

Impact of Change Granularity in Refactoring Detection



Lei Chen

Tokyo Institute of Technology
School of Computing



Shinpei Hayashi

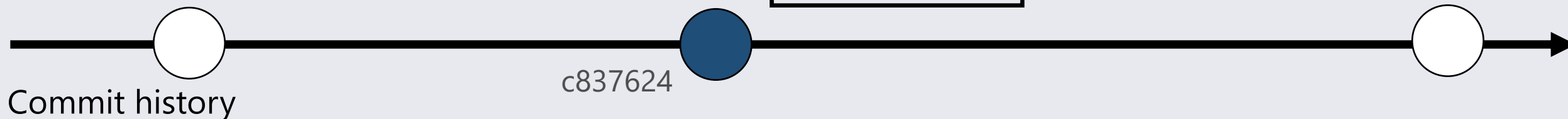
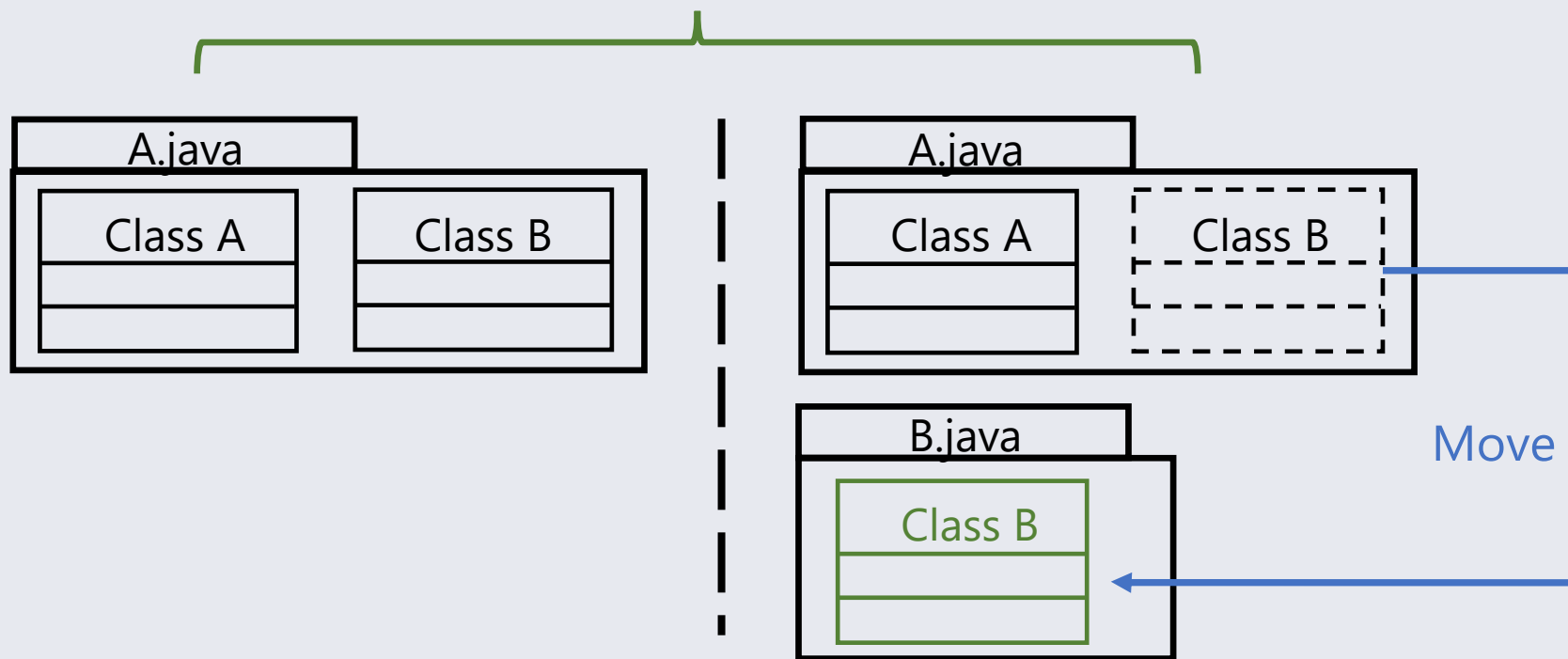
Tokyo Institute of Technology
School of Computing



