JQF: Coverage-Guided Property-Based Testing in Java

Rohan Padhye
University of California, Berkeley
USA
rohanpadhye@cs.berkeley.edu

Caroline Lemieux
University of California, Berkeley
USA
clemieux@cs.berkeley.edu

Koushik Sen
University of California, Berkeley
USA
ksen@cs.berkeley.edu

1 INTRODUCTION

Coverage-guided fuzzing (CGF) has recently become a very popular technique for automatic test-input generation. CGF tools like AFL [17] and libFuzzer [7] have discovered thousands of bugs and security vulnerabilities in programs that parse binary data, such as image decoders and media players. CGF works by first inserting lightweight instrumentation in a program under test for collecting code coverage. Then, the program is continuously executed with randomly generated inputs. If the program crashes, a bug is found. Instead of generating inputs from scratch, CGF evolves a set of newly generated inputs. If the program crashes, a bug is found.

COLLECTIONS-714) when fuzzing with ZestGuidance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISSTA '19, July 15–19, 2019, Beijing, China
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6224-5/19/07...
https://doi.org/10.1145/3293882.3339002

Figure 1: A sample property test using JQF that checks the construction of a Trie data structure in Apache Commons from an input JDK Map. Fires an assertion violation (bug COLLECTIONS-714) when fuzzing with ZestGuidance.

1 INTRODUCTION

Coverage-guided fuzzing (CGF) has recently become a very popular technique for automatic test-input generation. CGF tools like AFL [17] and libFuzzer [7] have discovered thousands of bugs and security vulnerabilities in programs that parse binary data, such as image decoders and media players. CGF works by first inserting lightweight instrumentation in a program under test for collecting code coverage. Then, the program is continuously executed with randomly generated inputs. If the program crashes, a bug is found. Instead of generating inputs from scratch, CGF evolves a set of newly generated inputs. If the program crashes, a bug is found.

COLLECTIONS-714) when fuzzing with ZestGuidance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISSTA '19, July 15–19, 2019, Beijing, China
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6224-5/19/07...
https://doi.org/10.1145/3293882.3339002

Figure 1: A sample property test using JQF that checks the construction of a Trie data structure in Apache Commons from an input JDK Map. Fires an assertion violation (bug COLLECTIONS-714) when fuzzing with ZestGuidance.

1 INTRODUCTION

Coverage-guided fuzzing (CGF) has recently become a very popular technique for automatic test-input generation. CGF tools like AFL [17] and libFuzzer [7] have discovered thousands of bugs and security vulnerabilities in programs that parse binary data, such as image decoders and media players. CGF works by first inserting lightweight instrumentation in a program under test for collecting code coverage. Then, the program is continuously executed with randomly generated inputs. If the program crashes, a bug is found. Instead of generating inputs from scratch, CGF evolves a set of newly generated inputs. If the program crashes, a bug is found.

COLLECTIONS-714) when fuzzing with ZestGuidance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISSTA '19, July 15–19, 2019, Beijing, China
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6224-5/19/07...
https://doi.org/10.1145/3293882.3339002

Figure 1: A sample property test using JQF that checks the construction of a Trie data structure in Apache Commons from an input JDK Map. Fires an assertion violation (bug COLLECTIONS-714) when fuzzing with ZestGuidance.

1 INTRODUCTION

Coverage-guided fuzzing (CGF) has recently become a very popular technique for automatic test-input generation. CGF tools like AFL [17] and libFuzzer [7] have discovered thousands of bugs and security vulnerabilities in programs that parse binary data, such as image decoders and media players. CGF works by first inserting lightweight instrumentation in a program under test for collecting code coverage. Then, the program is continuously executed with randomly generated inputs. If the program crashes, a bug is found. Instead of generating inputs from scratch, CGF evolves a set of newly generated inputs. If the program crashes, a bug is found.

COLLECTIONS-714) when fuzzing with ZestGuidance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISSTA '19, July 15–19, 2019, Beijing, China
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6224-5/19/07...
https://doi.org/10.1145/3293882.3339002

Figure 1: A sample property test using JQF that checks the construction of a Trie data structure in Apache Commons from an input JDK Map. Fires an assertion violation (bug COLLECTIONS-714) when fuzzing with ZestGuidance.

