Efficient Mutation Analysis by Propagating and Partitioning Infected Execution States

René Just\(^1\) & Michael D. Ernst\(^1\) & Gordon Fraser\(^2\)

\(^1\)University of Washington, USA
\(^2\)University of Sheffield, UK

July 25, 2014
A rough outline

Motivation

What is mutation analysis — is it useful?
A rough outline

**Motivation**
What is mutation analysis — is it useful?

**Problem**
Mutation analysis is expensive!
A rough outline

**Motivation**
What is mutation analysis — is it useful?

**Problem**
Mutation analysis is expensive!

**Solution**
Dynamic prepass analysis to make mutation practical!
Test suite quality

Why assess test suite quality?

- Selection: Given two test suites, which is better?
- Minimization: Are there redundant tests in a test suite?
- Prioritization: Which tests of a test suite should run first?
Test suite quality

Why assess test suite quality?

▶ Selection: Given two test suites, which is better?
▶ Minimization: Are there redundant tests in a test suite?
▶ Prioritization: Which tests of a test suite should run first?

How to assess test suite quality?

▶ A good test suite detects real faults
Test suite quality

Why assess test suite quality?

▶ Selection: Given two test suites, which is better?
▶ Minimization: Are there redundant tests in a test suite?
▶ Prioritization: Which tests of a test suite should run first?

How to assess test suite quality?

▶ A **good** test suite detects **real faults**
▶ **Problem**: Real faults in a program are unknown
▶ **Solution**: Seed **artificial faults** into the program
Test suite quality

Why assess test suite quality?

- Selection: Given two test suites, which is better?
- Minimization: Are there redundant tests in a test suite?
- Prioritization: Which tests of a test suite should run first?

How to assess test suite quality?

- A **good** test suite detects **real faults**
- **Problem:** Real faults in a program are unknown
- **Solution:** Seed **artificial faults** into the program

**Mutation analysis:** systemically seed artificial faults
Mutation analysis overview

Program

Test suite
Mutation analysis overview
Mutation analysis overview

Each mutant contains a small syntactic change

Program

Generate mutants

Mutants

Program

Generate mutants

Mutants

public int max(int a, int b){
    return (a > b) ? a : b;
}

Original

public int max(int a, int b){
    return (a >= b) ? a : b;
}

Mutant 1

public int max(int a, int b){
    return (a != b) ? a : b;
}

Mutant 2

public int max(int a, int b){
    return (a >= b) ? a : b;
}

Mutant 1

public int max(int a, int b){
    return (a != b) ? a : b;
}

Mutant 2
Mutation analysis overview

Program

Generate mutants

Mutants

Test suite

Execute test suite

Mutation score

Ratio of detected mutants
Where is the catch?

Many mutants can be generated!

\[
a > b \ ? \ a : b
\]
Where is the catch?

Many mutants can be generated!

\[
\begin{align*}
\text{a > b} & \quad ? \quad \text{a} : \quad \text{b} \\
\text{a >= b} & \quad ? \quad \text{a} : \quad \text{b} \\
\text{a < b} & \quad ? \quad \text{a} : \quad \text{b} \\
\text{a <= b} & \quad ? \quad \text{a} : \quad \text{b} \\
\text{a != b} & \quad ? \quad \text{a} : \quad \text{b} \\
\text{a == b} & \quad ? \quad \text{a} : \quad \text{b} \\
!(\text{a > b}) & \quad ? \quad \text{a} : \quad \text{b} \\
\text{true} & \quad ? \quad \text{a} : \quad \text{b} \\
\text{false} & \quad ? \quad \text{a} : \quad \text{b}
\end{align*}
\]
Where is the catch?

