditionals, we need to either

▶ explore every path
▶ merge information at join-points

For loops, we need to either

▶ unroll for a finite number of iterations
▶ capture variation using logical invariants
Using logical (or other) reasoning techniques, there are various different types of assertions that are useful for security checking, for example:

- **Bounds and range analysis**
- **Tainted data analysis**
- **Type state** and **Resource** tracking

**Exercise.** What kinds of security issues can these assertions help with? What kinds of security issues would need other assertions?
Bound/range Analysis

\textbf{alloc\_size}(dest) > strlen(src)

\textbf{array\_size}(a) > n before a[n] access

- Check integers are in required ranges
Taintedness

**tainted**(mypageinput)

**untainted**(newkey)

- Tracks whether data can be affected by adversary.
- Tainted input shouldn’t be used for security sensitive choices
- and should be sanitized before being output
- Taint analysis approximates information flow
  - information may be leaked *indirectly* as well as directly
Type State (Resource) Tracking

- `isnull(ptr)`, `nonnull(ptr)`
- `isopen_for_read(handle)`, `isclosed(handle)`
- `uninitialized(buffer)`, `terminatedstring(buffer)`
  - Tracks status of data value held by a variable
  - Helps enforce API usage contracts to avoid errors
    - e.g., DoS
  - Usage/lifecycle may be expressed with automaton
Example: avoiding double-free errors

1. **Start**
   - Initial state
   - Other operations

2. **Free(x)**
   - Freed
   - Other operations

3. **Free(x)**
   - Error
Null Pointers in CodeSonar

Not all null pointer analyses are equal! Some compilers spot only “obvious” null pointer risks, others perform deeper analysis like CodeSonar. IDE analysis may be in between.
For Java, there is a language called JML which adds similar pre- and post-conditions (requires/ensures). Open source JML toolsets have been through several versions but have had trouble keeping up with Java, Eclipse changes.
Outline

Overview

Program understanding

Program verification and property checking

Bug finding

How static analysis works
  Dynamic analysis: taint tracking

Summary
Bug finding tools look for suspicious patterns in code. **FindBugs** is an example:

- Finds possible Java bugs according to *rules*
  - rules are suspicious patterns in code
  - designed by experience of buggy programs
  - ... collected from real world and student(!) code

- Warnings are categorized by
  - **severity**: how serious in general the problem is
  - **confidence**: tool’s belief of true problem
Example bugs

Common accidents
An error found in Sun’s JDK 1.6:

```java
public String foundType() {
    return this.foundType();
}
```

Misunderstood APIs

```java
public String makeUserid(String s) {
    s.toLowerCase();
    return s;
}
```
if (this.fitz == null) {
    synchronized (mylock) {
        if (this.fitz == null) {
            this.fitz = new Fitzer();
        }
    }
}

[dice]da: findbugs Fitz.class
M M DC: Possible doublecheck on Fizz.fitz in Fitz.getFitz()
    At Fitz.java:[lines 1-3]
Findbugs GUI

Problem:

Method may fail to close stream

At Util.java [line 108]

In method edu.umd.cs.findbugs.util.Util.getXMLType(InputStream) [Lines 102 - 123]

Need to close java.io.Reader

Description:

The method creates an IO stream object, does not assign it to any fields, pass it to other methods that might close it, or return it, and does not appear to close the stream on all paths out of the method. This may result in a file descriptor leak. It is generally a good idea to use a finally block to ensure that streams are closed.

http://findbugs.sourceforge.net/
Clang Static Analyser

An open source tool for C, C++, Objective-C included in XCode.
Clang Static Analyser HTML reports

openssl-1.0.0 - scan-build results

<table>
<thead>
<tr>
<th>User:</th>
<th>user@localhost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working Directory:</td>
<td>/home/user/Exercise-4/openssl-1.0.0</td>
</tr>
<tr>
<td>Command Line:</td>
<td>make</td>
</tr>
<tr>
<td>Clang Version:</td>
<td>clang version 3.4 (tags/RELEASE_34/final)</td>
</tr>
<tr>
<td>Date:</td>
<td>Fri Jan 17 12:03:31 2014</td>
</tr>
</tbody>
</table>

Bug Summary

<table>
<thead>
<tr>
<th>Bug Type</th>
<th>Quantity</th>
<th>Display?</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Bugs</td>
<td>269</td>
<td>✔️</td>
</tr>
<tr>
<td>API</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argument with 'nonnull' attribute passed null</td>
<td>7</td>
<td>✔️</td>
</tr>
<tr>
<td>Dead store</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dead assignment</td>
<td>203</td>
<td>✔️</td>
</tr>
<tr>
<td>Dead increment</td>
<td>11</td>
<td>✔️</td>
</tr>
<tr>
<td>Dead initialization</td>
<td>2</td>
<td>✔️</td>
</tr>
<tr>
<td>Logic error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned value is garbage or undefined</td>
<td>3</td>
<td>✔️</td>
</tr>
<tr>
<td>Branch condition evaluates to a garbage value</td>
<td>1</td>
<td>✔️</td>
</tr>
<tr>
<td>Dereference of null pointer</td>
<td>30</td>
<td>✔️</td>
</tr>
<tr>
<td>Division by zero</td>
<td>1</td>
<td>✔️</td>
</tr>
<tr>
<td>Result of operation is garbage or undefined</td>
<td>7</td>
<td>✔️</td>
</tr>
<tr>
<td>Uninitialized argument value</td>
<td>4</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Reports

