

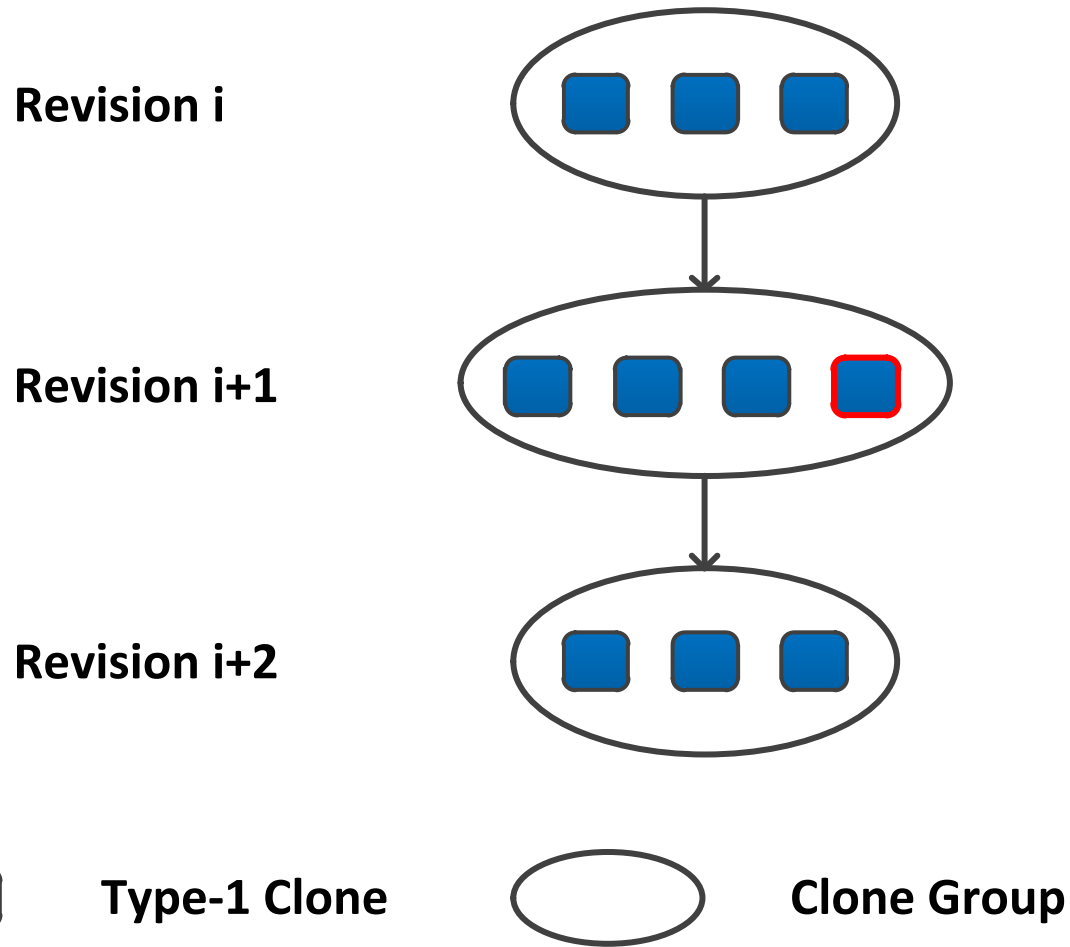
An Empirical Study of the Fault-Proneness of Clone Mutation and Clone Migration

Shuai Xie, Foutse Khomh, Ying Zou

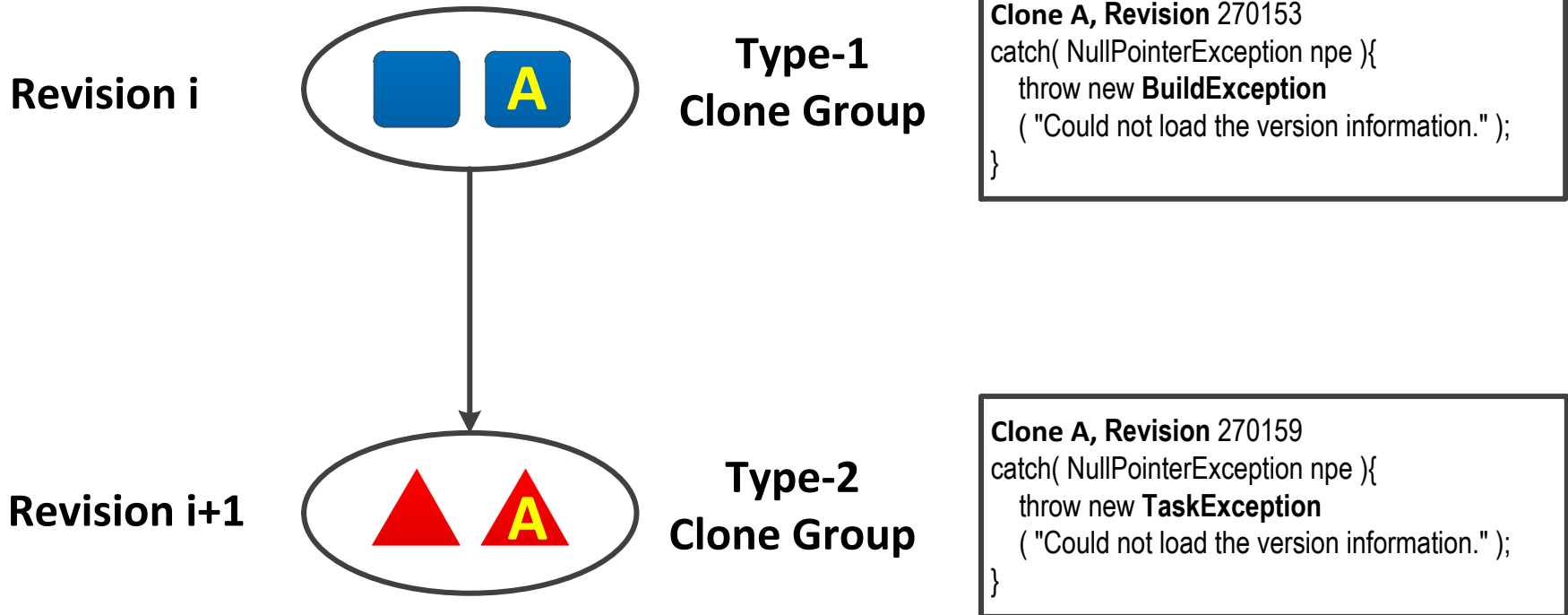
Department of Electrical and Computing
Engineering



