Constructing Subtle Higher Order Mutants from Java and AspectJ Programs

Elmahdi Omar, Sudipto Ghosh and Darrell Whitley
Department of Computer Science
Colorado State University
Problems with Mutation Testing

A fault-based testing technique that help testers measure and improve the ability of test suites to detect faults

Majority of First Order Mutants (FOMs) represent trivial faults that are often easily detected [Jia and Harman 2008]

Real faults are complex
- A large majority of real faults cannot be simulated with FOMs [Purush. and Perry 2005]
- A typical real fault involves about three to four tokens [Gopinath et al 2014]

Higher Order Mutants (HOMs) can be used to simulate real and complex faults
Subtle Higher Order Mutants

HOMs that are not killed by an existing test suite that kills all the FOMs of a given program

Can help researchers and practitioners gain a better understanding of the nature of faults and their interactions
Subtle Higher Order Mutants…

Can be costly to find:

- The search space of HOMs is (exponentially?) large

- *Coupling effect* makes subtle HOMs rare

- High computational cost of evaluating mutants
  - Involves compilation and execution
Contributions

Developed search techniques for finding subtle HOMs
- Search-based software engineering techniques
- Random search technique
- Enumeration search technique

Automated the process of finding subtle HOMs
- Developed a Higher Order Mutation Testing tool for AspectJ and Java programs (HOMAJ)

Performed a set of empirical studies
- Evaluated the relative effectiveness of the developed search techniques
- Investigated different factors that impact the creation of subtle HOMs
Objective Function

Provides a metric to measure the quality of HOMs

- \( \text{fitness (HOM)} = \alpha \times \text{difficulty of killing (HOM)} + (1 - \alpha) \times \text{fault detection difference (HOM)} \)

Classifies HOMs Based on their fitness value as follow:

\[
\text{fitness (HOM)} = \begin{cases} 
0 & \rightarrow \text{Entirely Coupled HOMs} \\
0 < \alpha < 1 & \rightarrow \text{Promising HOMs} \\
1 & \rightarrow \text{Subtle HOMs (optimal solutions)}
\end{cases}
\]
Objective Function

\[ DFF = \frac{|A| + |C|}{|A| + |B| + |C|} \]

\[ DOK = \frac{|A|}{|A| + |B| + |C|} \]

\[ DFF = 20/26 \]
\[ DOK = 18/26 \]

\[ DFF = 24/24 \]
\[ DOK = 24/24 \]
Objective Function

Note that every HOM not killed by the test set is a globally optimal solution.

So we are looking for all (or many) globally optimal solutions.

This is different than many other types of objective functions.
Genetic Algorithm

Create & evaluate first HOM population from FOMs

Create next HOM population

Evaluate HOMs in population

Store subtle HOMS from population
Local Search

Generate a new \textit{SOM} \( h \) \n\hspace{1cm} \rightarrow \hspace{1cm} \text{Evaluate } h \text{ and its neighbors} \rightarrow \hspace{1cm} \text{Store any subtle HOM found} \rightarrow \hspace{1cm} \text{Does a better neighbor } h' \text{ exist?} \hspace{1cm} \begin{cases} \text{Yes} & h := h' \\ \text{No} & \end{cases}
Local Search

Case #a

$$h_{j}^{n+1} = \{ f_1, f_2, f_3, f_4 \}$$

Case #b

$$h_{j}^{n-1} = \{ f_1, f_2 \}$$

Case #c

$$h_{i}^{n} = \{ f_1', f_2', f_3 \}$$

$$h_{j}^{n} = \{ f_1', f_2', f_4 \}$$
Data-Interaction Guided Local Search

Explores only neighboring HOMs that their mutated statements access the same variable(s)

Example:

Mutation (fom\textsubscript{1})= \{ return \texttt{movieType}; \Rightarrow return \texttt{movieType}++; \}\}
Mutation (fom\textsubscript{2})= \{ if (\texttt{movieType} == “C” \Rightarrow if (\texttt{movieType} ! = “C”) \}\}
Mutation (fom\textsubscript{3})= \{ if (\texttt{custName.equals}(name) \Rightarrow if (!\texttt{custName.equals}(name)) \}\}

Considered HOMs = \{(fom\textsubscript{1}, fom\textsubscript{2}), (fom\textsubscript{1}, fom\textsubscript{2}, fom\textsubscript{3})\}

Discarded HOMs = \{(fom\textsubscript{1}, fom\textsubscript{3}), (fom\textsubscript{2}, fom\textsubscript{3})\}
Test-Case Guided Local Search

Explores only neighboring HOMs that their constituent FOMs are killed by similar/common test cases