<table>
<thead>
<tr>
<th>Bug Group</th>
<th>Bug Type</th>
<th>File</th>
<th>Line</th>
<th>Path Length</th>
<th>View Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Argument with 'nonnull' attribute passed null</td>
<td>ssl/d1_both.c</td>
<td>1015</td>
<td>9</td>
<td>View Report</td>
</tr>
<tr>
<td>API</td>
<td>Argument with 'nonnull' attribute passed null</td>
<td>ssl/d1_srvr.c</td>
<td>1184</td>
<td>10</td>
<td>View Report</td>
</tr>
<tr>
<td>API</td>
<td>Argument with 'nonnull' attribute passed null</td>
<td>ssl/s3_srvr.c</td>
<td>1725</td>
<td>10</td>
<td>View Report</td>
</tr>
<tr>
<td>API</td>
<td>Argument with 'nonnull' attribute passed null</td>
<td>crypto/asn1/a_bytes.c</td>
<td>295</td>
<td>21</td>
<td>View Report</td>
</tr>
</tbody>
</table>
Outline

Overview

Program understanding

Program verification and property checking

Bug finding

How static analysis works

Dynamic analysis: taint tracking

Summary
Basic overview

Source Code → Build Model → Perform Analysis → Present Results

Security Knowledge
Building a program model

Starts off like a compiler, in stages. Simpler/older static analysis tools only use first stages.

1. **Lexical analysis**: tokenise input
2. **Parsing**: builds a *parse tree* from grammar
3. **Abstract Syntax Tree**: simplify parse tree
4. **Semantic analysis**
   - check program well-formedness
   - including *type-checking*
5. Produce an **Intermediate Representation (IR)**
   - higher level than for compiler
6. Produce **model** to capture control/data flows
   - *control-flow* and *call graphs*
   - variable-contains-data relationships
   - pointer analysis: aliasing, points-to
Control flow graphs

```java
if (a > b) {
    nConsec = 0;
} else {
    s1 = getHexChar(1);
    s2 = getHexChar(2);
}
return nConsec;
```

The CFG consists of *basic blocks* and the paths between them.
A trace is a possible sequence of basic blocks.
Above: [bb0,bb1,bb3] and [bb0,bb2,bb3].

Traces can be used to check against security constraints (e.g., as automata), to construct counterexamples. The CFG is also used to combine/chain assertions.
Call graphs

int a(int x) {
    if (x) { b(1); } else { c(); }
}
int b(int y) {
    if (y) { c(); b(0); } else { c(); }
}
int c() { /* empty */ }

Call graphs are used for *inter-procedural* analysis
- Check requires-ensures contracts connect together
Putting them together: local and global
Outline

Overview

Program understanding

Program verification and property checking

Bug finding

How static analysis works

  Dynamic analysis: taint tracking

Summary
Dynamic taint tracking

Idea: add security labels to data inputs (sources) and data outputs (sinks). Propagate labels during computation (cf dynamic typing).

Labels are:

**Tainted**

- Data from *taint sources* (e.g., user input)
- Data arising from or influenced by tainted data

**Untainted**

- Data that is safe to output or use in sensitive ways
Stopping tainted data being stored

Legend:
- Untainted
- Tainted

Source → \text{sanitization} → Sink

$db->query$ Security violation!
Preventing jumps to tainted addresses

See Schwartz, Avgerinos, Brumley, *All You Ever Wanted to Know About Dynamic Taint Analysis and Forward Symbolic Execution (but might have been afraid to ask)*, IEEE Security and Privacy, 2010. This paper explains tainting with a simple operational semantics.

<table>
<thead>
<tr>
<th>Line #</th>
<th>Statement</th>
<th>Δ</th>
<th>τΔ</th>
<th>Rule</th>
<th>pc</th>
</tr>
</thead>
<tbody>
<tr>
<td>start</td>
<td></td>
<td>{}</td>
<td>{}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>x := 2*get_input()</td>
<td>{x → 40}</td>
<td>{x → T}</td>
<td>T-ASSIGN</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>y := 5 + x</td>
<td>{x → 40, y → 45}</td>
<td>{x → T, y → T}</td>
<td>T-ASSIGN</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>goto y</td>
<td>{x → 40, y → 45}</td>
<td>{x → T, y → T}</td>
<td>T-GOTO</td>
<td>error</td>
</tr>
</tbody>
</table>
Taintdroid: notifying dynamic leaks on Android

Taintdroid uses a modification of the Android framework to track data flows at runtime. See the demo video.
Drawbacks of the dynamic method

Preventing code injection exploits using dynamic taint tracking is like letting a thief in your house and checking his bag for stolen goods at the very moment he tries to leave. It might work, but only if you never lose track of the gangster and if you really know your house. However, I would prefer a solution that does not let thieves in my house in the first place.

Analogy by Martin Johns used to explain dynamic taint tracking, 2007
Another drawback: implicit flows

- Simple dynamic tracking only captures direct flows
- To spot implicit flows, need to monitor every path
- Not only the ones actually taken by the program!
- Quickly impractical without severely pruning
  - special techniques like forward symbolic execution
- Partial solution: type checking for information flow
  - and recent hybrid dynamic-static methods
Outline

Overview

Program understanding

Program verification and property checking

Bug finding

How static analysis works
  Dynamic analysis: taint tracking

Summary
Take away points

Program analysis tools can help find security flaws.

- static: examine millions of lines, repeatedly
- dynamic: equip code with self-checking

Some tools are generic bug finding, built into IDE.
Others are specific to security, may include:

- risk analysis, including impact/likelihood
- issue/requirements tracking, metrics

Expect these to become mainstream

- current frequency of security errors unacceptable
- incentives will eventually affect priorities
Some of this lecture is based Chapters 2-4 of

- *Secure Programming With Static Analysis* by Brian Chess and Jacob West, Addison-Wesley 2007.

Recommended reading:

- Al Bessey et al. *A few billion lines of code later: using static analysis to find bugs in the real world*, CACM 53(2), 20101.