Subtractive Hardware Trojans

Mohamed Tarek Ibn Ziad and Simha Sethumadhavan
Department of Computer Science, Columbia University
New York, NY, USA
mtarek,simha@cs.columbia.edu

I. Introduction

Hardware Trojans are malicious additions or modifications to existing circuit elements, which can be inserted at any stage of integrated circuit (IC) life cycle. There is already significant work that illustrates trust problems with hardware supply chain security [1]; for instance, prior work has amply demonstrated how small additions, such as changing the dopant level in a transistor [2] or adding a capacitor [3] can impact full-system security.

In this work, we propose the concept of a *Subtractive* hardware Trojan. A Subtractive hardware Trojan is created by removing one wire from the circuit in order to make it generate wrong outputs under special rare input combinations. Removal of a single wire to create a Trojan reduces the area, timing and power footprint making it stealthier than additive Trojans. Further, Subtractive hardware Trojans also have higher deniability as they closely mimic failures that occur during manufacturing.

The security blind spot that enables Subtractive Trojans is incomplete fault coverage during manufacturing testing. Foundries depend on test patterns to ensure that a manufactured IC is free from defects. In real-world commodity ICs, in order to reduce manufacturing cost, and to enable power and area savings, manufacturers aim for very high (99%), but not perfect, fault coverage. This is for two reasons: 1) automatic test generation tools do not scale for for sequential circuits or very large combinational circuits and 2) in a security oblivious case, the cost of testing increases disproportionately with desired fault coverage. The above reasons provide more opportunities to Subtractive Trojans to survive.

II. Rules of The Game

Similar to recent work on manufacturing-level attacks [3], we adopt the threat model of *untrusted foundry*. It is strictly more challenging to implement attacks at the fabrication phase due to limited information and ability to modify the design compared to other back-end phases. We assume an adversary with access to the flattened (placed and routed) gate-level netlist in the form of a GDSII file. The adversary does not have any a priori knowledge of the design's internal workings. More precisely, the adversary has no information of module hierarchies, synthesis options, or names of gates and signals.

We assume that the adversary is aware of the manufacturing testing. However, s/he has no control over it (i.e., the adversary cannot add or remove test patterns). We also assume that testing is bound by the limits of practicality.

III. Results

We present an automated flow for creating Subtractive Trojans in a victim gate-level netlist by an untrusted foundry. Our framework uses available structural test patterns to identify possible Subtractive Trojan candidates by using post-synthesis simulations. Then, an open-source boolean satisfiability (SAT) solver is used to generate the corresponding malicious triggers.

We perform experiments to show the feasibility of constructing Subtractive Trojans by using two different sets of benchmarks, EPFL and ISCAS. Our results show that Subtractive Trojan insertion is possible whenever fault coverage tests are imperfect. We observe that vulnerability to Subtractive Trojans increases with the increase of circuit size and logic depth. Furthermore, we compare the side-channel overheads of Subtractive Trojans versus traditional additive hardware Trojans obtained from the Trust-Hub Trojan benchmark suite. We observe that our proposed Subtractive Trojan is more stealthy, while having almost zero area and power overheads.

Subtractive Trojans are effective against state-of-the-art hardware Trojan defensive techniques. The side-channel perturbations due to the removal of one wire are so small that they cannot be measured effectively by existing side channel based hardware Trojans defenses. For instance, the energy consumption of a typical `NAND` gate is in the order of nanojoules. The noise variability of each gate, due to manufacturing variations, needs to be within that nanojoule margin in order to allow for effectively identifying the presence or absence of one wire using side-channels in the best case. For example, the 20X leakage power and 30% delay variations in even a fairly old technology node (180 nm) can easily mask the side-channel effect of a single removed wire.

IV. Future Work

Subtractive Trojan sets a new lower bar on overhead from a hardware Trojan. Future work should include developing new methods to detect these sub-zero overheads hardware Trojans, most notably, VLSI testing should take in consideration trust concerns when setting the desired fault coverage.

References

[1] K. Xiao, D. Forte, Y. Jin, R. Karri, S. Bhunia, and M. Tehranipoor, "Hardware Trojans: Lessons learned after one decade of research," *ACM Transactions on Design Automation of Electronic Systems*, vol. 22, pp. 6:1–6:23, May 2016.

[2] G. T. Becker, F. Regazzoni, C. Paar, and W. P. Burleson, "Stealthy dopant-level hardware Trojans," in *CHES '13*, pp. 197–214, August 2013.

[3] K. Yang, M. Hicks, Q. Dong, T. Austin, and D. Sylvester, "A2: Analog malicious hardware," in *SP'16*, pp. 18–37, May 2016.