QA: Analysis Tools

17-313 Fall 2024

Foundations of Software Engineering

https://cmu-313.github.io

Michael Hilton and Rohan Padhye

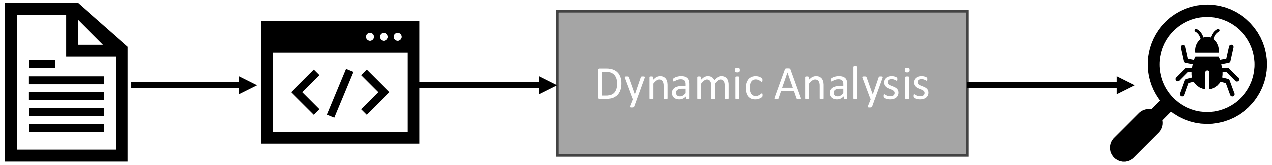
Learning Goals

- 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

Administrivia

- Midterm exam next week!
  - Practice exams released on website.
    - Not all topics are the same as previous semesters/years
  - Midterm review session tomorrow (Friday, Oct 4th) at 5pm TCS 358
    - Read old exams and come with questions or attempts prepared
- Project P2C (Second Sprint + Reflections) due next Thu, Oct 10th

What are Program Analysis Tools?



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 '-'?

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

The Bad News: Rice's Theorem
Every static analysis is necessarily incomplete, unsound, undecidable, or a combination thereof

"Any nontrivial property about the language recognized by a Turing machine is undecidable."

Henry Gordon Rice, 1953

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 dynamically

```python
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

print(n2s(12, 10))
```

1. What is the type of variable `u` during program execution?
2. Did the variable `u` ever contain a negative number?
3. For how many iterations did the while loop execute?
4. Was there ever be a division by zero?
5. Did 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

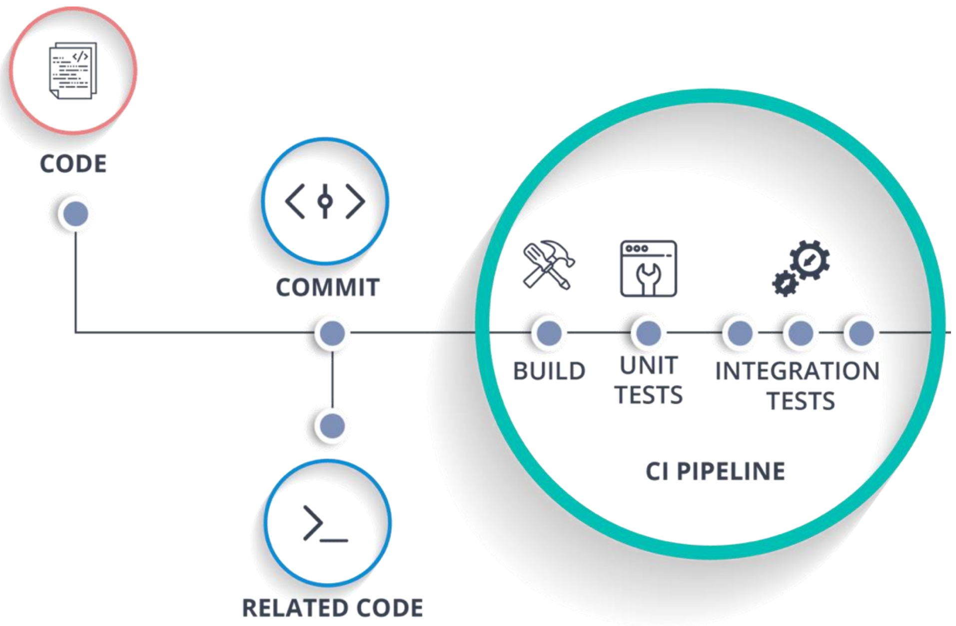
- Requires only source code
- Conservatively reasons about all possible inputs and program paths
- Reported warnings may contain false positives
- Can report all warnings of a particular class of problems
- Advanced techniques like verification can prove certain complex properties, but rarely run in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



Travis CI

GitHub Actions

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

GitHub acquires code analysis tool Semmle

Frederic Lardinois @fredericl / 1:30 pm EDT • September 18, 2019 Comment

Marketplace / Search results

Types

- Apps ×
- Actions

Categories

- API management
- Chat
- Code quality
- Code review
- Continuous integration
- Dependency management
- Deployment
- IDEs
- Learning
- Localization
- Mobile
- Monitoring
- Project management
- Publishing

Search for apps and actions

Apps

Build on your workflow with apps that integrate with GitHub.

