Client–Library Compatibility Testing with API Interaction Snapshots

Gustave Monce, Thomas Degueule, Jean-Rémy Falleri, Romain Robbes

New Ideas and Emerging Results

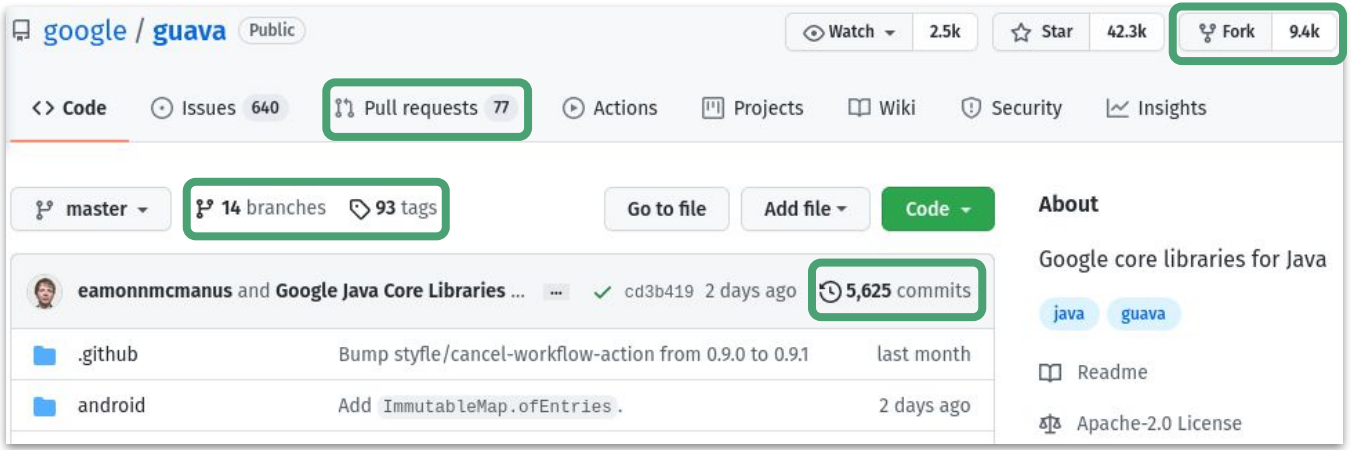
ICSME 2025 Auckland, New Zealand

université de BORDEAUX

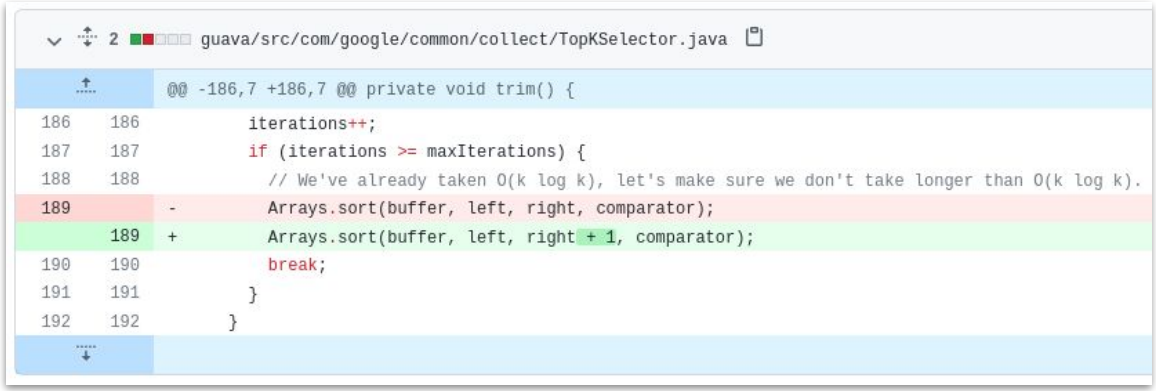
cnrs

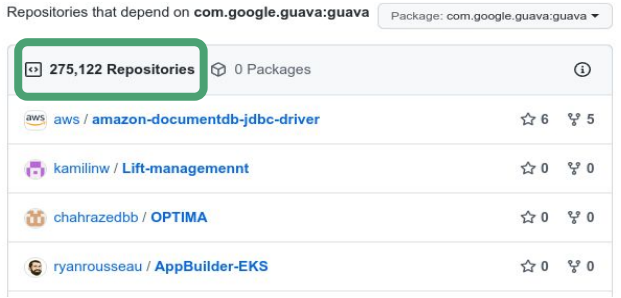
LaBRI

Software Libraries



Version	Repository	Usages	Date
30.1.1-jre	Central	2,839	Mar, 2021
30.1.1-android	Central	488	Mar, 2021
30.1-jre	Central	1,611	Dec, 2020
30.1-android	Central	424	Dec, 2020
30.0-jre	Central	1,490	Oct, 2020
30.0-android	Central	359	Oct, 2020
29.0-jre	Central	2,799	Apr, 2020
29.0-android	Central	382	Apr, 2020
28.2-jre	Central	1,785	Jan, 2020
28.2-android	Central	142	Jan, 2020
