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

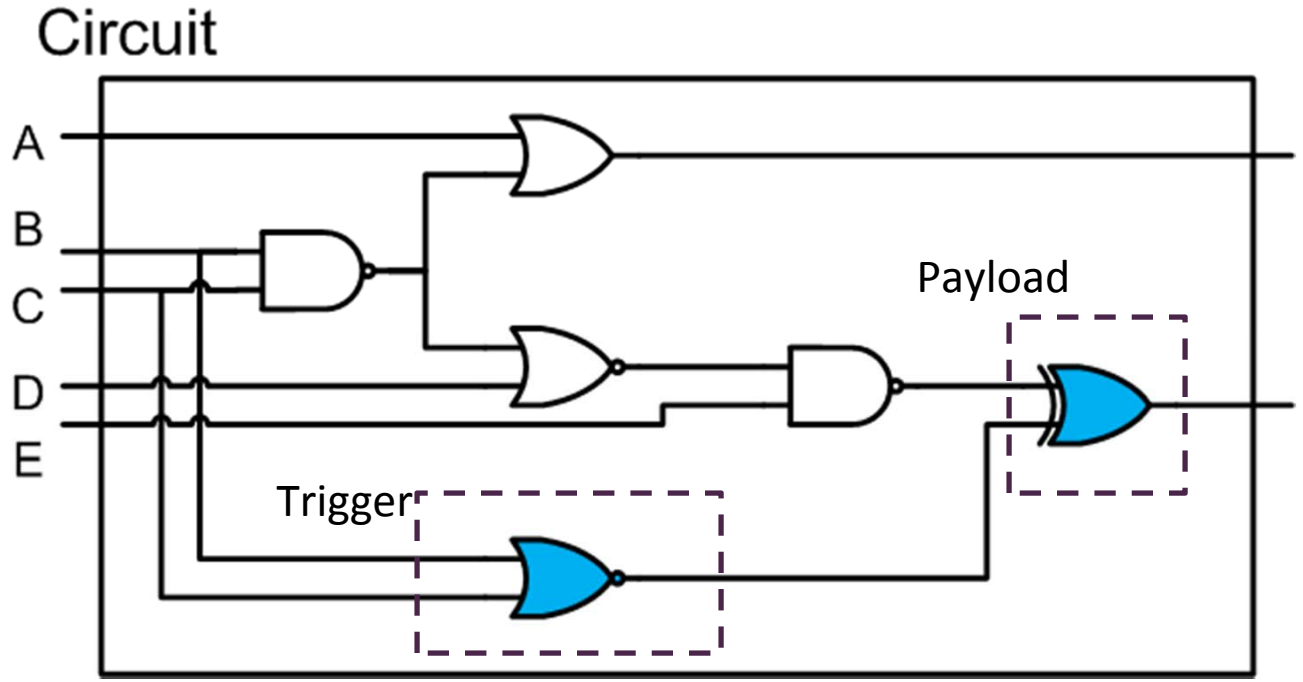
Best Paper Award in Application Design at the 2015 Design Automation and Test Conference

