Ending Hardware Attacks and IP Piracy

Chongxi Bao, Yang Xie, Domenic Forte and Ankur Srivastava

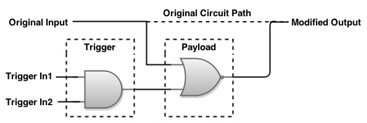
THIRTY YEARS OF SYSTEMS RESEARCH EXCELLENCE

INSTITUTE FOR SYSTEMS RESEARCH
A. JAMES CLARK SCHOOL OF ENGINEERING

Background

A Taxonomy of Hardware Attacks

IC Lifecycle	Attacks
Design/Synthesis	Trojan Insertion IP Theft/Piracy
Fabrication	Trojan Insertion Overbuilding Counterfeiting IP Theft/Piracy
Testing	Test Mode Reverse-engineering
Post-Deployment	Trojan Activation Reverse-engineering Recycling Side-channel Attacks



Hardware Trojans

They are malicious modifications to the circuit that can:

- Change the functionality
- Reduce circuit's reliability
- Leak valuable information

Prior Detection Methods

Test-time and Run-time

- Functional Testing
- Side-channel behavior

Design-time

- Formal Verification/Proofs
- Design-for-trust

Key Challenges:

- Golden model assumption
- Process variation
- Measurement noise
- Rare Trojan activation
- Resource overheads
- Automated proof generation

IP Piracy

A malicious foundry may reverse-engineering layout files to obtain original netlist and he can gain profits by

- Claiming the ownership of it
- Overbuilding the IC.

Prior Defense Approaches:

Split Manufacturing:

- A chip is split into metal layer and active layer, and the metal layer is fabricated in a trusted foundry to prevent the complete exposure of the design to offshore foundry

Hardware Obfuscation:

- Inserting additional key-based logics to obfuscate the net-list

Key challenges:

- Functionality obfuscation
- Layout obfuscation

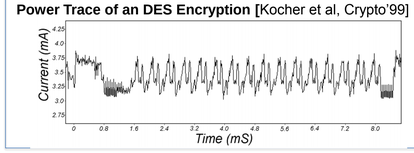
Future Work

Side-Channel Attack Resilient Architecture

Side-channel Attacks in Cryptography

- Recently, different side-channel attacks have been applied successfully to retrieve keys from DES, RSA and AES.
- Power analysis measures the power consumption of an encryption and correlates that with the underlying keys.
- Timing analysis measures the time of an encryption to retrieve the keys. Many cache timing side-channel attacks have been demonstrated very successful.

Power Trace of an DES Encryption [Kocher et al, Crypto'99]



Side-channel Attack Resilient Architecture

- Cache configurations that do not leak timing information
- Opportunities in 3D integration

Security Analysis and Enhancement of Split Manufacturing in 2D IC

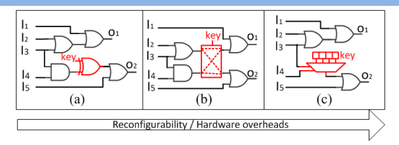
Split Manufacturing in 2D IC:

- A 2D IC can be split into active device layer and metal layer
- Active layer (gates) is outsourced for advanced fabrication
- Metal layer (interconnection) is performed by a trusted foundry

Attack Model: proximity attack + machine learning attack

- Proximity attack: two connected gates are usually placed close by
- Machine learning attack: learn the hidden interconnection based on the observed output difference between reconstructed and original ICs

Security Analysis and Enhancement of Three Hardware Obfuscation models



Three Hardware Obfuscation Models:

- XOR/XNOR locking
- Signal wire permutation
- Reconfigurable block

Analysis and Enhancement:

- Security analysis of three models based on a novel machine learning attack that intends to learn the secret keys
- Security enhancement techniques for each models to achieve the optimal defense results with low performance overheads

Current Research

Temperature-tracking Based Run-time Trojan Detection

Motivation

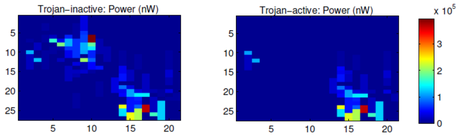
Test-time approaches: may miss some Trojan-infected ICs due to:

- Limited amount of test time
- Lack of Trojan activation
- Inactive Trojans have small impact on delay/power

Run-time monitoring approaches:

- Monitor the IC throughout its lifecycle
- Need a low-overhead solution

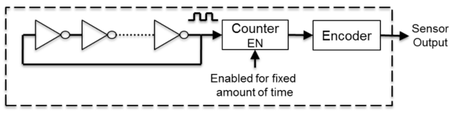
Avg. Power Consumption of RS232-T900 Benchmark



Key Observations:

- Active Trojans can cause large changes in power profile but monitoring power is expensive
- Changes in power will lead to changes in thermal profile
- Thermal profile can be measured with thermal sensors, which have low-overhead and are on most of the chips

Thermal Sensor [Zhang et al, ISPD'10**]**



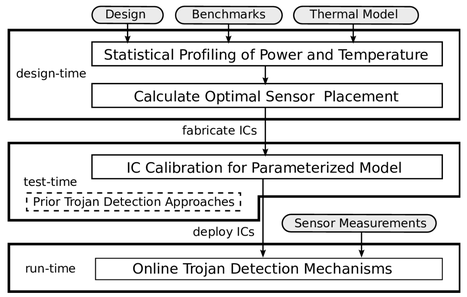
Problem Definition

Given two hypotheses of the system's state:

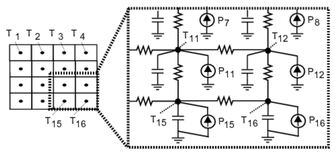
$\begin{cases} \boldsymbol{H_0} & \text{The state is Trojan} - \text{free or Trojan} - \text{inactive} \\ \boldsymbol{H_1} & \text{The state is Trojan} - \text{active} \end{cases}$

Use thermal sensor observations to determine if the IC's state (characteristics) correspond to $\boldsymbol{H_0}$ or $\boldsymbol{H_1}$

Overall Flow



RC Thermal Model (HOTSPOT) [Skadron et al, ISCA'03]



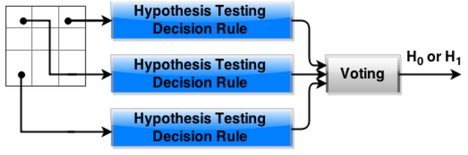
$$\sum_{\forall j \in N_i} \frac{1}{R_{ij}}\left(T_i(t) - T_j(t)\right) + \frac{C_i dT_i(t)}{dt} - P_i(t) = 0 \quad \forall i$$

- Sensors are placed in a way to minimize correlations between them following the work in [Zhang et al, ISPD'2010]
- IC parameter calibration is done to offset the effect of process variation following the steps in [Hu et al, DATE'2013]

Online Detection Methods

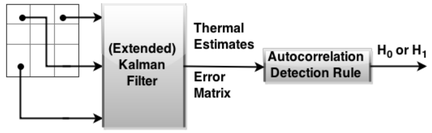
Local Approach

- Measurements are taken and processed locally with simple thresholding based on hypothesis testing
- Key feature: very low overhead



Global Approach

- Measurements are taken and processed together
- Use (Extended) Kalman Filter to track chip's thermal states
- Compare the autocorrelation of the error matrix with a threshold



Simulation Results

Approach	#sensors	Average (10 benchmarks)		
		t_+	f_+	Average detection time (ms)
Local	4	90%	61.4%	42.98
	16	90%	0%	101.9
	32	90%	0%	136.7
Global (Extended Kalman Filter)	4	100%	0%	44.67
	16	100%	0%	33.78
	32	100%	0%	27.14

t_+ -True positive rate f_+ -False positive rate

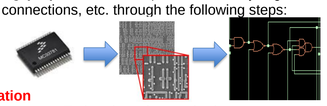
Reverse-engineering Based Hardware Trojan Detection

Motivation

Most of prior detection approaches require the existence of a golden (Trojan-free) chip, but golden chips are hard to acquire. We need to verify a chip is Trojan-free without the assumption of golden chip.