28.1-jre	Central	1,827	Aug, 2019
28.1-android	Central	80	Aug, 2019
28.0-jre	Central	1,482	Jun, 2019
28.0-android	Central	56	Jun, 2019
27.1-jre	Central	1,442	Mar, 2019





Library Upgrades

Bump org.revapi:revapi-java from 0.28.1 to 0.28.4 #58

Open **dependabot** wants to merge 1 commit into `main` from `dependabot/maven/org.revapi-revapi-java-0.28.4`

Conversation 0 | Commits 1 | Checks 5 | Files changed 1

dependabot bot commented on behalf of **github** on Aug 4 — Contributor

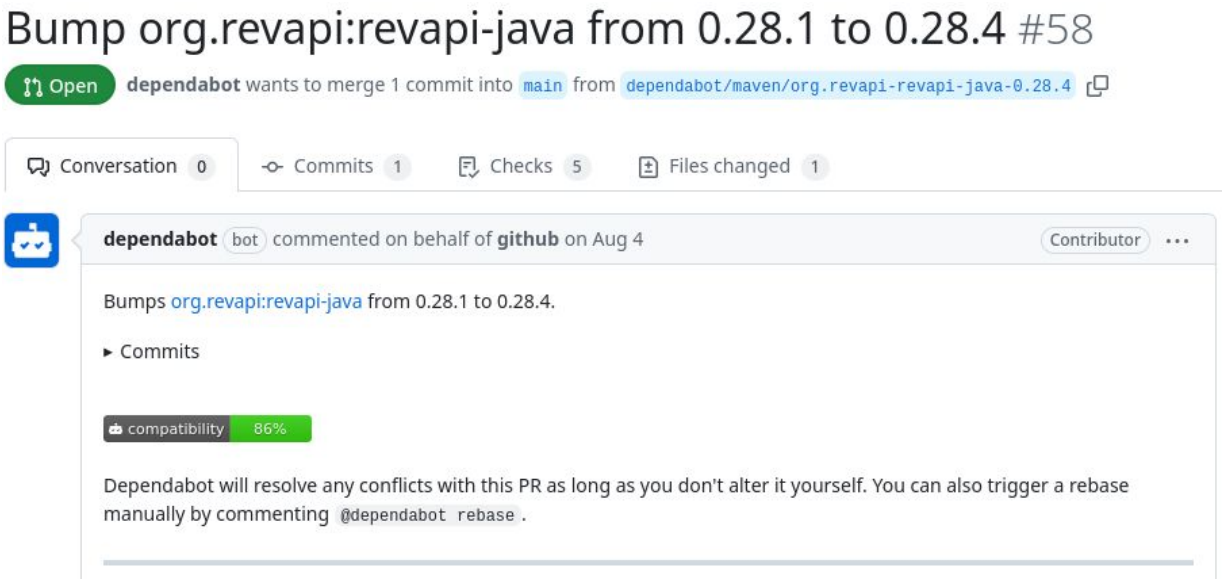
Bumps org.revapi:revapi-java from 0.28.1 to 0.28.4.

▸ Commits

compatibility 86%

Dependabot will resolve any conflicts with this PR as long as you don't alter it yourself. You can also trigger a rebase manually by commenting `@dependabot rebase`.

