Hardware Trojan Horses in Cryptographic IP Cores

Shivam Bhasin, **Jean-Luc Danger**, **Sylvain Guilley**, **Xuan Thuy Ngo** and **Laurent Sauvage**

TELECOM-ParisTech *and* Secure-IC S.A.S.

August 21, 2013 09:55 – 10:20

Presentation Outline

1. Introduction
2. Hardware Trojan Detection State-of-the-art
3. Layout-GDS II Comparison Technique
4. Results
5. Conclusion

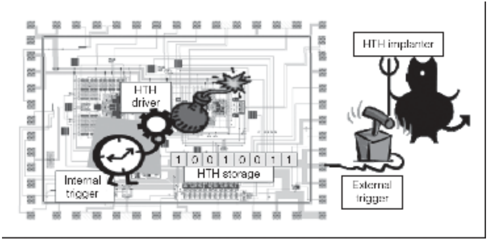
Presentation Outline

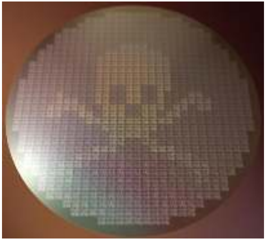
1. Introduction
2. Hardware Trojan Detection
   State-of-the-art
3. Layout-GDS II Comparison Technique
4. Results
5. Conclusion

Hardware Trojan Introduction

Hardware Trojan (HT) Definition

- Malicious modifications in Integrated Circuits (ICs).
- Realize malicious functions (Leakage of sensible information, alteration of IC behaviours, etc.).
- HT was born because of outsourcing design and fabrication process.

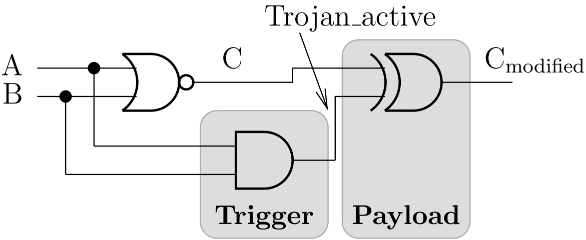




Hardware Trojan Structure

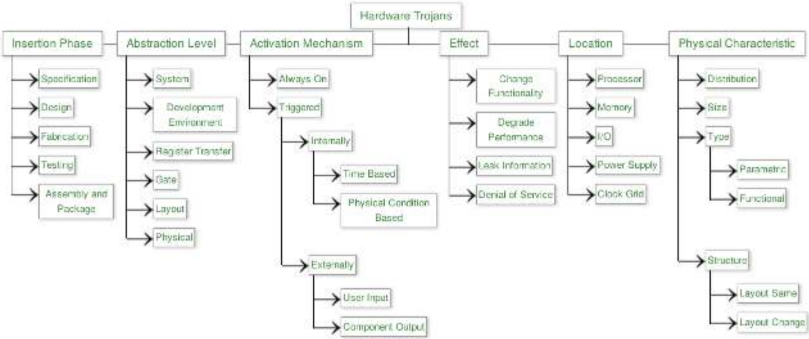
Any HT is composed of two main components

- **Trigger**: is the part of HTH used to activate the malicious activity.
- **Payload**: is the part of HTH used to realize/execute the malicious activity.



Hardware Trojan Taxonomy

- Classify all type of HT.
- Help to develop suitable detection techniques for each HT type.



Hardware Trojans

- Insertion Phase
  - Specification
  - Design
  - Fabrication
  - Testing
  - Assembly and Package
- Abstraction Level
  - System
  - Development Environment
  - Register Transfer
  - Gate
  - Layout
  - Physical
- Activation Mechanism
  - Always On
  - Triggered
    - Internally
      - Time Based
      - Physical Condition Based
    - Externally
      - User Input
      - Component Output
- Effect
  - Change Functionality
  - Degrade Performance
  - Leak Information
  - Denial of Service
- Location
  - Processor
  - Memory
  - I/O
  - Power Supply
  - Clock Grid
- Physical Characteristic
  - Distribution
  - Size
  - Type
    - Parametric
    - Functional
  - Structure
    - Layout Same
    - Layout Change

Presentation Outline

1 Introduction

2 Hardware Trojan Detection
State-of-the-art

3 Layout-GDS II Comparison Technique

4 Results

5 Conclusion

Hardware Trojan Detection

Classification of HT Detection techniques

- **Destructive reverse engineering**: try to reconstruct netlist and layout of ICs.
- **Invasive methods**: try to (prophylactically) modify the design of IC to prevent the HTH or to assist another detection technique.
- **Non-Invasive methods**: are done by comparing the performance characteristics of an IC with a known good copy also known as the “golden circuit”.