By:
Javad Bagherzadeh
Ameya Rane

Hardware Security Issue

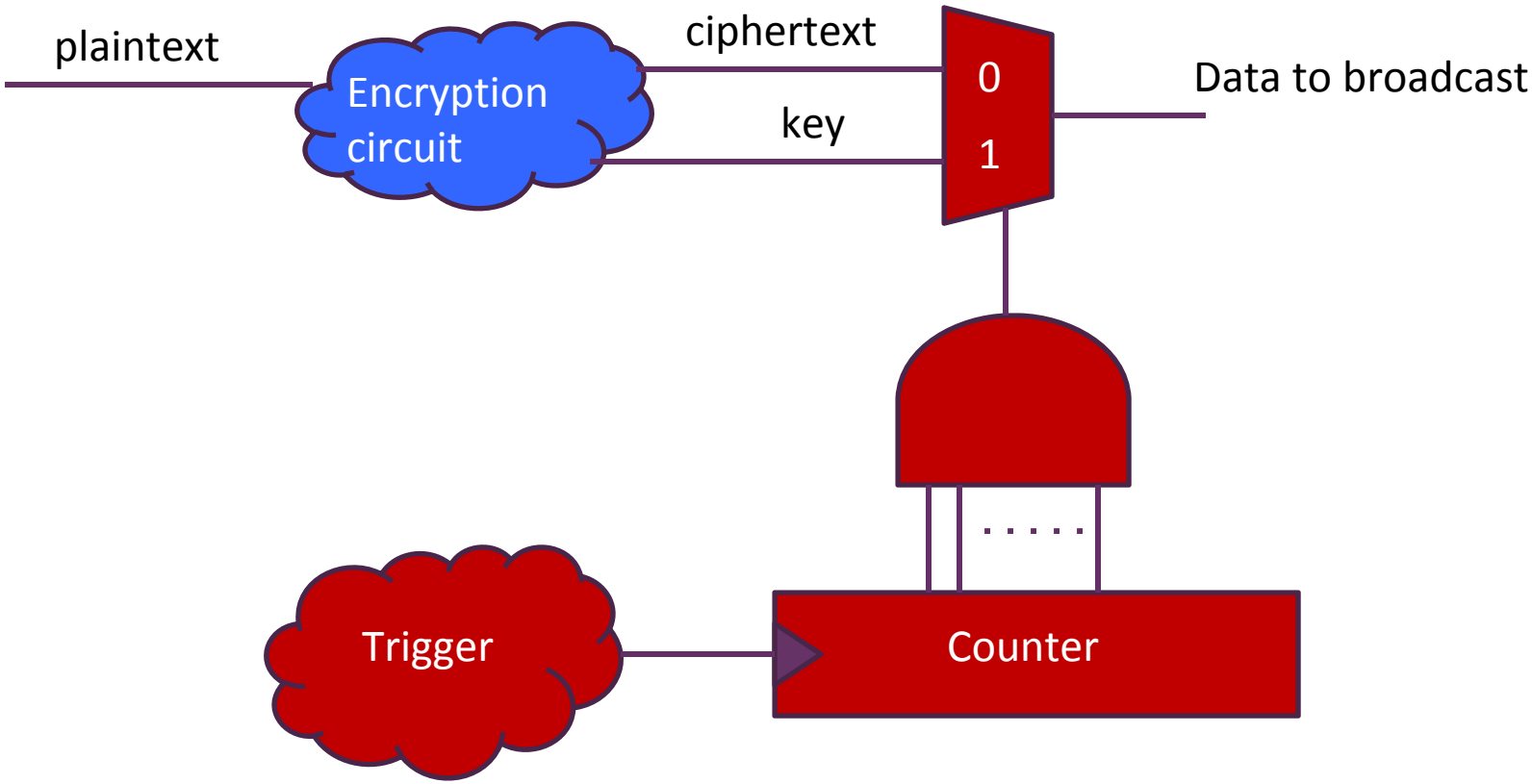
- Supply chain ICs spread globally to lower the cost
- More vulnerable to malicious modification
- **Hardware Trojan**: Inserted circuit by adversary which alters the intended circuit behavior
- **Trigger**: HT behaves like a monitor in chip and waits for
  - Certain events
  - Sequence of events
- **Payload:** Change the system behavior or leak confidential information

Circuit with Combinational Trojan

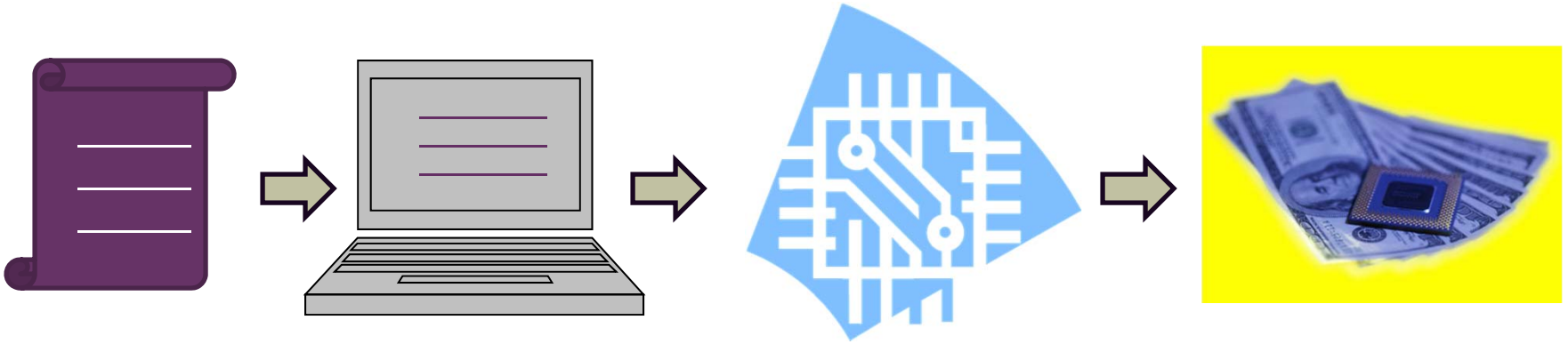


- Payload should affect something of functional importance to attack
  - Trigger should be stealthy
    - B=0, C=0 should be rare during functional operation
    - B=0, C=0 should not be targeted during structural test.

Sequential Trojan



Where can Trojans and Counterfeits be inserted?