1 INTRODUCTION

Coverage-guided fuzzing (CGF) has recently become a very popular technique for automatic test-input generation. CGF tools like AFL [17] and libFuzzer [7] have discovered thousands of bugs and security vulnerabilities in programs that parse binary data, such as image decoders and media players. CGF works by first inserting lightweight instrumentation in a program under test for collecting code coverage. Then, the program is continuously executed with randomly generated inputs. If the program crashes, a bug is found. Instead of generating inputs from scratch, CGF evolves a set of newly generated inputs. If the program crashes, a bug is found.

COLLECTIONS-714) when fuzzing with ZestGuidance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISSTA '19, July 15–19, 2019, Beijing, China
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6224-5/19/07...
https://doi.org/10.1145/3293882.3339002

Figure 1: A sample property test using JQF that checks the construction of a Trie data structure in Apache Commons from an input JDK Map. Fires an assertion violation (bug COLLECTIONS-714) when fuzzing with ZestGuidance.

1 INTRODUCTION

Coverage-guided fuzzing (CGF) has recently become a very popular technique for automatic test-input generation. CGF tools like AFL [17] and libFuzzer [7] have discovered thousands of bugs and security vulnerabilities in programs that parse binary data, such as image decoders and media players. CGF works by first inserting lightweight instrumentation in a program under test for collecting code coverage. Then, the program is continuously executed with randomly generated inputs. If the program crashes, a bug is found. Instead of generating inputs from scratch, CGF evolves a set of newly generated inputs. If the program crashes, a bug is found.

COLLECTIONS-714) when fuzzing with ZestGuidance.
2 FUZZING WITH JQF

Practitioners can use JQF to automatically generate test inputs for parameterized test methods using coverage-guided fuzzing. Figure 1 shows an example of a JQF test driver written in Java, which aims to check a basic property of the class PatriciaTrie from Apache Commons Collections. A trie data structure can be constructed from a pre-existing mapping of strings in a JDK Map object.

The test method testMap2Trie checks the following property: Given an arbitrary string key and a JDK map whose keys are strings, if key exists within map, then a trie constructed from this map should also contain the same key. The @Fuzz annotation on the test method enables JQF to automatically generate random instances of map and key to verify this property. The JUnit Assume API allows the user to specify preconditions on the generated inputs (e.g. Line 6). Test generation can be launched via JQF’s Apache Maven plugin1:

```java
mvn jqf:fuzz -Dclass=TestMap2Trie
```

By default, JQF uses the Zest algorithm (§4.2) to generate test inputs. Fuzzing continues either until it is explicitly stopped, until a user-specified timeout expires, or until a test failure is encountered.

For the test in Figure 1, the Zest fuzzing engine often finds a test failure in about 5 seconds, after executing about 5,000 test inputs (of which over 1,700 satisfy the precondition on Line 6). The failing test case leads to an assertion violation at Line 8 due to a very special case for Figure 1 even after 30 minutes (over 7 million executions). Pure random generation does not find a failing test case for Figure 1 even after 30 minutes (over 7 million executions).

3 FUZZING FRAMEWORK

We next explain how JQF generates random inputs, such as map and key in Figure 1, using coverage-guided algorithms called guidances.

3.1 The Guidance Interface

Figure 2 shows the Guidance interface. Researchers can implement this interface to specify a coverage-guided fuzzing algorithm. Guidance instances are stateful objects whose methods are invoked by the JQF framework in a fuzzing loop (depicted in Figure 3).

The Guidance method hasInput() returns whether a new input is available; the return value false ends fuzzing. The getInputStream() method returns the next input generated by the Guidance, as an InputStream. This stream is used to generate structured inputs such as Map objects (see §3.2). The structured inputs, called args in

```java
1 public interface Guidance {
2    boolean hasInput();
3    InputStream getInputStream();
4    void handleResult(Result result, Throwable error);
5    Consumer<TraceEvent> generateCallBack(Thread thread);
6 }
```

3.2 Parametric Generators

The arguments to a test method—such as map and key in Figure 1—are generated using the same mechanisms as supported by junit-quickcheck. In general, inputs of type T are generated by a backing Generator<T>, which provides a method to randomly sample a new instance of T. junit-quickcheck can either (1) implicitly pick a suitable generator from a library that it provides, (2) be directed to synthesize such a generator automatically, e.g. using the constructors or public fields of class T, or (3) be provided with a hand-written Generator<T>.

