Defeating Hardware Trojan through Software Obfuscation

Andrea Marcelli Marco Restifo Ernesto Sanchez Giovanni Squillero

Goal

Develop an obfuscation mechanism based on evolutionary algorithms to minimize the chance of activation of a multi- stage trigger Hardware Trojan.

It can be used to protect critical infrastructures and operations at a minimum and predictable loss of performances.

Hardware Trojan

HT conceals its malicious behavior during normal operations: it monitors specific circuit inputs or internal signals and activates the payload only when a predefined pattern, known as trigger, is detected.

Known countermeasures

The security research community proposed several techniques to discover hidden HT, such as functional tests, runtime monitoring, and side-channel based approaches. However, the effectiveness of existing methods is subject to a number of practical challenges.

The multi-stage trigger

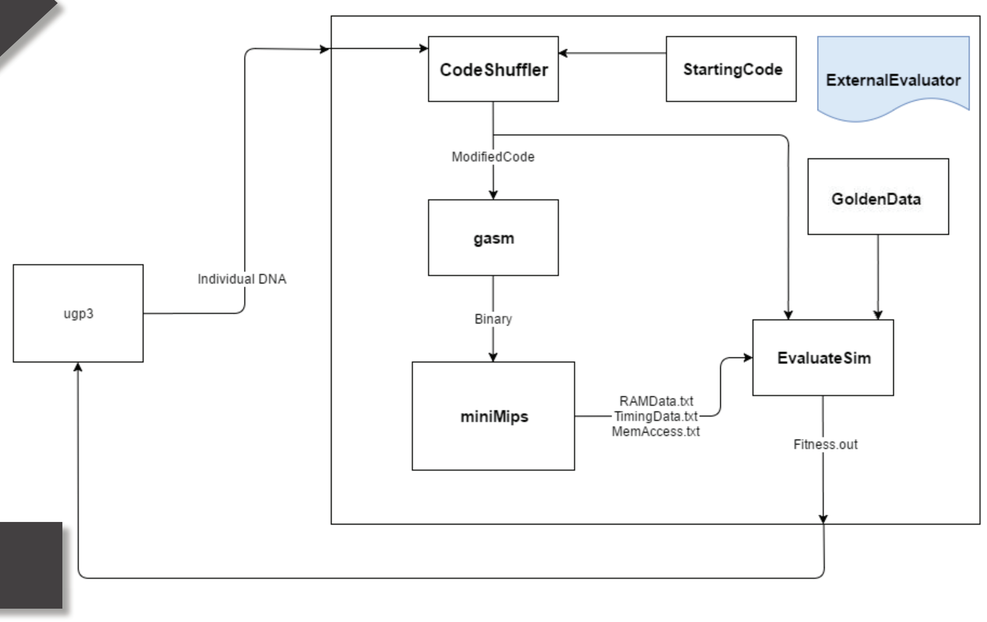
In 2016 Yang et al. [1] presented an interesting fabrication- time attack that could lead to a controlled privilege escalation attack. By injecting a single gate and a capacitor in the open spaces of an already placed and routed design, Yang et al. were able to force the flip -flop that holds the privilege bit to a desired value.

[1] K. Yang, M. Hicks, Q. Dong, T. Austin, and D. Sylvester, "A2: Analog malicious hardware," in Security and Privacy (SP), 2016 IEEE Symposium on. IEEE, 2016, pp. 18–37.-flop that holds the privilege bit to a desired value.

Our contribution

We introduce the idea of an evolutionary software-obfuscation of critical code to avoid the triggering of malicious hardware injected components.

Evolutionary process



Threat scenario

Due to the impossibility of verifying any piece of hardware to guarantee its trustworthiness, we conservatively consider any hardware platform as untrustable and already compromised by malicious parties.
In the same way, we define any software that could be executed on a particular hardware platform as untrustable and already compromised too.