Example:

\[
\text{KilledBy}(\text{fom}_1) = \{ \text{tc}_1, \text{tc}_3, \text{tc}_{13}, \ldots \}
\]

\[
\text{KilledBy}(\text{fom}_2) = \{ \text{tc}_1, \text{tc}_5, \text{tc}_{11}, \ldots \}
\]

\[
\text{KilledBy}(\text{fom}_3) = \{ \text{tc}_5, \text{tc}_{11}, \ldots \}
\]

Considered HOMs= \{ (\text{fom}_1, \text{fom}_2), (\text{fom}_2,\text{fom}_3), (\text{fom}_1, \text{fom}_2,\text{fom}_3) \}

Discarded HOMs= \{ (\text{fom}_1,\text{fom}_3) \}
Restricted Enumeration Search

Set the starting degree \( d \) at 2

Generate the next HOM \( h \) of degree \( d \)

Evaluate \( h \)

Keep \( h \) if subtle

If all HOMs of degree \( d \) have been evaluated \( d^+ \)
Restricted Random Search

Set the max degree $d$ of HOMs

Generate random HOM $h$ of degree $\leq d$

Evaluate $h$

Keep $h$ if subtle
Experimental Setup

Used 5 AspectJ and 5 Java programs of different sizes

Generated random test cases for each program that achieved statement coverage and killed all non-equivalent FOMs

Experiment steps:

- Ran each search technique 30 times per subject program
- The termination condition for each run was the exploration of 50,000 distinct HOMs
- Calculated the number of distinct, subtle HOMs that were found by each run
Experimental Setup

<table>
<thead>
<tr>
<th>Subject program</th>
<th>Type</th>
<th>LOC</th>
<th># of FOMs</th>
<th># of classes</th>
<th># of aspects</th>
<th># of advices</th>
<th># of pointcuts</th>
<th># of ITDs</th>
<th># of test cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinate</td>
<td>Java</td>
<td>121</td>
<td>242</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Roman Numbers</td>
<td>Java</td>
<td>179</td>
<td>208</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Cruise Control</td>
<td>Java</td>
<td>917</td>
<td>129</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Elevator</td>
<td>Java</td>
<td>1046</td>
<td>249</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>XStream</td>
<td>Java</td>
<td>14,388</td>
<td>1216</td>
<td>318</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
</tr>
<tr>
<td>Kettle</td>
<td>AspectJ</td>
<td>125</td>
<td>125</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Movie Rental</td>
<td>AspectJ</td>
<td>191</td>
<td>316</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Banking</td>
<td>AspectJ</td>
<td>243</td>
<td>92</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Telecom</td>
<td>AspectJ</td>
<td>928</td>
<td>152</td>
<td>10</td>
<td>3</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Cruise Control</td>
<td>AspectJ</td>
<td>1008</td>
<td>215</td>
<td>9</td>
<td>3</td>
<td>18</td>
<td>19</td>
<td>15</td>
<td>26</td>
</tr>
</tbody>
</table>
Measuring the Relative Effectiveness of the Search Techniques

**RQ1:** What is the relative effectiveness of the search technique in terms of their ability to find subtle HOMs?

- Effectiveness is measured in terms of the average number of subtle HOMs that can be found
- Restricted Random Search was used as a base line measure for the other five techniques
## Average Number of Subtle HOMs

<table>
<thead>
<tr>
<th>Program</th>
<th>Genetic</th>
<th>Local</th>
<th>Data Inter. Guided</th>
<th>Test Case Guided</th>
<th>Restricted Enumeration</th>
<th>Restricted Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise Java</td>
<td>76.1</td>
<td>77.8</td>
<td>80.7</td>
<td>77.1</td>
<td>34.8</td>
<td><strong>25.8</strong></td>
</tr>
<tr>
<td>Banking</td>
<td>30.9</td>
<td>29.2</td>
<td>30.8</td>
<td>28.9</td>
<td>27.1</td>
<td><strong>23.3</strong></td>
</tr>
<tr>
<td>Cruise AspectJ</td>
<td>20.3</td>
<td>29</td>
<td>39.9</td>
<td>33.2</td>
<td>22.1</td>
<td><strong>7</strong></td>
</tr>
<tr>
<td>Movie Rental</td>
<td>39.1</td>
<td>59.8</td>
<td>93.3</td>
<td>15.3</td>
<td>22</td>
<td><strong>4.7</strong></td>
</tr>
<tr>
<td>Kettle</td>
<td>35.3</td>
<td>55.1</td>
<td>56.1</td>
<td>57.7</td>
<td>31.5</td>
<td><strong>19.7</strong></td>
</tr>
<tr>
<td>Coordinate</td>
<td>72.4</td>
<td>200.9</td>
<td>213.8</td>
<td>223.3</td>
<td>84.4</td>
<td><strong>27.5</strong></td>
</tr>
<tr>
<td>Elevator</td>
<td>13.7</td>
<td>26</td>
<td>24.1</td>
<td>20.6</td>
<td>19</td>
<td><strong>5.7</strong></td>
</tr>
<tr>
<td>Telecom</td>
<td>10.3</td>
<td>20.5</td>
<td>19</td>
<td>19.9</td>
<td>6.8</td>
<td><strong>4</strong></td>
</tr>
<tr>
<td>XStream</td>
<td>0.4</td>
<td>20</td>
<td>11.4</td>
<td>12</td>
<td>13.4</td>
<td><strong>0.3</strong></td>
</tr>
<tr>
<td>Roman</td>
<td>28.6</td>
<td>30.4</td>
<td>35.4</td>
<td>37.9</td>
<td><strong>41</strong></td>
<td><strong>16.7</strong></td>
</tr>
</tbody>
</table>
Cost of Killing Subtle HOMs