Many mutants can be generated!

```
a > b ? a : b
```

```
a >= b ? a : b
0 > b ? a : b
-a > b ? a : b
a < b ? a : b
a > 0 ? a : b
a >-b ? a : b
a <= b ? a : b
a > b ? 0 : b
a > b ? -a : b
a != b ? a : b
a > b ? a : 0
a > b ? a :-b
a == b ? a : b
0
a > b ? a :-b
!(a > b) ? a : b
b > a ? a : b
-(a > b ? a : b)
true ? a : b
a > b ? b : a
a
false ? a : b
b
```
Mutation analysis overview

**Program**

- Generate mutants

**Test suite**

- Execute test suite

**Problem**
Test suite has to be executed for many mutants

**Mutation score**
Mutation analysis overview

Program

Generate mutants

Mutants

Filter mutants

Mutants

Execute test suite

Test suite

Mutation score

Problem
Test suite has to be executed for many mutants
Mutation analysis overview

Program → Generate mutants → Mutants

Test suite → Execute test suite → Mutation score

Related work: Mutant sampling (Jia and Harman, TSE’11)

Problem
Test suite has to be executed for many mutants
Mutation analysis overview

- **Program**
- **Test suite**

**Generate mutants**
- **Related work:** Selective mutation *(Namin et al., ICSE’08)*

**Filter mutants**
- **Problem**
  - Test suite has to be executed for many mutants

**Execute test suite**
- **Mutation score**
Mutation analysis overview

Program → Generate mutants → Mutants

Test suite

Related work: Code coverage analysis (Major, Javalanche, PIT, ...)

Mutants → Filter mutants → Mutants

Execute test suite → Mutation score

Problem
Test suite has to be executed for many mutants
Mutation analysis overview

Program

Generate mutants

Test suite

Mutants

Generate mutants

Filter mutants

Mutants

Execute test suite

Mutation score

Related work:
Test suite prioritization
(Just et al., ISSRE’12
Zhang et al., ISSTA’13)

Problem
Test suite has to be executed for many mutants
Mutation analysis overview

Program

Generate mutants

Mutants

Filter mutants

Our solution
Filter mutants with a dynamic analysis

Test suite

Execute test suite

Mutation score

Problem
Test suite has to be executed for many mutants

René Just, UW CSE Efficient Mutation Analysis by Propagating and Partitioning Infected Execution States
Wasted effort in mutation analysis

Example: testing triangle classification

```java
public TriangleType classify
    (int a, int b, int c) {
    ...
    if ( a + b <= c ) {
        return Invalid;
    }
    ...
} 
```

Original
Wasted effort in mutation analysis

Example: testing triangle classification

```java
public TriangleType classify (int a, int b, int c) {
    ...
    if (a + b <= c) {
        return Invalid;
    }
    ...
}
```

Original
Wasted effort in mutation analysis

Example: testing triangle classification

```
public TriangleType classify (int a, int b, int c) {
    ...
    if ( a + b <= c ) {
        return Invalid;
    }
    ...
}
```

```
public TriangleType classify (int a, int b, int c) {
    ...
    if ( a * b <= c ) {
        return Invalid;
    }
    ...
}
```

Mutant 1

Original
Wasted effort in mutation analysis

Example: testing triangle classification

Original

```java
public TriangleType classify(int a, int b, int c) {
    ...
    if (a + b <= c) {
        return Invalid;
    }
    ...
}
```

Mutant 1

```java
public TriangleType classify(int a, int b, int c) {
    ...
    if (a * b <= c) {
        return Invalid;
    }
    ...
}
```

Identical code surrounding mutation
Wasted effort in mutation analysis

Example: testing triangle classification

public TriangleType classify (int a, int b, int c) {
    ...
    if (a + b <= c) {
        return Invalid;
    }
    ...
}

Original

public TriangleType classify (int a, int b, int c) {
    ...
    if (a * b <= c) {
        return Invalid;
    }
    ...
}

Mutant 1

Optimizations:

- Infection
Wasted effort in mutation analysis

Example: testing triangle classification

```java
public TriangleType classify(int a, int b, int c) {
    ... if (a + b <= c) {
        return Invalid;
    } ... }
```

Original

```java
public TriangleType classify(int a, int b, int c) {
    ... if (a * b <= c) {
        return Invalid;
    } ... }
```