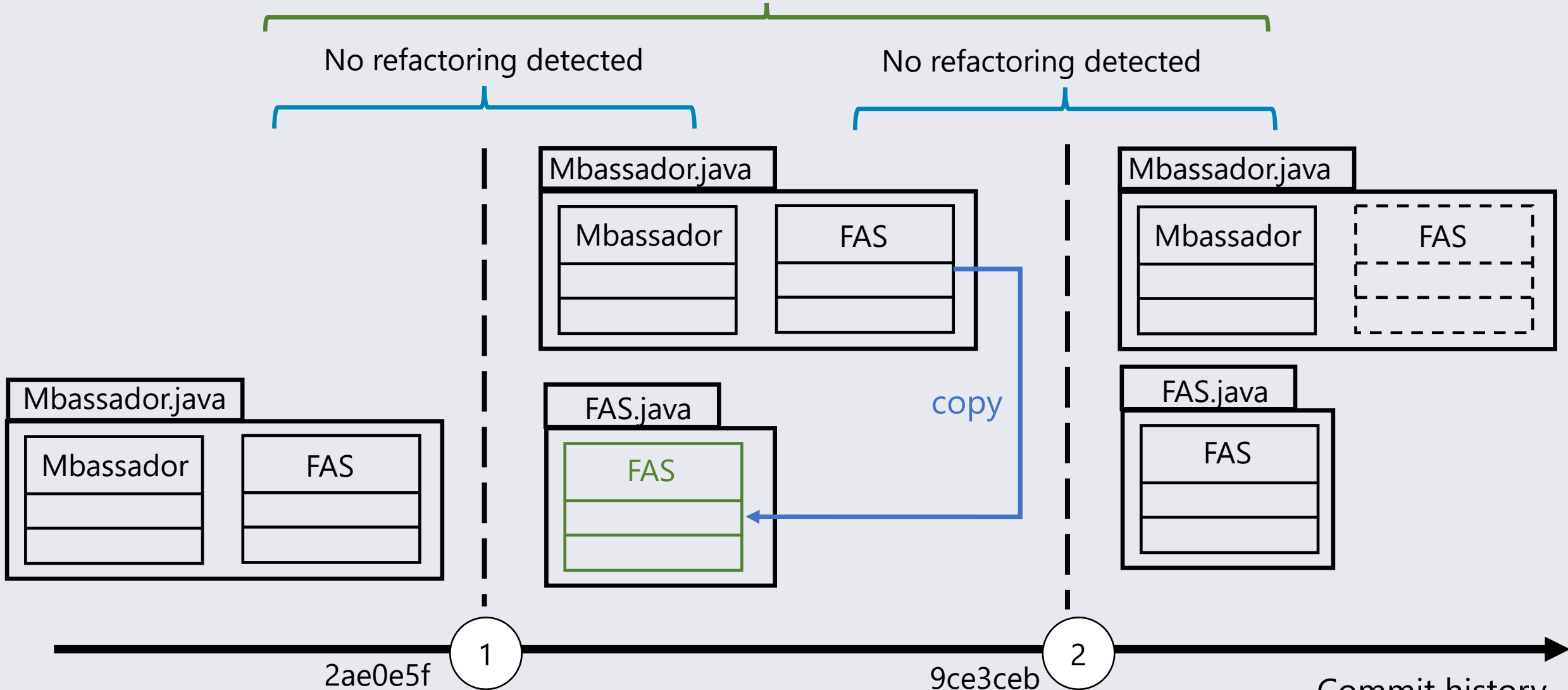
Tokyo Tech

Refactoring detectors detect refactorings by comparing two **source code snapshots**

Refactoring detected: Move Class



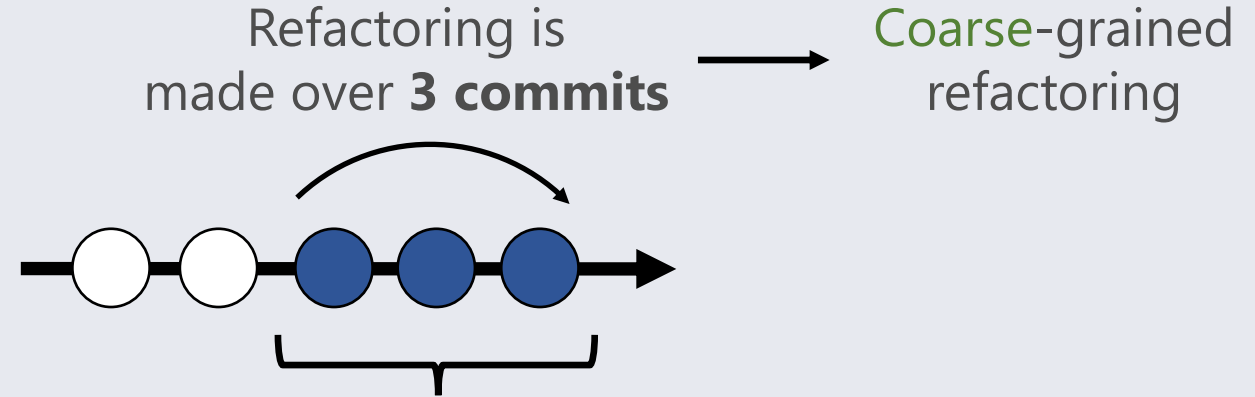
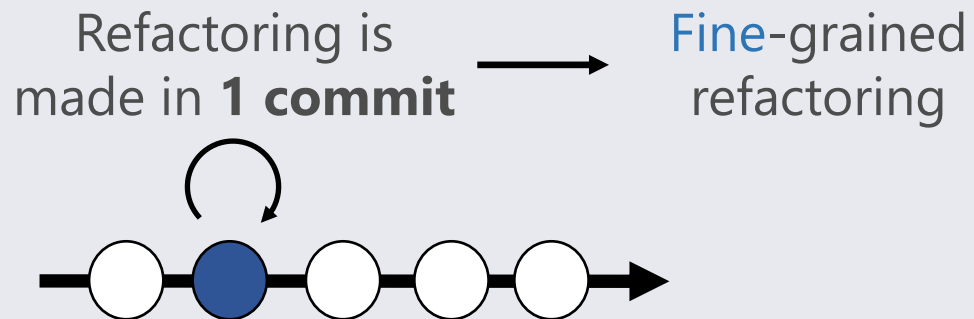
Refactoring detected: Move Class



Commit history
of Mbassador^[1]

[1] <https://github.com/bennidi/mbassador>

- Refactorings that are performed over multiple commits are referred as coarse-grained refactoring (CGR)



? What's the impact of CGR

RQ1: Frequency

RQ1: How frequently do CGRs appear?

A: CGRs are **common**, CGRs appear in all repositories in our dataset

RQ2: Type

RQ2: Which types of refactorings tend to be coarse-grained?

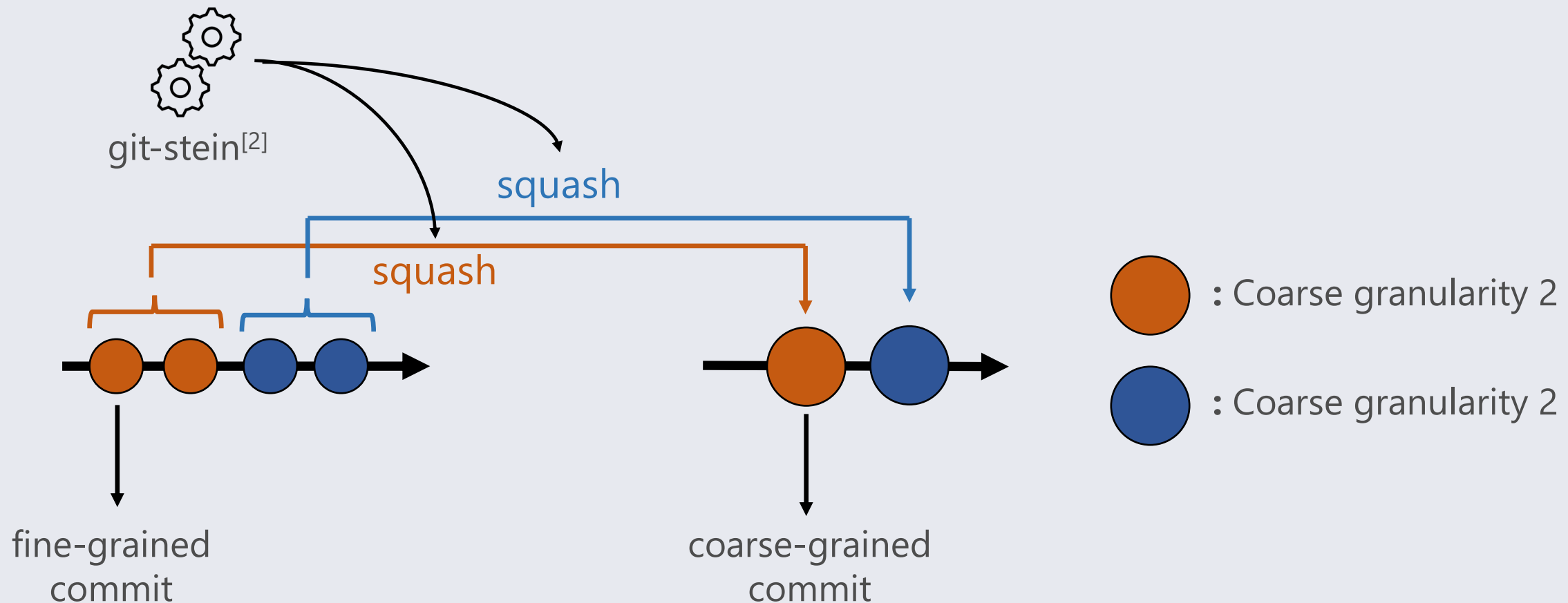
A: **Move-related refactoring** types tend to be coarse-grained

RQ3: Reasons

RQ3: What are the reasons for the occurrence of CGRs?

