A Testing Methodology for Hardware Trojan Detection


[1] KU Leuven ESAT/COSIC (BE)
[2] CEA Tech Region, DPACA/LSAS (FR)

TRUDEVICE 2015, 13 March 2015
Grenoble (France)
HINT Project – Overview (I)

- HINT = Holistic Approaches for Integrity of ICT-Systems
- Project Number: 317930
- Project Website: www.hint-project.eu
- Project start: October 1, 2012
- Project duration: 3 years
- Total Costs: € 5.103.893
- EC- Contribution: € 3.350.000
- Project is co-financed by the European Commission under Seventh Framework Programme
Motivation: ensure authenticity and integrity of hardware components in modern ICT systems
HINT Project – Overview (III)

WP6 – Project Management and Dissemination

WP1 – User Requirements and System Architecture

WP2 – Robust Energy-Optimized Nano Structures for Integrity-anchors

WP3 – Holistic Integrity Checking for Components in ICT Systems

WP4 – Integration, Prototyping, Validation

WP5 – Security Evaluation
Hardware Trojans – What, How and Why?

- Hardware Trojan (HT): **Malicious** modification of an Integrated Circuit during design flow
- Issue first raised by US Department of Defense
  - **Outsourcing** of IC fabrication questions **trust** in the final chip
- Very rich HT taxonomy
  - Insertion phase, infection level, effect, activation, location, ...

---

Diagram:
- Trigger
  - activation
- Payload
  - “sensing circuitry”
    - internal
    - external
  - “malicious activity”
    - information leaks
    - DoS
    - ...

---

13 March 2015

TRUDEVICE 2015
Our HT Detection Approach

- Fingerprinting side-channel characteristics
  - Learning phase: characterization of Golden circuit
  - Matching phase: comparison with Device under Test
- Realistic measurement scenario, no simulations
  - Target platform: FPGA Xilinx Spartan-6 LX75 (Sakura-G)
  - Side-channel: power consumption
- Golden circuit
  - AES-128 implementation
- HT Infected circuit
  - Many variants tested
  - This presentation: externally triggered, no payload!
HT Infected Circuit

- Insertion after PAR
  - Xilinx Native Circuit Description
- Externally triggered
  - 16-bit activation sequence
  - Only 2 slices
  - Close to occupied slices
Golden circuit: signal routings
HT Infected circuit: signal routings
T-test Distinguisher

- Welch’s two tailed T-test
  - Test the null hypothesis that the means of 2 populations are equal
    \[
    t = \frac{\mu_0 - \mu_1}{\sqrt{\frac{\sigma_0^2}{N_0} + \frac{\sigma_1^2}{N_1}}}
    \]
    - Robust, reliable, and with low computation effort
    - Quantify confidence in the result

- In our particular test scenario
  - Populations are sets of power measurements
    - set0 (Golden model), set1 (DuT)
  - Main idea:
    - DuT = Golden model, populations should have same means
    - DuT ≠ Golden model, populations should have different means
Experimental Results (I)

Single board / Single Measurement Setup

10,000 measurements per design
3 MHz clock, 1.25 GS/s
20,000 samples/ measurement

Population of power traces with random inputs

PC

SAKURA-G

measurement point J3

OSCILLOSCOPE

CONTROL FPGA

CRYPTO FPGA (AES)

30 dB Amp. 48 MHz LPF
How does a measurement look like?

- **input key**
- **input pt**
- **output ct**
- **quantized power measurement**
Golden vs. Golden

- Perfectly contained within $\pm 4.5 \rightarrow 99.999\%$ confidence
Golden vs. Golden’

- Environmental variations result in offset
Golden vs. HT Infected

- HT activity visible during transmission of input operands
Golden vs. Other HT infected
Conclusions

- First results towards evaluating the suitability of t-test for HT detection
- Real measurements, not simulations
- Ideal measurement conditions (1 board, 1 setup):
  - Good, stable results
  - But need to deal with environmental variations
- Non-ideal conditions (more boards, more setups):
  - Need to re-define decision thresholds
  - Currently under investigation in HINT
Thanks for your attention!

Questions?