Software Analysis Tools

17-313 Spring 2025

Foundations of Software Engineering

https://cmu-313.github.io

Michael Hilton, Austin Henley, and Nadia Nahar



Administrivia

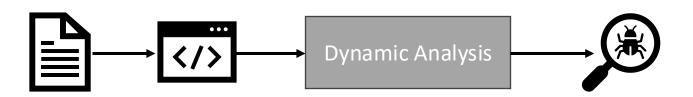
- P3A due Thursday
- Access your VMs?

Learning Goals

- Gain an understanding of the relative strengths and weaknesses of static and dynamic analysis
- Examine several popular analysis tools and understand their use cases
- Understand how analysis tools are used in large open-source software

What are Program Analysis Tools?





```
src/controllers/accounts/posts.is r
.f.. Show 135 more lines
137 ...
                    },
138 ...
139
140 ...
                  postsController.getBookmarks = async function (req, res, next) {
141 ...
                      await getPostsFromUserSet('account/bookmarks', req, res, o next);
                   This function expects 3 arguments, but 4 were provided
142 ...
143
144 ...
                  postsController.getPosts = async function (req, res, next) {
145 ...
                     await getPostsFromUserSet('account/posts', req, res, next);
```



What static analysis can and cannot do

- Type-checking is well established
 - Set of data types taken by variables at any point
 - Can be used to prevent type errors (e.g. Java) or warn about potential type errors (e.g. Python)
- Checking for problematic patterns in syntax is easy and fast
 - Is there a comparison of two Java strings using `==`?
 - Is there an array access `a[i]` without an enclosing bounds check for `i`?
- Reasoning about termination is impossible in general
 - Halting problem
- Reasoning about exact values is hard, but conservative analysis via abstraction is possible
 - Is the bounds check before `a[i]` guaranteeing that `i` is within bounds?
 - Can the divisor ever take on a zero value?
 - Could the result of a function call be `42`?
 - Will this multi-threaded program give me a deterministic result?
 - Be prepared for "MAYBE"
- Verifying some advanced properties is possible but expensive
 - CI-based static analysis usually over-approximates conservatively



Static Analysis is well suited to detecting certain defects

- Security: Buffer overruns, improperly validated input...
- Memory safety: Null dereference, uninitialized data...
- Resource leaks: Memory, OS resources...

Static Analysis: Broad classification

- Linters
 - Shallow syntax analysis for enforcing code styles and formatting
- Pattern-based bug detectors
 - Simple syntax or API-based rules for identifying common programming mistakes
- Type-annotation validators
 - Check conformance to user-defined types
 - Types can be complex (e.g., "Nullable")
- Data-flow analysis / Abstract interpretation)
 - Deep program analysis to find complex error conditions (e.g., "can array index be out of bounds?")



Static analysis can be applied to all attributes

- Find bugs
- Refactor code
- Keep your code stylish!
- Identify code smells
- Measure quality
- Find usability and accessibility issues
- Identify bottlenecks and improve performance



Activity: Analyze the Python program statically

```
def n2s(n: int, b: int):
 if n <= 0: return '0'
 r = ''
 while n > 0:
  u = n \% b
  if u >= 10:
   u = chr(ord('A') + u-10)
  n = n // b
  r = str(u) + r
 return r
```

- 1. What is the type of variable `u`?
- 2. Will the variable `u` be a negative number?
- 3. Will this function always return a value?
- 4. Will the program divide by zero?
- 5. Will the returned value ever contain a minus sign '-'?