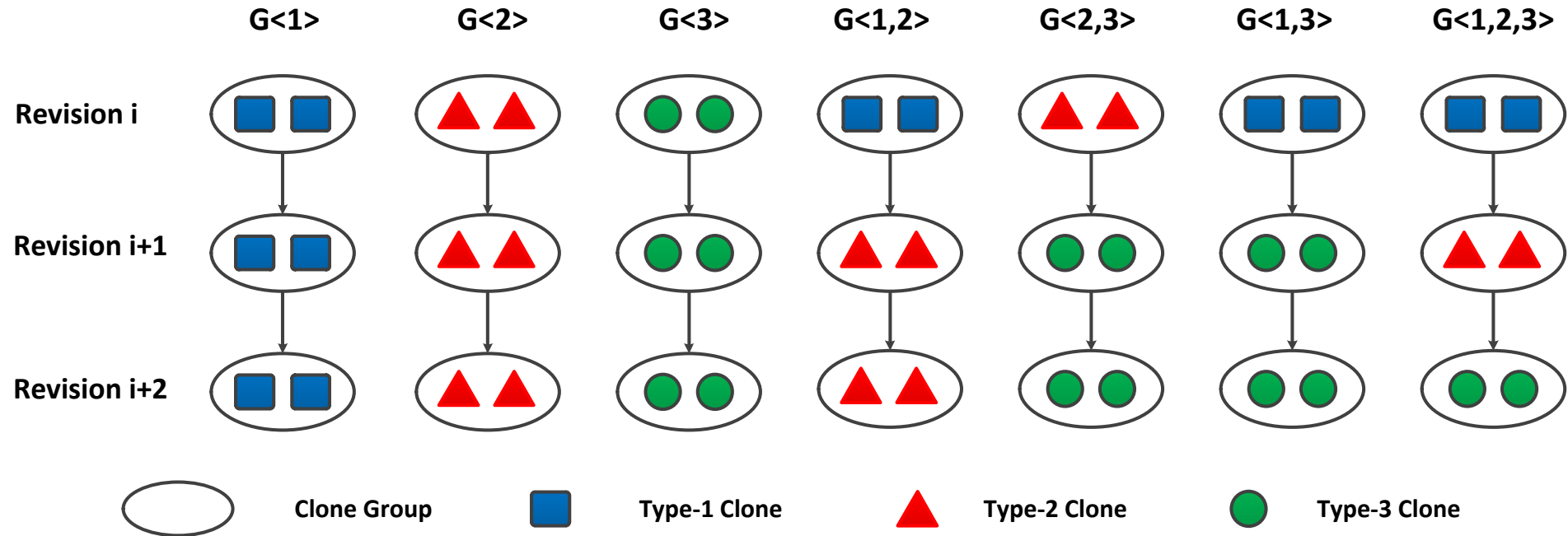
Clone Genealogies



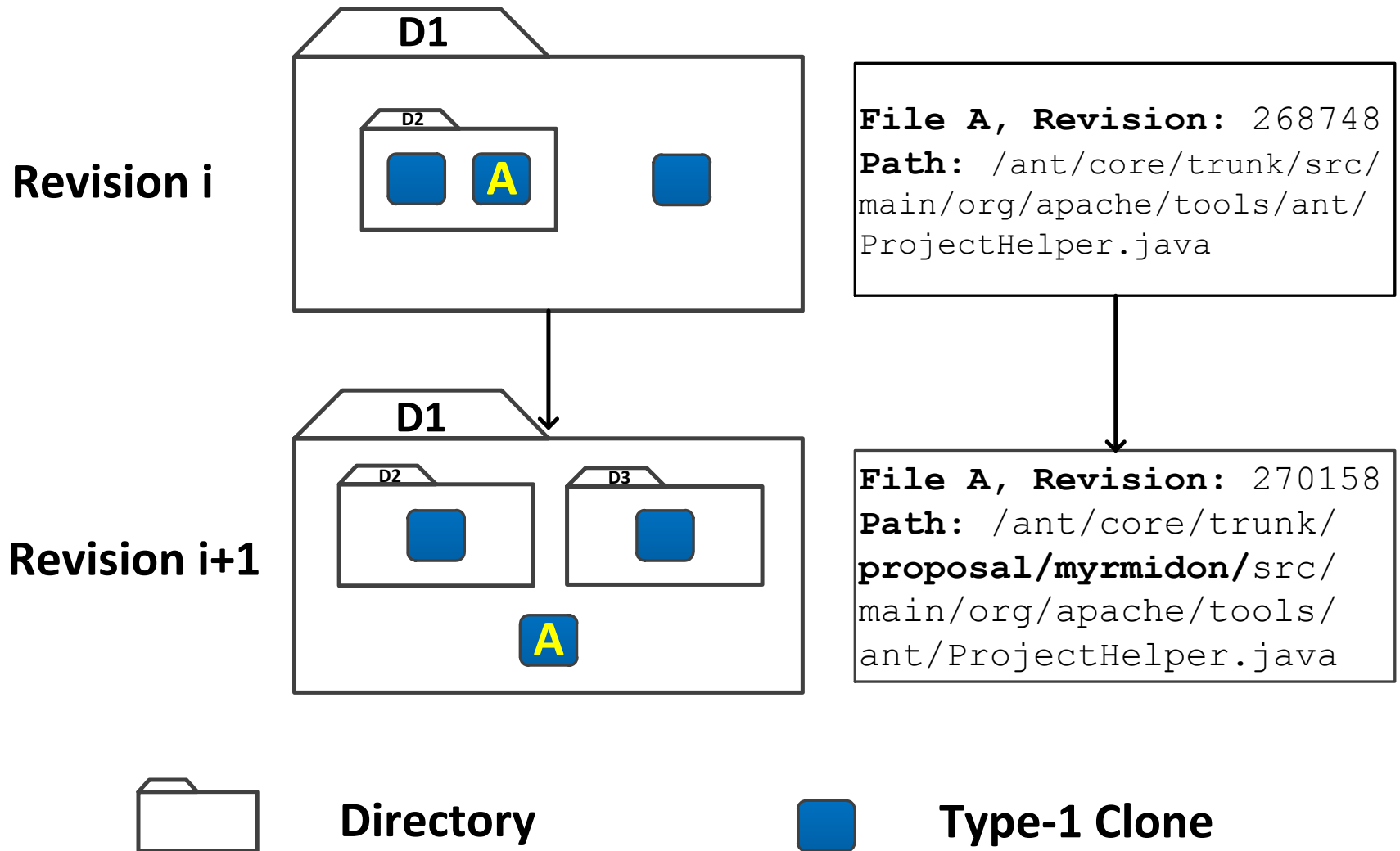
Clone Mutation





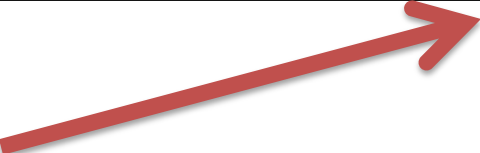
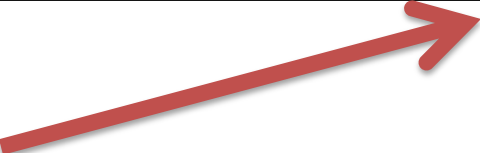
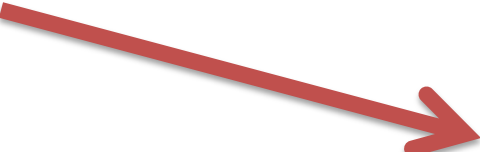
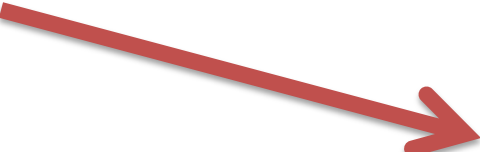


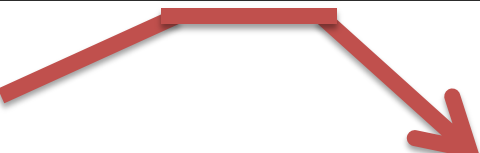
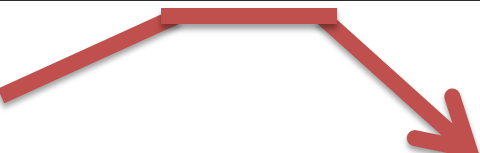
Clone Mutation Categories





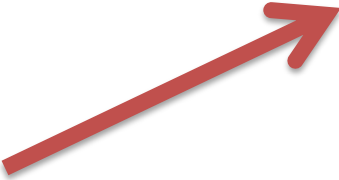
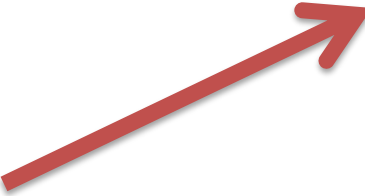
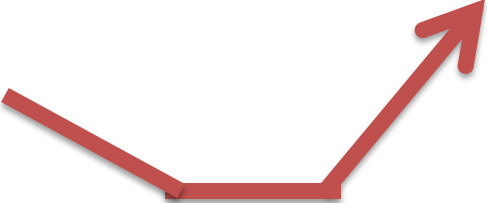
Clone Migration



Clone Migration Patterns

Evolution Trends	Evolution of Median Distance Among Clones in Clone Group	Evolution of Size of Clone Group
Constant		
Increase		
Decrease		
Wave Increase		
Wave Decrease		

Clone Migration Patterns




Patterns	Clone Migration?	Evolution of Median Distance Among Clones in Clone Group	Evolution of Size of Clone Group	
Constant	No			
High Density Strong Up	Yes			