306 results filtered by Apps ×

- **Zube** Agile project management that lets the entire team work with developers on GitHub
- **WhiteSource Bolt** Detect open source vulnerabilities in real time with suggested fixes for quick remediation
- **Crowdin** Agile localization for your projects
- **Slack + GitHub** Connect your code without leaving Slack
- **BackHub** Reliable GitHub repository backup, set up in minutes
- **GitLocalize** Continuous Localization for GitHub projects
- **Codacy** Automated code reviews to help developers ship better software, faster
- **Code Climate** Automated code review for technical debt and test coverage
- **Semaphore** Test and deploy at the push of a button
- **Flaptastic** Manage flaky unit tests. Click a checkbox to instantly disable any test on all branches. Works with your current test suite
- **DeepScan** Advanced static analysis for automatically finding runtime errors in JavaScript code
- **Depfu** Automated dependency updates done right

News

Snyk Secures $150M, Snags $1B Valuation

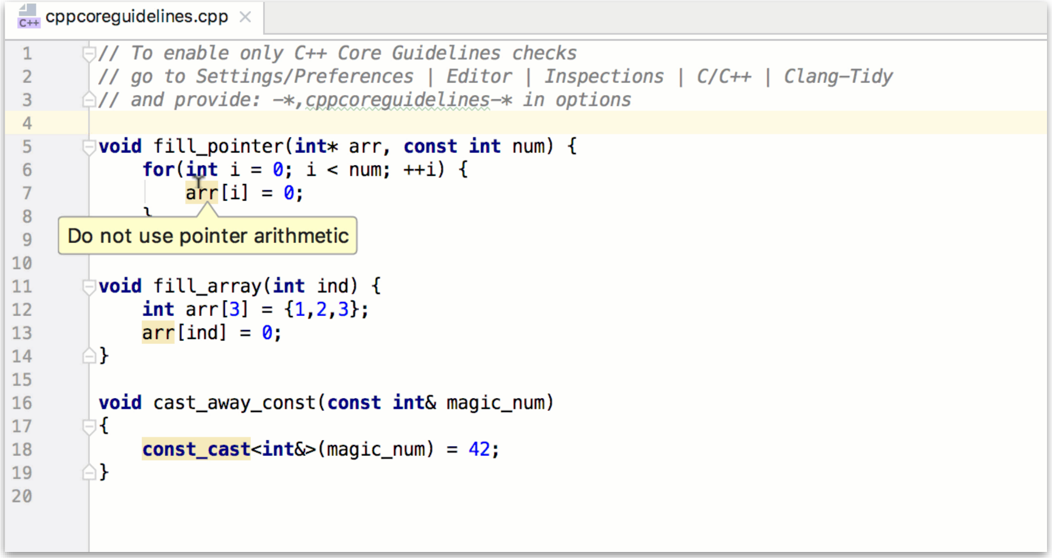
Sydney Sawaya | Associate Editor
January 21, 2020 1:12 PM