Dynamic analysis reasons about program executions

- Tells you properties of the program that were definitely observed
 - Code coverage
 - Performance profiling
 - Type profiling
 - Testing
- In practice, implemented by program instrumentation
 - Think "Automated logging"
 - Slows down execution speed by a small amount

Static Analysis vs Dynamic Analysis

- Requires only source code
- Conservatively reasons about all possible inputs and program paths
- Reported warnings may contain false positives
- Can report all warnings of a particular class of problems
- Advanced techniques like verification can prove certain complex properties, but rarely run in Cl due to cost

- Requires successful build + test inputs
- Observes individual executions
- Reported problems are real, as observed by a witness input
- Can only report problems that are seen. Highly dependent on test inputs. Subject to false negatives
- Advanced techniques like symbolic execution can prove certain complex properties, but rarely run in CI due to cost

Static Analysis



Tools for Static Analysis

















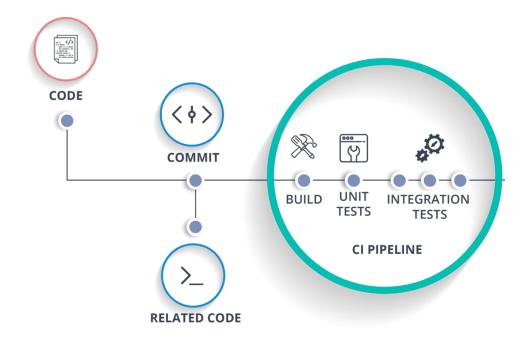








Static analysis is a key part of continuous integration









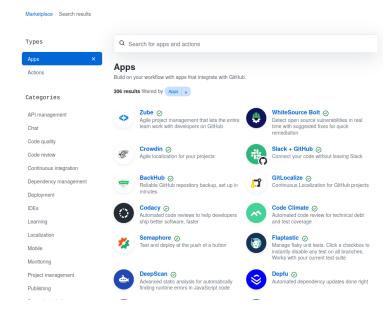


Static analysis used to be an academic amusement; now it's heavily commercialized

GitHub acquires code analysis tool Semmle









Snyk Secures \$150M, Snags \$1B Valuation



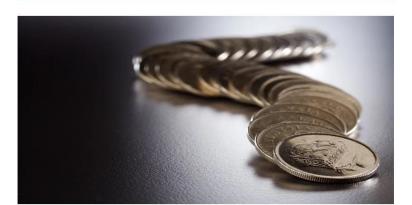












Snyk, a developer-focused security startup that and identifies vulnerabilities in open source applications, announced a \$150 million Series C funding round today. This brings the company's total investment to \$250 million alongside reports that put the company's valuation at more than \$1 billion.



Static analysis is also integrated into IDEs











```
cppcoreguidelines.cpp
      -// To enable only C++ Core Guidelines checks
       // go to Settings/Preferences | Editor | Inspections | C/C++ | Clang-Tidy
      △// and provide: -*,cppcoreguidelines-* in options
       void fill_pointer(int* arr, const int num) {
            for(int i = 0; i < num; ++i) {</pre>
                arr[i] = 0;
    Do not use pointer arithmetic
10
11
       void fill_array(int ind) {
12
            int arr[3] = \{1,2,3\};
13
            arr[ind] = 0;
14
15
        void cast_away_const(const int& magic_num)
16
17
18
            const_cast<int&>(magic_num) = 42;
19
20
```

```
content: item,
conten
```



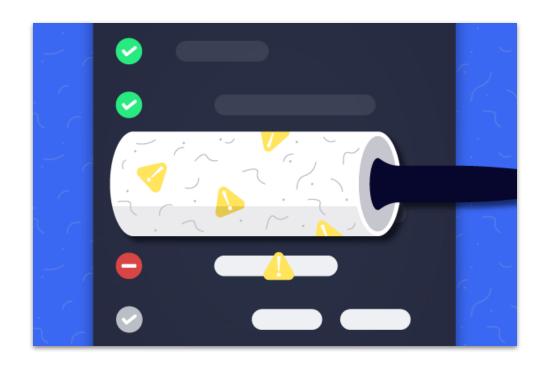
What makes a good static analysis tool?

- Static analysis should be fast
 - Don't hold up development velocity
 - This becomes more important as code scales
- Static analysis should report **few false positives**
 - Otherwise developers will start to ignore warnings and alerts, and quality will decline
- Static analysis should be continuous
 - Should be part of your continuous integration pipeline
 - Diff-based analysis is even better -- don't analyze the entire codebase; just the changes
- Static analysis should be **informative**
 - Messages that help the developer to quickly locate and address the issue
 - Ideally, it should suggest or automatically apply fixes





(1) Linters: Cheap, fast, and lightweight static source analysis





Use linters to enforce style guidelines

Don't rely on manual inspection during code review!