In all cases, the generator uses a SourceOfRandomness object, which provides an API for making non-deterministic decisions such as: choosing from a list of alternatives (e.g. whether to instantiate a TreeMap or HashMap for map in Figure 1), picking random sizes (e.g. how many entries to insert in map), or populating primitives (e.g. what keys and values to insert in map). In junit-quickcheck, the default SourceOfRandomness is backed a pseudo-random stream of bytes. JQF overrides this source to use the stream returned by Guidance.getInputStream() instead (ref. Line 5 in Figure 3), thereby making the generators deterministically dependent on the guidance.

1Non-Maven users can launch JQF programatically or via command-line scripts.

2We stumbled upon this bug while writing an example for this paper.
3.3 Code Coverage Events

When coverage-guided fuzzing is launched (e.g., via `mvn jqf:fuzz`), the test program’s classes are instrumented on-the-fly using the ASM bytecode manipulation library [13]. The instrumentation adds logic to generate `TraceEvents` during test execution. For example, a `BranchEvent` is generated when a test program executes a conditional branch, a `CallEvent` accompanies a method invocation, and an `AllocEvent` signals the creation of a new object or array on the heap. These event objects contain information about their source program locations as well other event-specific data. When a trace event e is generated in thread t, JQF invokes the function `handle_t(e)`, where `handle_t` is the callback returned by `Guidance.generateCallback(t)`. The guidance must choose how to update its internal state based on this coverage information, which will presumably be used to generate subsequent inputs.

4 GUIDANCES

JQF currently ships with the following `Guidance` implementations.

4.1 No Guidance

The most trivial guidance, called `NoGuidance`, returns an infinite stream of random values every time `getInput()` is called. This guidance completely ignores code coverage events. This guidance is almost equivalent to using vanilla JUnitQuickcheck.

4.2 Zest Guidance

JQF’s default guidance implements the Zest algorithm [15], which is specifically designed for coverage-guided property testing. The `ZestGuidance` returns dynamically sized parameter sequences via the `getInput()` method, which are generated randomly for the first iteration of the fuzzing loop. Zest maintains a set of saved parameter sequences. The `ZestGuidance` generates new inputs by randomly mutating previously saved parameter sequences. Byte-level mutations on these parameter sequences correspond to structural mutations in the generated test inputs. For example, a random mutation in the parameter sequence for map in Figure 1 may lead to the corresponding `Generator<Map>` to produce the next map with an additional entry. Dynamic sizing allows the parameter sequences to be lazily extended (if the `Generator` needs to make more choices than expected) or to be efficiently truncated (if the `Generator` makes fewer choices). Further, Zest separately tracks code coverage achieved by all test executions and code coverage by `valid` test executions (i.e., those whose result is `SUCCESS`). If a mutated parameter sequence leads to new code coverage overall, or if it leads to a valid test that covers code which has not been covered by any previous `valid` test, then the sequence is saved for subsequent mutation. Zest has been used to find complex semantic bugs, such as issues within compiler optimizations3.

4.3 AFL Guidance

JQF supports input generation using the popular AFL [17] tool, unmodified. This is possible because AFL, which is designed to fuzz C/C++ programs and x86 binaries, communicates with instrumented test programs via inter-process messages and a code coverage map in shared memory. The `AFLGuidance` in JQF implements this communication protocol via a proxy program. The proxy mocks an AFL-instrumented test target that reads input from a specific file. `AFLGuidance.getInput()` simply returns the contents of this file, which is continuously updated by AFL. During test execution, `AFLGuidance` collects code coverage information by handling `TraceEvents`. When `AFLGuidance.handleResult()` is invoked, the coverage information is written to AFL’s shared memory region via the proxy. Calls to `AFLGuidance.hasInput()` block until AFL is ready with the next input.

AFL’s mutation strategy uses various heuristics that are applicable to programs that parse fixed-size binary files (e.g., media players). Further, AFL does not explicitly distinguish between `INVALID` and `FAILURE` results. Due to these reasons, JQF’s `AFLGuidance` is most effective when used with test methods that take only one argument of type `InputStream` (since `Generator<InputStream>` returns the guidance-generated input stream as-is), and that do not use any assume statements. For example, AFLGuidance has been used to fuzz OpenJDK’s ImageIO library that reads PNG and JPEG files4, as well as Apache PDFBox’s processing of PDF documents5.