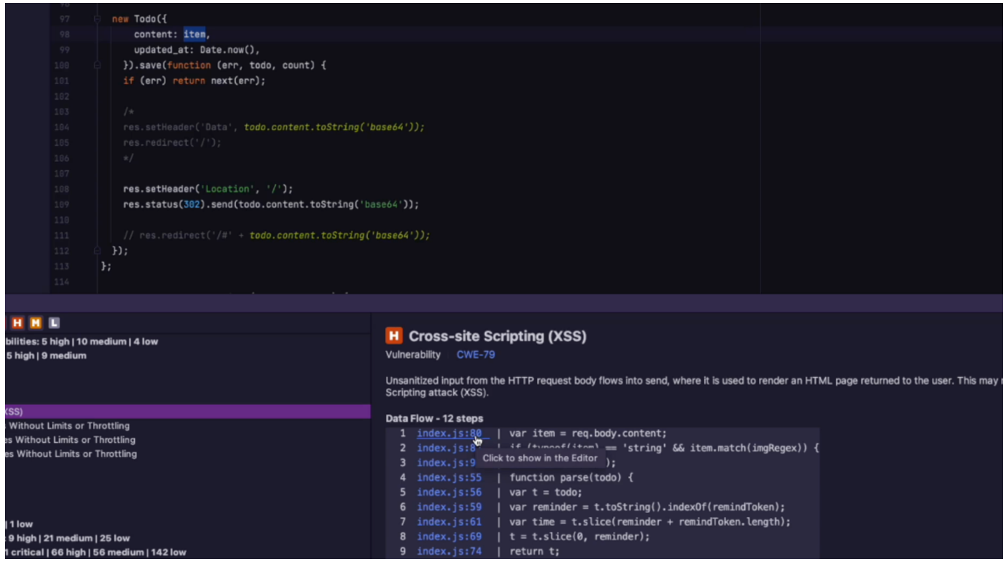
Share this article:

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.

https://www.sdxcentral.com/articles/news/snyk-secures-150m-snags-1b-valuation/2020/01/
https://techcrunch.com/2019/09/18/github-acquires-code-analysis-tool-semmle/
https://github.com/marketplace

Static analysis is also integrated into IDEs





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*

https://www.perforce.com/blog/qac/what-lint-code-and-why-linting-important

Use linters to enforce style guidelines

Don't rely on manual inspection during code review!

https://checkstyle.sourceforge.io/

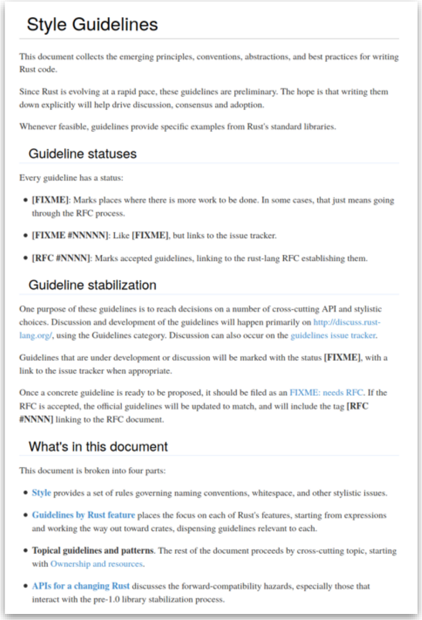
Linters use very “shallow” static analysis to enforce formatting rules

- 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


Style Guidelines

This document collects the emerging principles, conventions, abstractions, and best practices for writing Rust code.

Since Rust is evolving at a rapid pace, these guidelines are preliminary. The hope is that writing them down explicitly will help drive discussion, consensus and adoption.

Whenever feasible, guidelines provide specific examples from Rust's standard libraries.

Guideline statuses

Every guideline has a status:

- [FIXME]: Marks places where there is more work to be done. In some cases, that just means going through the RFC process.
- [FIXME #NNNNN]: Like [FIXME], but links to the issue tracker.
- [RFC #NNNN]: Marks accepted guidelines, linking to the rust-lang RFC establishing them.

Guideline stabilization

One purpose of these guidelines is to reach decisions on a number of cross-cutting API and stylistic choices. Discussion and development of the guidelines will happen primarily on http://discuss.rust-lang.org/, using the Guidelines category. Discussion can also occur on the guidelines issue tracker.

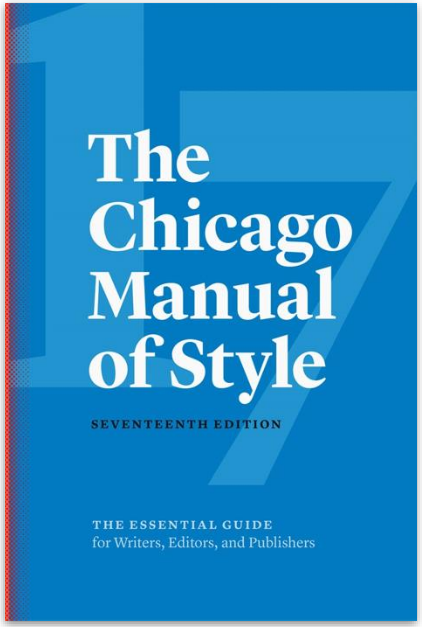
Guidelines that are under development or discussion will be marked with the status [FIXME], with a link to the issue tracker when appropriate.