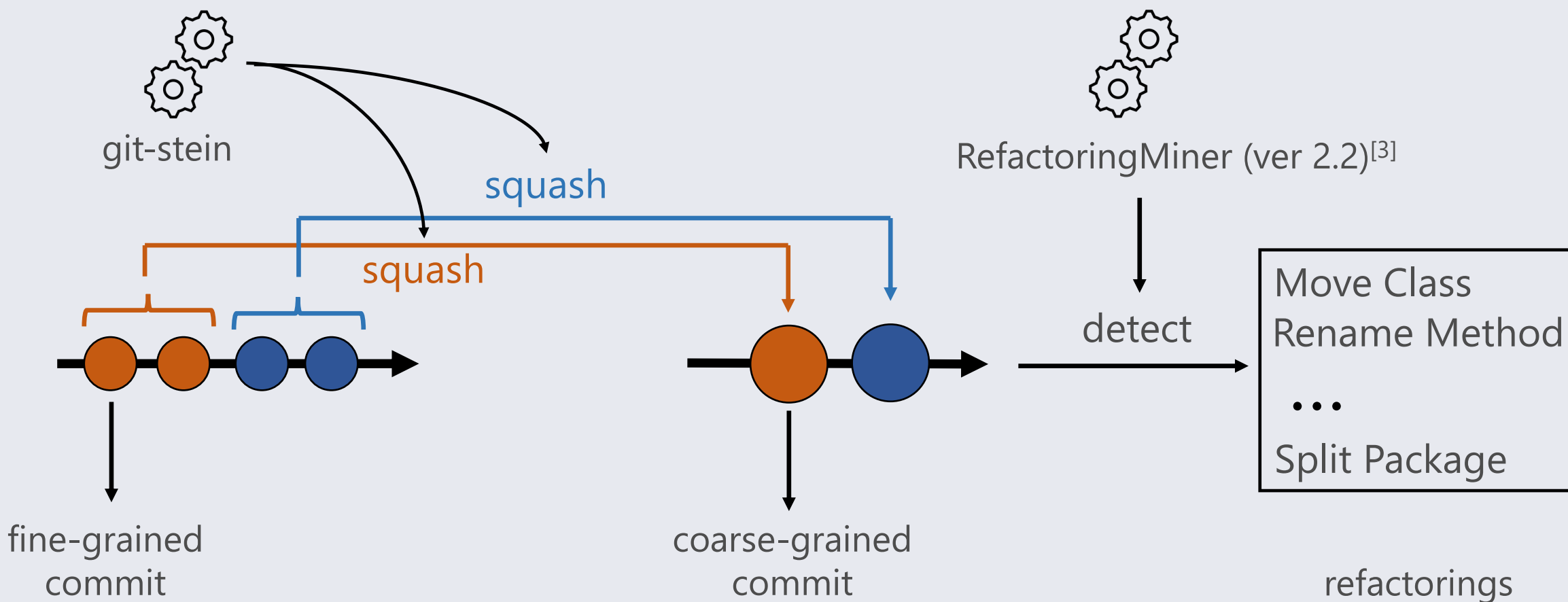
A: Reasons for the occurrence of CGRs can be categorized into **two types** according to their composition: **Generation** and **Combination**

- Fine-grained commits are squashed into coarse-grained ones
- Number of fine-grained commits squashed into one coarse-grained one is referred to as coarse granularity



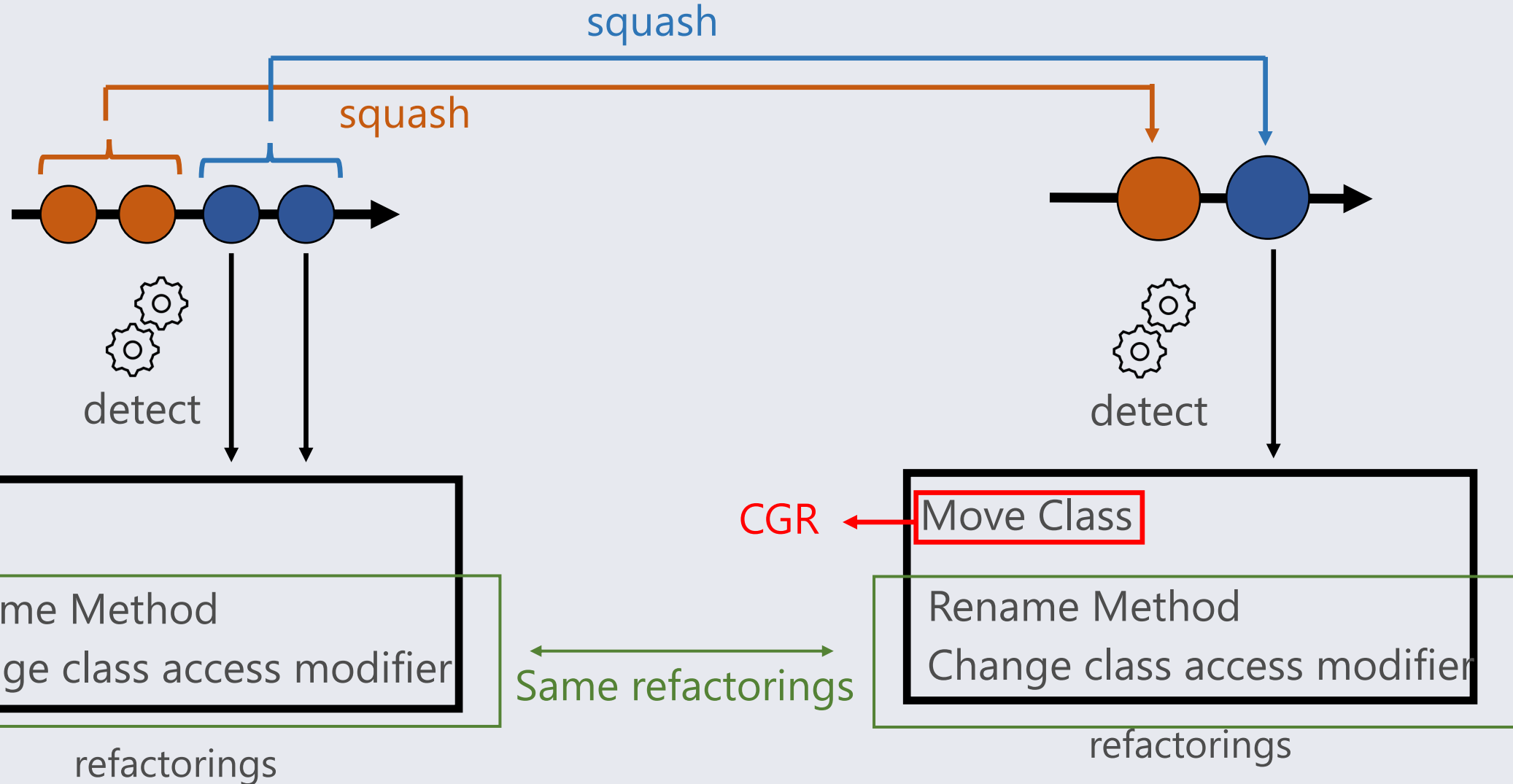
[2]<https://github.com/sh5i/git-stein>

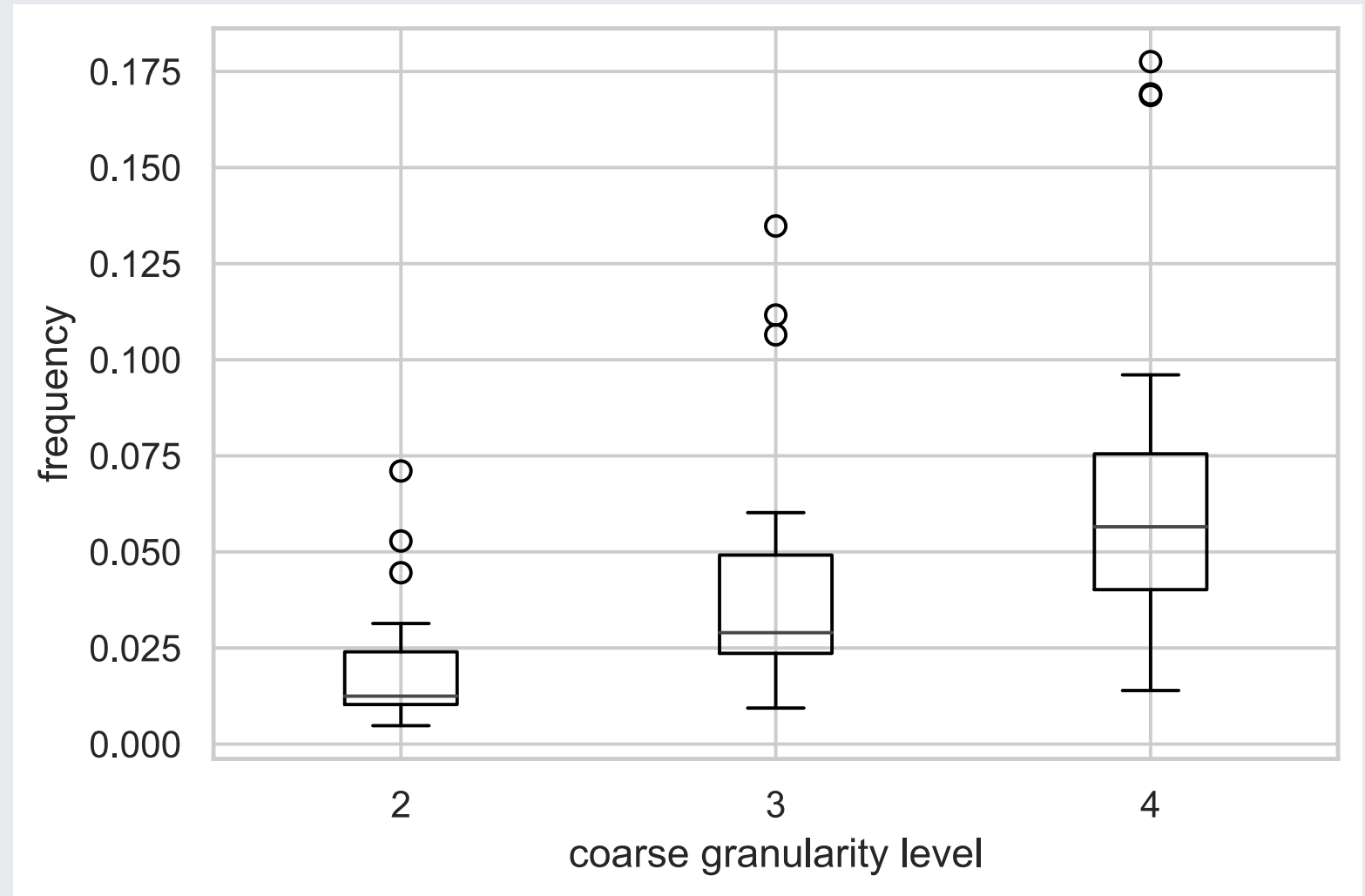
- CGRs can be detected by detectors in coarse-grained commits