Invasive Methods

- Chakraborty et al. propose a design with two operating modes (Normal and Transparent mode).
- Salmani et al. propose a procedure to insert dummy flip-flops into IC logic.
- Banga et al. propose using QN of D flip-flops.
- Other researchers also suggest logic additions that will make it easier to detect a HTH utilising side-channel analysis.

Non-Invasive Methods (1)

Non-Invasive methods can be done either at **runtime** or in the **testing phase**.

Non-invasive methods on runtime

- Bloom et al. detail a HTH detection approach that uses both hardware and software to detect HTs.
- Abramovici et al. propose real-time security monitors (**DEFENSE**).

Non-Invasive Methods (2)

Non-invasive methods on testing phase

Logic Testing:

- Compare the functionality of the design of the circuit with the implemented circuit.
- Chakraborty et al. suggest to test rare occurrences on an IC rather than testing for correctness.

Side Channel analysis use one or more side-channel parameters to obtain Fingerprint of ICs. We can cite:

- Rad et al. propose using power supply transient signal analysis.
- Banga and Hsiao propose the "sustained vector technique" that is able to magnify the side-channel.

Presentation Outline

1 Introduction

2 Hardware Trojan Detection
State-of-the-art

3 Layout-GDS II Comparison Technique

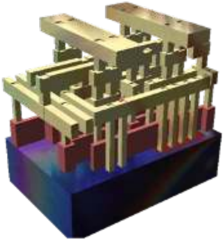
4 Results

5 Conclusion

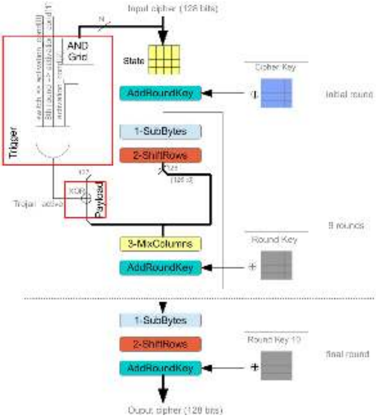
Introduction

Scenario

- Scenario: Attacker is founder.
- Study Hardware Trojan insertion in GDSII level.
- Impact of the insertion on the IC layout.
- The possibility to detect Hardware Trojan visually?



Case Study-AES 128 bits



Experiment setup

- Vary core utilization rate of AES (50% → 99%).
- Vary Hardware Trojan size (1 AND gate → 128 AND gates).
- Software used: Cadence / Encounter.

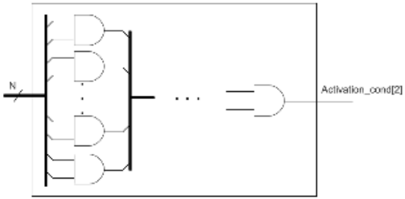


Figure: Hardware Trojan with $N-1$ AND gates

Presentation Outline

1 Introduction

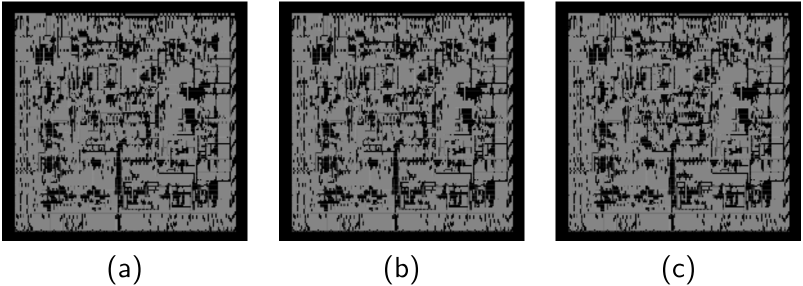
2 Hardware Trojan Detection
State-of-the-art

