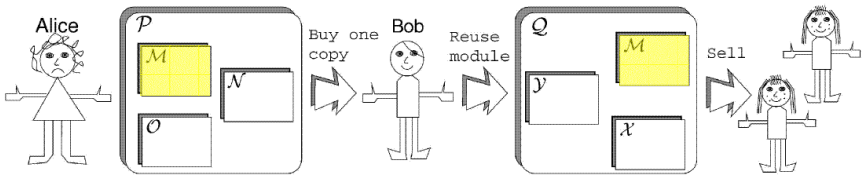
Preliminary Analysis of the Reverse Engineering Complexity

Ceccato Mariano
Fondazione Bruno Kessler-IRST, Trento, Italy

FONDAZIONE BRUNO KESSLER

Malicious reverse engineering

- Valuable piece of code is extracted from an application and incorporated into competitor's code.
- Software tampering



Obfuscation

Transforming a program into an equivalent one

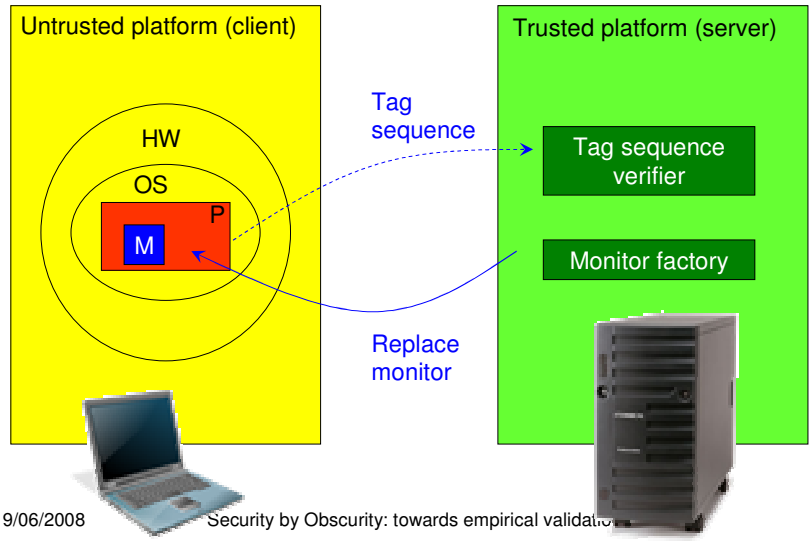
- Harder to reverse engineer
- Maintaining its semantics

```
Student guy = new Student();
String name = "Mathematics";
Course course = new Course(name);
guy.apply(course);
course.run();
name.match("jas");
```

T^1 T^2 T^3

```
y1 x1 = new y1();
String x2 = "Mathematics";
y2 x3 = new y2(x2);
x1.z1(x3);
x3.run();
x2.match("jas");
```


Reference architecture



FONDAZIONE BRUNO KESSLER

Research questions

RQ1: To what extent the obfuscation reduces the capability of subjects to comprehend decompiled source code?

RQ2: To what extent the obfuscation increases the time needed to perform a comprehension task?

RQ3: To what extent the obfuscation reduces the capability of subjects to perform an attack?

RQ4: To what extent the obfuscation increases the time needed to perform an attack?

FONDAZIONE BRUNO KESSLER

Experiment definition

Goal

Study is to analyze the effect of source code obfuscation techniques with the *purpose* of evaluating their effectiveness in making the code resilient to malicious attacks.

Quality focus

Capability of understanding the obfuscated code.

Capability to perform attacks on the obfuscated code

Treatments

Decompiled, obfuscated code vs. decompiled, clear code

Dependent variables

(i) Ability to perform comprehension tasks

(ii) Time required for comprehension

(iii) Ability to correctly perform an attack

(iv) Time required to perform an attack

FONDAZIONE BRUNO KESSLER

Null hypotheses

H01 The obfuscation does not significantly reduce source code comprehensibility.

H02 The obfuscation does not significantly increase the time needed to perform code comprehension tasks.

H03 The obfuscation does not significantly reduce the capability of subjects to correctly perform an attack.

H04 The obfuscation does not significantly increase the time needed to perform an attack.

FONDAZIONE BRUNO KESSLER

Balanced design

- Decompiled code
- Code browsing tools
- Debuggers
- API documentation
- Possibility to run the (modified) code

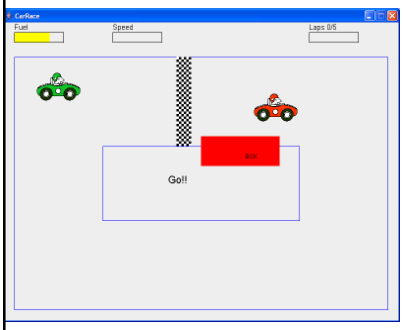
- Understanding tasks
- Change tasks

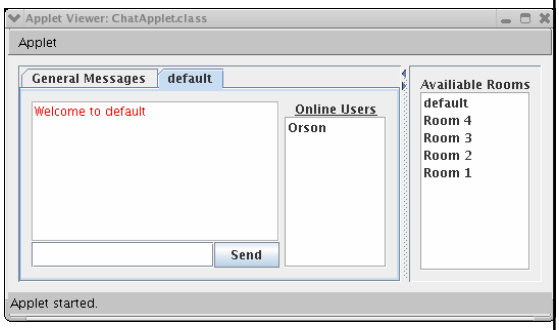
- Time/accuracy

1st session	Clear	Obfuscated
App1	**G1**	**G2**
App2	**G4**	**G3**

2nd session	Clear	Obfuscated
App1	**G3**	**G4**
App2	**G2**	**G1**

Objects





- 14 classes, for a total of 1215 LOC.
- 13 classes, for a total of 1030 LOC.