We also define other 8 clone migration patterns

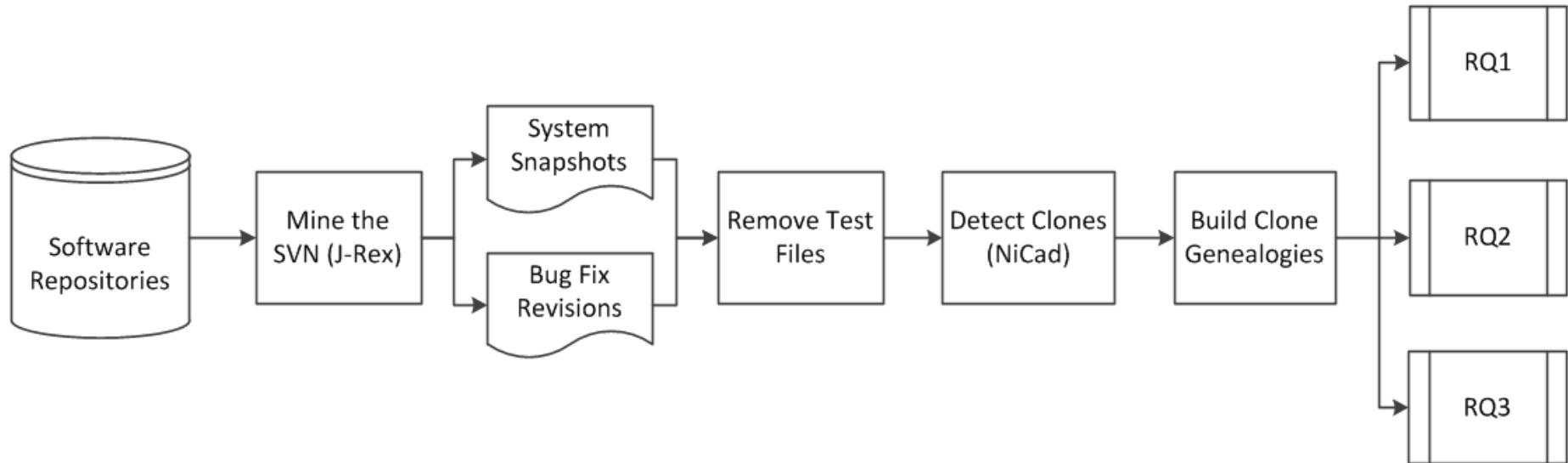
Research Questions

- **RQ1:** Do clone mutation and clone migration occur frequently in software systems?
- **RQ2:** Are some clone mutations more fault-prone than others?
- **RQ3:** Are some clone migrations more fault-prone than others?

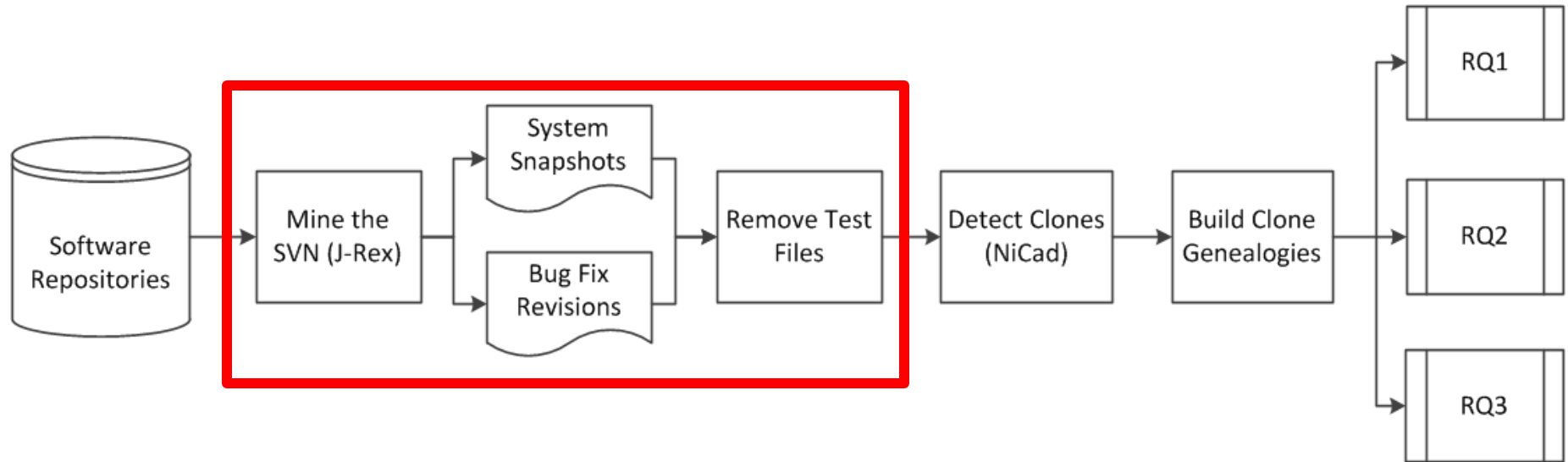
Subject Systems

Systems	# LOC	# Revisions	# Genealogies
 JBoss [®] by Red Hat	1.6M	109K	1.7K
 <APACHE ANT>	2.3M	1.0M	23K
 ArgoUML	3.1M	18k	15.6k