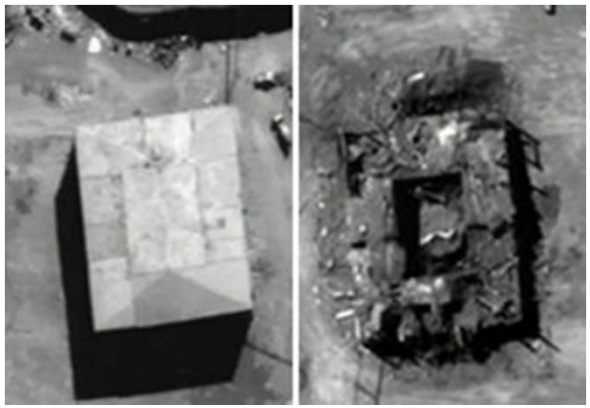


Spec **Design** **Manufacturing** **Supply Chain**

Why Hardware Trojans are Important

- Detection of Trojans has risen as a concern for possible threats
  - Military systems
  - Financial infrastructure
  - Medical devices
- Small in size and occupy only a small fraction of the circuit
- Can be inserted without changing
  - Circuit area
  - Die size
  - Pin count

Real Life Trojans….



before after

❖ On September 6, 2007, the Israeli Air Force carried out an airstrike on a Syrian nuclear reactor.

Hidden back door in microprocessors used in Syrian defense radar may have allowed them to be disabled remotely.

❖ French microprocessors used in military applications have remote “kill switches” to allow them to be disabled.

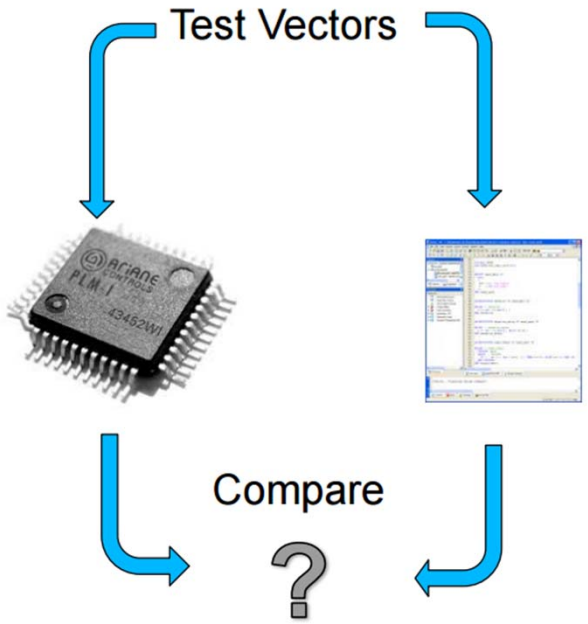
❖ During the Cold War, secret cameras were inserted inside Xerox 914 copy machines in the Soviet embassy to record copied documents.

Hardware Trojan Detection Methods

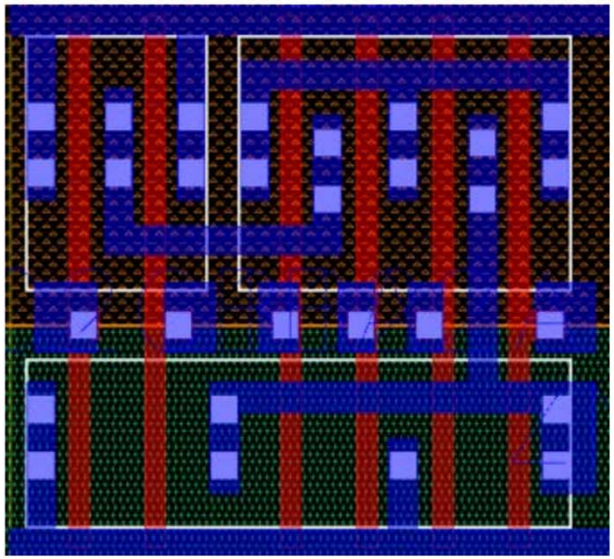
- Functional Testing
- Optical Inspection
- Side-Channels

Functional Testing

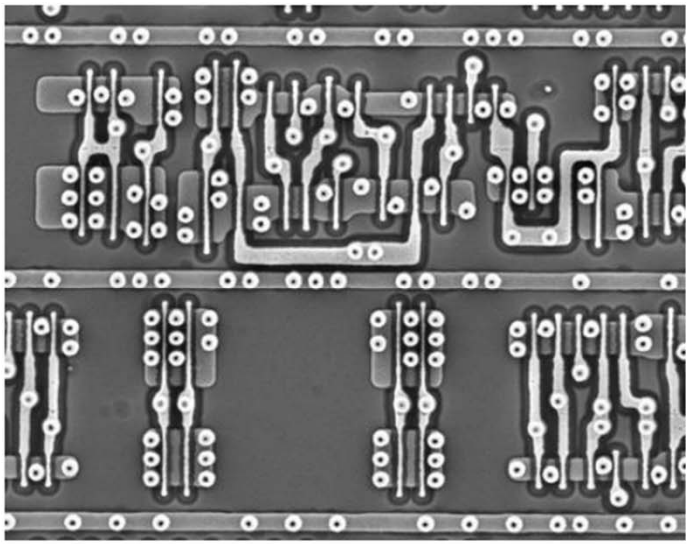
- It becomes almost impossible to exhaustively test the whole circuit
- Too hard to activate