Mutant 1

Optimizations:

- Infection
- Propagation
Wasted effort in mutation analysis

Example: testing triangle classification

```java
public TriangleType classify (int a, int b, int c) {
    ...
    if (a + b <= c) {
        return Invalid;
    }
    ...
}
```

Original

```java
public TriangleType classify (int a, int b, int c) {
    ...
    if (a * b <= c) {
        return Invalid;
    }
    ...
}
```

Mutant 1

```java
public TriangleType classify (int a, int b, int c) {
    ...
    if (a - b <= c) {
        return Invalid;
    }
    ...
}
```

Mutant 2

Optimizations:
- Infection
- Propagation
Wasted effort in mutation analysis

Example: testing triangle classification

```java
public TriangleType classify(int a, int b, int c) {
    ...
    if (a + b <= c) {
        return Invalid;
    }
    ...
}
```

```java
public TriangleType classify(int a, int b, int c) {
    ...
    if (a * b <= c) {
        return Invalid;
    }
    ...
}
```

```java
public TriangleType classify(int a, int b, int c) {
    ...
    if (a - b <= c) {
        return Invalid;
    }
    ...
}
```

Optimizations:

- Infection
- Propagation
- Partitioning
The big picture

Set of mutants

Execute test suite on 100% of mutants
The big picture

Set of mutants

Mutation score: 56%

Undetectable mutants

Detectable mutants

Execute test suite on 100% of mutants
The big picture

Set of mutants

Infection

Execute test suite on 90% of mutants
The big picture

Set of mutants

Infection

Propagation

Execute test suite on 84% of mutants
The big picture

Infection

Propagation

Partitioning

Set of mutants

Execute test suite on 70% of mutants

René Just, UW CSE  
Efficient Mutation Analysis by Propagating and Partitioning Infected Execution States
Dynamic analysis to filter mutants

Program

Generate mutants

Mutants

Test suite

Execute test suite

only once on an instrumented program version

Infection

Propagation

Partitioning

Execute test suite

Mutation score
Infection

A test infects the execution state of a mutant if the expression values of the mutation and the original version differ.
Infection

A test infects the execution state of a mutant if the expression values of the mutation and the original version differ.

**Example:** \( a=2, \ b=2, \ c=0 \)

\[
\text{if ( } a + b > c \text{ )}
\]
Infection

A test infects the execution state of a mutant if the expression values of the mutation and the original version differ.

Example: $a=2$, $b=2$, $c=0$

```latex
\begin{align*}
\text{if } (a + b > c) \\
\text{a * b} & \quad \text{a / b} & \quad \text{a % b} & \quad \text{a - b}
\end{align*}
```
Infection

A test infects the execution state of a mutant if the expression values of the mutation and the original version differ.

**Example:** $a=2$, $b=2$, $c=0$


define

\[
\text{if (} a + b > c \text{ )}
\]

- $a \times b = 4$
- $a / b = 1$
- $a \% b = 0$
- $a - b = 0$

**Optimization**

- Execute mutations and monitor infected execution states
- Filter mutants whose execution state is not infected
A test infects the execution state of a mutant if the expression values of the mutation and the original version differ.

**Example:** \( a=2, \ b=2, \ c=0 \)

\[
\text{if } (a + b > c)
\]

- \( a * b \): 4
- \( a / b \): 1
- \( a \% b \): 0
- \( a - b \): 0

**Optimization**
- Execute mutations and monitor infected execution states
- Filter mutants whose execution state is not infected
An infected execution state propagates if it leads to an infected execution state of a lexically enclosing expression.

**Example:** \( a = 2, \ b = 2, \ c = 0 \)
An infected execution state propagates if it leads to an infected execution state of a lexically enclosing expression.

**Example:** $a=2$, $b=2$, $c=0$

- $a + b > c$: true
- $a * b$: 4
- $a / b$: 1
- $a \% b$: 0
- $a - b$: 0

---

René Just, UW CSE
Efficient Mutation Analysis by Propagating and Partitioning Infected Execution States
**Propagation**

An infected execution state propagates if it leads to an infected execution state of a lexically enclosing expression.