Approach Overview

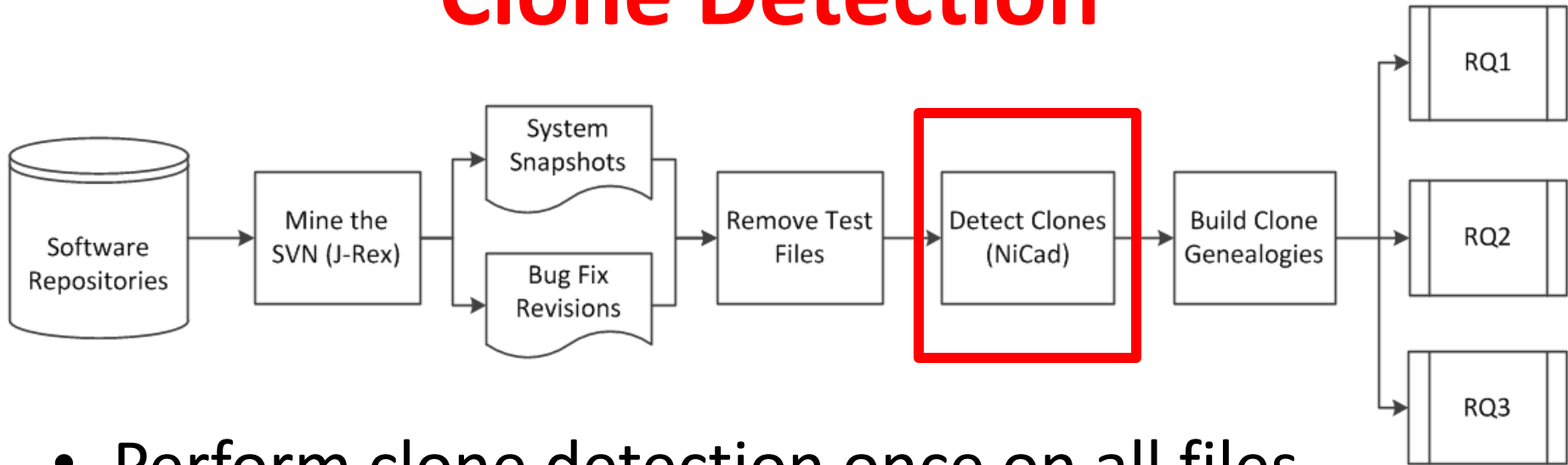


Mine the SVN



- Mine the SVN using J-Rex
 - Identify fault fixing changes
 - Snapshots for all changed files
- Remove the test files

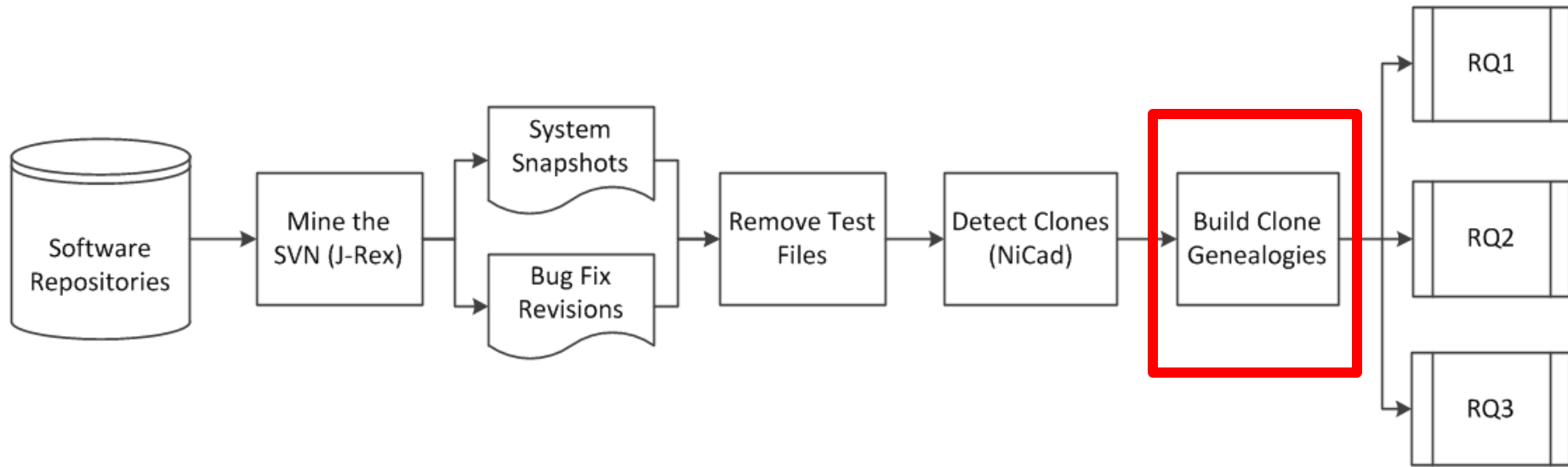
Clone Detection



- Perform clone detection once on all files
- Use NiCad to detect clone
- NiCad parameters:

Clone Type	Similarity	Blind-rename
Type-1	100%	No
Type-2	100%	Yes
Type-3	80%	Yes

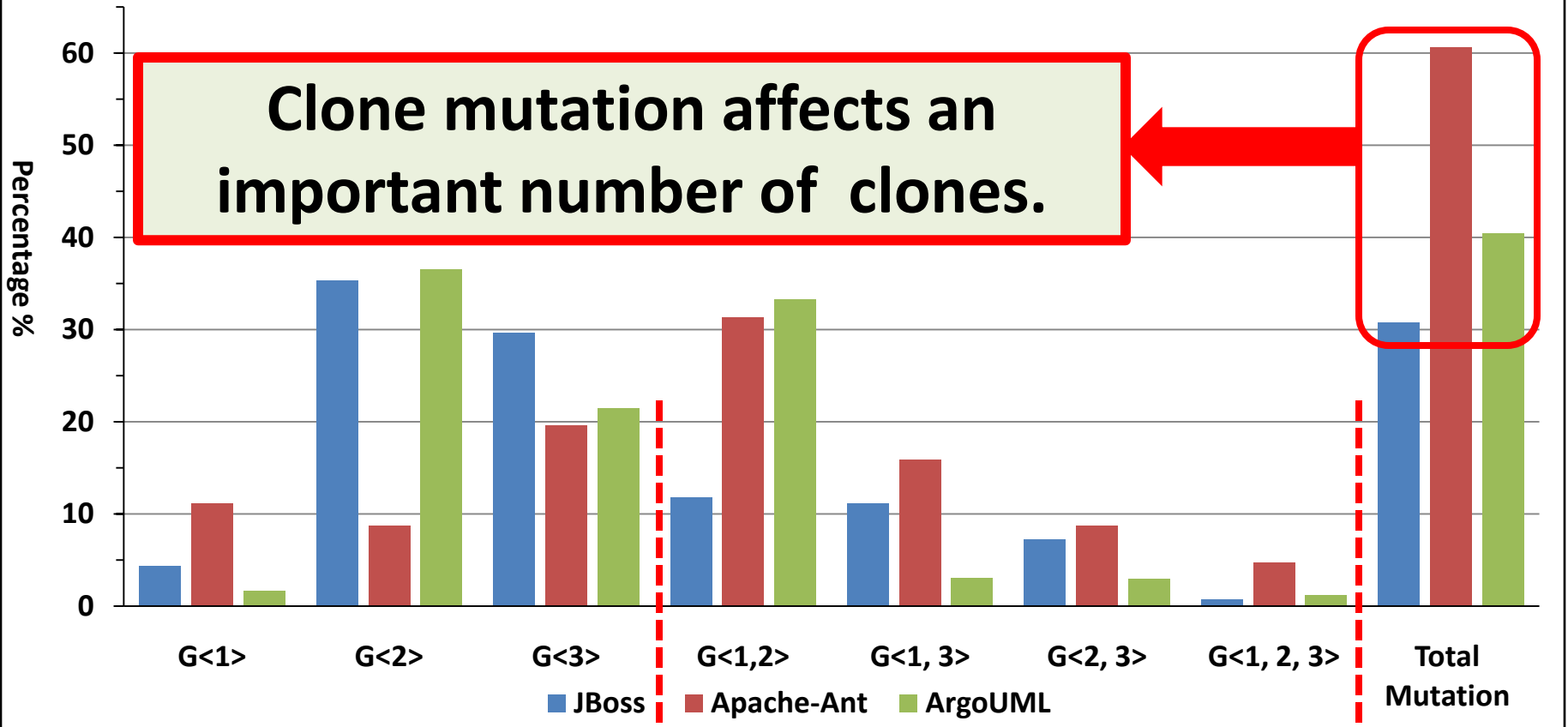
Build Clone Genealogies



- Build clone genealogies by existing clone groups
- Use *diff* to track changes to each clone
- Connect clone groups that share clones

RQ1: Do clone mutation and clone migration occur frequently in software systems?

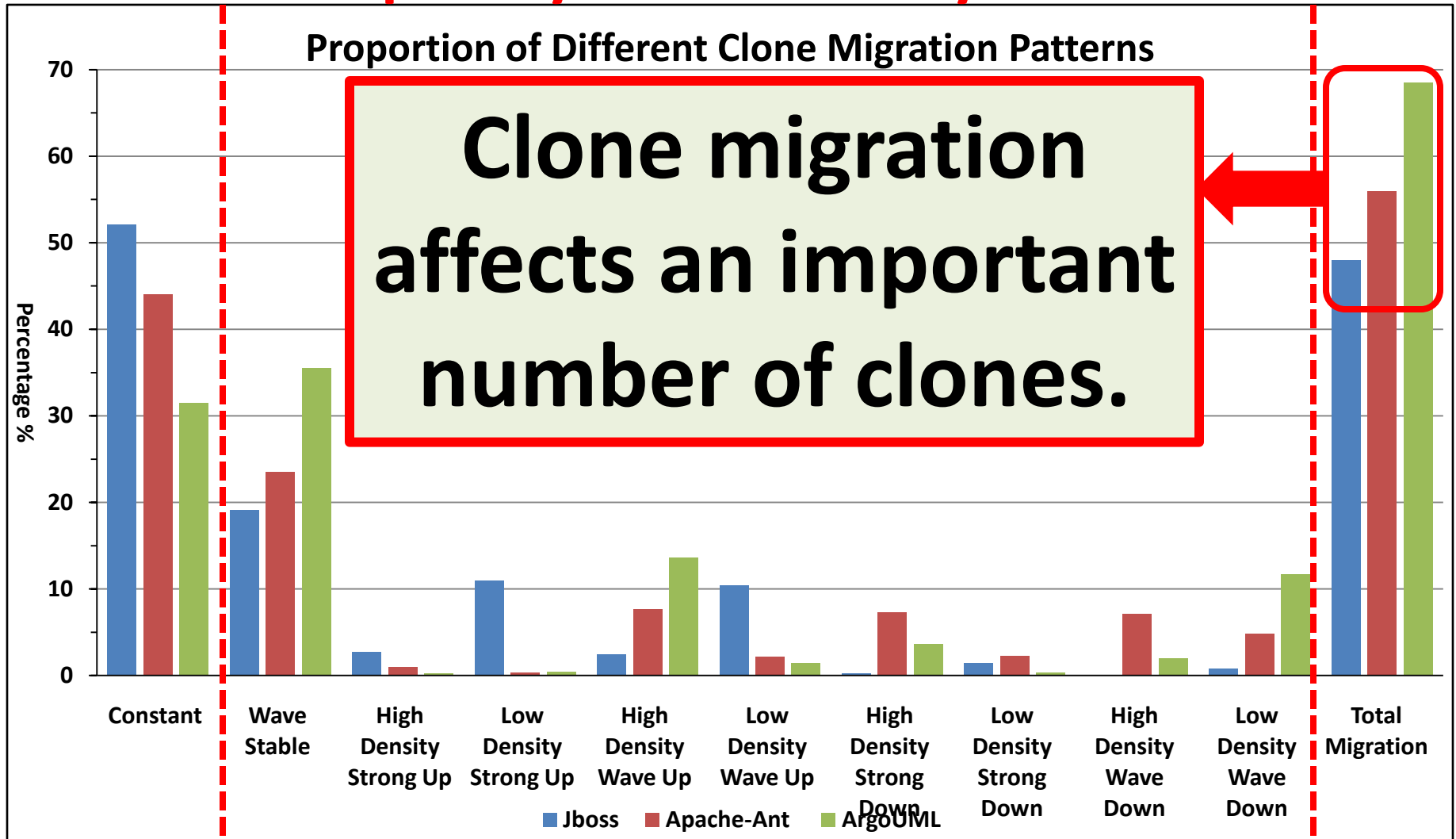
Proportion of Different Clone Mutation Categories



