Markus Wagner

Heuristically Creating Test Cases for Program Verification Systems

Joint work with Bernard Beckert and Thorsten Bormer
from the Karlsruhe Institute of Technology

Who guards the guardians?

How to improve trust in verification systems?

$$a = b \vdash 2 = 1$$

Modern verification systems are large and complex systems

- Soundness bugs are not rare
- Such bugs are often hard to detect in a real proof

“Auto-active” Verification Systems

Validating verification systems by

- Formal methods
- Code inspection
- Testing
- …

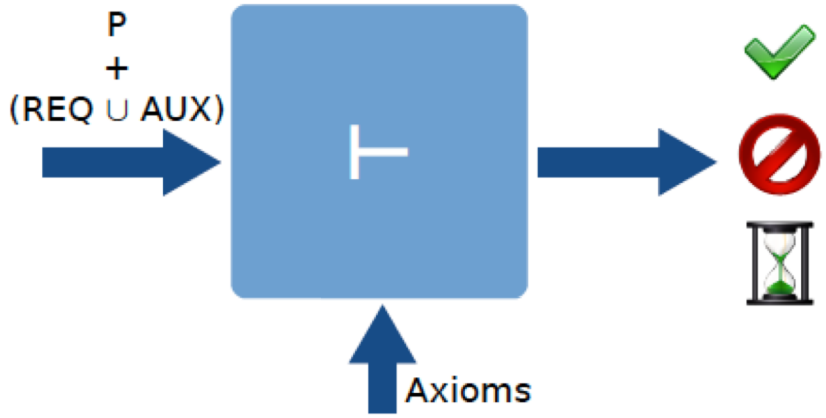
Program Language Semantics

Static checkers Verifying compilers Logic frameworks

We have to test both!

But how to determine the quality of the test cases?

Test Cases



A test case is a program *P*, together with requirement and auxiliary specifications.

Computing coverage for the test cases takes from a few minutes to several hours.