**Example:** \( a=2, \ b=2, \ c=0 \)

\[
\begin{align*}
\text{true} & \quad \text{if} \ (a + b > c) \\
\text{true} & \quad a * b \\
\text{false} & \quad a / b > c \\
\text{false} & \quad a \% b > c \\
\text{true} & \quad a - b > c \\
\end{align*}
\]

**Optimization**

- Propagate infected execution states in composed expressions
- Filter mutants whose infected state does not propagate
An infected execution state propagates if it leads to an infected execution state of a lexically enclosing expression.

Example: $a=2$, $b=2$, $c=0$

```
true
if ( a + b > c )
```

- $a \times b = 4$
- $a \div b > c = true$
- $a \% b > c = false$
- $a - b > c = false$

Optimization

- Propagate infected execution states in composed expressions
- Filter mutants whose infected state does not propagate
Partitioning

Build partition of identically infected execution states.

**Example:** \( a=2, \ b=2, \ c=0 \)

```
true
if ( \( a + b > c \) )
```

- \( a * b \):
  - 4
  - true

- \( a / b > c \):
  - true

- \( a \% b > c \):
  - false

- \( a - b > c \):
  - false
Partitioning

Build partition of identically infected execution states.

Example: \( a=2, \ b=2, \ c=0 \)

\[
\text{true}
\]

\[
\text{if (} a + b > c \text{)}
\]

\[
\begin{align*}
& a * b \quad \quad 4 \\
& a / b > c \quad \text{true} \\
& a \% b > c \quad \text{false} \\
& a - b > c \quad \text{false}
\end{align*}
\]

Optimization

- **Partition** mutants based on their expression values
- Only execute a test for one mutant per **partition cell**
Partitioning

Build partition of identically infected execution states.

Example: \(a=2, \ b=2, \ c=0\)

- \(a + b > c\) \(\text{true}\)
- \(a \ast b = 4\)
- \(a / b > c\) \(\text{true}\)
- \(a \% b > c\) \(\text{false}\)
- \(a - b > c\) \(\text{false}\)

Optimization

- **Partition** mutants based on their expression values
- Only execute a test for one mutant per **partition cell**
Implementation details

\[ a + b > c \]
Implementation details

\[
\begin{align*}
3_e & \mapsto +, 1_e, 2_e \\
5_e & \mapsto >, 3_e, 4_e
\end{align*}
\]

Instrumentation

\[\text{eval}(5, \text{eval}(3, a, b), c)\]
Implementation details

Expressions

\[ 3_e \mapsto +, \ 1_e, \ 2_e \]
\[ 5_e \mapsto >, \ 3_e, \ 4_e \]

Mutations

\[ 1_m \mapsto 3_e, \ - \]
\[ 2_m \mapsto 3_e, \ * \]
\[ 3_m \mapsto 5_e, \ >= \]
\[ 4_m \mapsto 5_e, \ == \]

\[ 1_m: \ a - b > c \]
\[ 2_m: \ a * b > c \]
\[ 3_m: \ a + b >= c \]
\[ 4_m: \ a + b == c \]

\( a + b > c \)

eval(5, eval(3, a, b), c)
Implementation details

\[
\begin{align*}
&\quad 5_e \\
&\quad \quad 3_e \\
&\quad \quad \quad 1_e \quad 2_e \\
&\quad a + b > c \\
&\quad \quad \quad \quad \quad \text{instrumentation} \\
&\quad \quad \quad \quad \quad \text{eval(5, eval(3, a, b), c)} \\
&\quad \quad \quad \quad \quad \text{mapping} \\
&\quad \quad \quad \quad \quad \text{Expressions} \\
&\quad \quad \quad \quad 3_e \leftrightarrow +, 1_e, 2_e \\
&\quad \quad \quad \quad 5_e \leftrightarrow >, 3_e, 4_e \\
&\quad \quad \quad \quad \text{Mutations} \\
&\quad 1_m \leftrightarrow 3_e, - \\
&\quad 2_m \leftrightarrow 3_e, * \\
&\quad 3_m \leftrightarrow 5_e, >= \\
&\quad 4_m \leftrightarrow 5_e, == \\
&\quad \quad \text{Example: \(a=2, b=2, c=1\)} \\
&\quad \quad \quad \text{InfectedValues}
\end{align*}
\]
Implementation details

