Dynamic K-gram Software Birthmark

Yameng Bai Xingming Sun Guang Sun
Xiaohong Deng Xiaoming Zhou

19th Australian Conference on Software Engineering, 2008

Summary: Brief overview

Abstract

"A birthmark can help to prove software theft by identifying intrinsic properties of a program. Two programs with the same birthmark are likely to share a common origin. In this paper, we propose a novel dynamic birthmark. ... To evaluate the strength of the birthmarking technique, we compare static k-gram based software with dynamic approach from similarity with academic obfuscation tools."

Strengths: Detailed descriptions (1)

Article describes the existing technique:

- Defines the problem formally
- Explains how static birthmarks work
- Analyzes the technique pointing out where an improvement could be made

Valuable to the reader:

- Gives a background to someone not familiar with previous developments in the area
- Introduces the reader to terminology and notation used in the rest of the article

Strengths: Detailed descriptions (2)

Then then the article describes its contribution:

- Introduces the idea of considering input as well as the code to make birthmarks harder to manipulate
- Gives information on how the system is implemented by providing pseudo-code snippets
- Illustrates the process flow with worked examples

Valuable to the reader:

- Gives grounds for deciding whether the contribution was significant
- Provides enough information to actually implement the system

Weaknesses: Experimental design (1)

Although the article explains what the test program – Conzilla – is, it does not give any justification as to why it was the only one selected or why it can be considered representative of others

Claims

"The result shows that the new birthmark provides both high credibility and resilience. In particular, it proves that the dynamic birthmark is more resilient to semantics-preserving transformations than the static k-gram birthmark."

Is this true **only** for Conzilla?
There is nothing in the article that suggests otherwise!

Weaknesses: Experimental design (2)

Readers are left guessing why there was just one test subject:

- Performance
  - Dynamic birthmarking takes too long to test more programs
  - But this raises a question: is the added accuracy worth the added complexity?
- Aiming for better results
  - Other programs gave worse results, so they were discarded
  - But then the results are selective
- Focus on obfuscation
  - Main focus was on different obfuscation techniques, not on different input programs
  - But without exploring the cases that the system works on, one cannot claim that it is more accurate than the other

Discussion topic

What do you value more in a security paper: detailed descriptions or rigorous testing?