3 Layout-GDS II Comparison Technique

4 Results

5 Conclusion

AES layouts for 6th metal layer



(a) (b) (c)

Figure: 6th metal layer AES layouts (1200 μm × 1200 μm) with 50% core utilization rate for (a) Original AES, (b) AES with 1 AND gate HTH, (c) AES with 128 AND gate HTH

Cross correlation between original AES layout and affected AES layout

		Hardware Trojan size (Nb of AND gates)							
		1	2	4	8	16	32	64	128
Core utilization rate	50%	0.9991	0.9972	0.9981	0.9950	0.9933	0.9918	0.9815	0.9668
	60%	0.9987	0.9968	0.9959	0.9955	0.9944	0.9893	0.9788	0.9670
	70%	0.9989	0.9981	0.9918	0.9941	0.9881	0.9850	0.9594	0.9067
	80%	0.9999	0.9965	0.9898	0.9957	0.9780	0.9711	0.8970	0.8509
	90%	0.9988	0.9990	0.9983	0.9962	0.9832	0.9572	0.8858	0.4010
	95%	0.9997	0.9984	0.9980	0.9889	0.9589	0.9115	0.8824	0.8202
	99%	0.9917	0.938	0.9714	0.9527	0.3798	NC	NC	NC

- In black: ECO route works
- In red: total rerouting
- NC: routing impossible (not placement)

AES layout for all metal layer

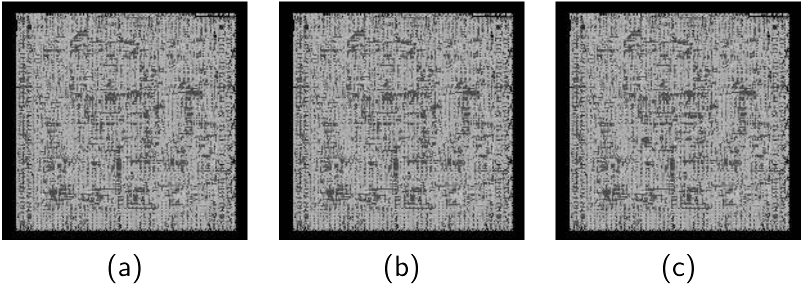
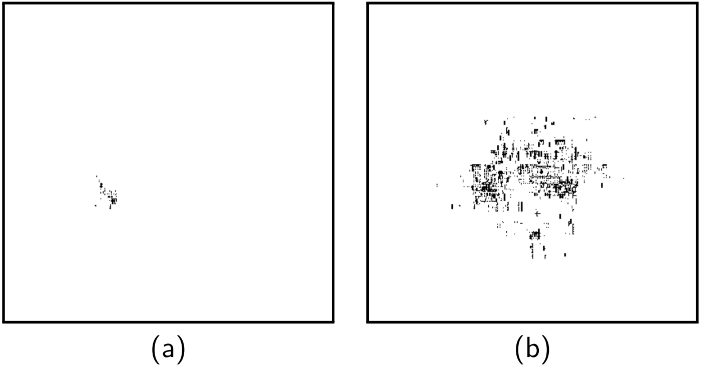

(a)
(b)
(c)

Figure: AES layouts (1200 μm × 1200 μm) with 50% core utilization rate for (a) Original AES, (b) AES with 1 AND gate HTH, (c) AES with 128 AND gate HTH

Cross correlation between original AES layout and affected AES layout

		Hardware Trojan size (Nb of AND gates)							
		1	2	4	8	16	32	64	128
Core utilization rate	50%	0.9996	0.9788	0.9794	0.9770	0.9766	0.9760	0.9719	0.9874
	60%	0.9993	0.9975	0.9973	0.9970	0.9964	0.9923	0.9843	0.9783
	70%	0.9991	0.9987	0.9947	0.9952	0.9924	0.9889	0.9655	0.9272
	80%	0.9997	0.9976	0.9942	0.9969	0.9868	0.9790	0.9291	0.8915
	90%	0.9990	0.9988	0.9981	0.9964	0.9878	0.9709	0.9255	0.5162
	95%	0.9995	0.9978	0.9974	0.9927	0.9742	0.9387	0.9159	0.8661
	99%	0.994	0.9965	0.9803	0.962	0.4905	NC	NC	NC

Pixelwise difference of AES layouts


(a)

(b)

Figure: Pixelwise difference of AES layouts with 50% core utilization rate for Original layout and Infected Layout with (a) 1 AND gate, (b) 128 AND gate.