Expression mapping:

$\text{eval}(5, \text{eval}(3, a, b), c)$

Evaluation:

- $3_e \mapsto +, 1_e, 2_e$
- $5_e \mapsto >, 3_e, 4_e$

Example: $a=2$, $b=2$, $c=1$

Infected Values:

- $3_e: a + b = 4$
- $1_m: a - b = 0$
- $2_m: a \times b = 4$

Mutations:

- $1_m \mapsto 3_e, -$
- $2_m \mapsto 3_e, *$
- $3_m \mapsto 5_e, >=$
- $4_m \mapsto 5_e, ==$
Implementation details

\[ a + b > c \]

Expressions:
- \( 3_e \mapsto +, 1_e, 2_e \)
- \( 5_e \mapsto >, 3_e, 4_e \)

Mutations:
- \( 1_m \mapsto 3_e, - \)
- \( 2_m \mapsto 3_e, * \)
- \( 3_m \mapsto 5_e, >= \)
- \( 4_m \mapsto 5_e, == \)

Example: \( a=2, b=2, c=1 \)

InfectedValues:
- \( 3_e: a + b = 4 \)
- \( 1_m: a - b = 0 \)
- \( 1_m: 0 \)
Implementation details

Expressions

\[
\begin{align*}
3_e & \leftrightarrow +, 1_e, 2_e \\
5_e & \leftrightarrow >, 3_e, 4_e
\end{align*}
\]

Mutations

\[
\begin{align*}
1_m & \mapsto 3_e, - \\
2_m & \mapsto 3_e, * \\
3_m & \mapsto 5_e, >= \\
4_m & \mapsto 5_e, ==
\end{align*}
\]

Example: \(a=2, b=2, c=1\)

\[
\begin{align*}
5_e : 4 > c & = \text{true} \\
3_m : 4 >= c & = \text{true} \\
4_m : 4 == c & = \text{false}
\end{align*}
\]

InfectedValues

\[
\begin{align*}
1_m : 0
\end{align*}
\]
Implementation details

\[
\begin{align*}
5_e & \quad \text{mapping} \\
3_e & \\
1_e & \quad 2_e & \quad 4_e
\end{align*}
\]

Expressions

\[
\begin{align*}
3_e & \leftrightarrow +, 1_e, 2_e \\
5_e & \leftrightarrow >, 3_e, 4_e
\end{align*}
\]

Mutations

\[
\begin{align*}
1_m & \leftrightarrow 3_e, - \\
2_m & \leftrightarrow 3_e, * \\
3_m & \leftrightarrow 5_e, >= \\
4_m & \leftrightarrow 5_e, ==
\end{align*}
\]

Example: a=2, b=2, c=1

InfectedValues

\[
\begin{align*}
5_e: & \quad 4 > c = \text{true} \\
3_m: & \quad 4 >= c = \text{true} \\
4_m: & \quad 4 == c = \text{false} \\
1_m: & \quad 0 > c = \text{false} \\
1_m: & \quad 0
\end{align*}
\]
Implementation details

Expressions

<table>
<thead>
<tr>
<th>3e</th>
<th>4e</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>&gt;</td>
</tr>
</tbody>
</table>

Mutations

<table>
<thead>
<tr>
<th>1m</th>
<th>2m</th>
<th>3m</th>
<th>4m</th>
</tr>
</thead>
<tbody>
<tr>
<td>3e, -</td>
<td>3e, *</td>
<td>5e, &gt;=</td>
<td>5e, ==</td>
</tr>
</tbody>
</table>