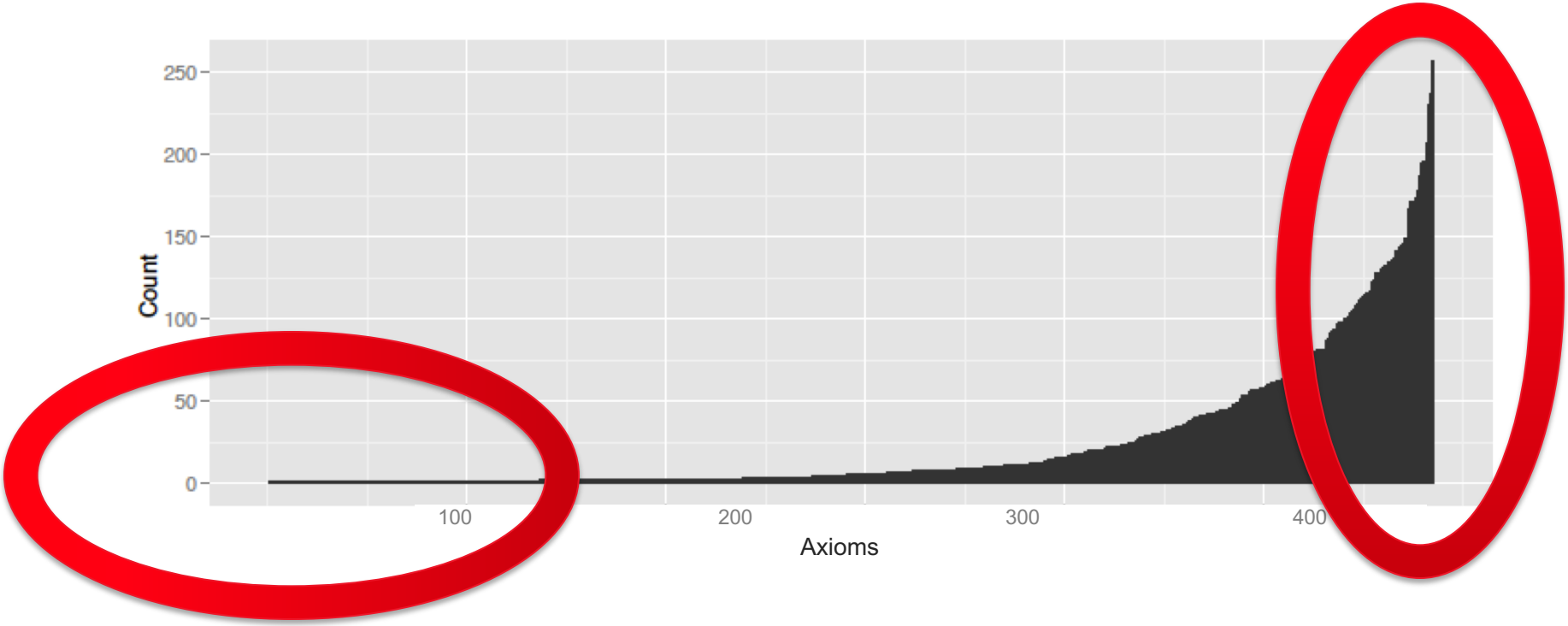
Case study: KeY

The KeY System

- Deductive verification system for JavaCard
- Sequent calculus for Java Dynamic Logic, uses symbolic execution for Java programs
- Interactive verification with automatic proof mode

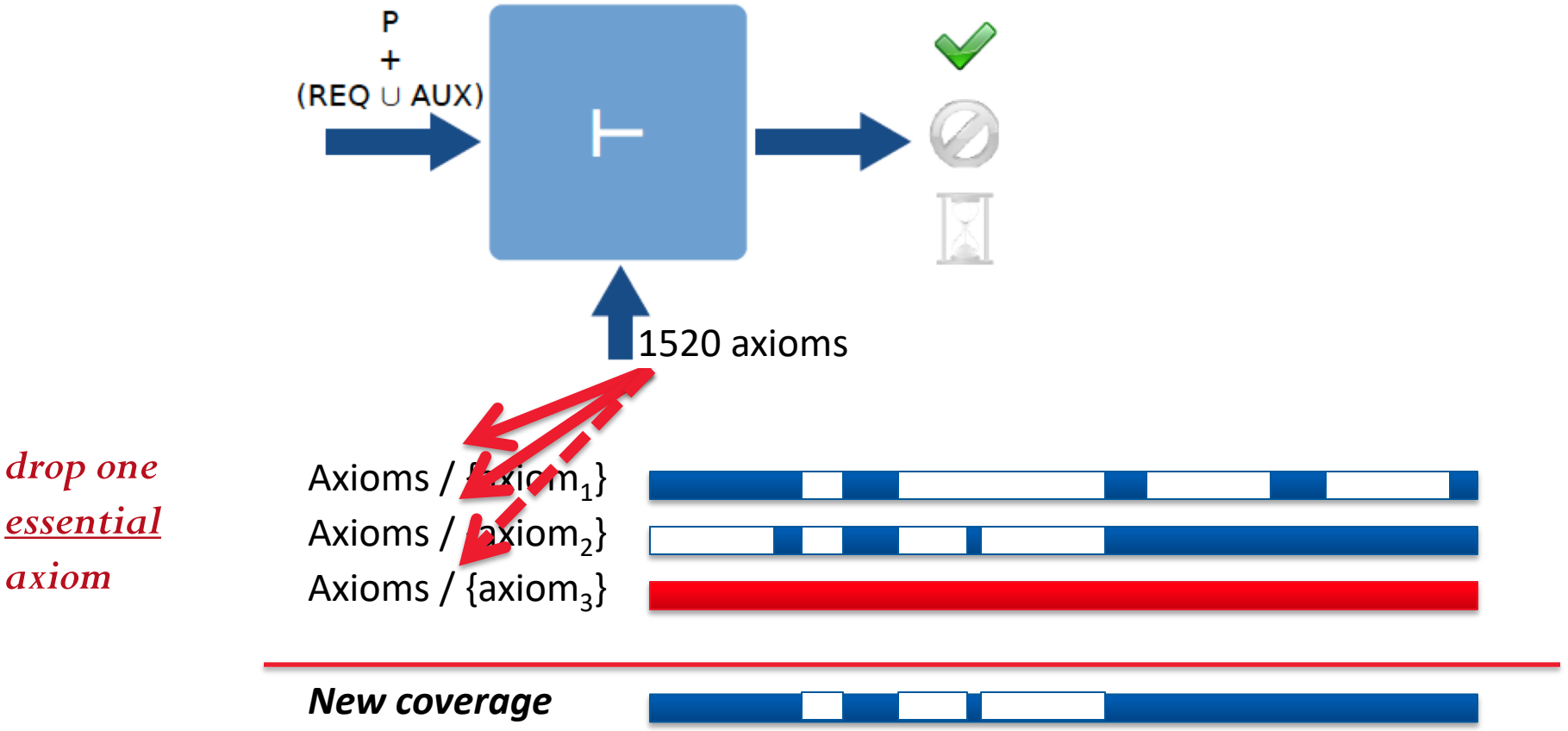
Coverage Results (naïve, TAP 2013)