Once a concrete guideline is ready to be proposed, it should be filed as an FIXME: needs RFC. If the RFC is accepted, the official guidelines will be updated to match, and will include the tag [RFC #NNNN] linking to the RFC document.

What's in this document

This document is broken into four parts:

- Style provides a set of rules governing naming conventions, whitespace, and other stylistic issues.
- Guidelines by Rust feature places the focus on each of Rust's features, starting from expressions and working the way out toward crates, dispensing guidelines relevant to each.
- Topical guidelines and patterns. The rest of the document proceeds by cross-cutting topic, starting with Ownership and resources.
- APIs for a changing Rust discusses the forward-compatibility hazards, especially those that interact with the pre-1.0 library stabilization process.



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

SpotBugs

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

FindBugs because it's easy

Docs and Info
FindBugs 2.0
Demo and data
Users and supporters
FindBugs blog
Fact sheet
Manual
Manual(ja/日本語)
FAQ
Bug descriptions
Bug descriptions(ja/日本語)
Bug descriptions(fr)
Mailing lists
Documents and Publications
Links

Downloads

FindBugs Swag

Development
Open bugs
Reporting bugs
Contributing
Dev team
API [no frames]
Change log
SF project page
Browse source
Latest code changes

FindBugs Bug Descriptions

This document lists the standard bug patterns reported by FindBugs version 3.0.1.

Summary

Description	Category
BC: Equals method should not assume anything about the type of its argument	Bad practice
BIT: Check for sign of bitwise operation	Bad practice
CN: Class implements Cloneable but does not define or use clone method	Bad practice
CN: clone method does not call super.clone()	Bad practice
CN: Class defines clone() but doesn't implement Cloneable	Bad practice
CNT: Rough value of known constant found	Bad practice
Co: Abstract class defines covariant compareTo() method	Bad practice
Co: compareTo()/compare() incorrectly handles float or double value	Bad practice
Co: compareTo()/compare() returns Integer.MIN_VALUE	Bad practice
Co: Covariant compareTo() method defined	Bad practice
DE: Method might drop exception	Bad practice
DE: Method might ignore exception	Bad practice
DMI: Adding elements of an entry set may fail due to reuse of Entry objects	Bad practice
DMI: Random object created and used only once	Bad practice
DMI: Don't use removeAll to clear a collection	Bad practice
Dm: Method invokes System.exit(...)	Bad practice
Dm: Method invokes dangerous method runFinalizersOnExit	Bad practice
ES: Comparison of String parameter using == or !=	Bad practice
ES: Comparison of String objects using == or !=	Bad practice
Eq: Abstract class defines covariant equals() method	Bad practice
Eq: Equals checks for incompatible operand	Bad practice
Eq: Class defines compareTo(...) and uses Object.equals()	Bad practice
Eq: equals method fails for subtypes	Bad practice
Eq: Covariant equals() method defined	Bad practice
FI: Empty finalizer should be deleted	Bad practice
FI: Explicit invocation of finalizer	Bad practice
FI: Finalizer nulls fields	Bad practice
FI: Finalizer only nulls fields	Bad practice
FI: Finalizer does not call superclass finalizer	Bad practice
FI: Finalizer nullifies superclass finalizer	Bad practice
FI: Finalizer does nothing but call superclass finalizer	Bad practice
FS: Format string should use %n rather than \n	Bad practice
GC: Unchecked type in generic call	Bad practice
HE: Class defines equals() but not hashCode()	Bad practice
HE: Class defines equals() and uses Object.hashCode()	Bad practice
HE: Class defines hashCode() but not equals()	Bad practice
HE: Class defines hashCode() and uses Object.equals()	Bad practice
HE: Class inherits equals() and uses Object.hashCode()	Bad practice
IC: Superclass uses subclass during initialization	Bad practice
IMSE: Dubious catching of IllegalMonitorStateException	Bad practice
ISC: Needless instantiation of class that only supplies static methods	Bad practice
It: Iterator next() method can't throw NoSuchElementException	Bad practice
J2EE: Store of non serializable object into HttpSession	Bad practice
JCIP: Fields of immutable classes should be final	Bad practice
ME: Public enum method unconditionally sets its field	Bad practice