Example: a=2, b=2, c=1

InfectedValues

<table>
<thead>
<tr>
<th>4m</th>
<th>1m</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
</tr>
</tbody>
</table>
Implementation details

\[
\begin{align*}
1_e & \rightarrow 3_e, - \\
2_e & \rightarrow 3_e, * \\
3_e & \rightarrow +, 1_e, 2_e \\
4_e & \rightarrow >, 3_e, 4_e \\
5_e & \rightarrow \geq, 3_e, 4_e
\end{align*}
\]

Example: \( a=2, b=2, c=1 \)

\[
\begin{align*}
4_m &: \quad \text{false} \\
1_m &: \quad \text{false}
\end{align*}
\]
Implementation details

\[
\begin{align*}
3_e & \mapsto +, 1_e, 2_e \\
5_e & \mapsto >, 3_e, 4_e
\end{align*}
\]

Examples:
- \(1_m \mapsto 3_e, -\)
- \(2_m \mapsto 3_e, *\)
- \(3_m \mapsto 5_e, >=\)
- \(4_m \mapsto 5_e, ==\)

Example: \(a=2, b=2, c=1\)

InfectedValues
- \(4_m: \text{false}\)
- \(1_m: \text{false}\)

Implemented in Major
- Compact instrumentation
- Soundly handles side effects and short-circuit operators
Experimental setup

14 subject programs

- Open-source programs from different application domains
- 670,000 lines of code
- 540,000 generated mutants
Experimental setup

14 subject programs

▶ Open-source programs from different application domains
▶ 670,000 lines of code
▶ 540,000 generated mutants

4 test suites for each program

▶ 1 developer-written test suite (released with program)
▶ 3 generated test suites (EvoSuite)
  ▶ Weak-mutation
  ▶ Branch coverage
  ▶ Random
Experimental setup

14 subject programs
▶ Open-source programs from different application domains
▶ 670,000 lines of code
▶ 540,000 generated mutants

4 test suites for each program
▶ 1 developer-written test suite (released with program)
▶ 3 generated test suites (EvoSuite)
  ▶ Weak-mutation
  ▶ Branch coverage
  ▶ Random

Coverage optimization is baseline
**Ratio of analyzed mutants**

The graph shows the ratio of analyzed mutants for different combinations of techniques:

- **Infect**: 80% of mutants need to be analyzed.
- **Infect+Prop**: Similar ratio as Infect.
- **Infect+Part**: Lower ratio for weaker test suites (e.g., Random).
- **Infect+Prop+Part**: The most effective after Propagation.

Each bar represents a different technique, with manual, weak-mutant, branch, and random test suites indicated by different colors.
Ratio of analyzed mutants

Findings

- Only 70% of covered mutants need to be analyzed
- Similar ratio for Propagation and Partitioning
- Partitioning is (most) effective after Propagation
- Lower ratio for weaker test suites (e.g., Random)
Ratio of total runtime

Findings

- Total run time reduced by 40%
- Filtering costs are almost negligible
- Partitioning is (most) effective after Propagation
- Run-time improvements similar for all test suites
**Ratio of total runtime**

Findings

- Total run time reduced by 40%
- Filtering costs are almost negligible
- Partitioning is (most) effective after Propagation
- Run-time improvements similar for all test suites
Future work

Equivalent mutant detection

- Can propagation predict equivalent mutants?
- Solve constraints necessary to achieve propagation
Future work

Equivalent mutant detection

- Can propagation predict equivalent mutants?
- Solve constraints necessary to achieve propagation

Test generation

- Generate tests that achieve propagation
- Improve mutation-driven test generation
  (Zhang et al., ICSM’10, Fraser and Zeller, TSE’12)
Contributions

Dynamic prepass analysis

- Three new optimizations that significantly improve efficiency
- Filter mutants with single test execution on instrumented program

Empirical evaluation

- 14 programs and 540,000 mutants
- Total run time reduced by 40%
- Propagation and Partitioning should be combined

http://www.mutation-testing.org