The 319 completeness tests of KeY covered 31% of all axioms (474 out of 1520).



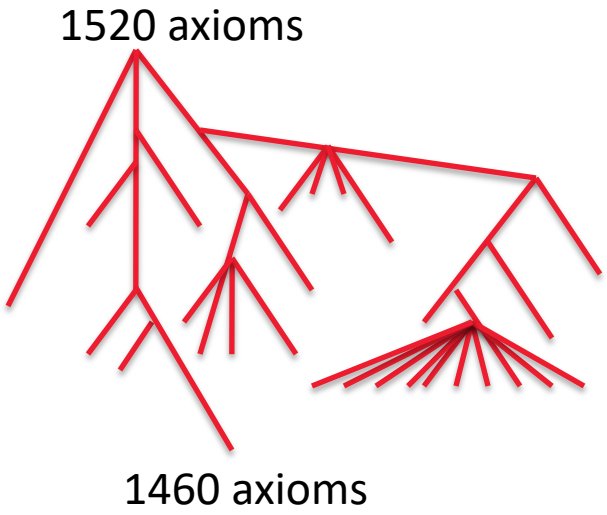
Heuristic Approaches

Reusing Test Cases



Idea: given a test case *T*, run the tool with just a subset of the 1520 axioms.

Reusing Test Cases

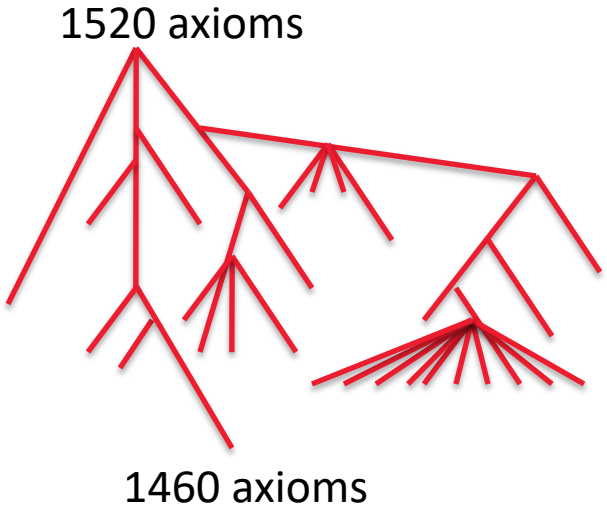


Three simple heuristics to pick the “next axiom to drop”:

1. Depth-first
2. Random selection
3. Greedy (try to remove groups)

Complimentary by design, verified by experiments (see Table 3).