<table>
<thead>
<tr>
<th>Subject program</th>
<th># of test cases that killed all FOMs</th>
<th># of test cases that were generated to kill subtle HOMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinate</td>
<td>14</td>
<td>1290</td>
</tr>
<tr>
<td>Roman</td>
<td>11</td>
<td>876</td>
</tr>
<tr>
<td>Cruise (Java)</td>
<td>18</td>
<td>818</td>
</tr>
<tr>
<td>Elevator</td>
<td>14</td>
<td>1017</td>
</tr>
<tr>
<td>XStream</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>Kettle</td>
<td>12</td>
<td>912</td>
</tr>
<tr>
<td>Movie Rental</td>
<td>15</td>
<td>1015</td>
</tr>
<tr>
<td>Banking</td>
<td>9</td>
<td>1012</td>
</tr>
<tr>
<td>Telecom</td>
<td>10</td>
<td>1115</td>
</tr>
<tr>
<td>Cruise (AspectJ)</td>
<td>26</td>
<td>908</td>
</tr>
</tbody>
</table>
Measuring the Relative Effectiveness of the Search Techniques

RQ2: How does the relative effectiveness of the search techniques compare over time?

- Investigated the growth in the average number of distinct, subtle HOMs
- The number of explored, distinct HOMs is considered a quasi-representation of the time
Growth in the Average Number of Subtle HOMs Over Time

Data from the Kettle program
Growth in the Average Number of Subtle HOMs Over Time

Line Chart 7.17: Growth in the average number of subtle HOMs that were found over the number of explored HOMs for Coordinate
Comparing Sets of Subtle HOMs Found by Different Search Techniques

RQ1: What set of subtle HOMs is found by all techniques and what set of subtle HOMs is uniquely found by each technique?

Subtle HOMs were classified into:

- **Easiest-to-find subtle HOM**: can be found by all the search techniques
- **Hardest-to-find subtle HOM**: can be uniquely found by only one search technique
Easiest-to-find and Hardest-to-find Subtle HOM

Data from the Kettle program
RQ1: What is the computational cost of finding subtle HOMs using the search techniques?

- The cost is measured in terms of the time taken to find subtle HOMs

Answer:

- On average, exploring and evaluating 50,000 HOMs requires around 19 hours
- The compilation and execution process of HOMs represented 98% of the computational cost of finding subtle HOMs
- Optimizing the compilation process of HOMs reduced the computational cost of finding subtle HOMs by 32%
Composition and Decomposition
Relationships between Subtle HOMs

RQ1: Can subtle HOMs be composed to create new subtle HOMs of higher degrees?

- Investigated composing subtle HOMs that were found by the Restricted Enumeration Search to create new subtle HOMs of higher degrees
Composing HOMs: Variable Interaction

The Variable Interaction Graph
Composing Subtle HOMs

- Found by the search techniques
- Found by composing subtle HOMs

Number of subtle HOMs

- Elevator
- Cruise
- Roman
- Xstream
- Telecom
- Banking
- Kettle
- Cruise (AspectJ)
- Movie Rental
## Average Number of Subtle HOMs