Syntactic vs Semantic Breaking Changes



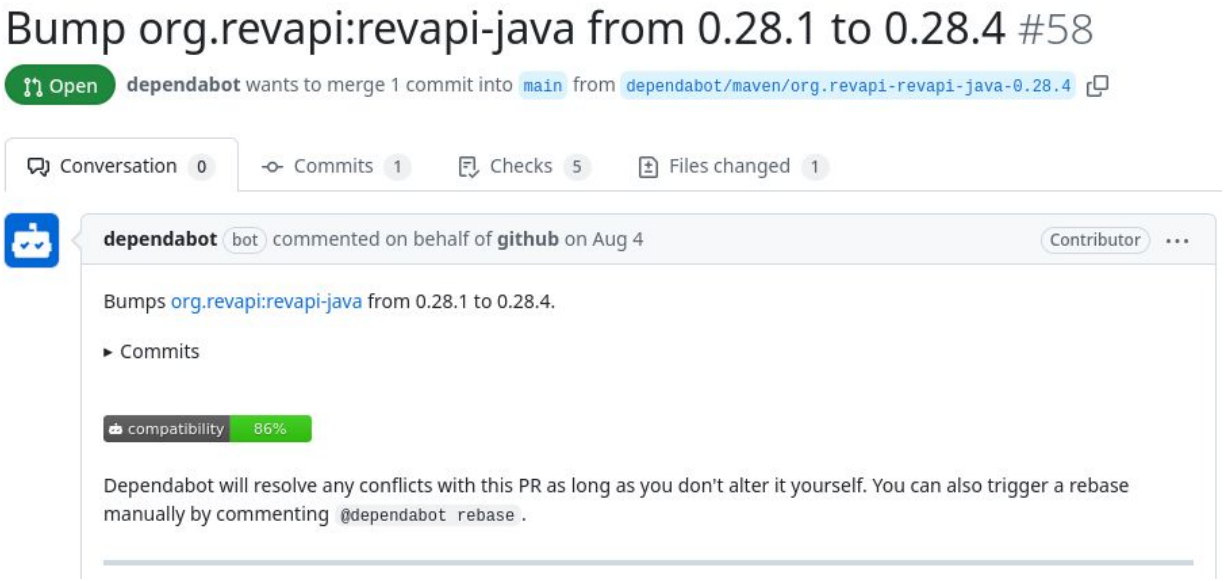

```
- public String  fetch(String productId) {
+ public Product fetch(String productId) {
```

Syntactic BCs

- Affect API signatures
- Trigger compile/link-time errors
- Extensive literature
- Can be detected statically with good accuracy (e.g., japicmp, RevApi, **Roseau**)

Also @ ICSME'25!