Reusing Test Cases



Resources:
24h per heuristic
per test case

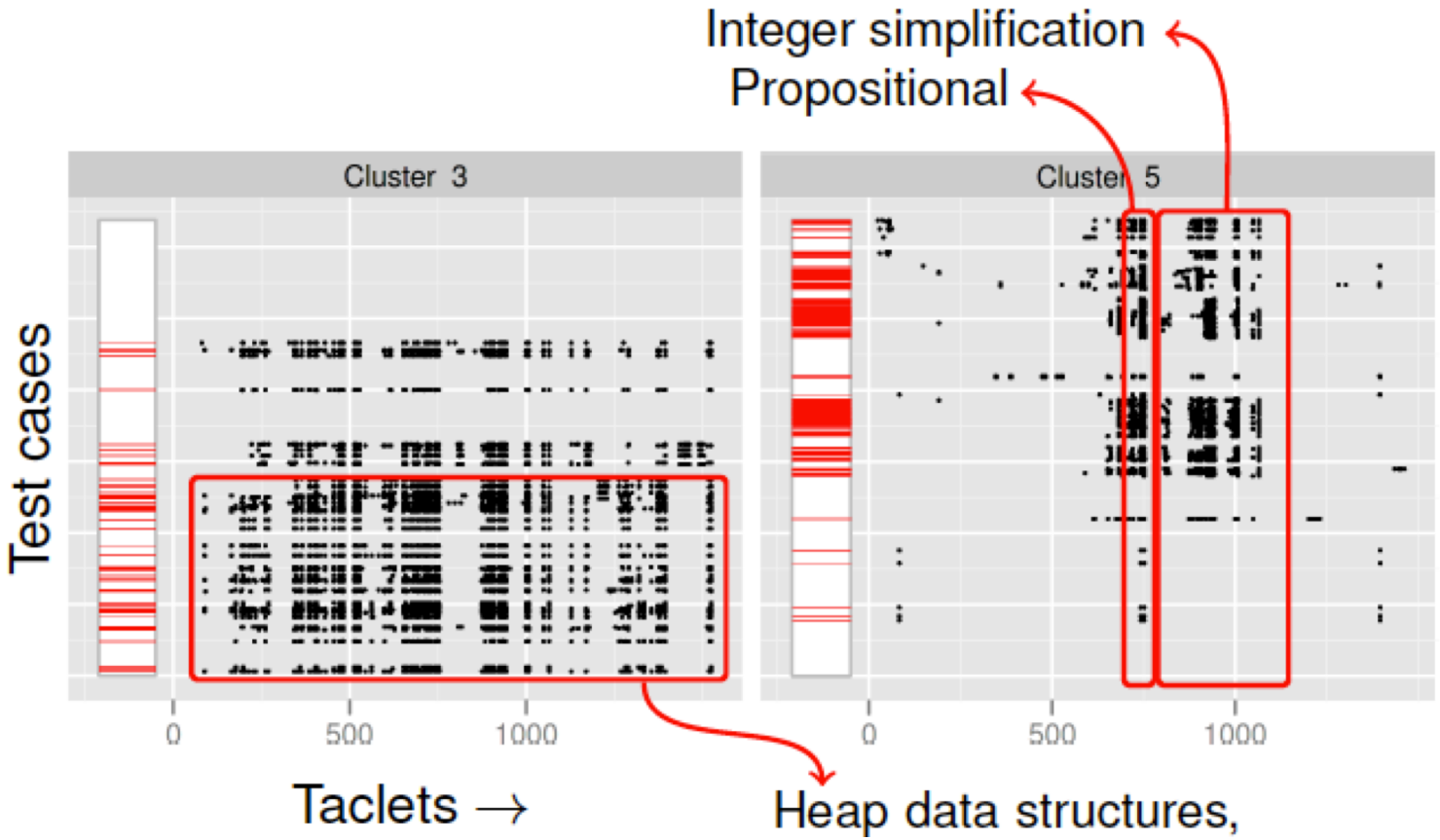
Three simple heuristics to pick the "next axiom to drop":

[0. Base case] 474 (31%)

1. Depth-first
2. Random selection
3. Greedy (try to remove groups)

Complimentary by design, verified by experiments (see Table 3).