Linters use very "shallow" static analysis to enforce formatting rules

- Ensure proper indentation
- Naming convention
- Line sizes
- Class nesting
- Documenting public functions
- Parenthesis around expressions
- What else?



Use linters to improve maintainability

- Why? We spend more time reading code than writing it.
 - Various estimates of the exact %, some as high as 80%
- Code is ownership is usually shared
- The original owner of some code may move on
- Code conventions make it easier for other developers to quickly understand your code



Use Style Guidelines to facilitate communication







Guidelines are inherently opinionated, but **consistency** is the important point. Agree to a set of conventions and stick to them.

https://www.chicagomanualofstyle.org/ https://google.github.io/styleguide/ https://www.python.org/dev/peps/pep-0008





Take Home Message:

Style is an easy way to improve readability

- Everyone has their own opinion (e.g., tabs vs. spaces)
- Agree to a convention and stick to it
 - Use continuous integration to enforce it
- Use automated tools to fix issues in existing code



(2) Patten-based Static Analysis Tools



- Bad Practice
- Correctness
- Performance
- Internationalization
- Malicious Code
- Multithreaded Correctness
- Security
- Dodgy Code



Bad Practice:

```
String x = new String("Foo");
String y = new String("Foo");

if (x == y) {
    System.out.println("x and y are the same!");
} else {
    System.out.println("x and y are different!");
}
```

Bad Practice: ES_COMPARING_STRINGS_WITH_EQ Comparing strings with ==

```
String x = new String("Foo");
String y = new String("Foo");

if (x == y) {
  if (x.equals(y)) {
    System.out.println("x and y are the same!");
  } else {
    System.out.println("x and y are different!");
}
```

How do we check for equality in JS?

```
if (a = b) {
if (a == b) {
if (a === b) {
if (Object.is(a, b)) {
```

Performance:

```
public static String repeat(String string, int times)
{
   String output = string;
   for (int i = 1; i < times; ++i) {
      output = output + string;
   }
   return output;
}</pre>
```

Performance: SBSC_USE_STRINGBUFFER_CONCATENATION Method concatenates strings using + in a loop

```
public static String repeat(String string, int times)
{
   String output = string;
   for (int i = 1; i < times; ++i) {
      output = output + string;
   }
   return output;
   The method seems to be building a String iteration, the String is converted to a converted back to a String. This can let
      converted back to a String this can let
      converted back to
```

The method seems to be building a String using concatenation in a loop. In each iteration, the String is converted to a StringBuffer/StringBuilder, appended to, and converted back to a String. This can lead to a cost quadratic in the number of iterations, as the growing string is recopied in each iteration.

Performance: SBSC_USE_STRINGBUFFER_CONCATENATION Method concatenates strings using + in a loop

```
public static String repeat(String string, int times)
{
  int length = string.length() * times;
  StringBuffer output = new StringBuffer(length);
  for (int i = 0; i < times; ++i) {
    output.append(string);
  }
  return output.toString();
}</pre>
```

Correctness:

```
@Override
public Connection getConnection() throws SQLException {
QwicsConnection con = new QwicsConnection(host, port);
try {
 con.open();
} catch (Exception e) {
  new SQLException(e);
 return con;
```

Correctness: Missing "throw" before "new Exception"

```
@Override
public Connection getConnection() throws SQLException {
QwicsConnection con = new QwicsConnection(host, port);
try {
 con.open();
} catch (Exception e) {
 throw new SQLException(e);
return con;
```

Challenges with pattern-based static analysis

- The analysis must produce zero false positives
 - Otherwise developers won't be able to build the code!
- The analysis needs to be really fast
 - Ideally < 100 ms
 - If it takes longer, developers will become irritated and lose productivity
 - Practically, this means the analysis needs to focus on "shallow" bugs rather than verifying some complex logic spanning multiple functions/classes.
- You can't just "turn on" a particular check
 - Every instance where that check fails will prevent existing code from building
 - There could be thousands of violations for a single check across large codebases