Syntactic vs Semantic Breaking Changes

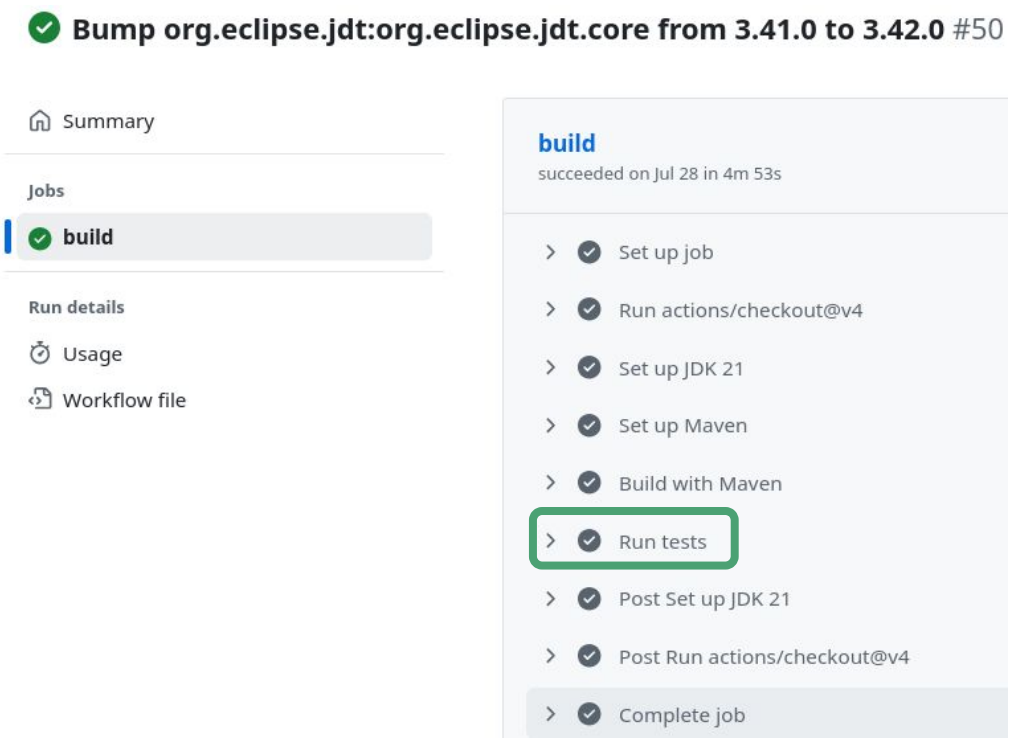


Behavioral BCs

- Undecidable
- Manifest as run-time errors, exceptions, etc.
- Hard to estimate impact on client code
- Very few approaches

```
- return Arrays.sort(buf, left, right, comparator)
+ return Arrays.sort(buf, left, right + 1, comparator)
```