Experimental Evaluation

Case 1

unrolled fibonacci implementation

96.4% drop*

27 out of 28 triplets removed

Case 2

matrix multiplication

77.6% drop*

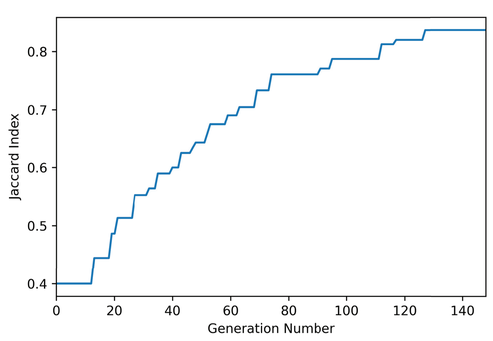
52 out of 67 triplets removed

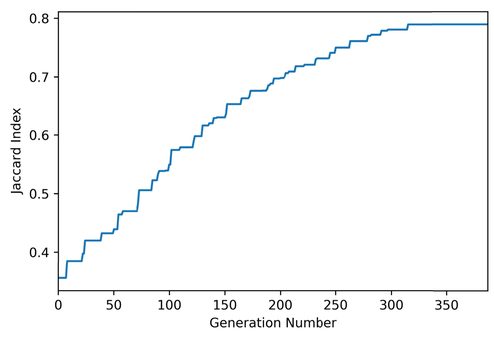
fitness 1

Jaccard Index

It is used to evaluate the similarity between the obfuscated code and the original one.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

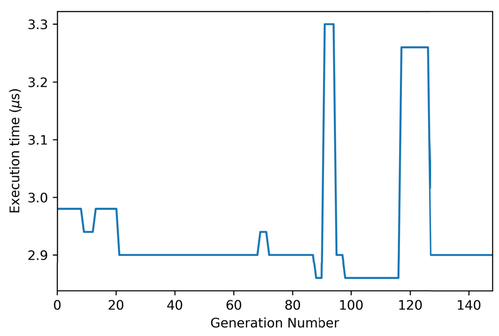


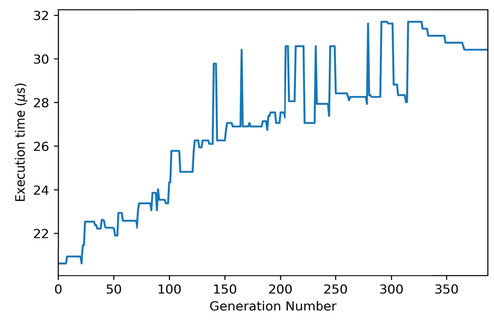


fitness 2

Execution Time

Exection time could be affected by the offuscation algorithm. The evolutive process tries to minimize possible overheads

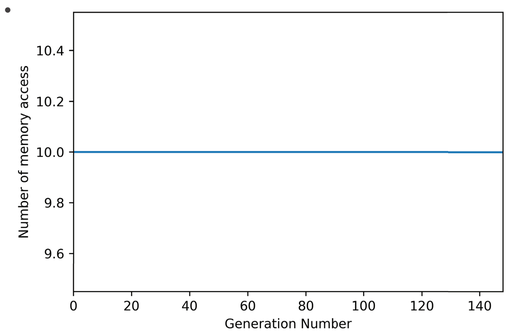


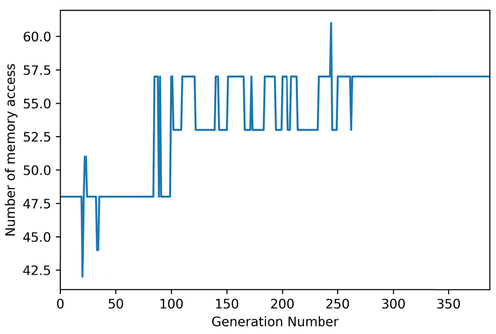


fitness 3

Memory access

The number of memory accesses could be affected by the offuscation algorithm. The evolutive process tries to minimize possible overheads





* Percentage of probability-drop of HW trojan activation, assuming each program has been injected one three-stage HT trigger sequence.

cad.polito.it