http://findbugs.sourceforge.net/bugDescriptions.html
https://spotbugs.readthedocs.io/en/latest/bugDescriptions.html

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!");
}
```

Performance:

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

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 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();
}
```

Correctness:

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

https://github.com/pbrune1973/qwics/blob/3b4ec904468eaed89ea890b5ae97ebaaa8e4a3e6/workspace/QwicsJDBCDriver/src/org/qwics/jdbc/QwicsXid.java

https://github.com/pbrune1973/qwics/issues/7

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;
}
```

https://github.com/pbrune1973/qwics/blob/3b4ec904468eaed89ea890b5ae97ebaaa8e4a3e6/workspace/QwicsJDBCDriver/src/org/qwics/jdbc/QwicsXid.java

https://github.com/pbrune1973/qwics/issues/7

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 or MyPy

CHECKER framework

: my[py]

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)
{ … }
```

```java
import org.checkerframework.checker.nullness.qual.*;

public class NullnessExampleWithWarnings {
   public void example() {
      @NonNull String foo = "foo";
      String bar = null;
      foo = bar;
      println(foo.length());
   }
}
```

`@Nullable` **is applied by default**

Error: [assignment.type.incompatible] incompatible types in assignment.
  found   : @Initialized **@Nullable** String
  required: @UnknownInitialization **@NonNull** String

```java
import org.checkerframework.checker.nullness.qual.*;

public class NullnessExampleWithWarnings {
  public void example() {
    @NonNull String foo = "foo";
    String bar = null;   // @Nullable

    if (bar != null) {
      foo = bar;
    }
    println(foo.length());


  }

}
```