[3] <https://github.com/tsantalis/RefactoringMiner>

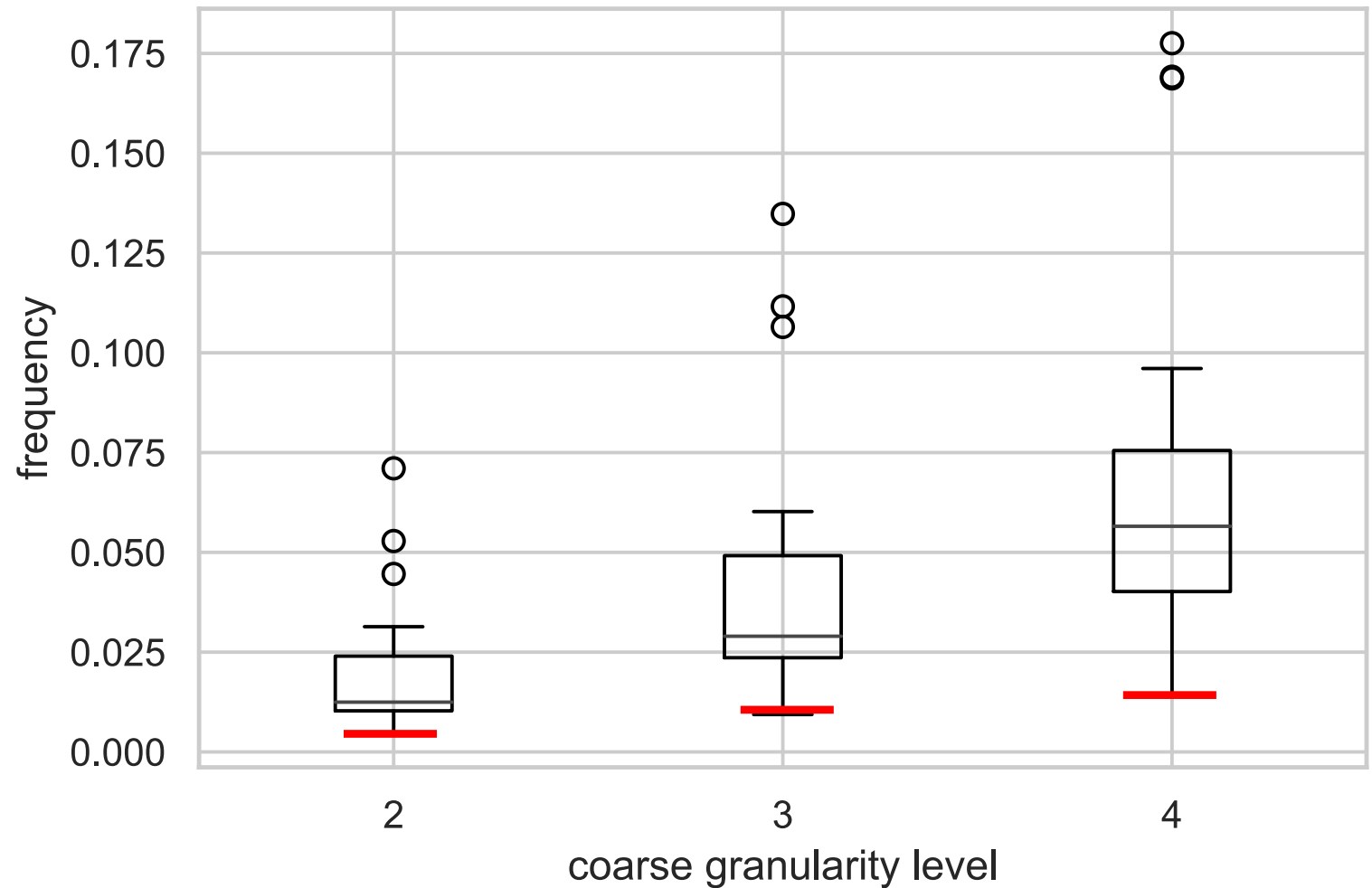
- CGR can be detected only in coarse-grained commit