Optical Reverse-Engineering



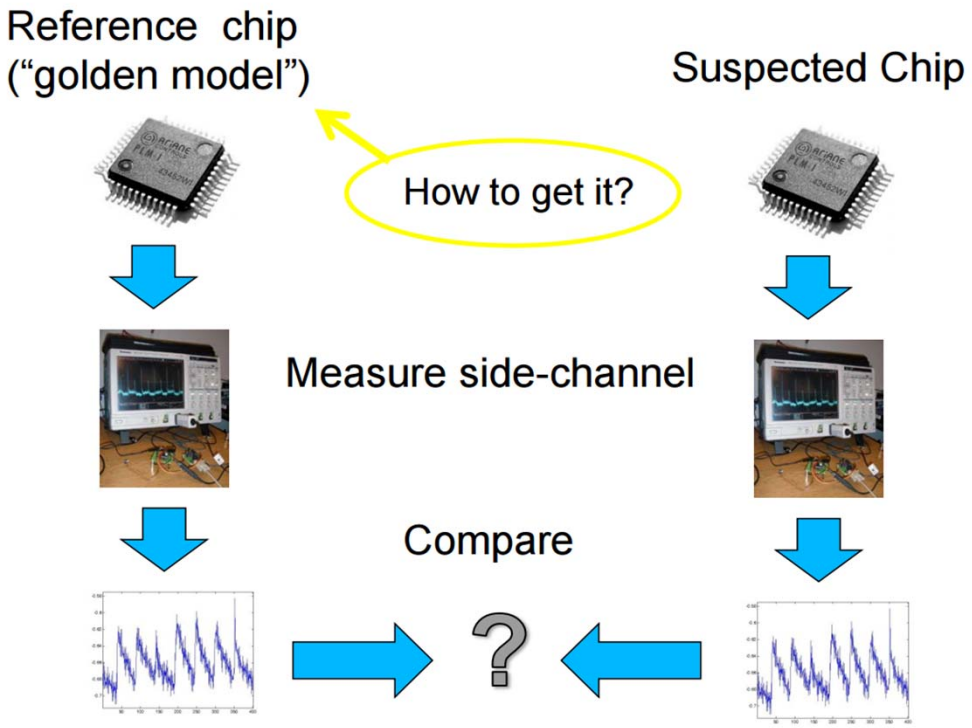
VS



Used to detect Trojans inserted during manufacturing stage

Side-Channel Comparison

- Side channels affected by even inactive Trojans
  - Delay
  - Power
- Process variations make comparison difficult



Overview of Paper

- Main idea :

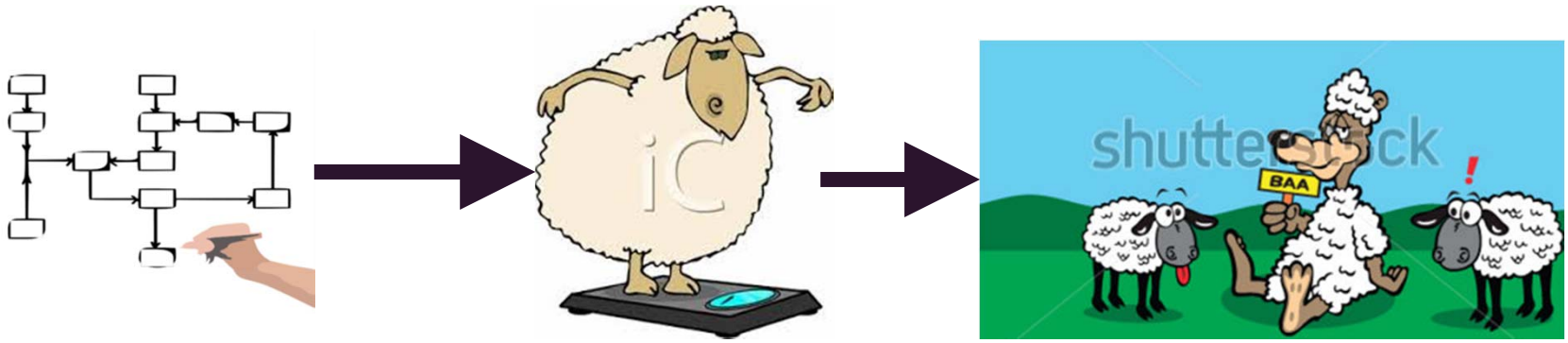
Trojan logic has weak statistical correlation with the rest of the circuit.