Reverse-engineering (RE) of an IC is the process of analyzing an IC's internal structures, connections, etc. through the following steps:

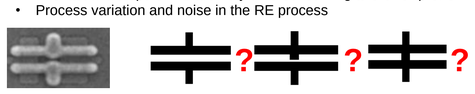
- Decapsulation
- Delayering
- Imaging
- **Annotation**
- **Schematic creation**



RE is arguably the strongest Trojan detection methods.

Challenges

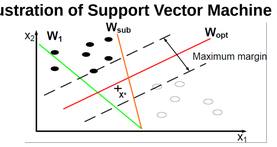
- The last two steps of RE are very time-consuming and error-prone
- Process variation and noise in the RE process



Our Proposed Approach

- We compare the images obtained by RE a golden layout (which is assumed to be handy with the designers)
- We apply **one-class SVM** to learn the noise induced by fabrication and RE process while sensitive to malicious modifications.

Illustration of Support Vector Machine (SVM)



- One-class SVM is used when only samples from one class is known, since we cannot know what modifications will be made.
- Several features are extracted by comparing the golden layout with images obtained from RE: centroid difference, area difference, etc.

Simulation Results

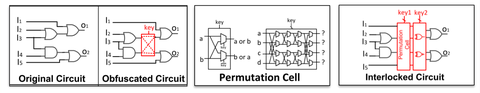
Detection Accuracy of Different Types of Trojans

Benchmark	#of gates	Trojan-free	TA	TD	TP
S27	57	100%	100%	100%	99.8%
S298	283	100%	100%	100%	99.6%
S5378	3455	100%	100%	100%	100%
S15850	10984	100%	100%	100%	99.8%
S38417	30347	100%	100%	100%	100%
b18	122559	99.4%	100%	100%	100%

Summary of Merits

- Our method eliminates RE steps 4 and 5 which require lots of manual effort and are error-prone
- Our method does not rely on golden IC
- Our method accounts for noise in the RE steps
- Our method is fully automated

IP Piracy Prevention Based on Signal Wire Permutation



Signal Wire Permutation:

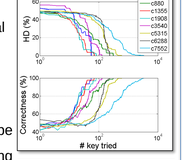
- A key-based hardware obfuscation technique that hides and permutes multiple wires of a netlist

Machine Learning Attack:

- Gradually improve a random key guess based on the observed Hamming Distance (HD) between a obfuscated circuit *F'(key)* and a functional circuit *F* obtained from open market

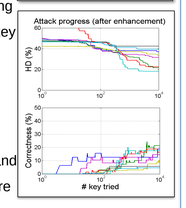
Why is the Attack Successful?

- The correct output response from a functional circuit can be utilized as a reference to determine how good a key is
- The effective key space is not large enough since 50% of the 2-bit permutation cells can be correctly permuted by simple random guessing
- The correlation between output HD and the key correctness is strong



Two security enhancement approaches:

- Key space enlarging using 8-bit permutation cells
- Reducing the correlation between output HD and key correctness by creating a interlock structure



IP Piracy Prevention Based on 2.5D IC Technology

2.5D IC Security Benefits:

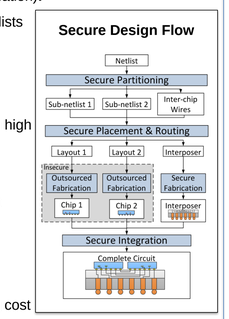
- The interposer can be fabricated in a trusted foundry to prevent complete exposure of a design and thus prevent IP piracy

Secure Partitioning (Functionality obfuscation)**:**

- Partitioning a netlist *F* into two sub-netlists F_1 and F_2 and hide cut-wires
- Goal: HD(*F*, *F'(F1,F2)*) = 50%
- Normal approach: min-cut partitioning
- Secure approach: select cut-wires with high controllability and observability

Secure placement (Layout obfuscation):

- Reduce the correlation between layout proximity and connection of two gates
- Goal: correctness = 0 (percentage of correctly guessed hidden wire)
- Secure Approach: simulated annealing cost function $Cost = f(Area, Wirelength, \boldsymbol{Correctness})$



Normal Partitioning + Normal Placement		Secure Partitioning + Secure Placement			
HD	Correctness	HD	Correctness	Area Overhead	Wire-length Overhead
13.62%	21.84%	47.75%	0%	6.17%	12.86%

This work is supported by NSF Grant 1223233