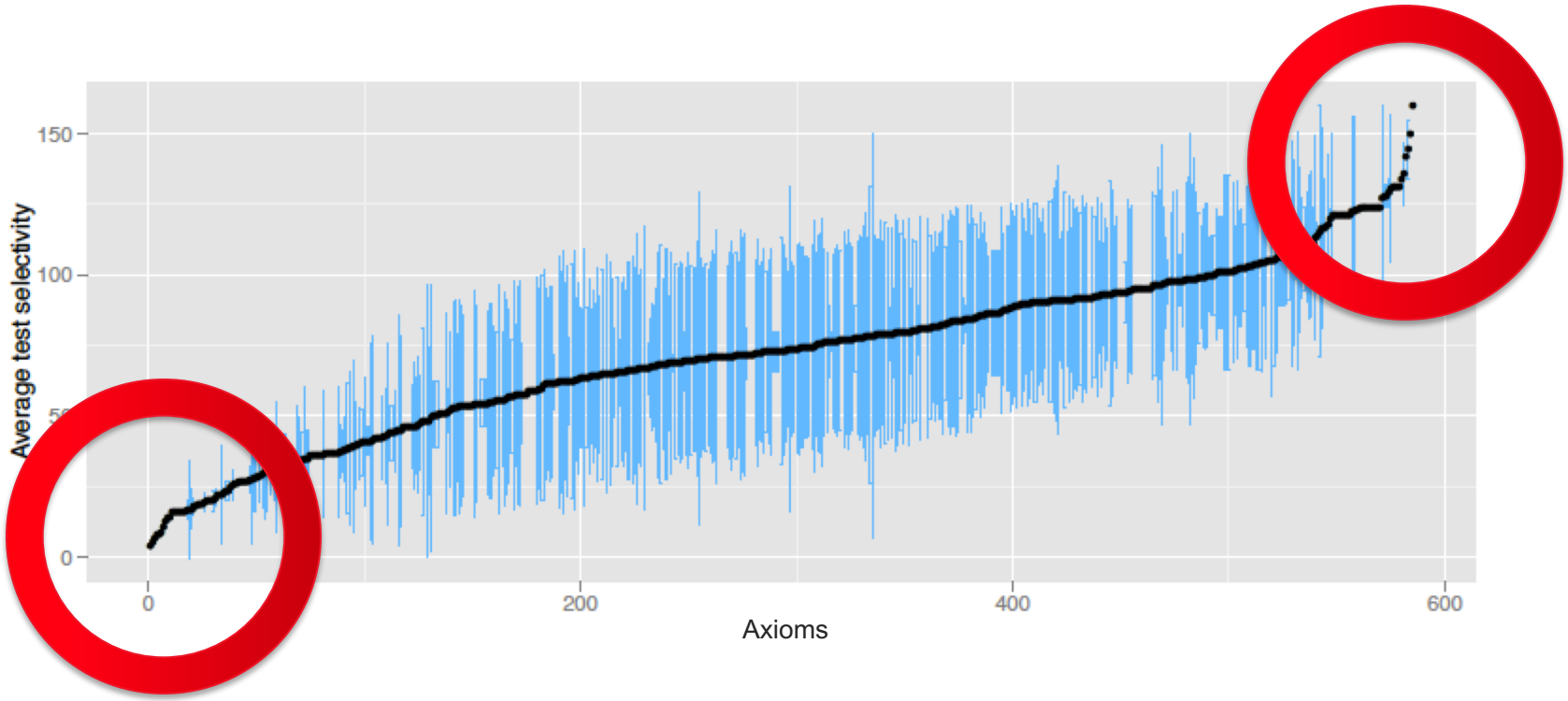
Clustering Analysis (excerpt)



MIC 2015 NOTTINGHAM

Dario Silva & Per Kristian Lehre — University of Nottingham
Markus Wagner — University of Adelaide

Test Case Selectivity



Only specific test cases, or test cases with broad coverage for an axiom may not be sufficient.