- Find statistically uncorrelated group of signals as this indicates some level of functional divergence

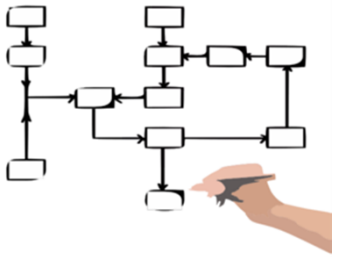
1- Circuit graph and weight calculation

2- Weight Normalization

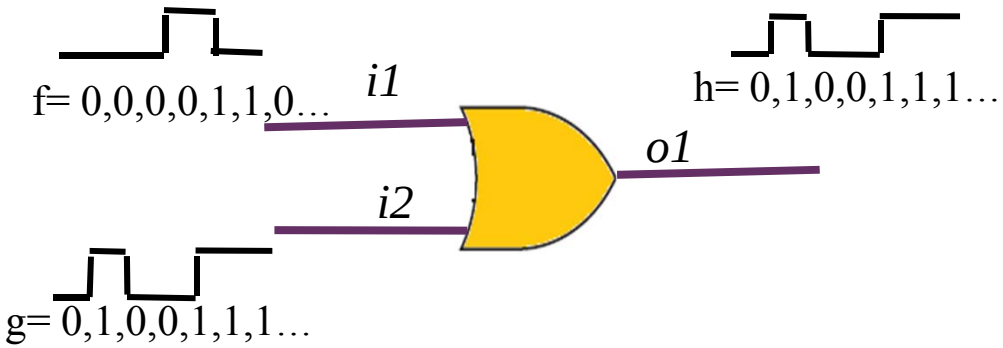
3- Trojan Detection

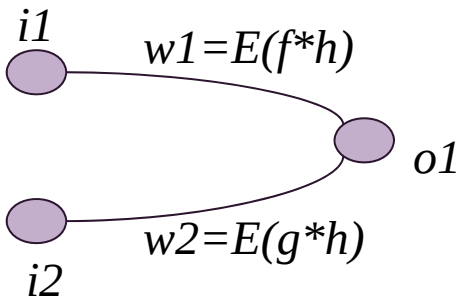


Step 1: Computation of Edge Weights on the Circuit Graph



- Creating Graph representation of circuit
- Simulating circuit with test sets developed during design verification
- Metric: **Cross-Correlation** of all adjacent signal pairs

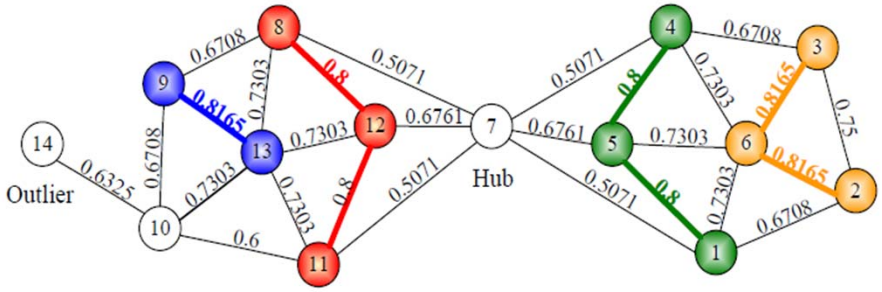




Step 2 - Weight Normalization:

- For the clustering algorithm, the structural connectivity of the graph is needed
- **Notion of structural similarity:** local connectivity density of two adjacent nodes in a weighted graph

$$\sigma(u,v) = \frac{\sum_{x \epsilon \Gamma(u) \cap \Gamma(v)} w(u,x) \cdot w(v,x)}{\sqrt{\sum_{x \epsilon \Gamma(u)} w^2(u,x)} \sqrt{\sum_{x \epsilon \Gamma(v)} w^2(v,x)}}$$



- The distance between each vertexes is inversely proportional to the σ(u, x).

OPTICS-Ordering Points To Identify Clustering Structure

- Hierarchical clustering algorithm for density based clusters
- OPTICS- ordering based on the smallest reachability-distance wrt object processed before it
- Reachability-distance: max(core-distance, actual distance)
- Low reachability-distance: object part of dense region
- High reachability-distance: noise or start of a new cluster

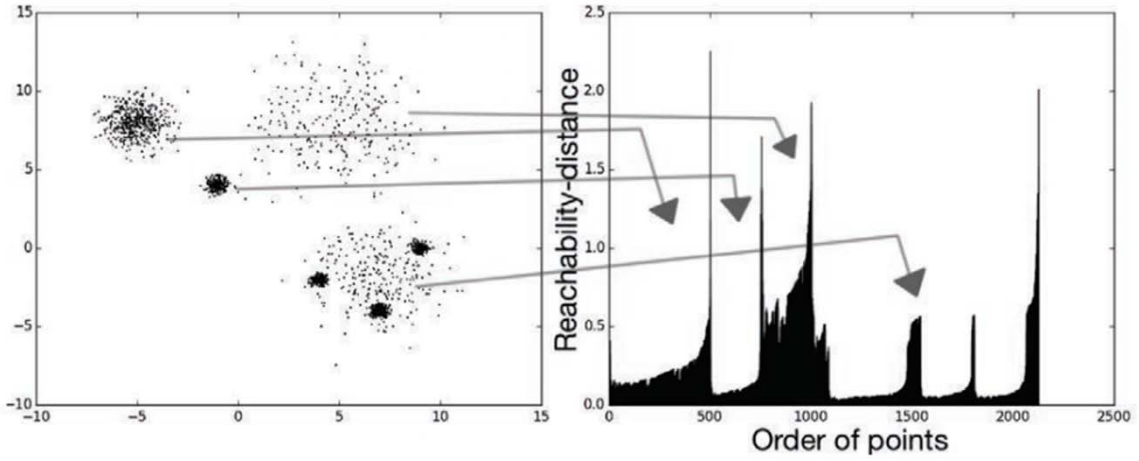


Fig: An example of application of OPTICS

Step 3: Trojan detection Overview

- Trojans - Functional divergence from the rest of the circuit
- Cluster placement and shape dependent on size of Trojan
  - Node wires - low controllability (backdoor)
  - Payload - low observability (keeping invisible)
- Implies Trojan activity will appear as noise
- Compute local density, find regions of similar density, vertices having lower density
- A probability value for being an outlier or not by fitting the Gaussian model

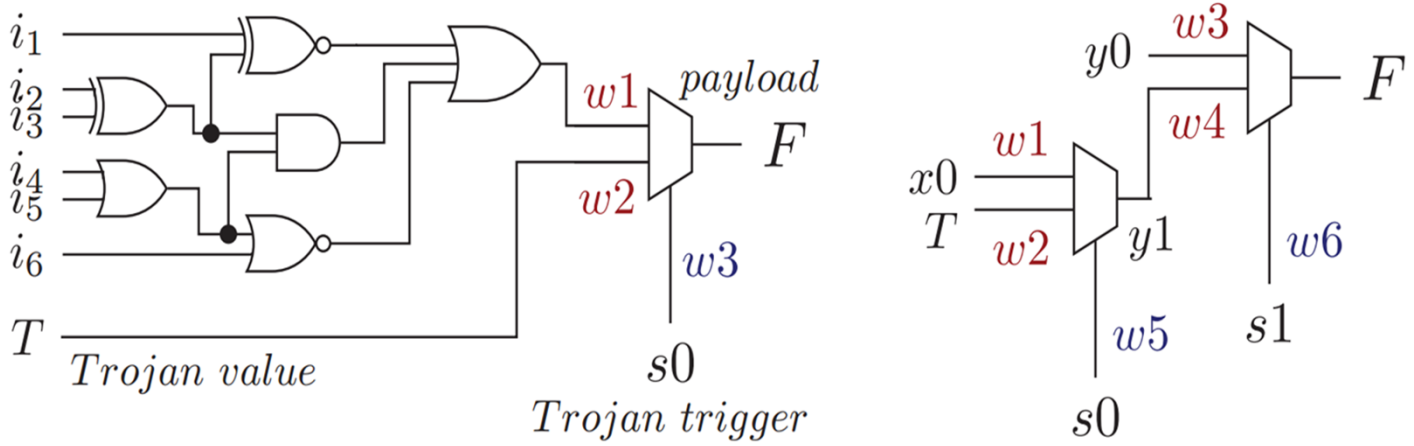
Step 3: Trojan detection based on the Plots

- Trojan Logic - Rise in reachability distance along triggering path
- Separate cluster if the malicious logic larger than a few gates

Detecting peak points

$$\downarrow lof(p) = \frac{\sum_{o \epsilon N_{cd}(p)} lrd(o)}{|N_{cd}(p)|} / lrd(p) \uparrow$$

Step 3: An Example



(a) Simple Trojan payload with weighted edges

(b) Distributed Trojan

(a) Low lof values for Trojan nodes
(b) w1, w3>>w2 due to stealthy nature of Trojan

- Behaves like outliner between two clusters, huge lof values
- No need to activate the Trojan in the circuit to detect the Trojan footprint

Plots for different benchmarks

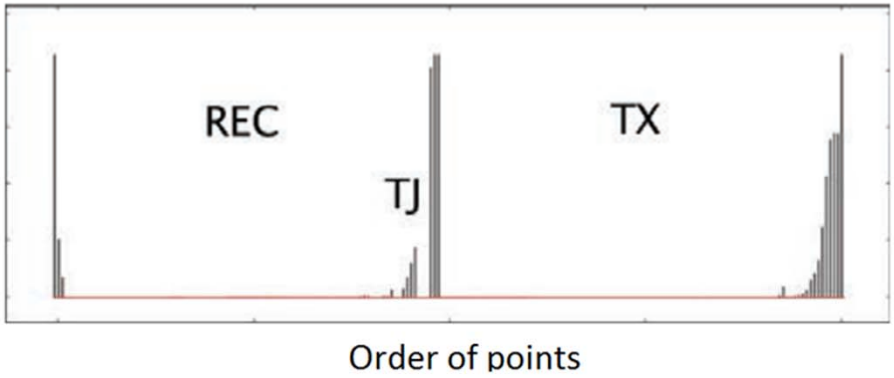


Fig a: Reachability Plot for RS 232 UART

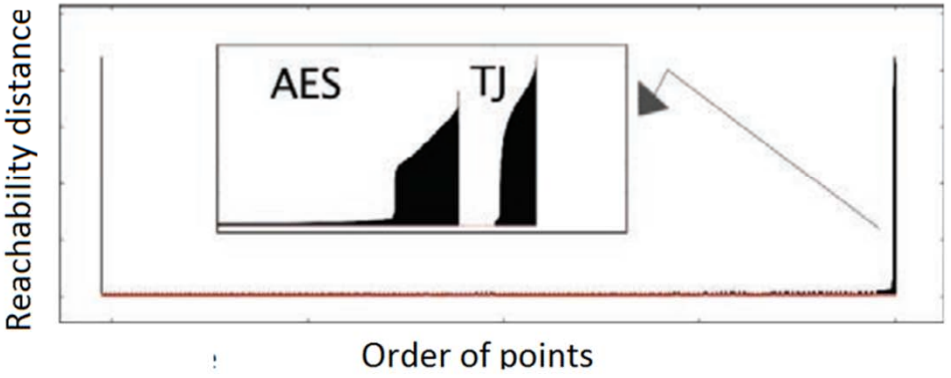


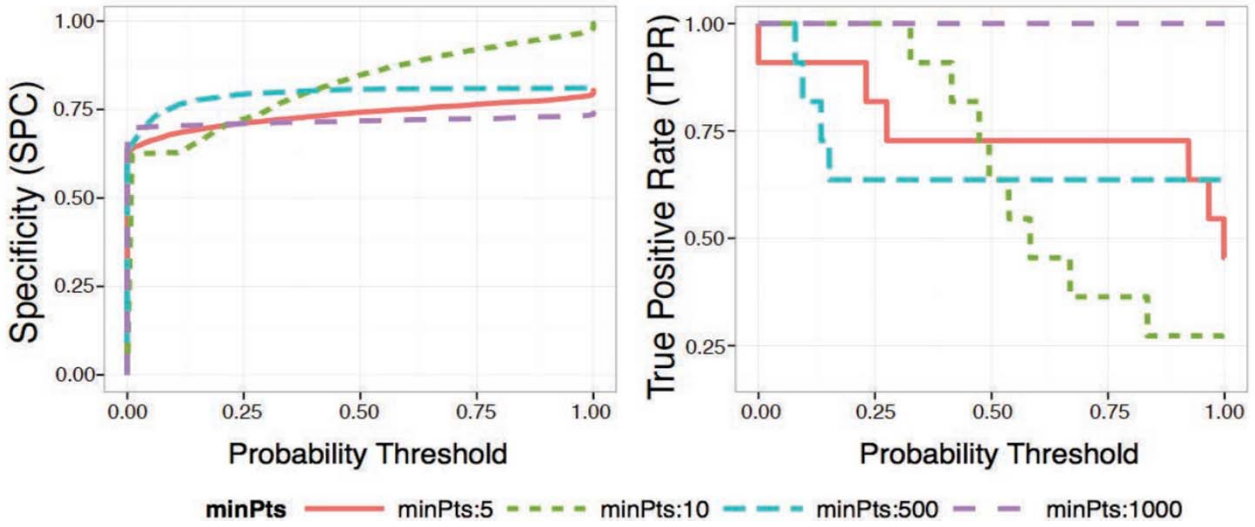
Fig b: Reachability Plot for AES-1800 Crypto IP

(a) Trojan cluster - couple of gates, triggering logic still pushed to the boundary

(b) Trojan causes excess power consumption after activation

Analysis of Results

- Exact nature of Trojans determined through code review of the nodes
- TPR – (Number of Trojan nodes correctly detected) /(Total number of Trojan gates) ↑
- SPC – Ratio of true negatives over the number of non-Trojan gates ↑
- (1-SPC) – Fraction of gates falsely flagged as being suspicious ↓



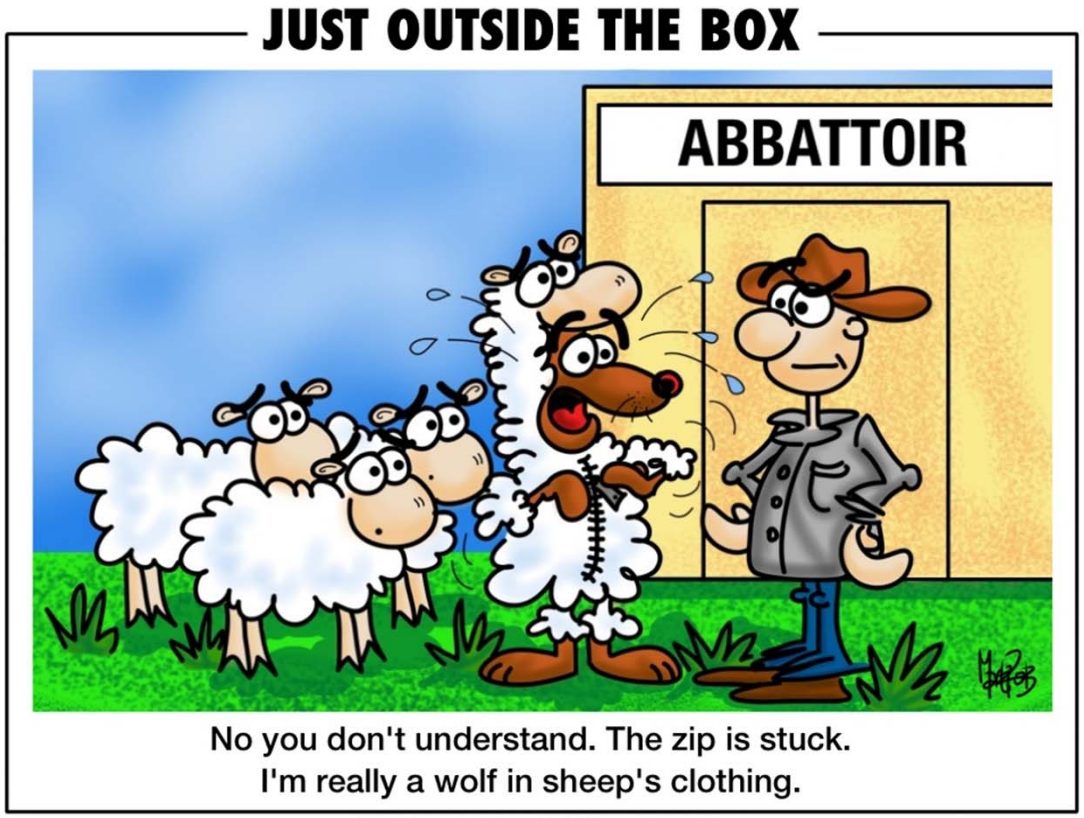
Experimental Take-away

- Choosing MinPts: Estimate size of smallest circuit module/cluster in dataset
- Target: Minimizing (1-SPC) while still detecting some part of Trojan circuitry
- Convention: Choosing MinPts values such that number of nodes flagged suspicious is manageable within the chosen threshold for manual inspection

Conclusion

- Investment on Hardware security increasing
- Paper proposes a simulation-based clustering technique
- Focus on reviewing a small subset, helps reduce authentication time
- Belief that the process will yield better results for real chips
- Concept still in testing stage, more work needed to make it foolproof

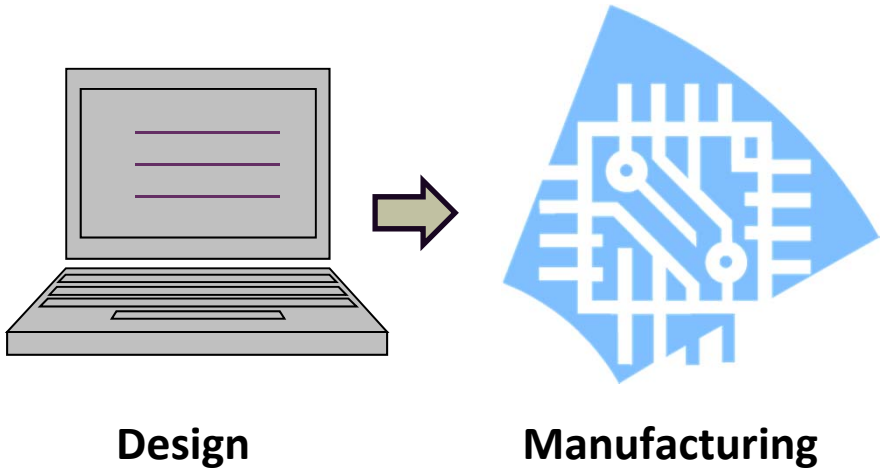
Questions?!


JUST OUTSIDE THE BOX
ABBATTOIR
No you don't understand. The zip is stuck.
I'm really a wolf in sheep's clothing.
Copyright www.justoutsidetheboxcartoon.com

Debate:

Is this method able to detect Hardware Trojans inserted in manufacturing stage?

Is it able to detect all kinds of Trojans with different type of functionalities of payloads (reveal information, causing error, etc)?



Design **Manufacturing**

Backup Slides

Experimental Setup

- TrustHub Verilog circuits, Synopsys Design Compiler, 45nm SOI
- Replaced the buffers and inverters with a vertex
- Ignored the clock and reset signals to reduce noise in the circuit
- Synopsys TetraMAX ATPG tool to generate tests for manufacturing faults
- Treated scan-chain as regular circuits