Subject

- 8 Master students form the University of Trento (computer science)
- Good knowledge of Java programming
- Knowledge of software engineering topics
  - Design
  - Testing
  - Software evolution
  - Code analysis

Treatment

- Identifier renaming
- Decompiled code
- Typical attack shenario

```
Student guy = new Student();
String name = "Mathematics";
Course course = new Course(name);
guy.apply(course);
course.run();
name.match("jas");
```

T1 T2 T3

```
y1 x1 = new y1();
String x2 = "Mathematics";
y2 x3 = new y2(x2);
x1.z1(x3);
x3.run();
x2.match("jas");
```


Preliminary lecture

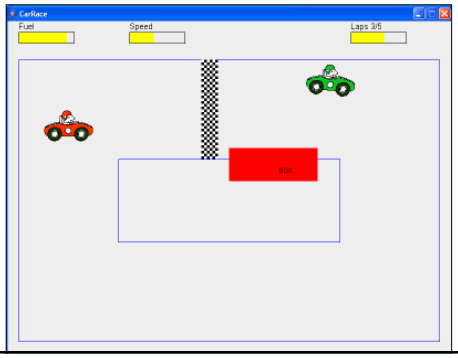
- Preliminary lecture to make the subjects aware of the experimental environment
  - IDE
  - Obfuscation
  - Debugging facilities
  - Pre questionnaire
  - Informed consent
  - Exercise on an application
    - To practice with the environment and mitigate the learning effect.

Experimental sessions

- 2 experimental sessions
  - Description of the application
  - Either clear or obfuscated source code
  - Possibility to run the (modified) code
  - Four paper sheets (each one contains a task)
  - A post questionnaire

Kinds of attacks

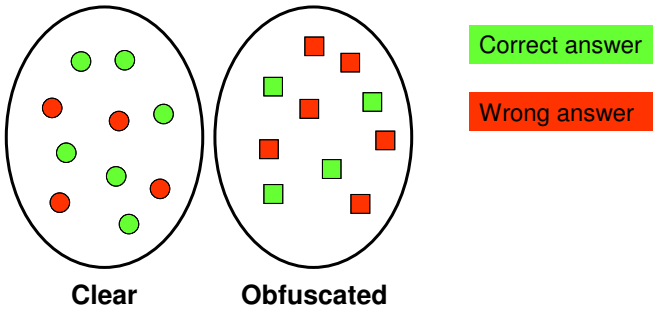
- Spotting specific functionalities
  - Observable features
- Tampering with the application
  - Make the application do something that is not available is the original code



Survey questionnaire

- Clarity of task and objective
- Difficulties experienced when performing the tasks
- Confidence in using the development environment and the debugger
- Percentage of time spent looking at the code or executing the system

Descriptive statistics



Is the distribution of correct and wrong answers statistically correlated with the treatment (obfuscation)?

Accuracy

	Comprehension		Attack		Overall	
Treatment	**Wrong**	**Correct**	**Wrong**	**Correct**	**Wrong**	**Correct**
Clear	7	11	3	15	10	26
Obfuscated	12	8	12	8	24	16
Fisher test	0.33		**0.009**		**0.006**	
Odds	2.3		**7.1**		**3.8**	

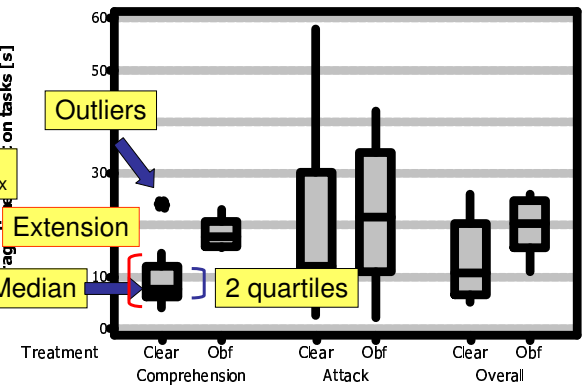
$$OR = \frac{q/(1-p)}{p/(1-q)}$$

An odds indicate how much likely is that an event will occur as opposed to it not occurring.

Time

$$d = (M_1 - M_2)/\sigma$$

The Cohen *d* effect size ind... magnitu... factor treatment effect on the dependent variables



	Comprehension	Attack	Overall
Mann-Whitney	**0.002**	0.19	**0.02**
Effect Size	**1.8**	0.2	**1.03**

FONDAZIONE BRUNO KESSLER

Null hypotheses

- **H01** The obfuscation does not significantly reduce source code comprehensibility.