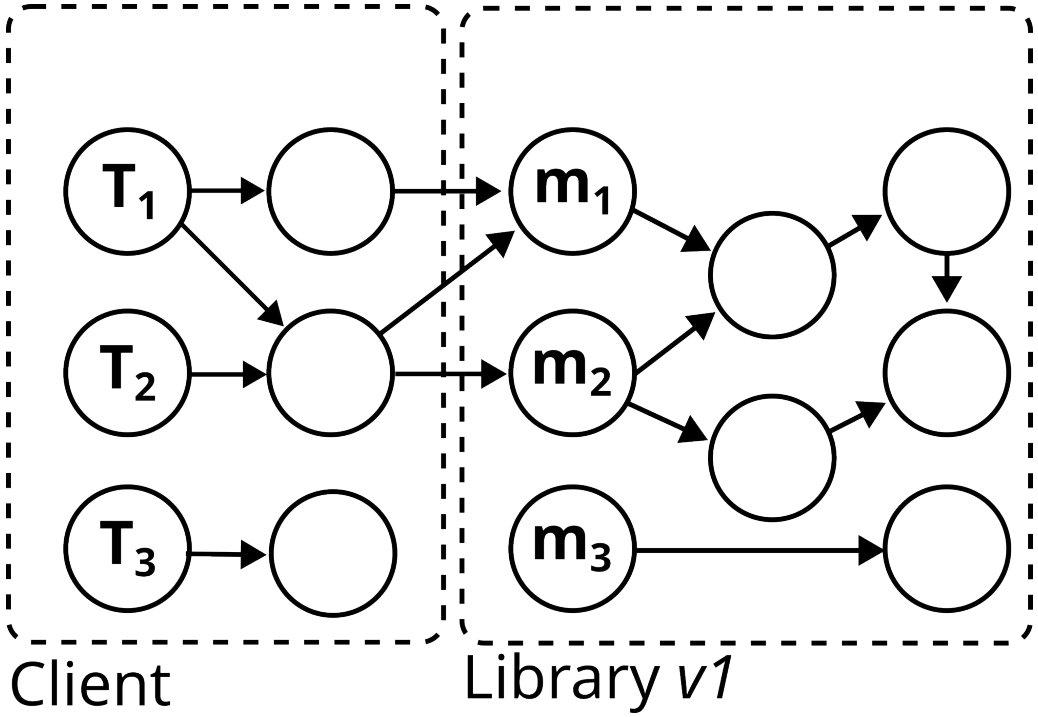
Regression Testing



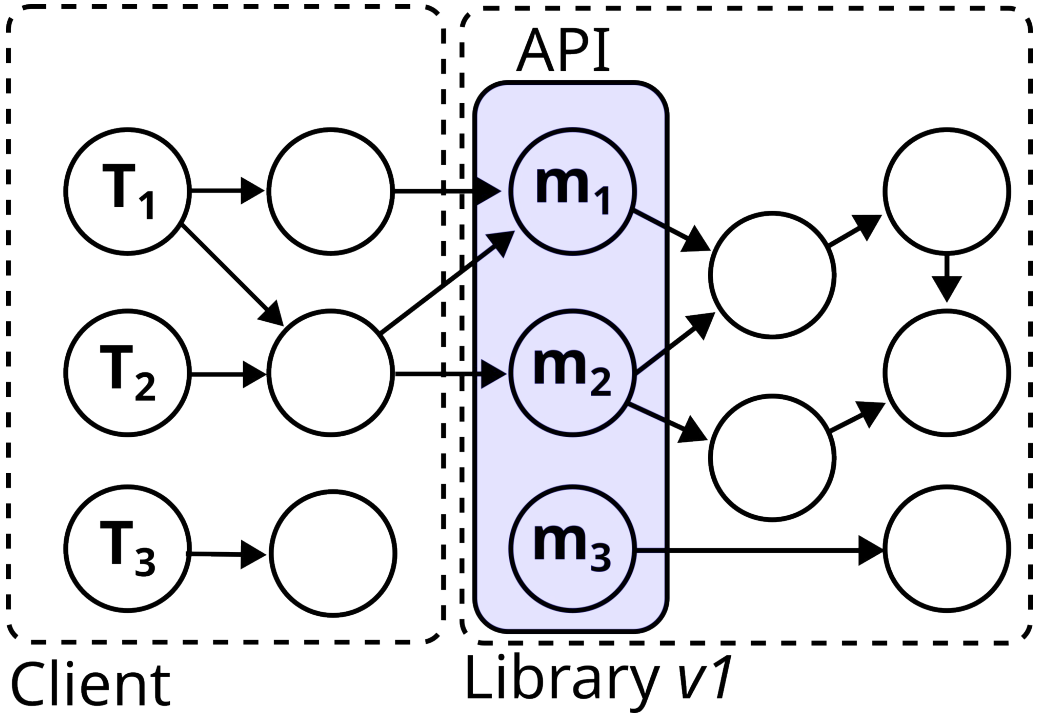
- Client tests cannot reliably detect BBCs[†]
  - Insufficient coverage
  - Weak assertions
  - Distance to fault
  - Lenient/exception-swallowing code
- Even when detected, the distance between the root cause and the failing assertions severely hurts diagnosis and remediation

[†] Jayasuriya et al. *"Understanding the impact of APIs behavioral breaking changes on client applications"* (FSE'24)
[†] Hejderup et al. *"Can we trust tests to automate dependency updates? A case study of Java projects"* (JSS'22)
[†] Gyori et al. *"Evaluating regression test selection opportunities in a very large open-source ecosystem"* (ISSRE'18)