No Clone Mutation

Clone Mutation Categories

RQ1: Do clone mutation and clone migration occur frequently in software systems?

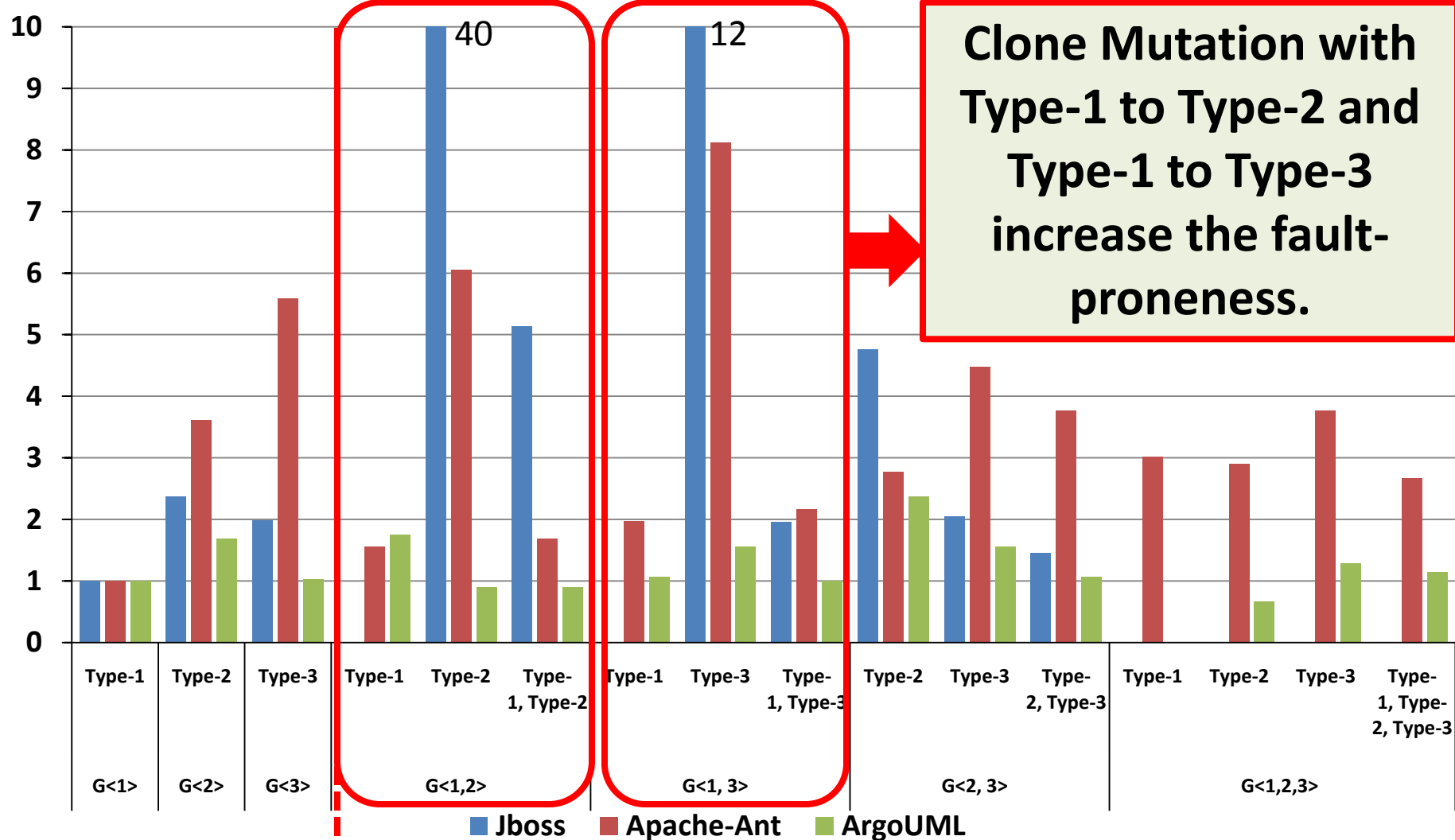


No Clone Migration

Clone Migration Patterns

RQ2: Are some clone mutations more fault-prone than others?