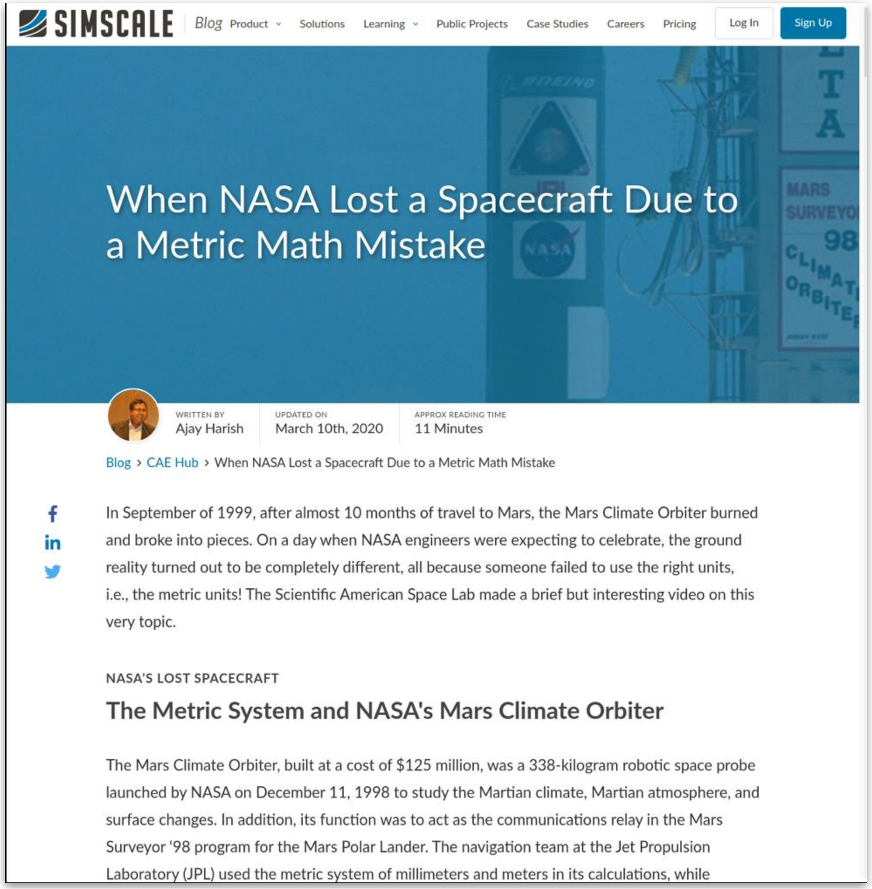
bar is refined to @NonNull

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, …

https://www.nist.gov/pml/weights-and-measures/metric-si/si-units

Remember the Mars Climate Orbiter incident from 1999?


SIMSCALE Blog Product Solutions Learning Public Projects Case Studies Careers Pricing Log In Sign Up

When NASA Lost a Spacecraft Due to a Metric Math Mistake

WRITTEN BY Ajay Harish | UPDATED ON March 10th, 2020 | APPROX READING TIME 11 Minutes

Blog › CAE Hub › When NASA Lost a Spacecraft Due to a Metric Math Mistake

In September of 1999, after almost 10 months of travel to Mars, the Mars Climate Orbiter burned and broke into pieces. On a day when NASA engineers were expecting to celebrate, the ground reality turned out to be completely different, all because someone failed to use the right units, i.e., the metric units! The Scientific American Space Lab made a brief but interesting video on this very topic.

NASA'S LOST SPACECRAFT

The Metric System and NASA's Mars Climate Orbiter

The Mars Climate Orbiter, built at a cost of $125 million, was a 338-kilogram robotic space probe launched by NASA on December 11, 1998 to study the Martian climate, Martian atmosphere, and surface changes. In addition, its function was to act as the communications relay in the Mars Surveyor '98 program for the Mars Polar Lander. The navigation team at the Jet Propulsion Laboratory (JPL) used the metric system of millimeters and meters in its calculations, while

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;

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;
}
```

`@m` indicates that `x` represents meters

To assign a unit, multiply appropriate unit constant from `UnitTools`

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;

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;
}
```

`@m` indicates that `x` represents meters

To assign a unit, multiply appropriate unit constant from `UnitTools`

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;
}
```

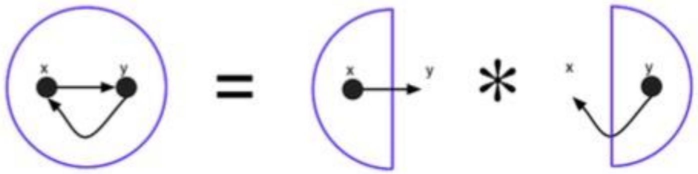
Addition and subtraction between meters and seconds is physically meaningless

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

(Alternative) *Infer*: Type-checking without the annotations

- Focused on memory safety bugs
  - Null pointer dereferences, memory leaks, resource leaks, ...
- Compositional interprocedural reasoning
  - Based on separation logic and bi-abduction
- Scalable and fast
  - Can run incremental analysis on changed code
- Does not require annotations
- Supports multiple languages
  - Java, C, C++, Objective-C
  - Programs are compiled to an intermediate representation



NULLPTR_DEREFERENCE

Reported as "Nullptr Dereference" by pulse.

Infer reports null dereference bugs in Java, C, C++, and Objective-C when it is possible that the null pointer is dereferenced, leading to a crash.

Null dereference in Java

Many of Infer's reports of potential Null Pointer Exceptions (NPE) come from code of the form

```
p = foo(); // foo() might return null
stuff();
p.goo();   // dereferencing p, potential NPE
```

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

Michael Pradel
michael@binaervarianz.de
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.

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	

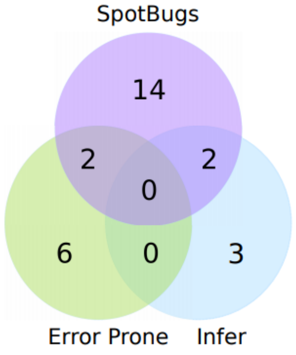


Figure 4: Total number of bugs found by all three static checkers and their overlap.

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!