Saint Petersburg Research Center

INTERNSHIPS

Encoding of Java Bytecode to CIRCUIT-SAT
To Solve Programs Equivalence Problem

Daniil Lyubaev

Project background

Code Clone Detection

- 4 types of clones

- Interested in 4th, the semantic clones

- Enough tools to detect types 1, 2, 3, but few to detect 4

- Mostly textual, lexical, AST-, Tree-, Metric- based, ML

- Good decision is to abstract from source code

Problem statement

- Why not use SAT-solvers?

- But we need Boolean Schemes…

Solution: translate Java Bytecode to a Boolean circuit
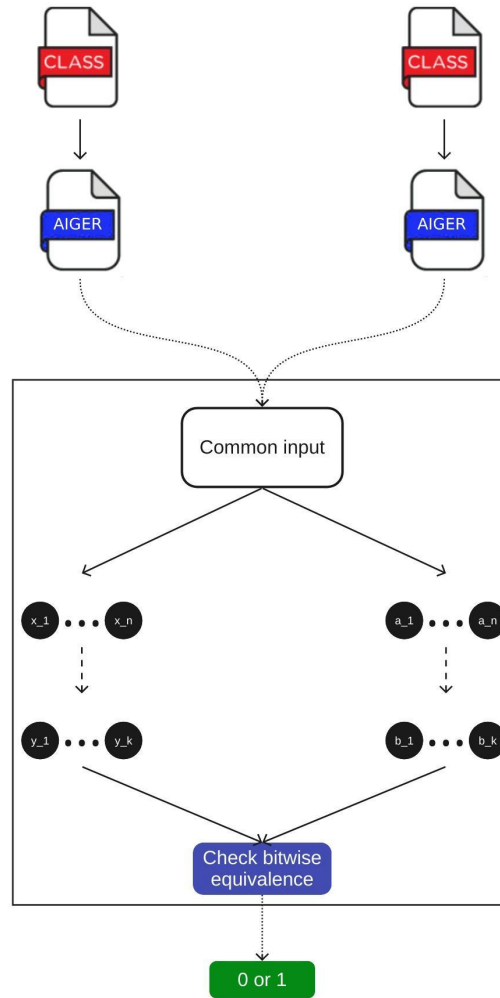
Competitors Vs proposed solution

My project	Translation + SAT-solvers	Java Bytecode	Detects type 4
DroidCC	Textual + Token-Based, MD5 Hash	Java (Android)	Detects types 1, 2, 3
Detection Framework	Tree-based + Semantic, SVM, Forests, etc..	Java	Detects types 1, 2, 3, 4
LWH	Textual + Metric	C, Java	Detects types 1, 2, 3, 4

Conclusion: usage of translation and SAT-solvers can be a very good approach to detect semantic clones

Different users – different scenarios

Programmers	Researchers
Detect bugs in duplicated code	Encode domain-specific things to circuits
If bug occurs in one part of code, it can occur in an equivalent part as well	Write in high-level language, get encoding like you wrote in any language
Use our tool, Fix bugs	Use our tool, write in Java / Kotlin using primitive variables, get encoding

What has been done?

- Created console utility called **transbyte;**
- Used Kotlin, Apache BCEL;
- Encoded simple instructions: sum, multiply, XOR, AND, OR;
- Encoded some cryptographic functions: LFSR, A5/1, Wolfram Generator, checked on equivalence;
- Encoded Bubble, Selection, Insertion, Pancake Sortings, checked for equivalence;
- **Sortings are checked for pairwise equivalence (so, checked to be semantic clones)**

Demo

Next step

- Add full classes support

- Add multidimensional arrays

- Encode DFS / BFS

- Add more bytecode instructions...

Team

- Daniil Lyubaev
- Alexander Semenov
- Ilya Otpuschennikov

With help from

- Dmitry Ivanov
- Konstantin Chukharev
- Stepan Kochemazov

Daniil Lyubaev

Contacts for questions

E-mail: lyubaevda@gmail.com

Telegram: @eqimd

Project link: