# Hardware Trojan Detection for Gate-level ICs Using Signal Correlation Based Clustering

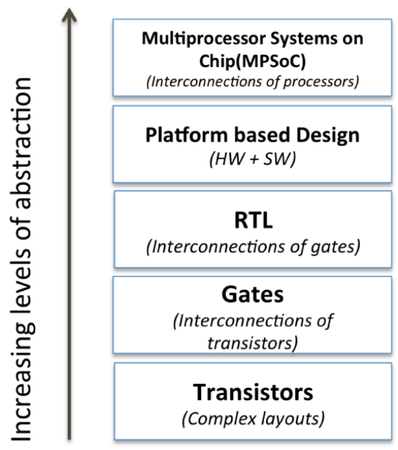
Burcin Cakir, Sharad Malik
Princeton University

Center for Future Architectures — CFAR — STARnet



## Malicious tampering of the internal circuits

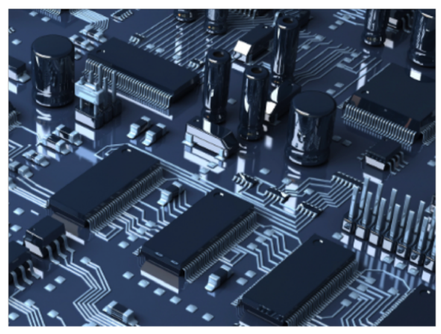
- DIGITAL CIRCUITS are commonly designed at multiple levels of abstraction.
  - Malicious behaviors can be inserted at various abstraction levels.
  - Often only post-synthesis gate-level netlists or actual silicon chips are available.
  - Difficult to reason about malicious design at gate-level.






## Problem Addressed

**Problem Addressed**
An information-theoretic approach for **Trojan detection**



- Estimate the statistical correlation between the signals in a design
- Explore how this estimation can be used in a clustering algorithm to detect the Trojan logic.



## Malicious tampering of the internal circuits

What is a hardware Trojan?

**`Hardware Trojans?`**
Malicious modifications of an integrated circuit (IC) that aims to compromise the integrity, security and reliability of the IC.

- Stealthy nature
- Small in size
- No change in IC physical characteristics
- A monitor in the chip
  - Wait for certain events or a sequence of events
  - ***Trigger the malicious circuitry***



## Overview

Detection of Trojans employing a statistical-correlation-based clustering

Using the simulation data

- Correlation-based similarity weight for each input-output pairs
- Gate-level design -> Circuit graph
- Weigh each edge based on similarity values

**Main Idea**
Trojan logic has weak statistical correlation with the rest of the circuit.

- Use the weights to obtain a *local connectivity distance*
- Apply a *density-based clustering algorithm* called (OPTICS)
- Output a *special kind of dendrogram*, called a *reachability plot*



## Step 1 - Functional Simulation based Statistical Correlation

Weight Computation

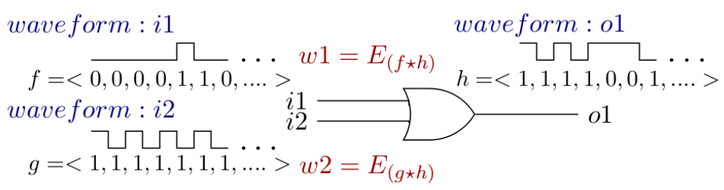



Figure: Weight calculation for the input-output pairs of an OR gate from the simulation waveforms by calculating the energy of the cross-correlation signal

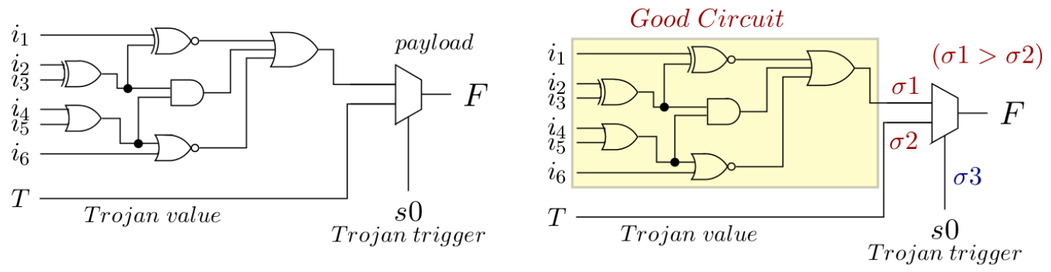


## Step 2 - Weight Normalization & Clustering

**Weight Normalization:**
The structural connectivity of the graph is needed. Important to identify the hubs and outliers.

Structural Similarity
Local connectivity density of two adjacent nodes in a weighted graph.

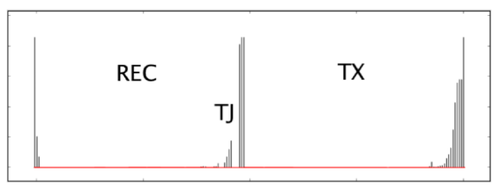
**Clustering:**



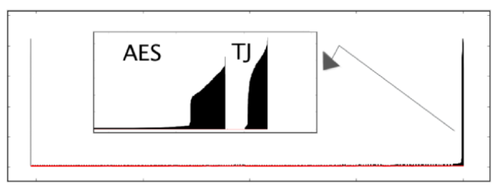



## Step 3 - Trojan Detection based on Reachability Plots

Trojan Logic on Reachability Plots




(a) Reachability plot for RS232-800 showing the receiver (REC) and the transmitter (TX) modules of the uart circuit with Trojan (TJ) logic pushed to the border of the REC cluster




(b) Reachability plot for AES-1800 with the Trojan (TJ) logic appearing as a separate cluster at the end of the ordered list

Figure: Types of reachability plots observed with TrustHub Trojan benchmarks



## Sensitivity and Specificity Analysis

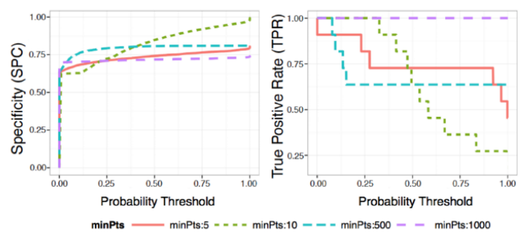



Figure: The TPR (*sensitivity*) and SPC (*specificity*) values for s35932-200 bechmark with different *MinPts*

Our tool,

- Effective finding suspicious nodes
- Estimates statistical distributions of the circuit

Difficult to activate a Trojan behavior,
- Not fully activated
- Activated but not propagated

Triggered, yet go undetected during logic testing,
- Not change any of the ports in the circuit
- Has invisible action

Table: Results for TrustHub Benchmarks

| Design Information | | Trojan Detection | | |
|---|---|---|---|---|
| name | gate/latch | *MinPts* | TPR(%) | SPC(%) |
| s15850-100 | 3478 | 50 | 61 | 99 |
| s35932-200 | 8107 | 10 | 27 | 99 |
| s38417-100 | 8422 | 50 | 100 | 99 |
| s38584-200 | 9548 | 50 | 99 | 98 |
| AES-1800 | 164800 | 50 | 92 | 99 |
| wb-conmax-200 | 20224 | 50 | 28 | 96 |
| PIC16F84-100 | 1616 | 20 | 75 | 96 |
| RS232-800 | 205 | 5 | 80 | 94 |

† As seen from TPR values, in each case, at least a quarter of the nodes of each Trojan have been identified.

In all cases, the reachability plots help

- Trojan logic as a rise in reachability-distance along the triggering path.
- Even be seen as a separate cluster