Odds Ratio for Different Categories of Clone Genealogies



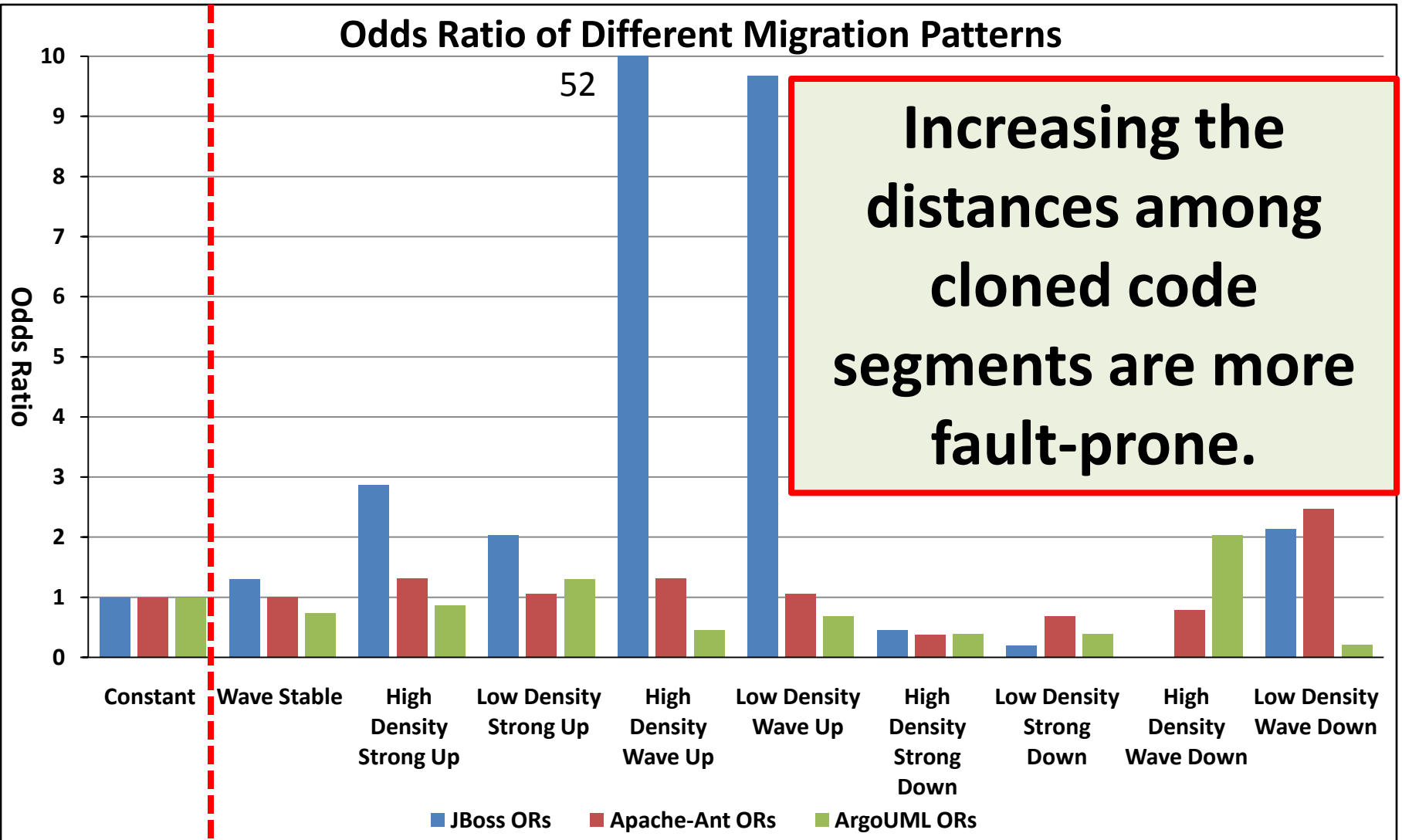
Clone Mutation with Type-1 to Type-2 and Type-1 to Type-3 increase the fault-proneness.

No Mutation

Clone Mutation Categories

RQ3: Are some clone migrations more fault-prone than others?

Odds Ratio of Different Migration Patterns



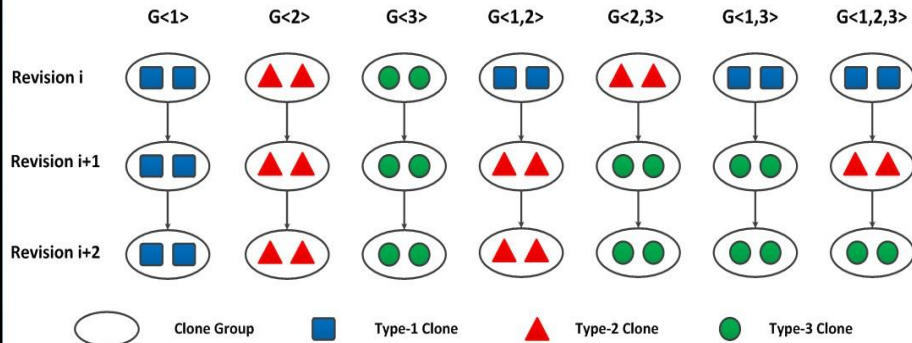
Increasing the distances among cloned code segments are more fault-prone.

No Clone Migration

Clone Migration Patterns

Clone Mutation and Clone Migration

Clone Mutation Categories



Clone Migration Patterns

Pattern	Clone Migration?	Evolution of Median Distance Among Clones in Clone Group	Evolution of Size of Clone Group
Constant	No	→	→
High Density Strong Up	Yes	↗	↗ → ↗

We also define other 8 clone migration patterns

RQ1

- **RQ1: Do clone mutation and clone migration occur frequently in software systems?**

- **Main Findings:**

Both clone mutation and clone migration affect a important number of clones in clone genealogies.

RQ2

- **RQ2: Are some clone mutations more fault-prone than others?**

- **Main Findings:**

Clone Mutation with Type-1 to Type-2 and Type-1 to Type-3 in clone genealogies increase the fault-proneness.

RQ3

- **RQ3: Are some clone migrations more fault-prone than others?**

- **Main Findings:**

Increasing the distances among cloned code segments make the clone genealogies more fault-prone.