4.4 PerfFuzz Guidance

PerfFuzz [10] is a technique for automatically generating test inputs that maximize performance counters, such as loop execution counts. PerfFuzz’s goal is to automatically discover hot spots and performance bottlenecks. PerfFuzz is a fork of AFL that extends its code coverage map with performance feedback in the form of (k, v) pairs where v is a value to be maximized for every key k. PerfFuzz saves a mutated input either if it leads to new code coverage, or if it maximizes the value of v for some key k.

JQF’s `PerfFuzzGuidance` is a sub-class of `AFLGuidance` which overrides `handleResult()` to communicate this additional performance map via the proxy program. `PerfFuzzGuidance` can be configured either to find hot spots (where keys are branch locations and values are execution counts for the corresponding branch) or to find memory consumption issues (where keys are allocation sites and values are number of bytes allocated at the corresponding site). For example, we used `PerfFuzzGuidance` to find an algorithmic complexity bug in the Google Closure Compiler, where reporting a specific case of syntax error in a JavaScript program can take time that is exponential in the size of the input program6. With the memory allocation feedback, we found an issue in OpenJDK’s handling of PNG images that specify very large dimensions7.

4.5 Repro Guidance

Finally, the `ReproGuidance` is a trivial guidance whose `getInput()` method returns the contents of a given file on disk, and then ends the loop. This guidance enables debugging of saved test failures.

5 EVALUATION AND IMPACT

Table 1 summarizes the impact that JQF has had in discovering previously unknown bugs in widely used Java software. These

---

3https://github.com/google/closure-compiler/issues/2842
4https://bugs.openjdk.java.net/browse/JDK-8191073
5https://issues.apache.org/jira/browse/PDFBOX-4333
6https://github.com/google/closure-compiler/issues/3173
7https://bugs.openjdk.java.net/browse/JDK-8190332
bugs were found over the course of various experiments performed throughout 2017–2019. The Zest paper [15] describes a systematic study involving five of the projects from this table and three different guidances. The study showed that NoGuidance is not very reliable, that AFLGuidance is effective in finding bugs in syntax parsers, and that ZestGuidance excels at finding semantic bugs. This study resulted in the discovery of 20 of the bugs in Table 1.

Of the total 42 bugs found using JQF, 11 semantic bugs were found with ZestGuidance, 29 syntax parsing bugs were found with AFLGuidance, and 2 bugs were found with PerfFuzzGuidance. 24 of the 42 reported bugs have been fixed at the time of writing, while the rest await patches.

Notably, 7 of the 42 bugs (including 4 security vulnerabilities with assigned CVEs) were discovered by two independent practitioners who are not affiliated with the authors of this paper. We were made aware of JQF’s success via social media [1] and blog posts [12]. All 7 of these bugs have been fixed. We are encouraged by these findings, and believe that they provide evidence to support JQF’s usefulness to the software testing community at large.

### 6 RELATED WORK

JQF is one of few tools that enable coverage-guided fuzzing of Java programs. Kelinci [9] is a wrapper around afl-fuzz that targets Java programs. Unlike JQF’s extensible guidance mechanism, Kelinci’s instrumentation directly updates AFL-specific coverage feedback; therefore, it cannot easily be adapted to work with tools like PerfFuzz. Kelinci also expects a test driver with a main method that reads inputs as files, in contrast to JQF’s property-based testing approach. Thus, Kelinci does not support structured input fuzzing algorithms such as Zest. Further comparison between Kelinci and JQF can be found in a blog post by an independent software company [12]. Barro [8] has implemented another wrapper around AFL for fuzzing Java main programs, similar to Kelinci. This tool actually borrows its dynamic instrumentation logic from JQF itself.


### ACKNOWLEDGMENTS

We would like to thank Tobias Ospelt of moderno AG for reporting various issues with alpha versions of JQF. We appreciate the feedback of Tim Allison, VP of Apache Tika, on early bugs found by JQF. We are also grateful to Paul Holser for developing the very extensible junit-t-quickcheck library, using which JQF is built. This research is supported in part by gifts from Samsung, Facebook, and Futurewei, and by NSF grants CCF-1409872 and CNS-1817122.

### REFERENCES


### Table 1: Number of new bugs discovered using JQF.

<table>
<thead>
<tr>
<th>Project</th>
<th>Bugs Found</th>
<th>Bugs Fixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mozilla Rhino</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Apache Commons - Compress</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Apache Commons - Lang</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Apache Commons - Collections</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Apache Maven</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Apache Ant</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Apache BCEL</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Aphona PDFbox</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Apache Tika</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Google Closure Compiler</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>24</td>
</tr>
</tbody>
</table>