API Interaction Snapshots with Gilesi

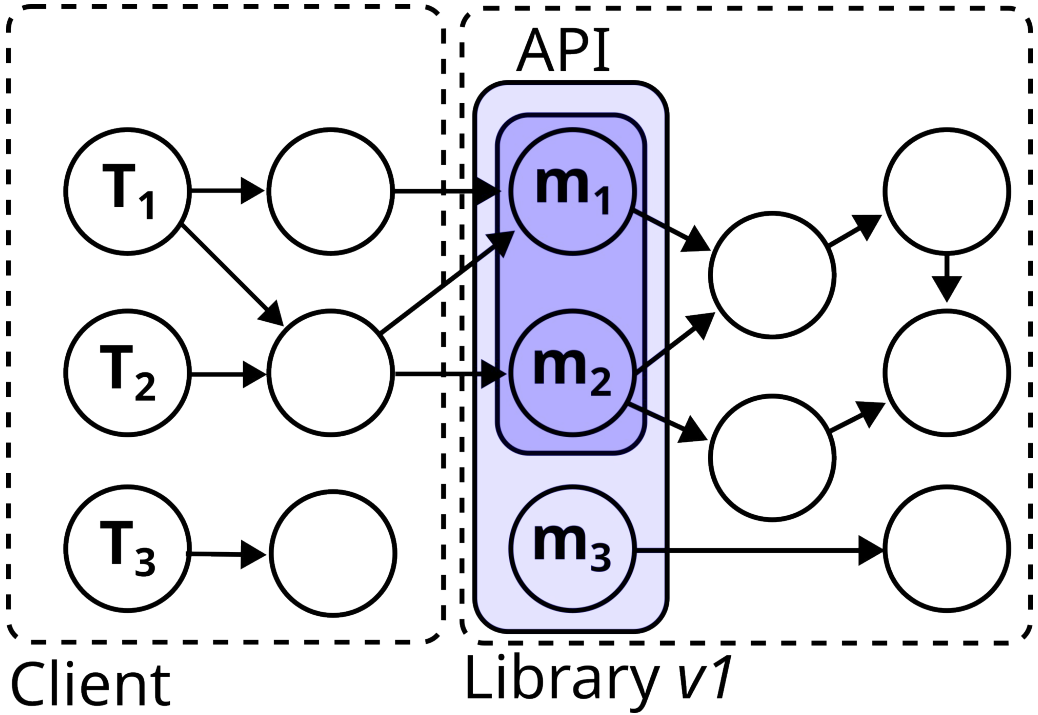


API Interaction Snapshots with Gilesi



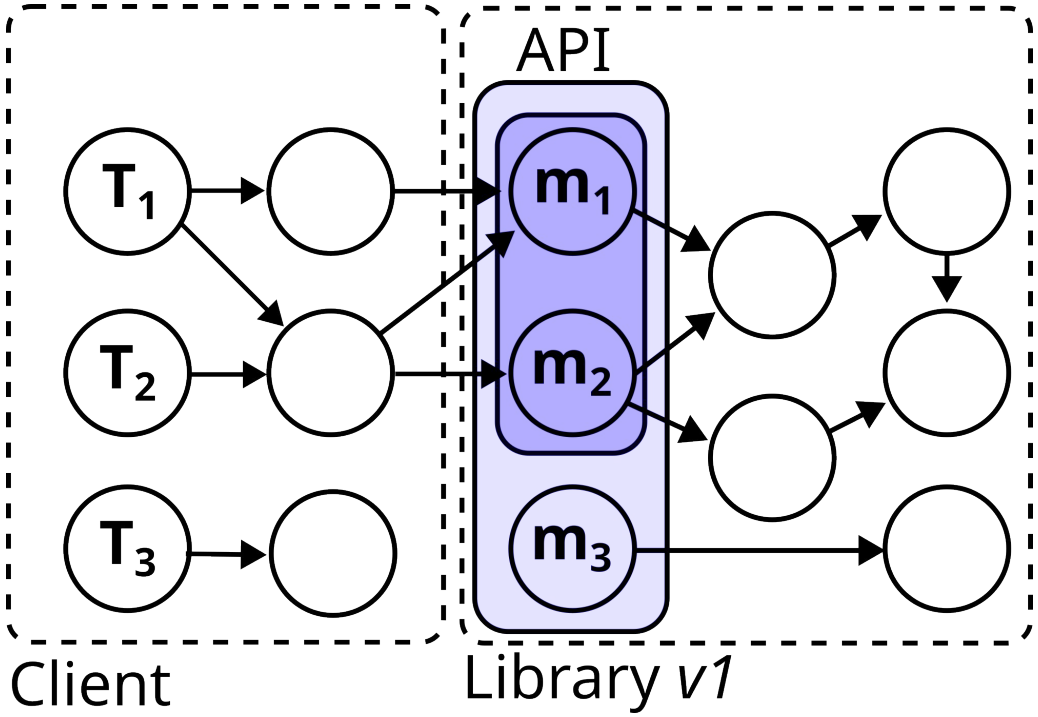
1. Identify the library's API

API Interaction Snapshots with Gilesi



1. Identify the library's API
2. Identify the library's footprint

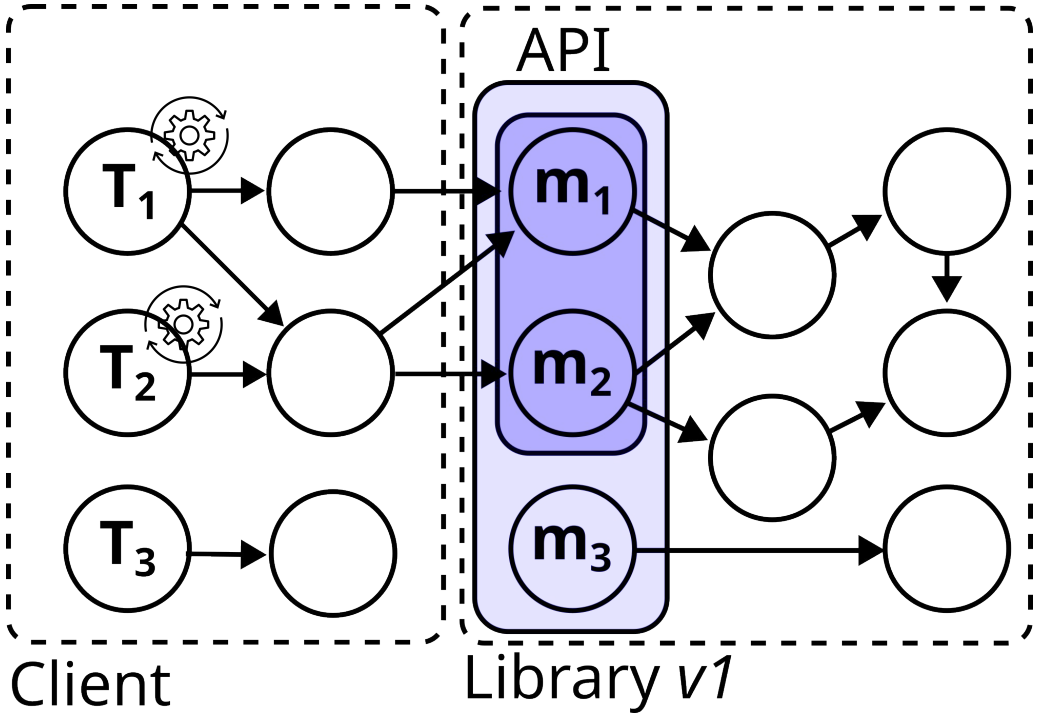
API Interaction Snapshots with Gilesi



1. Identify the library's API
2. Identify the library's footprint
3. Instrument the footprint with probes, recording:
   a. Invoked method
   b. Receiving object
   c. In-parameter values
   d. Out-values/exceptions

$$I_1 = \langle m_1, o_1, \langle p_1, \ldots, p_n \rangle, r_1 \rangle$$

API Interaction Snapshots with Gilesi



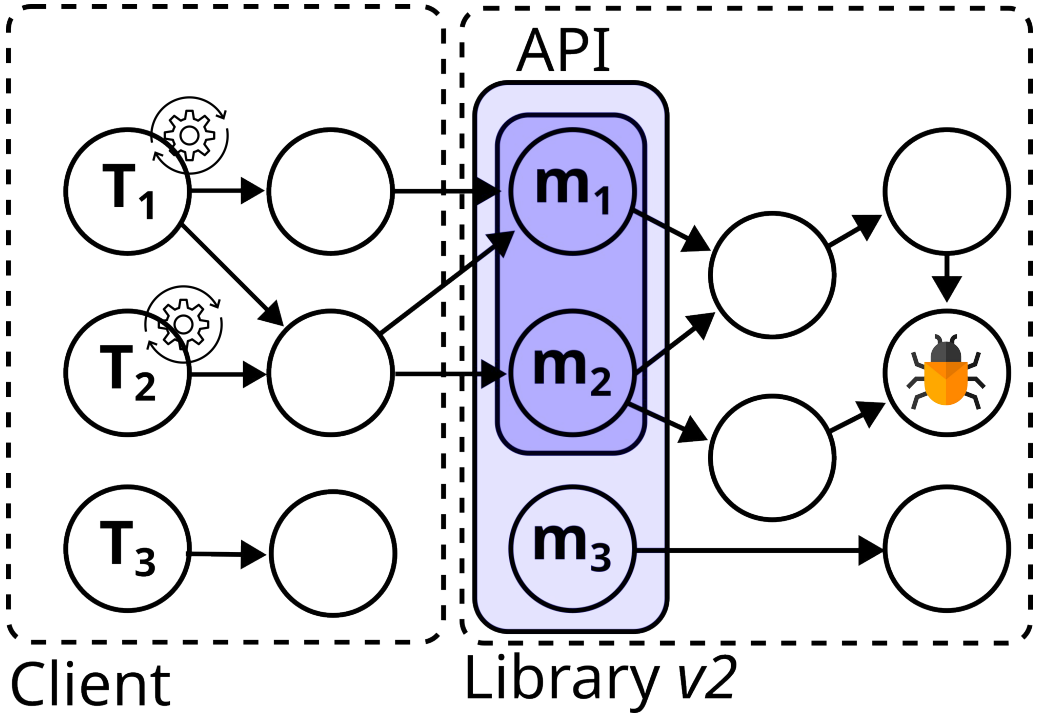
T1's Snapshot (v1)

$$I_1 = \langle m_1, o_1, \langle p_1, \ldots, p_n \rangle, r_1 \rangle$$

T2's Snapshot (v1)

$$I_1 = \langle m_1, o_2, \langle p_1, \ldots, p_n \rangle, r_2 \rangle$$
$$I_2 = \langle m_2, o_2, \langle p_2, \ldots, p_n \rangle, r_3 \rangle$$

