Combinatorial Methods Enable the Detection and Location of Hardware Trojan Horses

Dimitris E. Simos, Ludwig Kampel, Paris Kitsos

COMET Competence Centers for Excellent Technologies www.ffg.at/comet

SBA Research MATRIS

Problem & Motivation: Cryptographic Hardware Trojan Testing

Problem Description

- Hardware security needs to be addressed when treating the security of an electronic system holistically.
- Globalization of semiconductor industry additionally raises concerns about the authenticity and security of fabricated Integrated Circuits.
- The integration of a Hardware Trojan (HT), i.e. a malicious modification to FPGAs, microprocessors or IoT devices, is one of the most threatening attacks.
- A HT is a small circuit that integrates a logic that was not intended in the circuits design and generally consists of:
  - **trigger**: activates the HT when specific rare events appear in the input of the HT
  - **payload**: executes the malicious function of the HT when the activating input is recognized

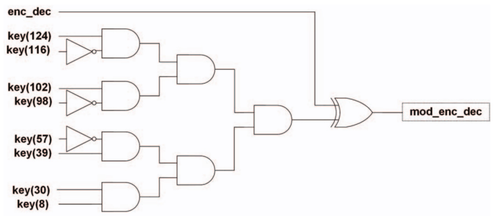


Figure 1: Example of a HT circuit.

Attack Scenario/Threat Model

- We consider an Integrated Circuit implementing the AES cryptographic algorithm in ECB mode for 128 bit keys.
- Attacker has means to integrate a HT, activated by a small ℓ-bit pattern "hidden" in the plaintext input space (2^{128}).
- The HT affects the logic of the circuit such that the output is changed, e.g. switch the encryption/decryption mode of the circuit $\Rightarrow$ DoS attack.

Detecting Hardware Trojans using CT

Using CT as logic testing approach for combinatorial HT detection

- Demonstrate the applicability and efficiency of CT for HT detection by drawing the analogy with combinatorial black-box software testing (see Figure 4):
  - The input space is modelled by 128 binary parameters
  - The desired length ℓ for which HT testing should be applied, maps to the strength t guiding combinatorial test suite generation
  - A Covering Array of strength t is generated; the rows serve as test suite and are used as input for the plaintext
  - The output of the AES module is compared against the output of a trusted implementation of the algorithm which provides an oracle
  - When the outputs disagree for some test a HT has been detected

Results

- CT provides the **theoretical guarantees** for exciting a HT with a triggering pattern of specific length ℓ.

Length	Positions	Pattern	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
$\ell=2$	21-114	01	3	10	30	63	308	616	4,125
$\ell=3$	21-79-114	101	1	4	15	32	156	308	2,063
$\ell=4$	21-79-97-119	0101	0	3	7	17	82	160	987
$\ell=5$	3-23-89-107-124	10100	0	2	7	9	38	80	532
$\ell=6$	3-23-89-95-117-124	001101	0	0	2	6	21	44	200
$\ell=7$	3-23-63-90-96-118-122	1010110	0	0	2	3	12	23	107
$\ell=8$	3-23-63-79-90-96-118-122	01100100	0	0	0	1	7	10	67

Figure 2: Activations of variable length ℓ Trojans per strength t test suite.

- CT methods reduce the number of needed tests by orders of magnitude:

k	t	exhaust. t-bit	CWV	CTdetect	CTlocate
128	2	2^7	129	11	54
128	3	-	256	37	135
128	4	2^{13}	8,256	112	346
128	5	-	16,256	252	5,921
128	6	-	349,504	720	29,830
128	7	-	682,752	2,462	103,691
128	8	2^{23}	11,009,376	17,544	595,979

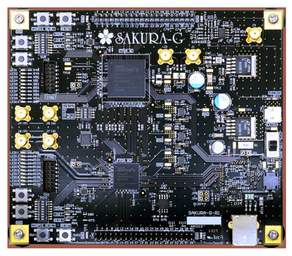


Figure 3: Left: Comparison of sizes of test suites coming from CT methods (CTdetect, CTlocate) against other state-of-the-art logic testing techniques for combinatorial HT detection. Right: The SAKURA-G FPGA used for the experiments.