Boxplots for frequency of CGRs

1. CGR is common

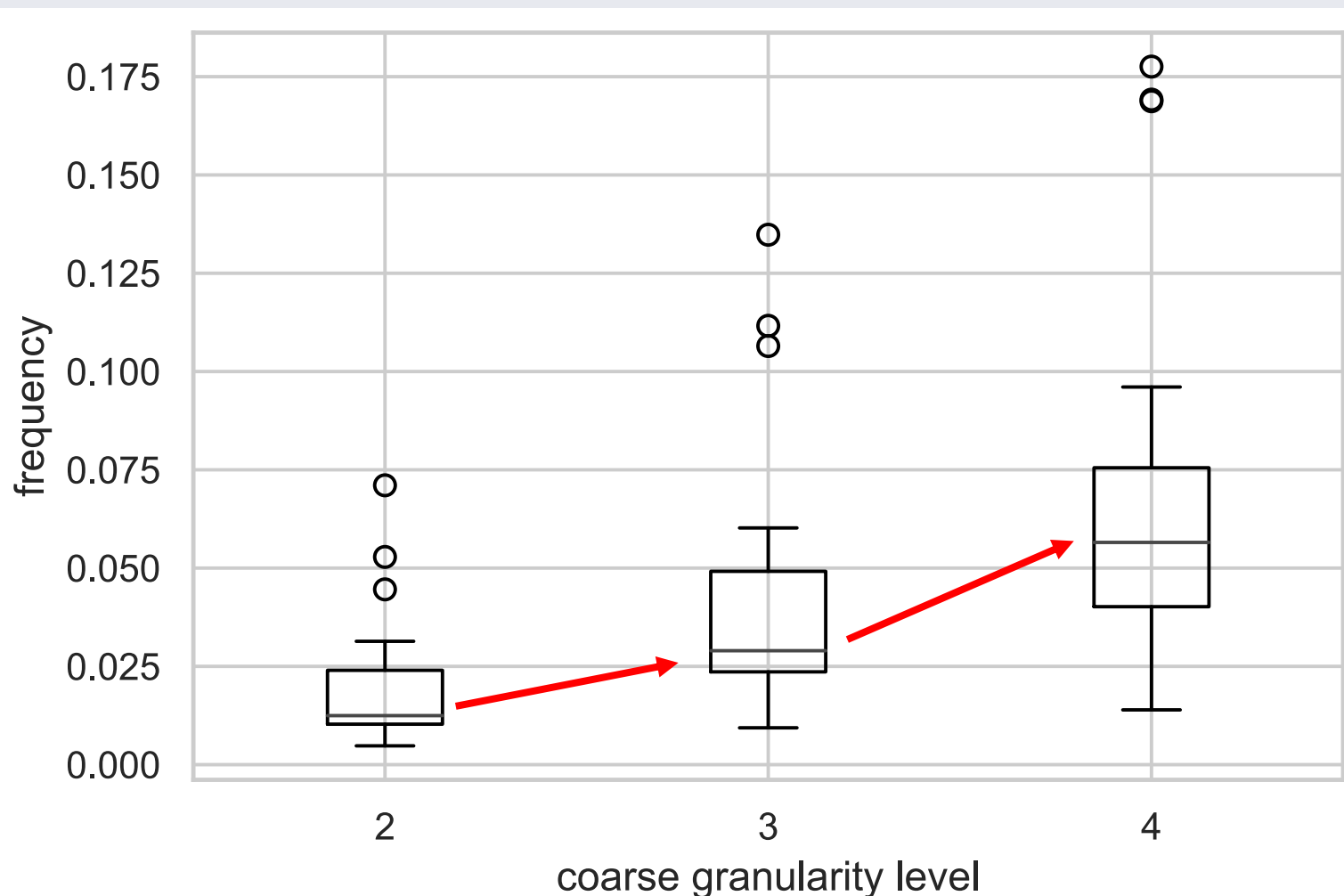


Boxplots for frequency of CGRs

1. CGR is common
2. CGRs are more frequent when coarse granularity increases



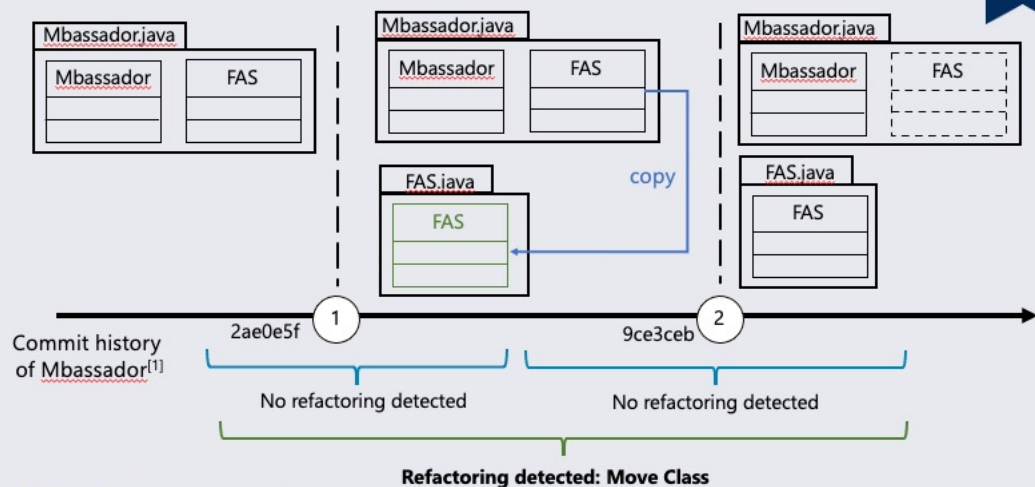
**Refactoring detectors
should cover CGRs**



Boxplots for frequency of CGRs

Motivation

2



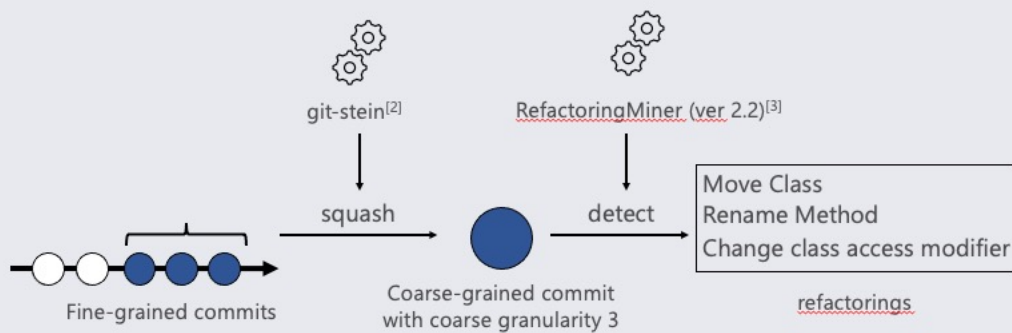
[1] <https://github.com/bennidi/mbassador>

Study Design

5



➤ CGR can be detected by detectors if code changes are merged



[2] <https://github.com/sh5i/git-stein>

[3] <https://github.com/tsantalts/RefactoringMiner>

Research Questions

4



RQ1: Frequency