API Interaction Snapshots with Gilesi



T1's Snapshot (v2)

$$I_1 = \langle m_1, o_1, \langle p_1, \ldots, p_n \rangle, r_1 \rangle$$

T2's Snapshot (v2)

$$I_1 = \langle m_1, o_2, \langle p_1, \ldots, p_n \rangle, r_2 \rangle$$
$$I_2 = \langle m_2, o_2, \langle p_2, \ldots, p_n \rangle, r_3 \rangle$$

API Interaction Snapshots with Gilesi

```
Test1_Snapshot1 = (
  ⟨StrTokenizer.<init>, ∅, “apple,banana”, o1⟩,
  ⟨setDelimiter, o1, “,”, ∅⟩,
  ⟨getTokenList, o1, ∅, ArrayList(“apple”, “banana”)⟩
)
```

An example test’s snapshot on commons-text:1.9

```
Test1_Snapshot2 = (
  ⟨StrTokenizer.<init>, ∅, “apple,banana”, o1⟩,
  ⟨setDelimiter, o1, “,”, ∅⟩,
  ⟨getTokenList, o1, ∅, Arrays$List(“apple”, “banana”)⟩
)
```

An example test’s snapshot on commons-text:1.10

- Pairwise comparison of old and new test snapshots
- Any discrepancy in
  - Protocol
  - Types
  - Exchanged values
  - etc.
- Signals a behavioral perturbation that *might* be a BBC and warrants investigation

An example BBC introduced in `commons-text` commit `2d1ab7`. Identified in `TEXT-219` and later fixed in commit `f9846b`.

Evaluation: Case Study

- `Jsoup` & `commons-lang3`, 27 clients
- Injecting artificial faults in the library through extreme mutations

Library	Mutants	Killed by	
		Tests	GILESI
COMMONS-LANG3	106	100	105
JSOUP	52	41	46

- Client tests miss even extreme mutations
- Surviving mutants are i) equivalent or ii) unobservable I/O side effects
- Realistic/subtle mutations more likely to be caught by Gilesi

From New Idea to ~~Emerging~~ Results?

- Concurrency, non-determinism, snapshot flakiness
- Test generation/amplification for greater coverage and observability
- Behavioral changes vs. behavioral *breaking* changes; *what's breaking?*
- Scale from small-scale case study to large-scale datasets
- Extensive comparison (Uppdatera, SemBid, DeBBI, CompCheck)

BUMP: A Benchmark of Reproducible Breaking Dependency Updates

1st Frank Reyes
KTH Royal Institute of Technology
Stockholm, Sweden
frankrg@kth.se

2nd Yogya Gamage
KTH Royal Institute of Technology
Stockholm, Sweden
yogya@kth.se

3rd Gabriel Skoglund
KTH Royal Institute of Technology
Stockholm, Sweden
gabsko@kth.se

CompSuite: A Dataset of Java Library Upgrade Incompatibility Issues

Xiufeng Xu
xiufeng001@e.ntu.edu.sg
Nanyang Technological University
Singapore

Chenguang Zhu
cgzhu@utexas.edu
The University of Texas at Austin
USA

Yi Li
yi_li@ntu.edu.sg
Nanyang Technological University
Singapore

DUETS: A Dataset of Reproducible Pairs of Java Library-Clients

Thomas Durieux, César Soto-Valero, and Benoit Baudry

Awesome community work <3

Some Key Insights

- Snapshots embody the **exact expectations** of a particular client towards a specific version of a library
- Leveraging the execution paths offered by client tests' with stronger snapshot-based assertions increases the **sensitivity** of the tests and the **observability** of behavioral changes
- Asserting directly at the API–client boundary **eases diagnosis and remediation**

ICSME 2025 Auckland, New Zealand

université de BORDEAUX

cnrs

LaBRI