<table>
<thead>
<tr>
<th>Program</th>
<th>Genetic</th>
<th>Local</th>
<th>Data Inter. Guided</th>
<th>Test Case Guided</th>
<th>Restricted Enumeration</th>
<th>Restricted Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise Java</td>
<td>76.1</td>
<td>77.8</td>
<td>80.7</td>
<td>77.1</td>
<td>34.8</td>
<td>25.8</td>
</tr>
<tr>
<td>Banking</td>
<td>30.9</td>
<td>29.2</td>
<td>30.8</td>
<td>28.9</td>
<td>27.1</td>
<td>23.3</td>
</tr>
<tr>
<td>Cruise AspectJ</td>
<td>20.3</td>
<td>29</td>
<td>39.9</td>
<td>33.2</td>
<td>22.1</td>
<td>7</td>
</tr>
<tr>
<td>Movie Rental</td>
<td>39.1</td>
<td>59.8</td>
<td>93.3</td>
<td>15.3</td>
<td>22</td>
<td>4.7</td>
</tr>
<tr>
<td>Kettle</td>
<td>35.3</td>
<td>55.1</td>
<td>56.1</td>
<td>57.7</td>
<td>31.5</td>
<td>19.7</td>
</tr>
<tr>
<td>Coordinate</td>
<td>72.4</td>
<td>200.9</td>
<td>213.8</td>
<td>223.3</td>
<td>84.4</td>
<td>27.5</td>
</tr>
<tr>
<td>Elevator</td>
<td>13.7</td>
<td>26</td>
<td>24.1</td>
<td>20.6</td>
<td>19</td>
<td>5.7</td>
</tr>
<tr>
<td>Telecom</td>
<td>10.3</td>
<td>20.5</td>
<td>19</td>
<td>19.9</td>
<td>6.8</td>
<td>4</td>
</tr>
<tr>
<td>XStream</td>
<td>0.4</td>
<td>20</td>
<td>11.4</td>
<td>12</td>
<td>13.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Roman</td>
<td>28.6</td>
<td>30.4</td>
<td>35.4</td>
<td>37.9</td>
<td>41</td>
<td>16.7</td>
</tr>
</tbody>
</table>
Composing HOMs: Variable Interaction

The Variable Interaction Graph
Composing HOMs: Variable Interaction

Table 11.2: Comparing the number of subtle HOMs that were found by the search techniques and by composing subtle HOMs that were found by Restricted Enumeration Search

<table>
<thead>
<tr>
<th>Program</th>
<th># of all subtle HOMs that were found by the search techniques</th>
<th># of all subtle HOMs that were found by composing subtle HOMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevator</td>
<td>31</td>
<td>962</td>
</tr>
<tr>
<td>Cruise (Java)</td>
<td>355</td>
<td>3464</td>
</tr>
<tr>
<td>Roman</td>
<td>105</td>
<td>6717</td>
</tr>
<tr>
<td>XStream</td>
<td>216</td>
<td>291</td>
</tr>
<tr>
<td>Telecom</td>
<td>30</td>
<td>44</td>
</tr>
<tr>
<td>Banking</td>
<td>79</td>
<td>650</td>
</tr>
<tr>
<td>Kettle</td>
<td>150</td>
<td>724</td>
</tr>
<tr>
<td>Cruise (AspectJ)</td>
<td>227</td>
<td>558</td>
</tr>
<tr>
<td>Movie Rental</td>
<td>283</td>
<td>764</td>
</tr>
</tbody>
</table>
**Composition and Decomposition Relationships between Subtle HOMs**

**RQ2:** To what extent do subtle HOMs of higher degrees represent a composition of subtle HOMs of lower degrees?

- Investigated the number of subtle HOMs that were found by each search technique with respect to their decomposition type
  - Fully decomposable into other subtle HOMs
  - Partially decomposable into other subtle HOMs
  - Not decomposable into other subtle HOMs
Decomposing Subtle HOMs

- **Fully Decomposable**
- **Partially Decomposable**
- **Not Decomposable**

Number of subtle HOMs

<table>
<thead>
<tr>
<th>Number</th>
<th>Restricted Enumeration</th>
<th>Local</th>
<th>Data-Interaction Guided</th>
<th>Test-Case Guided</th>
<th>Restricted Random</th>
<th>Genetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>5+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Gold: Fully Decomposable
- Green: Partially Decomposable
- Orange: Not Decomposable
Conclusions

The search-based software engineering techniques can produce a large number of distinct, subtle HOMs.

Local Search and both the Guided Local Search techniques were more effective than the other techniques in terms of their ability to find subtle HOMs.

Combining FOMs that are closer to each other in terms of their location is more likely to create subtle HOMs.
Subtle HOMs of higher degrees are likely to exist as compositions of multiple subtle HOMs of lower degrees.

Subtle HOMs of higher degrees can be effectively found by composing subtle HOMs of lower degrees.

The search-based software engineering techniques were able to find subtle HOMs of higher degrees that could not be found by composing subtle HOMs of lower degrees.
Publications

1. Higher Order Mutation Testing Tool For Java and AspectJ Programs, ICST, proceedings of the 7th IEEE International Conference on Software Testing, Verification and Validation, Mutation, 2014