- Compared to random methods, CT stands out by covering all triggering patterns of length $\leq t$:

Length	Total patterns	CTdetect	Random	Missing
$\ell=2$	32,512	100%	94, 92%	1,649
$\ell=3$	2,731,008	100%	99, 20%	21,718
$\ell=4$	170,688,000	100%	99, 92%	129,882
$\ell=5$	8,466,124,800	100%	99, 96%	3,295,565
$\ell=6$	347,111,116,800	100%	99, 998%	4,268,479

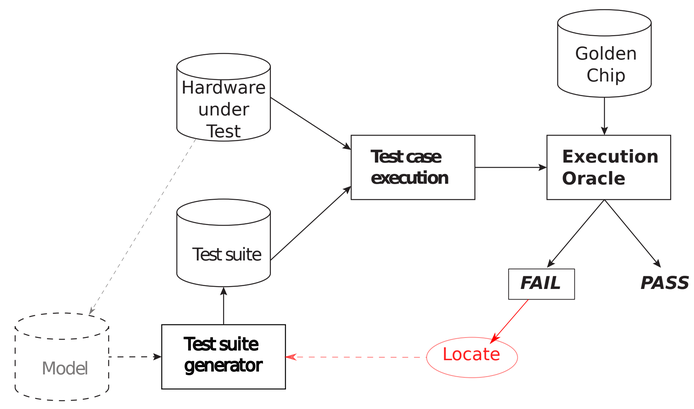
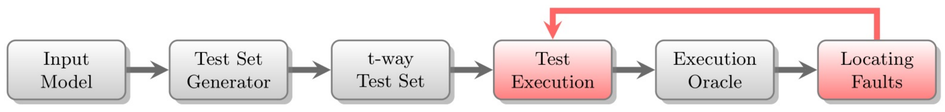


Figure 4: Combinatorial Testing Cycle adopted for hardware testing. The extension to HT location, i.e., identification of trigger patterns by means of their input bits and values, is highlighted in red.

Combinatorial Methods for HT Location

- From mere **detection** of the presence of a HT to HT **location**:
  - Identify bits and values of the triggering pattern
- Combinatorial fault localization has the means for HT location.



- Adaptive and non-adaptive fault localization methods can be applied:
  - Experiments show that non-adaptive methods are highly applicable, producing test suites of competitive size (see Figure 3)
- Precisely locating the HTs triggering pattern in the input space:
  - Allowing for post-analysis to understand the purpose of the attack

Conclusion & Outlook

- CT provides a complementary method for HT detection and location:
  - below an "undetectability level" ℓ HTs can be detected and located
  - beyond ℓ established hardware testing techniques can be applied
- A closer collaboration between researchers and practitioners of the fields of CT and hardware testing seems fruitful.
- Augment established hardware testing methods with CT methods.

L. Kampel, P. Kitsos, and D. E. Simos. Combinatorial fault localization for hardware trojan location. In *Internal Report*, 2020.
P. Kitsos, D. E. Simos, J. Torres-Jimenez, and A. G. Voyiatzis. Exciting fpga cryptographic trojans using combinatorial testing. In *2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE)*, pages 69–76, Nov 2015.

Bundesministerium Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie
Bundesministerium Digitalisierung und Wirtschaftsstandort
FFG Forschung wirkt.
wirtschaftsagentur wien Ein Fonds der Stadt Wien
Europäische Kommission
FWF Der Wissenschaftsfonds.
netidee OPEN INNOVATIONS

SBA Research (SBA-K1) is a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the federal state of Vienna. The COMET Programme is managed by FFG.