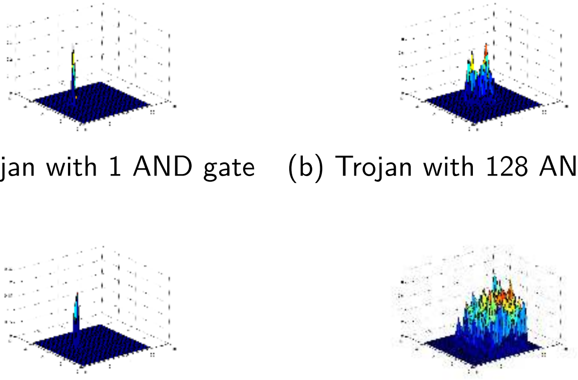
Grid-correlation between layouts

Grid-correlation definition

- Improve Cross correlation coefficients.
- Split images on different pieces
- Compare theses pieces of one with these corresponding pieces of others.
- Reverse Cross correlation coefficients are computed to visual improvement.

Grid-correlation examples

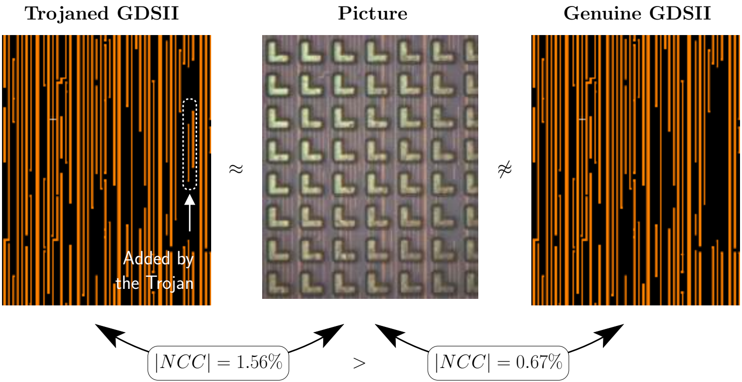
Grid-correlation exemple for CUR of 50% (a, b) and 95% (c, d)



(a)Trojan with 1 AND gate (b) Trojan with 128 AND gates

(c)Trojan with 1 AND gate (d) Trojan with 128 AND gates

Experiment on SECMAT circuit



Trojaned GDSII **Picture** **Genuine GDSII**

$\approx$ $\not\approx$

Added by the Trojan

$|NCC| = 1.56\%$ $>$ $|NCC| = 0.67\%$

Figure: Cross correlation based comparison between trojaned (*lefthand side*) / genuine (*righthand side*) GDSII and an actual picture, a microscope image of an AES area where the inserted HTH shows up (*center*).

Presentation Outline

1. Introduction
2. Hardware Trojan Detection
   - State-of-the-art
3. Layout-GDS II Comparison Technique
4. Results
5. Conclusion

Conclusion

Hardware Trojan threat

- Become a serious threat in military, financial fields.
- For now, there is NO technique which can detect all type of Hardware Trojan.

GDSII comparison technique conclusion

- Can detect Hardware Trojan at layout level.
- No need golden model for detection.
- With a CUR superior than 80%, designer can prevent Hardware Trojan insertion.

Future Works

- Improve this technique with minutæ analyses.
- Insertion of Hardware Trojan in processors.

- We wish to collaborate with https://www.trust-hub.org/

Acknowledgements: HOMERE (2011-2013)

- Collaborative project sponsored by the French government:
  - Partners from industry are: Cassidian, Gemalto, Secure-IC,
  - Partners from academia are: ENSMSE, Telecom-ParisTech, CEA LETI, Université de Montpellier,
  - Institutional partner: Agence Nationale de Sécurité des Systèmes d'Information (ANSSI).

Hardware Trojan Horses in Cryptographic IP Cores

Shivam Bhasin, **Jean-Luc Danger**, **Sylvain Guilley**, **Xuan Thuy Ngo** and **Laurent Sauvage**

TELECOM-ParisTech *and* Secure-IC S.A.S.

August 21, 2013 09:55 – 10:20