HA2 The obfuscation significantly increases the time needed to perform code comprehension tasks
Effect size = 1.8

HA3 The obfuscation significantly reduces the capability of subjects to correctly perform an attack.
Odds ratio = 7.1

- **H04** The obfuscation does not significantly increase the time needed to perform an attack.

FONDAZIONE BRUNO KESSLER

Threat to validity

Construction validity

- Measurements were as objective as possible
  - Comprehension tasks had only one correct solution
  - Change tasks evaluated with test cases

Internal validity

- Full factorial design with random assignments to balance individual factors and to limit learning effect

Conclusion validity

- Non parametric tests are used, we do not assume data normality

External validity

- The subject are students, only further studies can confirm that our results can be generalized to professional developers

Ongoing work

Consider the impact of other factors

- Subjects' ability
- System
- Lab

Evaluate feedback after the experiment

- Clarity of objectives/tasks
- Difficulties
- Confidence with the environment
- Allocation of time code browsing/execution

Ongoing work

Torino:

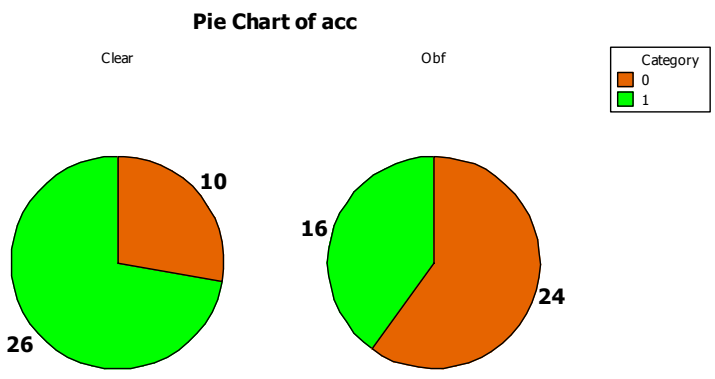
- 22 PhD students
- Same obfuscation

Benevento:

- 16 master students,
- Different obfuscation

What with multiple obfuscations?

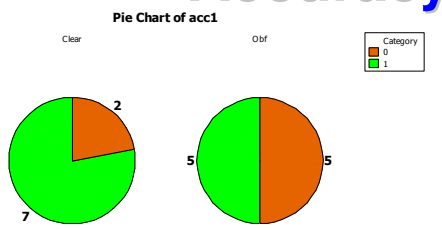
Accuracy



Fisher test
p-value = 0.005977
odds ratio = 0.2613782

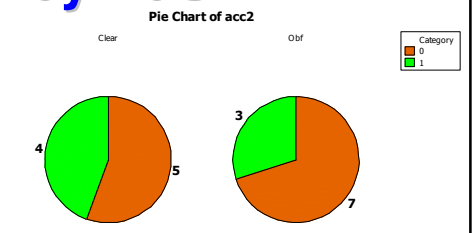
Panel variable: treat

Accuracy by Task



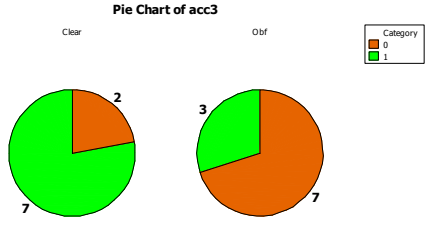
Panel variable: treat

Fisher test
p-value = 0.3498
odds ratio = 0.3059173



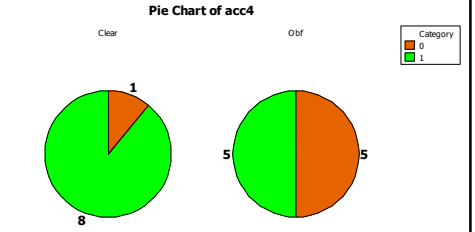
Panel variable: treat

Fisher test
p-value = 0.6499
odds ratio = 0.5539091



Panel variable: treat

Fisher test
p-value = 0.06978
odds ratio = 0.1395424



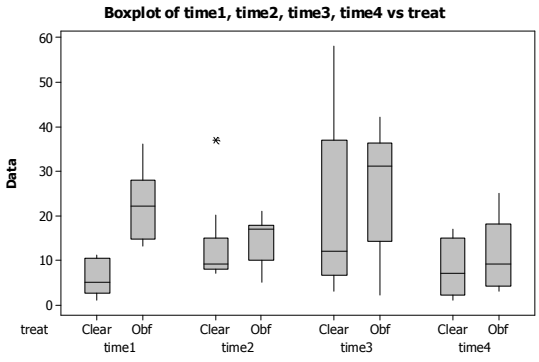
Panel variable: treat

Fisher test
p-value = 0.1409
odds ratio = 0.1399176

Time

Wilcox test unpaired one-tailed
P-value 0.02487

Time by task



Wilcox test unpaired one-tailed
P-value t_1:0.0001373 t_2:0.1421 t_3:0.1733 t_4:0.3418