RQ1: How frequently do CGRs appear because of granularity change?

A: CGR is common, and CGRs are more frequent when coarse granularity increases

RQ2: Type

RQ2: Which types of refactorings tend to be coarse-grained?

A: Move-related refactoring types tend to be coarse-grained

RQ3: Reasons

RQ3: What are the reasons for the occurrence of CGRs?

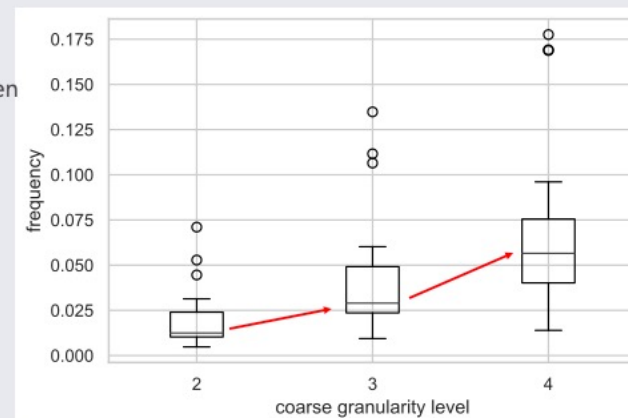
A: Reasons for the occurrence of CGRs can be categorized into two types according to their composition: Generation and Combination

Frequency of CGR

9



1. CGR is common
2. CGRs are more frequent when coarse granularity increases



Boxplots for frequency of CGRs