(3) Use type annotations to detect common errors

- Uses static types to prevent meaningless operations from executing in the first place (instead of dealing with bad results later)
- Annotations can enhance type system already in the language
- Examples: Java Checker Framework, MyPy, TypeScript







Example: Detecting null pointer exceptions

- **@Nullable** indicates that an expression may be null
- @NonNull indicates that an expression must never be null
- Guarantees that expressions annotated with @NonNull will never evaluate to null, forbids other expressions from being dereferenced

```
// return value
@NonNull String toString() { ... }

// parameter
int compareTo(@NonNull String other)
{ ... }
```

```
import org.checkerframework.checker.nullness.qual.*;
public class NullnessExampleWithWarnings {
  public void example() {
     @NonNull String foo = "foo";
                                        @Nullable is applied by
                                        default
     String bar = null;
     foo = bar; -
                                Error: [assignment.type.incompatible] incompatible types in assignment.
    println(foo.length());
                                 found: @Initialized @Nullable String
                                  required: @UnknownInitialization @NonNull String
```

```
import org.checkerframework.checker.nullness.qual.*;
public class NullnessExampleWithWarnings {
 public void example() {
  @NonNull String foo = "foo";
  String bar = null; // @Nullable
                                bar is refined to
  if (bar != null) {
                                @NonNull
   foo = bar;
  println(foo.length());
```

Another example: Units checker

- Guarantees that operations are performed on the same kinds and units
- Kind annotations
 - @Acceleration, @Angle, @Area, @Current, @Length, @Luminance, @Mass, @Speed, @Substance, @Temperature, @Time
- SI unit annotation
 - @m, @km, @mm, @kg, @mPERs, @mPERs2, @radians, @degrees @A, ...







Remember the Mars Climate Orbiter incident from 1999?



NASA's Mars Climate Orbiter (cost of \$327 million) was lost because of a discrepancy between use of metric unit Newtons and imperial measure Pound-force.

```
import static org.checkerframework.checker.units.UnitsTools.m;
import static org.checkerframework.checker.units.UnitsTools.mPERs;
import static org.checkerframework.checker.units.UnitsTools.s;
void demo() {
 @m int x;
x = 5 * m;
 @m int meters = 5 * m;
 @s int seconds = 2 * s;
 @mPERs int speed = meters / seconds;
 @m int foo = meters + seconds;
 @s int bar = seconds - meters;
```

```
import static org.checkerframework.checker.units.UnitsTools.m;
import static org.checkerframework.checker.units.UnitsTools.mPERs;
import static org.checkerframework.checker.units.UnitsTools.s;
                                         @m indicates that x represents meters
void demo() {
 @m int x;
x = 5 * m;
                                                        To assign a unit, multiply appropriate
                                                        unit constant from UnitTools
 @m int meters = 5 * m;
 @s int seconds = 2 * s;
 @mPERs int speed = meters / seconds;
 @m int foo = meters + seconds;
 @s int bar = seconds - meters;
```

Does this program compile?

```
import static org.checkerframework.checker.units.UnitsTools.m;
import static org.checkerframework.checker.units.UnitsTools.mPERs;
import static org.checkerframework.checker.units.UnitsTools.s;
                                         @m indicates that x represents meters
void demo() {
 @m int x;
x = 5 * m;
                                                        To assign a unit, multiply appropriate
                                                        unit constant from UnitTools
 @m int meters = 5 * m;
 @s int seconds = 2 * s;
 @mPERs int speed = meters / seconds;
 @m int foo = meters + seconds;
 @s int bar = seconds - meters;
```

Does this program compile? No.

```
import static org.checkerframework.checker.units.UnitsTools.m;
import static org.checkerframework.checker.units.UnitsTools.mPERs;
import static org.checkerframework.checker.units.UnitsTools.s;
```

```
void demo() {
  @m int x;
  x = 5 * m;

@m int meters = 5 * m;
  @s int seconds = 2 * s;

@mPERs int speed = meters / seconds;
  @m int foo = meters + seconds;
  @s int bar = seconds - meters;
```

Addition and subtraction between meters and seconds is physically meaningless

Limitations of Type-based Static Analysis

- Can only analyze code that is annotated
 - Requires that dependent libraries are also annotated
 - Can be tricky to retrofit annotations into existing codebases
- Only considers the signature and annotations of methods
 - · Doesn't look at the implementation of methods that are being called
- Can't handle dynamically generated code well
 - Examples: Spring Framework, Templates
- Can produce false positives!
 - Byproduct of necessary approximations



The best QA strategies employ a combination of tools

How Many of All Bugs Do We Find? A Study of Static Bug Detectors

Andrew Habib andrew.a.habib@gmail.com Department of Computer Science TU Darmstadt Germany

ABSTRACT

Static bug detectors are becoming increasingly popular and are widely used by professional software developers. While most work on bug detectors focuses on whether they find bugs at all, and on how many false positives they report in addition to legitimate warnings, the inverse question is often neglected: How many of all real-world bugs do static bug detectors find? This paper addresses this question by studying the results of applying three widely used static bug detectors to an extended version of the Defects4J dataset that consists of 15 Java projects with 594 known bugs. To decide which of these bugs the tools detect, we use a novel methodology that combines an automatic analysis of warnings and bugs with a manual validation of each candidate of a detected bug. The results of the study show that: (i) static bug detectors find a non-negligible amount of all bugs, (ii) different tools are mostly complementary to each other, and (iii) current bug detectors miss the large majority of the studied bugs. A detailed analysis of bugs missed by the static detectors shows that some bugs could have been found by variants of the existing detectors, while others are domain-specific problems that do not match any existing bug pattern. These findings help potential users of such tools to assess their utility, motivate and outline directions for future work on static bug detection, and provide a basis for future comparisons of static bug detection with other bug finding techniques, such as manual and automated testing.

Michael Pradel michael@binaervarianz.de Department of Computer Science TU Darmstadt Germany

International Conference on Automated Software Engineering (ASE '18), September 3–7, 2018, Montpellier, France. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3238147.3238213

1 INTRODUCTION

Finding software bugs is an important but difficult task. For average industry code, the number of bugs per 1,000 lines of code has been estimated to range between 0.5 and 25 [21]. Even after years of deployment, software still contains unnoticed bugs. For example, studies of the Linux kernel show that the average bug remains in the kernel for a surprisingly long period of 1.5 to 1.8 years [8, 24]. Unfortunately, a single bug can cause serious harm, even if it has been subsisting for a long time without doing so, as evidenced by examples of software bugs that have caused huge economic loses and even killed people [17, 28, 46].

Given the importance of finding software bugs, developers rely on several approaches to reveal programming mistakes. One approach is to identify bugs during the development process, e.g., through pair programming or code review. Another direction is testing, ranging from purely manual testing over semi-automated testing, e.g., via manually written but automatically executed unit tests, to fully automated testing, e.g., with UI-level testing tools. Once the software is deployed, runtime monitoring can reveal so far missed bugs. e.g., collect information about abnormal runtime

Tool	Bugs
Error Prone	8
Infer	5
SpotBugs	18
Total:	31
Total of 27 unique bugs	

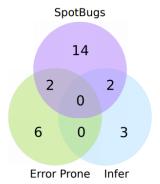


Figure 4: Total number of bugs found by all three static checkers and their overlap.

Which tool to use?

- Depends on use case, available resources
- Linters: Fast, cheap, easy to address issues or set ignore rules
- Pattern-based bugs: Intuitive, but need to deal with false positives.
- **Type-annotation-based checkers**: More manual effort required; needs overall project commitment. But good payoff once adopted.
- **Deep analysis tools**: Can find tricky issues, but can be costly. Might need some awareness of the analysis to deal with false positives.
- The best QA strategy involves multiple analysis and testing techniques!

